Pattern-Driven Design of Visualizations for High-Dimensional Data

Dissertation zur Erlangung des

akademischen Grades eines Doktors der

Naturwissenschaften (Dr.rer.nat.)

vorgelegt von

Michael Blumenschein (geb. Hund)

an der

Universität Konstanz

Mathematisch-Naturwissenschaftliche Sektion

Informatik und Informationswissenschaft

Konstanz, 2020

Tag der mündlichen Prüfung: 12. November 2020

1. Referent: Prof. Dr. Daniel A. Keim

2. Referent: Prof. Dr. Tobias Schreck

Acknowledgement

This thesis would not have been possible without the help and continuous support of *many* people – all of which contributed in their way to finish this thesis!

I want to sincerely thank my advisor Daniel Keim for the excellent opportunity to be part of an amazing (research) group. You gave me the scientific freedom to choose projects and papers of my interest and supported me with valuable advice when necessary. I also would like to thank my second advisor, Tobias Schreck, for his support during my journey, particularly at the beginning of my research career. You showed me how to think papers in terms of contributions and research questions. I would also like to say a big "Thank You" to Harald Reiterer for taking the time and effort of being part of the thesis committee, but also for quite a few critical questions and comments, which helped to shape my research and publications.

Before starting as a Ph.D. student, and during my journey, I had amazing mentors and people who showed me the beauty of research. Thank you, Michael Behrisch, Christian Rohrdantz, Milos Krstajic, Dominik Jäckle, as well as Wei-Nchih Lee (from HP Research Labs) for your constructive feedback, sharing your knowledge, your experience, and your thoughts. Your support significantly improved this thesis and helped me develop my scientific and personal skills on a more general level.

Mennatallah El-Assady – you are probably one of the main reasons why this thesis is finally printed. As a real friend and a fantastic mentor, you helped me keep up my motivation, make the right decisions, and pushed me to submit a paper when I thought it would be easier to wait for the next deadline.

I want to thank the whole Data Analysis and Visualization Group for your help and support, the interesting discussions, and the small talk around the coffee machine, making busy days still a nice and sometimes funny experience. You have become a family to me. In particular, I would like to thank Rita Sevastjanova for our deep conversations apart from work, Frederik Dennig, for the regular coffee breaks, and Alexander Jäger for being a valuable companion. Thanks to our support team (Matt Sharinghousen, Benedikt Bäumle, and Timon Kilian) for the help with all my technical issues, and Johannes Fuchs for sharing your expertise on designing user studies. Evanthia Dimara, in the last few meters, you helped me to improve my writing by teaching me to think about each and every word, sentence, and paragraph critically. And finally, Sabine Kuhr, who helped with me all administrative issues, particularly when it came to reconciling family and career.

Moreover, I would like to thank Luka Debbeler, Deborah Wahl, Karoline Villinger, Nadine Lages, and Britta Renner of the Psychological Assessment and Health Psychology Group. I have very much enjoyed our regular coffee breaks, collaborating interdiciplinaryly, exchanging views, and contributing to our both reasearch fields.

During the past years, I had the privilege to work with amazing students. Stefanie Schmid, Xuan Zhang, David Pomerenke, Jagrut Kosti, Alpin Türkoglu, and Yannick Schlosser: you have not only helped with some of the implementations but also shaped my research with fruitful discussions and exciting new ideas.

This thesis would not have been possible without the funding and amazing research network from SFB-TRR 161.

Finally, I want to say "Thank You" to my family and Moni for your love, support, and encouragement. Thank you for your help in finding my way in life and by sometimes having critical attitudes on my decisions. Last but not least, thanks to our latest family member, Tom, who encouraged me to let go of my perfection and finally write up this thesis.

Abstract

Data-informed decision-making processes play a fundamental role across disciplines. To support these processes, knowledge needs to be extracted from high-dimensional (HD) and complex datasets. Visualizations play hereby a key role in identifying and understanding patterns within the data. However, the choice of visual mapping heavily influences the effectiveness of the visualization. While one design choice is useful for a particular task, the very same design can make another analysis task more difficult, or even impossible. This doctoral thesis advances the quality and pattern-driven optimization of visualizations in two core areas by addressing the research question: *"How can we effectively design visualizations to highlight patterns – using automatic and user-driven approaches?"*

The first part of the thesis deals with the question *"how can we automatically measure the quality of a particular design to optimize the layout?"* We summarize the state-of-the-art in quality-metrics research, describe the underlying concepts, optimization goals, constraints, and discuss the requirements of the algorithms. While numerous quality metrics exist for all major HD visualizations, research lacks empirical studies to choose a particular technique for a given analysis task. In particular for parallel coordinates (PCP) and star glyphs, two frequently used techniques for high-dimensional data, no study exists which evaluates the impact of different axes orderings. Therefore, this thesis contributes an empirical study and a novel quality metric for both techniques. Based on our findings in the PCP study, we also contribute a formalization of how standard parallel coordinates distort the perception of patterns, in particular clusters. To minimize the effect, we propose an automatic rendering technique.

The second part of the thesis is user-centered and addresses the question *"how can analysts support the design of visualization to highlight particular patterns?"* We contribute two techniques: The *v-plot designer* is a chart authoring tool to design custom hybrid charts for the comparative analysis of data distributions. It automatically recommends basic charts (e.g., box plots, violin-typed visualizations, and bar charts) and optimizes a custom hybrid chart called v-plot based on a set of analysis tasks. *SMARTexplore* uses a table metaphor and combines easy-to-apply interaction with pattern-driven layouts of rows and columns and an automatically computed reliability analysis based on statistical measures.

In summary, this thesis contributes quality-metrics and user-driven approaches to advance the quality- and pattern-driven optimization of high-dimensional data visualizations. The quality metrics and the grounding of the user-centered techniques are derived from empirical user studies while the effectiveness of the implemented tools is shown by domain expert evaluations.

Zusammenfassung

Dateninformierte Entscheidungsprozesse spielen eine grundlegende Rolle in verschiedensten Disziplinen. Um diese Prozesse zu unterstützen, muss Wissen aus hochdimensionalen (HD) und komplexen Daten extrahiert werden. Visualisierungen spielen dabei eine Schlüsselrolle beim Erkennen und Verstehen von Mustern innerhalb der Daten. Die Wahl des visuellen Mappings beeinflusst jedoch stark die Effektivität der Visualisierung. Während ein Design für eine bestimmte Aufgabe nützlich ist, kann dasselbe Design eine andere Analyseaufgabe erschweren oder sogar unmöglich machen. Diese Doktorarbeit bringt die Qualität und mustergetriebene Optimierung von Visualisierungen in zwei Kernbereichen voran - und befasst sich dabei mit der Forschungsfrag: *"Wie können wir Visualisierungen automatisch und benutzergesteuert so gestalten, dass sie Muster hervorheben?"*

Der erste Teil der Dissertation befasst sich mit der Frage: *"Wie können wir die Qualität eines bestimmten Designs automatisch messen, um das Layout zu optimieren?"* Wir fassen den aktuellen Stand der Forschung im Bereich der Qualitätsmessung zusammen, beschreiben die zugrunde liegenden Konzepte, Optimierungsziele und Randbedingungen und diskutieren die Anforderungen an die Algorithmen. Während für alle wichtigen HD-Visualisierungen zahlreiche Qualitätsmetriken existieren, fehlen der Forschung empirische Studien zur Auswahl einer bestimmten Technik für eine bestimmte Analyseaufgabe. Insbesondere für Parallel Coordinates (PCP) und Star Glyphs, zwei häufig verwendete Techniken für hochdimensionale Daten, gibt es keine Studie, die die Auswirkungen verschiedener Achsenanordnungen bewertet. Deshalb trägt diese Arbeit je eine empirische Studie und eine neue Qualitätsmetrik für beide Techniken bei. Auf der Grundlage der Ergebnisse der Studie tragen wir auch eine Formalisierung bei, wie Standard PCPs die Wahrnehmung von Mustern, insbesondere von Clustern, verzerren. Um den Effekt zu minimieren, schlagen wir eine automatische Renderingtechnik vor.

Der zweite Teil der Arbeit ist benutzerzentriert und befasst sich mit der Frage, *"wie können Analysten den Designprozess von Visualisierung unterstützen, um bestimmte Muster hervorzuheben?"* Wir steuern zwei Techniken bei: Der *v-plot-designer* ist ein Tool zur Erstellung von Diagrammen. Dieses Tool erlaubt es, benutzerdefinierte Hybridcharts für die vergleichende Analyse von Datenverteilungen zu entwerfen. Der v-plot designer empfiehlt automatisch grundlegende Diagramme (z.B. Box-Plots, violin-chart typische Visualisierungen und Balkendiagramme) und optimiert ein benutzerdefiniertes Hybridchart mit dem Namen v-plot auf der Grundlage einer Auswahl von Analyseaufgaben. *SMARTexplore* verwendet eine Tabellenmetapher und kombiniert einfach anzuwendende Interaktion mit mustergesteuerten Layouts von Zeilen und Spalten und einer automatisch berechneten Zuverlässigkeitsanalyse auf der Grundlage statistischer Maße.

Zusammenfassend lässt sich sagen, dass diese Arbeit einen Beitrag zur Qualitätsmessung und zu nutzergesteuerten Ansätzen leistet, um die qualitäts- und mustergesteuerte Optimierung von hochdimensionalen Datenvisualisierungen voranzutreiben. Die Qualitätsmetriken und die Grundlagen der nutzerzentrierten Techniken werden aus empirischen Benutzerstudien abgeleitet, während die Wirksamkeit der implementierten Tools durch Expertenevaluierungen aufgezeigt wird.

Contents

1 Introduction

Knowledge derived from data is the foundation for data-informed decision-making across research and economic disciplines. These disciplines and their applications range from insurance companies, health care systems, and emergency response teams to logistics analysis, weather forecasting as well as psychology applications where large-scale human studies lead to insights that can help improve our life.

Due to the increasing awareness that data collection is the new fuel for smart decisions of huge economic value, more data than ever is measured, recorded, and generated on a daily basis. Data management becomes complex due to a huge number of data records (=observations) and dimensions (=attributes). Such data, often of dynamic nature, comes from heterogeneous data sources and contains anomalies, contradicting information, and missing values. The availability of cheap storage hardware leads to massive data being stored for potential usage in the future without any prior filtering or refinement. Analytic tools emerge as a solution to extract highly valuable information from those blindfold data lakes.

One of the key methods to successfully analyze such complex datasets are static and interactive visualizations. Abstract data and information are mapped to visual elements. Humans are visual creatures, and visualizations “augment human capacity by allowing us to surpass the limitations of our own internal cognition and memory” [Mun14, p.1]. Thereby, humans can make sense of large and abstract data [CMS99], for example, estimating the correlation from a sequence of numbers, or identifying the similarity of attributes.

Exploratory data analysis (EDA) [Tuk77] has been proven to be effective in getting an overview of an unknown dataset, or for ill-defined analysis problems. Analysts start without a concrete hypothesis and follow an explorative, interactive, and often undirected search for structures and trends in the data.

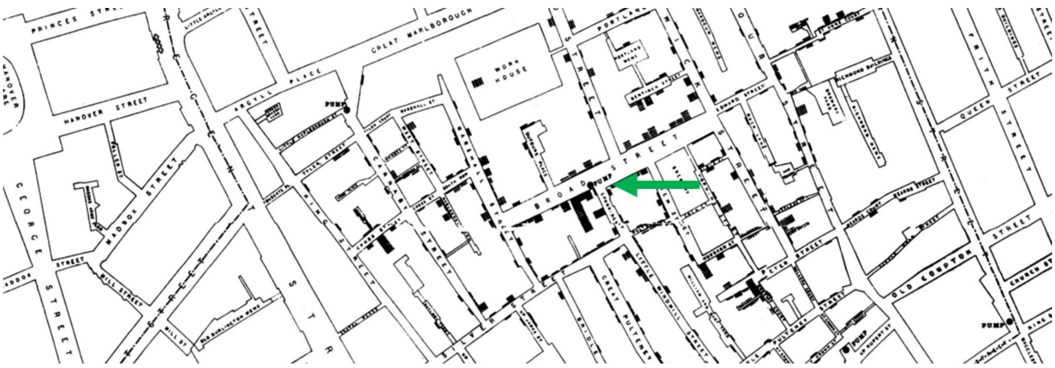


Fig. 1.1. Part of John Snows’s cholera map [Sno55]. Cholera cases are marked in black, revealing a high number of cases around a water pump at Broad Street.

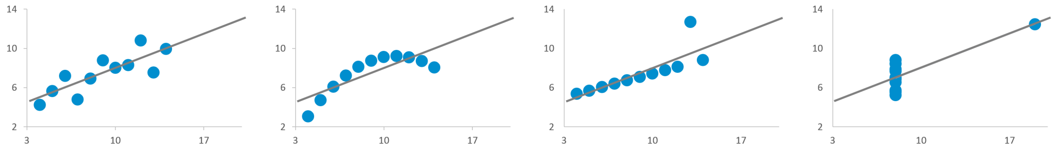


Fig. 1.2. Anscombe's quartet showing same summary statistics, but huge differences in their distribution. Charts recreated with original data from [Ans73].

The Cholera map from 1854 [Sno55] (see Figure 1.1) is a prominent example how visualizations can help to explore patterns in data. John Snow (1813–1858) investigated the Broad Street cholera outbreak in London. By talking to residents, Snow marked each death's location with a small black rectangle on a map. With the help of the visualization he identified the primary source of cholera as a public water pump.

Automatic data mining, machine learning algorithms, and statistical methods can also be used to extract knowledge from data. Their main advantage is that they are typically faster than manual exploration. However, automatic algorithms are often considered a black box, hiding potentially relevant data characteristics. A popular example is shown in Figure 1.2. Anscombe's quartet [Ans73] represents four different datasets with almost identical summary statistics (i.e., mean, variance, correlation, and regression), but significant differences in their actual distribution. If an analyst "relies only on the hard numbers", wrong hypotheses may lead to bad (business) decisions. Visualizations are the "interfaces" for human analysts. Hence, they can be effectively combined with automatic approaches for result analysis and verification, and parameter tuning of automatic algorithms.

In summary, visualizations play an essential role when knowledge is derived from large and complex datasets. They can be used as a primary analysis method, or support the understanding of automatic analysis methods. The question is, however, how to design a visualization such that it is most effective?

1.1 Design Challenges of Visualizations

The design space of visualizations is huge, making it hard to ultimately come up with a design that fits a particular analysis task and corresponds to the characteristics of the data. Further design challenges are the limitations of computers, displays, and the human analyst. Hence, many designed visualizations are, therefore, full of trade-offs and often not effective for a particular task [Mun14].

To create a visualization, there is a complex iterative process involved as described by Card et al. [CMS99] (see Figure 1.3). During that process, the data analyst is overwhelmed with numerous choices, such as selecting the data transformation and visual mapping to be applied, and deciding for a view transformation. To ensure that the final visualization will be indeed effective, the steps of "visual mapping selection" and "view transformation" require expert visualization design skills.

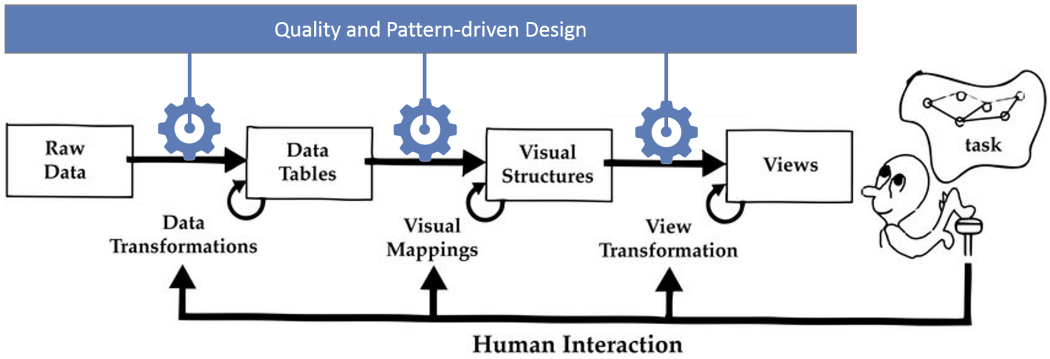


Fig. 1.3. Visualization reference model by Card et al. [CMS99]. Original figure is extended by a quality- and pattern-driven automation, supporting to design good views.

Non-visualization experts are particularly affected by the complex design process. Grammel et al. [GTS10] shows that information visualization novices face three barriers when designing effective charts: selecting data attributes, designing visual mappings, and interpreting the visualizations. A poor selection of attributes and visual mappings can hinder the data analysis workflow often results in misleading conclusions [Hee+08].

From a general perspective, there are two categories of design choices which have to be made: First, an analyst needs to select a *visualization type* for a given dataset and analysis question. Second, the *properties of the respective visualization* need to be adjusted. Regarding the visualization type, Saket et al. [SED19] recently investigated the performance of five common visualization techniques (Table, Line Chart, Bar Chart, Scatterplot, and Pie Chart) for ten different analysis tasks and conclude "that the effectiveness of [...] visualization types often significantly varies across tasks." Earlier studies for other visualizations types and analysis tasks (e.g., [CG14; Har+14; SL91; Kos19]) also found significant performance differences. Hence, as a first step, it is crucial to choose an appropriate visualization type for an analysis scenario.

Once a particular visualization is being chosen, the properties of visualization need to be selected. For example, mapping dimensions to visual variables, choosing colors and colormaps, arranging dimensions, and scaling distributions. Many user studies exist which measure the performance of particular visual designs for different analysis tasks. For example, Harrower & Brewer [HB03] investigate the usage of different colormaps, Johansson & Johansson [JJ09] discuss that the order of axes in parallel coordinates influence how analysts perceive patterns in the data, Kosara & Skau [KS16] judge errors in different pie chart variations, and Skau et al. [SHK15] analyze the impact of embellishments in bar charts. These and other studies highlight that it is essential to choose particular design variations carefully.

1.2 Support for Visualization Design and Open Research Questions

The goal of this thesis is to support users in the design of visualizations to extract patterns of interest. In the following, we summarize the state-of-the-art to support the design of effective visualizations. Based on the results, we highlight three open research challenges tackled in this doctoral thesis.

To be of broad usage, this thesis focuses on *multi-* and *high-dimensional* (HD datasets as commonly given in many applications. Optimizing visualizations for high-dimensional data are particularly challenging due to the high number of data records and dimensions.

(1) Limited Visualization Studies on Design Variations

In many cases, selecting a visualization type and finalizing the design choices are common sense and well-studied. For example, using a line chart instead of a bar chart to represent a value that is changing over time. However, many design choices are not obvious, difficult to select, and dependent on the analysis task and characteristics of the data. As a fundamental step, empirical studies are necessary to understand and justify design choices in various settings. Such studies can then be provided as general guidelines, or be encoded as rules and suggestions into chart authoring tools.

Many user studies have already been conducted for various design variations. Most of these studies are linked to particular applications or analysis tasks. Hence, there are clear guidelines for a large number of techniques and design variations. However, there are still many empirical studies for important visualizations missing. For example, parallel coordinates and star glyphs are, among others, the most commonly used visualizations for high-dimensional data. Many authors claim that the ordering of axes in both techniques plays an important role when designing the visualization and can either highlight or hide interesting patterns. For parallel coordinates more than 30, and for star glyphs, more than ten ordering algorithms have been proposed. However, none of the techniques have been evaluated in an empirical study.

As a result, we, as a research community, do not know which design variation is best for a particular task and a specific data characteristic. Therefore, we cannot provide justified guidelines to practitioners. It is essential to identify the design choices which have not been evaluated yet, compare the approaches, and conduct corresponding user studies. This thesis contributes empirical studies for axes orderings for parallel coordinates and star glyphs.

(2) Lack of Chart Authoring Tools for Concurrent Tasks

When visualization designers want to create useful charts, they are faced with one of the following two choices: Either rely on their expertise, read information visualization books (e.g., [Spe14; War20; Mun14; TS20]), and keep studying the most recent results of user studies in which design alternatives are compared based on some study constraints. Then use one of the many tools and user interfaces to build the final visualization. Or, rely on automatic chart recommendation engines that encode the knowledge of the literature into re-usable information. The first option is very time-consuming; it is also not feasible for a general audience to keep up with the latest research in different fields. Hence, there is a need for automatic support in the design process of visualizations.

The general idea of such automatic support is shown on top of the visualization reference model in Figure 1.3. Quality-driven automation helps with the entire design process of visualization while allowing users to encode their own decisions to choose, for example, domain-specific requirements. These (semi-)automatic chart authoring tools typically use a combination of established rules, design guidelines (e.g., Bertin's work on visual variables [Ber83], Cleveland & McGill work [CM84], or the GestaltLaws by Wertheimer [Wer23]), the results of empirical user studies, and metrics which measures properties of a visualization.

However, the usefulness of a particular visualization design depends on the user's selection of tasks. While a design can be useful for a particular task, the very same design can make another task more difficult, or even impossible. There is often a trade-off if multiple tasks are relevant at the same time. Take, for example, the comparative analysis of data distributions. One of the most commonly used charts that you find in the literature are box plots. These charts easily support the identification or comparison of the median values and the general spread of the data. However, box plots are not a good design choice if the frequency of individual data records is of interest (for example, in discrete distributions), or if the shape of the distribution needs to be described. In many application scenarios, it is not possible to select one chart which supports many tasks at the same time - and all with the same focus and quality.

Most chart recommendation engines focus on proposing a visualization type for a given task. While this may be sufficient for some basic tasks, there is still a lack of automating the design of a particular visualization. This is especially true when multiple tasks are relevant at the same time. More work needs to be done to develop such chart recommendation engines, particularly for a large set of concurrent tasks. This may also involve to extend or combine existing visualizations which are then capable of supporting a multi-task analysis.

(3) Lack of Tools Combining User & Pattern-driven Designs

Many patterns in complex data can only be detected and characterized during an exploratory analysis. The user needs to tell the system how interesting patterns for a particular application (may) look like. Then the system can help to find these patterns in the data and propose similar patterns that are potentially interesting as well. Guidance can go in two directions: the system provides recommendations for views and to optimize the design choices, while the user provides input to the system, tailoring the analysis in a particular direction [CGM19].

Many tools have been developed to navigate high-dimensional datasets and support a user-guided exploration. To name a few, recently, LDSScanner [Xia+18], Subspace Voyager [WM18], Dimension Projection Matrix/Tree [Yua+13] and a quality-metric guided framework for exploratory dimensionality reduction by Fernstad et al. [FSJ13] have been presented. However, these tools are designed for expert users due to its complexity. Furthermore, most approaches miss real support to identify patterns and are based on a dimensionality reduction technique, which hides the actual data records and values.

Instead, there is a need for tools with interaction possibilities and visual representations that are easy to understand or shallow learning curves to be useful for a broad audience ranging from novice users to InfoVis experts.

1.3 Contributions and Structure of the Thesis

To tackle these limitations, this thesis addresses the following research question: *"How can we effectively design visualizations to highlight patterns – using automatic and user-driven approaches?"* The thesis thereby advances the quality and pattern-driven design and optimization of visualizations in two core areas, which structure the two parts of this thesis.

Part I: Quality Metric-Driven Design for Pattern Analysis

Part I focuses on *quality metrics* and addresses the question *"how can we automatically measure the quality of a particular design to optimize the layout?"*. A large body of research has been done to develop quality metrics for different visualizations and design variations. Many of these metrics follow similar concepts (also across visualization types) but differ in their vocabulary, or in their understanding of what quality means. Chapter 2 **contributes a survey of quality metrics research for visualizations for high-dimensional data**. In this survey, we unify the vocabulary, enumerate on the different metrics, and highlight research gaps - in particular with respect to (empirical) user evaluation. In particular, we focus on the following visualization techniques: scatter plots, scatter plot matrices, parallel coordinates, pixel-based techniques, radial visualizations, and glyph representations.

Based on this survey, we identified that for two of the most common visualizations for high-dimensional data, parallel coordinates, and star glyphs, necessary user studies are missing. Many axes orderings have been proposed in the literature, but no empirical validation has been conducted yet. Therefore, in this thesis, we push axes reordering for the two visualization approaches towards empirical guidance by **conducting a user study for parallel coordinates (Chapter 4) and star glyphs (Chapter 3) for cluster identification tasks**. We choose cluster analysis as the primary focus as the majority of strategies are design for this task. Our main findings are that ordering dimensions based on dissimilarity (place dimensions with a high dissimilarity next to each other) outperform the often proposed similarity-based arrangement in different settings.

While experimenting with different axes orderings in parallel coordinates, we found out that standard parallel coordinates distort the perception of patterns, in particular clusters. This problem is inherent to the technique itself: diagonal line segments are rendered longer (=need more pixels) and closer to each other (=less background color), compared to horizontal lines. As a consequence, clusters are distorted, and ghost clusters (fake clusters, not existing in the data) can emerge. In Chapter 5, **we contribute a formalization of this problem an provide and automatic method to adjust the rendering of the polylines based on their slope** to reduce these effects.

Part II: User- and Task-Driven Design for Pattern Analysis

Part II of this thesis provides *user- and task-driven* approaches to (semi-)automatically optimize visualizations. This second part addresses the question *"how can analysts support the design of visualization to highlight particular patterns?"* In many applications, the design of visualizations and the selection of visual elements depends on the underlying analysis tasks and may even need a highly iterative approach to describe and identify the patterns of interest. The second part of this thesis, therefore, contributes two analysis techniques that advance the automatic design of visualizations from a user-centered research perspective.

The *v-plot designer* (Chapter 6) is build for the comparative analysis of data distributions. Based on the selection of one or multiple analysis tasks, the v-plot designer **proposes an automatic recommendation of basic charts (e.g., box plots, violin-typed visualizations, and bar charts), along with a customized hybrid chart which is called a *v-plot*.** v-plots are automatically optimized to support all selected analysis tasks, and highlight required distribution properties. The automatic recommendations and the system design are grounded in a user study of 20 InfoVis and statistic practitioners, providing a solid foundation for the automation of the v-plot designer.

The second technique, *SMARTexplore* (Chapter 7), uses a table-based representation to simplify the analysis of a high-dimensional dataset for both novice and expert users, alike. Rows of a table can be aggregated manually, or with the help of clustering algorithms. Dimensions can be grouped into semantically meaningful subspaces, or automatically into groups of similar dimension patterns. **SMARTexplore combines easy-to-apply interaction concepts with the automatic and pattern-driven layout of rows and columns of the table.** The reliability of the perceived patterns can be verified by an automatic performed statistical analysis, which is encoded as possible overly in the visualization.

In summary, this thesis contributes quality metrics and user-driven approaches to advance the pattern-driven design of high-dimensional data visualizations. The quality metrics and the majority of the support for the user-centered approaches are derived from empirical user studies. The effectiveness of the user-centered approaches is shown by domain expert evaluations, typically conducted in pair-analytic sessions.

Based on the different techniques, this thesis has a strong InfoVis focus. However, there are core contributions in other computer science fields. To give an overview, these contributions are summarized in the following table:

Tab. 1.1. Relative importance of thesis chapters' contributions for computer science subfields. Rating schema: some relevance ○○●, largely relevant ○●●, highly relevant ●●●.

Computer Science Fields	Ch. 2	Ch. 3	Ch. 4	Ch. 5	Ch. 6	Ch. 7
Information Visualization	○●●	●●●	●●●	●●●	●●●	○○●
Visual Analytics	○●●	○○●	○○●	○○●	○●●	●●●
Evaluation	○○●	●●●	●●●	●●●	○●●	○○●
Applications	○○●	○○●	○○●	○○●	○●●	●●●

1.4 Citation Rules and Contribution Clarification

As it is the accepted scientific practice and guidelines of the research community in computer science, all major contributions of this thesis have been previously published in journals and conference proceedings. I retain the copyright of my publications that are the basis for this thesis. Parts of thesis chapters, which appear verbatim in my publications, were either written by myself or were rephrased by myself during the paper or thesis writing process.

To avoid any suspicion about plagiarism and self-plagiarism, I try to be as transparent as possible concerning the origin of all chapters of my thesis. In Section 1.5, I list all publications that I authored or co-authored. I specify the contribution and work distribution among all authors for each paper.

At the beginning of each chapter, I state the publication from which texts and figures are taken or adapted. For these integrated publications, I use the following rules:

- Quoted paragraphs are not written by myself and contain contributions of other authors.
- Chapters "taken from" my publications are copied and differ only in slight wording changes. These chapters contain my own contributions, and I did all writing myself or rephrased the chapters during the paper writing process.
- Chapters "based on" a publication are mostly rephrased, and the content has been modified. These chapters contain my own contributions but were changed to fit nicely into this thesis.

This resulting thesis is a trade-off between a nicely readable dissertation (rewriting of all my peer-reviewed articles) and a thesis following the strictest citation rules (quoting all sections being related to a publication). I decided to focus on the content, contributions, and the reader, as I believe these to be most important.

1.5 Publications

During my time as a doctoral researcher, I published several publications in high-level journals and conferences. These publications are foundation of this thesis:

[Beh+18] Michael Behrisch, **Michael Blumenschein**, Nam Wook Kim, Lin Shao, Mennatallah El-Assady, Johannes Fuchs, Daniel Seebacher, Alexandra Diehl, Ulrik Brandes, Hanspeter Pfister, Tobias Schreck, Daniel Weiskopf, and Daniel A. Keim. *"Quality Metrics for Information Visualization"*. In: Computer Graphics Forum 37.3 (2018), pp. 625–662.

Contribution clarification. This paper is a collaborative effort between many authors, in particular, Michael Behrisch and myself. M. Behrisch and I contributed equally to the paper. I initiated the project and designed the overall content and structure of the different chapters. During the writing process, M. Behrisch took over the lead to finalize the paper due to other obligations on my part. Ulrik Brandes, Hanspeter Pfister, Tobias Schreck, Daniel Weiskopf, and Daniel A. Keim supervised the project and regularly provided feedback on paper drafts. Nam Wook Kim, Lin Shao, Mennatallah El-Assady, Johannes Fuchs, Daniel Seebacher, and Alexandra Diehl provided references and material for the different visualization techniques and their quality metrics. In particular, for the visualizations for high-dimensional data, material and drafts for parallel coordinates were written by myself, scatter plot and scatter plot matrix was provided by L. Shao, for radial and pixel-based visualizations by D. Seebacher, and glyphs by Johannes Fuchs. I was responsible for unifying the content of all HD visualizations and writing large parts of the introduction, background, methodology, and discussions. M. Behrisch was responsible for finalizing the visualizations of the other data types, i.e., visualizations for relational data, geospatial data, sequential, temporal, and text data. All sections used in this thesis were either written by myself or revised by myself several times during the writing process. Hence, I use the material without any citation marks in Chapter 2 and Chapter 8.

[Mil+19] Matthias Miller, Xuan Zhang, Johannes Fuchs, and **Michael Blumenschein**. *"Evaluating Ordering Strategies of Star Glyph Axes"*. In: IEEE Visualization Conference (VIS). 2019, pp. 91–95.

Contribution clarification. This paper is a close collaboration between Xuan Zhang and myself (I supervised her BA thesis). I had the idea to compare and analyze different axes ordering strategies of star glyphs in an empirical user study. I also defined the research question and contribution. Daniel A. Keim and Johannes Fuchs provided feedback on the general idea, the study design, and regularly commented on paper drafts. X. Zhang implemented the tool and conducted the user study based on my input. The statistical analysis was done by myself. I was responsible for the writing of the paper. Matthias Miller provided initial drafts of the abstract, introduction, related work, and the figures. I revised these paragraphs several times during the writing process. The other parts of the paper were written entirely by myself. Hence, I use the text without citation marks in Chapter 3.

[Blu+20b] **Michael Blumenschein**, Xuan Zhang, David Pomerenke, Daniel A. Keim, and Johannes Fuchs. *"Evaluating Reordering Strategies for Cluster Identification in Parallel Coordinates"*. In: Computer Graphics Forum 39.3 (2020), pp. 537–549.

Contribution clarification. This paper is also a close collaboration between Xuan Zhang and myself (I supervised her BA thesis). Based on the result of X. Zhang's thesis, we designed the user study conducted in this paper. Johannes Fuchs supervised this paper project, and Daniel A. Keim provided feedback on the general idea and commented on paper drafts. I had the idea to summarize and categorize different reordering approaches for parallel coordinates, and to conduct a user study to evaluate two particular reordering approaches. I also defined the research question and contribution. The user study design and hypothesis were a result of regular discussions among J. Fuchs, X. Zhang, and myself. X. Zhang implemented the reordering techniques and an initial version of the interface for the user study. Based on my input, David Pomerenke also contributed to the study implementation and supported the analysis of the study results. X. Zhang and D. Pomerenke helped to finalize the supplementary material. All writing was done by myself, or I revised paragraphs several times during the writing process. Thus, I use the text without citation marks in Chapter 4.

[Pom+19] David Pomerenke, Frederik L. Dennig, Daniel A. Keim, Johannes Fuchs, and **Michael Blumenschein**. *"Slope-Dependent Rendering of Parallel Coordinates to Reduce Density Distortion and Ghost Clusters"*. In: IEEE Visualization Conference (VIS). 2019, pp. 86–90.

Contribution clarification. During the design of the benchmark dataset in our parallel coordinates study [Blu+20b] (see above), I identified the problem that parallel coordinates may distort the perception of clusters. David Pomerenke had the idea to draw the line segments of the parallel coordinates plot with a different width to overcome the problem. He also implemented different versions of the adjustment algorithm based on the suggestions by myself. The final rendering formula is a collaborative effort of D. Pomerenke, F. Dennig, and myself. I was responsible for the writing of the paper and identified the research question and contribution. D. Pomerenke and F. Dennig supported the structure of the paper with fruitful discussions and provided initial drafts for the related work and the description of the rendering algorithm. Daniel A. Keim and Johannes Fuchs provided feedback on the general idea and commented on paper drafts. All sections of the paper were written by myself, or I revised them several times during the writing process. Thus, I use the text without citation marks in Chapter 5. In this thesis, I extended the content of this chapter by a more detailed description of the rendering formula. This text is based on the supplementary material of the corresponding paper [Pom+19] (available at `https://osf.io/sy3dv`) and also authored / revised by myself. Hence, I also use this material without citation marks.

[Blu+20a] **Michael Blumenschein**, Luka J. Debbeler, Nadine C. Lages, Britta Renner, Daniel A. Keim, and Mennatallah El-Assady. *"v-plots: Designing Hybrid Charts for the Comparative Analysis of Data Distributions"*. In: Computer Graphics Forum 39.3 (2020), pp. 565–577.

Contribution clarification. The general idea of the v-plot's design was done by myself in a previous paper [Deb+18]. Based on these initial ideas, we extended the manual design of a v-plot into a chart authoring tool, which recommends basic charts and automatically adjusts v-plots based on their selected analysis tasks. The contribution and research question was a collaborative effort between Mennatallah El-Assady and myself. I designed the structured overview of the analysis tasks and visualization techniques for comparative analysis of data distributions. Furthermore, I conducted the design study, was responsible for the implementation, the design of the recommendation engine, and the v-plot matrix. Luka J. Debbeler contributed concrete examples for the classification of analysis tasks and provided domain-specific examples for the use case. Nadine C. Lages conducted the expert user study and provided an initial draft of the corresponding text. L. Debbeler and N. Lages were also involved in many discussions shaping the overall structure of the paper. M. El-Assady supervised the project and provided textual drafts for different paragraphs during the writing. Britta Renner and Daniel A. Keim provided feedback on the general idea and paper drafts. All writing was done by myself, or I revised paragraphs several times during the writing process. Thus I use the text without citation marks in Chapter 6.

[Blu+18] **Michael Blumenschein**, Michael Behrisch, Stefanie Schmid, Simon Butscher, Deborah R. Wahl, Karoline Villinger, Britta Renner, Harald Reiterer, and Daniel A. Keim. *"SMARTexplore: Simplifying High-Dimensional Data Analysis through a Table-Based Visual Analytics Approach"*. In: IEEE Conference on Visual Analytics Science and Technology. 2018, pp. 36–47.

Contribution clarification. This paper is the result of a close collaboration between Stefanie Schmid and myself (I supervised her BA thesis, and this paper builds on top of her thesis). I came up with the research question and contribution of the paper. Furthermore, I contributed the requirement analysis, the visual design, the user-guided analysis, the automatic pattern detection, and the actual implementation of the technique. I also conducted the user study. M. Behrisch helped to shape the structure of the paper and provided drafts of several sections. Simon Butscher helped with discussions on user interaction. Deborah R. Wahl and Karoline Villinger provided concrete application examples for the evaluation and examples throughout the paper. Furthermore, they were very much involved during the design process of the SMARTexplore technique and provided data, analysis questions, and feedback from a user-centered perspective. Britta Renner, Harald Reiterer, and Daniel A. Keim supervised the project and commented on paper drafts. All writing was done by myself, or I revised paragraphs several times during the writing process. Thus I use the text without citation marks in Chapter 7.

Additionally, I authored and contributed to 19 publications which inspired the needs and contributions of this thesis, but are not included therein. These publications are listed in the following:

[Wah+20] Deborah R. Wahl, Karoline Villinger, **Michael Blumenschein**, Laura M. König, Katrin Ziesemer, Gudrun Sproesser, Harald T. Schupp, and Britta Renner. *"Why We Eat What We Eat: Assessing Dispositional and In-the-Moment Eating Motives by Using Ecological Momentary Assessment"*. In: JMIR mHealth and uHealth 8.1 (2020), pp. e13191.

[Sch+19] Christin Schätzle, Frederik L. Dennig, **Michael Blumenschein**, Daniel A. Keim, and Miriam Butt. *"Visualizing Linguistic Change as Dimension Interactions"*. In: Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change. Florence, Italy: Association for Computational Linguistics, 2019, pp. 272–278.

[Ben+18] Houssem Ben Lahmar, Melanie Herschel, **Michael Blumenschein**, and Daniel A. Keim. *"Provenance-Based Visual Data Exploration with EVLIN"*. In: Proceed- ings of the 21th International Conference on Extending Database Technology, EDBT 2018, Vienna, Austria, March 26-29, 2018. 2018, pp. 686–689.

[Deb+18] Luka J. Debbeler, Martina Gamp, **Michael Blumenschein**, Daniel A. Keim, and Britta Renner. *"Polarized but illusory beliefs about tap and bottled water: A product- and consumer-oriented survey and blind tasting experiment"*. In: Science of The Total Environment 643 (2018), pp. 1400–1410

[Jäc+17] Dominik Jäckle, **Michael Hund**, Michael Behrisch, Daniel A. Keim, and Tobias Schreck. *"Pattern Trails: Visual Analysis of Pattern Transitions in Subspaces"*. In: IEEE Conference on Visual Analytics Science and Technology. 2017, pp. 1–12.

[Sch+17] Christin Schätzle, **Michael Hund**, Frederik L. Dennig, Miriam Butt, and Daniel A. Keim. *"HistoBankVis: Detecting Language Change via Data Visualization"*. In: Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language. Linköping University Electronic Press, 2017, pp. 32–39.

[Mer+17] Leonel Merino, Johannes Fuchs, **Michael Blumenschein**, Craig Anslow, Mohammad Ghafari, Oscar Nierstrasz, Michael Behrisch, and Daniel A. Keim. *"On the Impact of the Medium in the Effectiveness of 3D Software Visualizations"*. In: VISSOFT'17: Proceedings of the 5th IEEE Working Conference on Software Visualization. 2017, pp. 11–21.

[Hun+16b] **Michael Hund**, Ines Färber, Michael Behrisch, Andrada Tatu, Tobias Schreck, Daniel A. Keim, and Thomas Seidl. *"Visual Quality Assessment of Subspace Clusterings"*. In: KDD Workshop on Interactive Data Exploration and Analytics (IDEA'16). 2016.

[Zha+16] Leishi Zhang, Chris Rooney, Lev Nachmanson, William Wong, Bum Chul Kwon, Florian Stoffel, **Michael Hund**, Nadeem Qazi, Uchit Singh, and Daniel A. Keim. *"Spherical Similarity Explorer for Comparative Case Analysis"*. In: IS&T Electronic Imaging Conference on Visualization and Data Analysis. 2016, pp. 1–10.

[Hun+16a] **Michael Hund**, Dominic Böhm, Werner Sturm, Michael Sedlmair, Tobias Schreck, Torsten Ullrich, Daniel A. Keim, Ljiljana Majnaric, and Andreas Holzinger. *"Visual Analytics for Concept Exploration in Subspaces of Patient Groups"*. In: Brain Informatics 3.4 (2016), pp. 233–247.

[Beh+16a] Michael Behrisch, Benjamin Bach, **Michael Hund**, Laura von Rüden, Michael Delz, Jean-Daniel Fekete, and Schreck, Tobias. *"Magnostics: Image-based Search of Interesting Matrix Views for Guided Network Exploration"*. In: IEEE Transactions on Visualization and Computer Graphics 23.1 (2016), pp. 31–40.

[Sch+16] Christoph Schulz, Arlind Nocaj, Mennatallah El-Assady, Steffen Frey, Marcel Hlawatsch, **Michael Hund**, Grzegorz Karch, Rudolf Netzel, Christin Schätzle, Miriam Butt, Daniel A. Keim, Thomas Ertl, Ulrik Brandes, and Daniel Weiskopf. *"Generative Data Models for Validation and Evaluation of Visualization Techniques"*. In: Beyond Time And Errors: Novel Evaluation Methods For Visualization. 2016, pp. 112–124.

[Hun+15b] **Michael Hund**, Werner Sturm, Tobias Schreck, Torsten Ullrich, Daniel A. Keim, Ljiljana Majnaric, Andreas Holzinger. *"Analysis of Patient Groups and Immunization Results Based on Subspace Clustering"*. In: Brain Informatics and Health. Vol. 9250. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 358–368.

[Hun+15a] **Michael Hund**, Michael Behrisch, Ines Färber, Michael Sedlmair, Tobias Schreck, Thomas Seidl, and Daniel A. Keim. *"Subspace Nearest Neighbor Search - Problem Statement, Approaches, and Discussion"*. In: Similarity Search and Applications. Vol. 9371. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 307–313.

[May+14] Thomas Mayer, Bernhard Wälchli, **Michael Hund**, and Christian Rohrdantz. *"From the Extraction of Continuous Features in Parallel Texts to Visual Analytics of Heterogeneous Areal-typological Datasets"*. In:

Language Processing and Grammars. The Role of Functionally oriented Computational Models (2014), pp. 13–38.

[Hao+13] Ming C. Hao, Manish Marwah, Sebastian Mittelstädt, Halldór Janetzko, Daniel A. Keim, Umeshwar Dayal, Cullen Bash, Carlos J. Felix, Chandrakant D. Patel, Meichun Hsu, Yuan Chen, **Michael Hund**. *"Visual Analytics of Cyber Physical Data Streams using Spatio-temporal Radial Pixel Visualization"*. In: In Proceedings of Visualization and Data Analysis (2013), pp. 865404–86541.

[ElA+13] Mennatallah El-Assady, Daniel Hafner, **Michael Hund**, Alexander Jäger, Wolfgang Jentner, Christian Rohrdantz, Fabian Fischer, Svenja Simon, Tobias Schreck, and Daniel A. Keim. *"Visual Analytics for the Prediction of Movie Rating and Box Office Performance"*. In: VAST Challenge 2013 - Award for Effective Analytics. 2013.

[Roh+12] Christian Rohrdantz, **Michael Hund**, Thomas Mayer, Bernhard Wälchli, and Daniel A. Keim. *"The World's Languages Explorer: Visual Analysis of Language Features in Genealogical and Areal Contexts"*. In: Computer Graphics Forum 31.3 (2012), pp. 935–944.

[Krs+12] Milos Krstajic, Christian Rohrdantz, **Michael Hund**, and Andreas Weiler. *"Getting There First: Real-Time Detection of Real-World Incidents on Twitter"*. In: IEEE Workshop on Interactive Visual Text Analytics. 2012.

Part I

Quality Metric-Driven Design for Pattern Analysis

Quality Metrics for High-Dimensional Data

2

Summary

The visualization community has developed to date many intuitions and understandings of how to judge the *quality* of views in visualizing high-dimensional data. The computation of a visualization's quality and usefulness ranges from measuring clutter and overlap, up to the existence and perception of specific (visual) patterns. This chapter attempts to report, categorize and unify the diverse understandings, and aims to establish a common vocabulary that will enable a wide audience to understand their differences and subtleties. For this purpose, we present a commonly applicable quality metric formalization that should detail and relate all constituting parts of a quality metric. We organize our corpus of reviewed research papers along the most commonly used visualization techniques for multi- and high-dimensioanl data. For each technique, we survey the quality metrics, report their findings, reason on the underlying concepts, describe goals and outline the constraints and requirements. One central goal of this chapter is to provide guidance on future research opportunities for the field and motivate the visualization community to compare computed measures to the perception of humans.

This chapter is *based on* the following publication. Please refer to Sections 1.4 and 1.5 for the contribution clarification and general citation rules.

[Beh+18] Michael Behrisch, **Michael Blumenschein**, Nam Wook Kim, Lin Shao, Mennatallah El-Assady, Johannes Fuchs, Daniel Seebacher, Alexandra Diehl, Ulrik Brandes, Hanspeter Pfister, Tobias Schreck, Daniel Weiskopf, and Daniel A. Keim. *"Quality Metrics for Information Visualization"*. In: Computer Graphics Forum 37.3 (2018), pp. 625–662.

2.1 Introduction

The idea of measuring the quality of a visualization is as old as the information visualization community itself. Early work in the field can be traced back to the work of Bertin [Ber81], although the notion and importance of quality were developed far earlier in cartography. Undoubtedly, Tufte was the first research pioneer formalizing the quality metric idea to a simple, thus understandable quality metric: the *data-to-ink ratio* [TG83]; a metric to convey the core principles of an effective and efficient, crisp design.

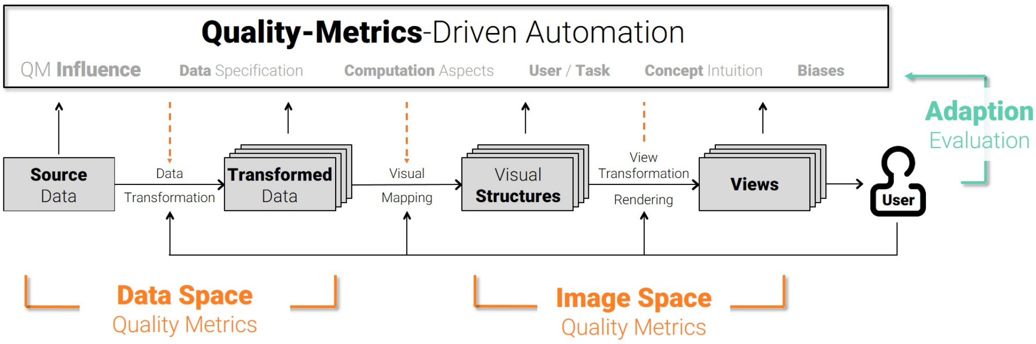


Fig. 2.1. Quality Metrics-driven Visual Analytics pipeline. The pipeline adds an additional layer named Quality Metrics-driven Automation on top of the traditional information visualization pipeline [CMS99]. The layer could obtain information about the several stages of the pipeline (the boxes) and influences the processes of the pipeline through the quality metrics it calculates. The user is always in control. Image and text adapted from [BTK11].

Generally, effective and efficient visualizations follow a simple mantra: They show the most information in the simplest possible form. However, the current data to be visualized puts more and more challenges on visualization designers: high-dimensional spaces, complex relationships, or the sheer amount of data to be visualized demand a careful choice of the visual variables for a *faithful* representation of the underlying dataset.

Following the accepted *information visualization pipeline* of Card et al.[CMS99] –as one possible example– a visualization designer will inevitably be confronted with the dilemma of choosing from a multitude of data processing possibilities and an even greater choice of potential visualization options. To give a practical example: If a user wishes to visualize a 20-dimensional dataset, not only data-specific questions, such as normalization and outlier removal, play a critical role, but also which data characteristic should be highlighted first. In case that a visualization designer decides for a scatter plot, which fixes most of the choices of the visual variables, $n \times (n-1)/2$ potentially meaningful dimension combinations can be depicted. Each of these 190 views needs to be evaluated independently for its usefulness by analyzing its effectiveness concerning other visual encodings, such as color mapping, visual marks, and axis ranges.

In the general case, the number of visual mappings for an arbitrary data type grows exponentially with the number of mapping options, thus making information visualization design to a *trial-and-error process*. More importantly, however, is that only those visualizations can be considered effective that support the building of mental models for the underlying dataset [Nor06]. Hence, the essence of effectiveness resides in the identification of *interpretable visual patterns* that contribute to the overarching analysis goal.

The research field of Quality Metrics (QMs) has devoted its efforts to develop *quantitative measures* for detecting visualizations that contain one or multiple interpretable visual patterns. Applied to exploration and navigation contexts, quality metrics can help to guide the user to views of interest or can help to mitigate the cognitive overload by filtering cluttered or uninteresting views. In general, quality metrics stand as an umbrella term for quantifying the (visual) quality and such the effectiveness and interestingness of a visualization. These approaches find broad applications in the visualization of high-dimensional, relational, or geospatial data. Over the last 30 years, a myriad of approaches, techniques, and concepts have been developed to help the user find a suitable data transformation and visual mapping by iterating and evaluating every possible visualization design combination.

Our motivation for this report is two-fold. First, we recognize that by now the most recent quality metrics surveys date back several years [BTK11; ED07]. In the meantime, the field was undergoing an important development from quality metrics that heuristically quantify the amount of clutter toward a pattern- and analysis task-driven exploration. Therefore, we aim to provide an update by adding more recent publications to the body of work presented in these earlier surveys. Second, we noticed that, although a wide range of approaches was presented under the headline of quality metrics, only little effort has been devoted to describing the methodological and conceptual background of these approaches. Consequently, this work aims to bring depth into the discussion, by consistently enumerating, describing, and relating the underlying concepts with the same vocabulary. As the third motivation point, we claim that most approaches have not yet been evaluated for their perceptual relationships. However, novel and innovative evaluation approaches, such as crowdsourcing and hardware developments (eye trackers in a sub 100$ range) are opening new potentials for this research field.

In summary, the contribution of this chapter is to give a comprehensive overview of existing quality metrics for different multi- and high-dimensional information visualizations techniques, particularly scatter plots (Section 2.6.1) and scatter plot matrices (Section 2.6.2), parallel coordinates (Section 2.6.3), pixel-based techniques (Section 2.6.4), radial visualizations (Section 2.6.5), and glyphs (Section 2.6.6). Our selection is targeted towards visualizations in which QMs are in focus of the research, but we also outline a potential usefulness of QMs for other visualization techniques. As a guiding theme, we not only concentrate on a pure enumeration of techniques but focus more on a detailed description of the underlying *concepts* and *models* and their variety of different implementation possibilities. We also survey how QMs are evaluated and whether results are compared to the human perceptiveness.

2.2 Background and Conceptualization

This section introduces definitions and concepts that we rely upon to describe quality metrics approaches. We discuss common concepts and methodologies across different visualization domains. As one of the core motivations of this survey, we plan to unify the vocabulary and understanding of quality metrics. To achieve this bold goal we gradually increase the level of formalism in the following section. To ease the readability we decided to begin with a purely informal description of our *quality metric vision*. Then we present our attempt to formalize the problem and describe thoroughly constituents and facets influencing the understanding of QMs.

2.2.1 Quality Metric Vision

The grand and sketchy vision behind the visual quality metrics research is the following: Imagine a visual analysis would be based on a black-box that that is fed with your current analysis task(s), user preferences, and the dataset at hand. This black-box would "auto-magically" derive a recommendation of the best possible visualization type and visualization instantiation; would derive the most effective visual variable settings (e.g., color map, shape, texture) and all necessary data preprocessing steps depending on multitude of soft and hard influencing factors; and would finally present the most interesting view on the data that reveals most information.

However, while this vision sounds overarchingly promising, parts of the questions can already be tackled with current technologies: More or less sophisticated "Show Me" buttons (e.g., [MHS07]) decide for the user which visualization is appropriate based on data types. Other approaches even add considerations about the underlying data distribution into their recommendations of a visualization type and visual mapping [Won+16; Won+17].

Other approaches start from the constraint that the visualization type is fixed, e.g., scatter plots for projections of high-dimensional data and tackle the question which views can be discarded due to the high overlap or *visual clutter* [BS04; Tat+10]. Again other approaches, such as the so-called *-gnostics [WAG05; SSK06; DK10; Leh+15; Beh+16a], focus on the quantification of visual patterns for their specific visualization type, following the core idea of promoting only views containing interpretable visual patterns and thus helping build mental models about the dataset and task relationships.

But, while we are seeing more and more advanced research for supporting the user in the exploration process, the current research is struggling with the definition, categorization, and labeling of the current exploration task in place. Partially this problem arises because exploration tasks are not necessarily separable in terms of their temporal characteristics and oftentimes even nested in nature. To make matters worse, most users do not follow a structured exploration path but conduct several exploration tasks in parallel with a more or less prominent specificity. While basic research has been presented in this field, such as various task taxonomies with different levels-of-details [BM13; Kei+08; Lee+06; Shn96], only a few works focused on automatically quantifying the current exploration task at hand. Sacha et al. [Sac+17] list a range of works following this research stream. Additionally, quality measures should approximate the users' perception and cognition. Yet, only a few approaches have been evaluated with user studies and only a few evaluations compare the usefulness of multiple different metrics.

The aforementioned consideration sketches outline a far-reaching and extensive research field with multi-faceted foci and research potential for at least the next decade. Consequently, we will not be able to report on all developments. Rather, we decided to put emphasis on what we denote as *Mid-level Perceptual Quality Metrics*. This emerging field focuses on perceptually-inspired quality metrics that try to mimic parts of the human perception/cognition in order to ease the exploration process.

These approaches not only reduce the cognitive overload by separating the "wheat from the chaff", i.e., by removing noise, but also facilitate building task-related mental models by mimicking the humans' ability to recognize and differentiate between visual patterns.

2.2.2 Definitions

We use the following definitions. Formally, measuring the quality of a visualization V consists of computing one visualization definition $\phi \in \Phi$ from a universe Φ of potential instantiations that maximizes or minimizes a specified *quality criterion* $q(D, U, T)$, such that:

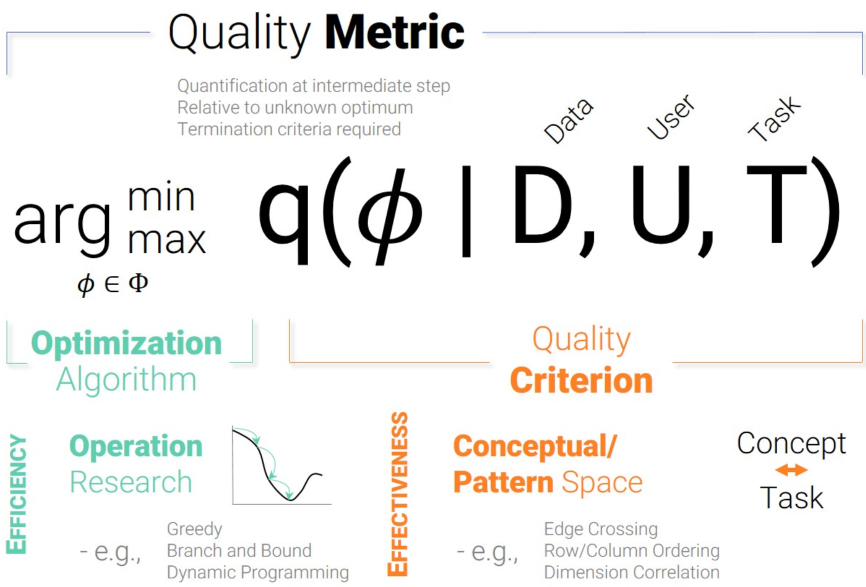


Fig. 2.2. Quality Metrics (QM) formalization. QMs are composed of an algorithmic part and a quality criterion. A potential multi-objective optimization algorithm tries to find efficiently a valid visualization configuration (ϕ) that optimizes the designed quality criterion $q(\phi \mid D, U, T)$. The quality criterion tries to heuristically capture how an effective visualization instance might look like. This intuition is bound and influenced by the task T at hand (defines the to-be-expected visual appearance), the dataset characteristics D (defines if a visual pattern is producible), and the user preferences U. Consequently, a QM *arg min/max* $q(\phi \mid D, U, T)$ determines a perceptually preferable visualization configuration ϕ for a given quality criterion *q(...)* given the influencing factors D, U, T.

To illustrate our formalism let us imagine the following scenario: We describe our user U as a statistically knowledgeable person with average attention potential whose task T is to understand data/dimension (dis-)similarities in a high-dimensional dataset D. Our Quality Metric-driven recommendation system could decide that a scatter plot display is a suitable choice to show (dis-)similarities for this kind of user.

The quality criterion $q(\dots)$ could then compute the sum of pairwise distances over all displayed points in D with respect to a chosen distance function while taking the data specifics into consideration (i.e., needs outlier cleaning). The Equation in Figure 2.2 would find for a specific task $t_1 \in T$, a $\phi_1 \in \Phi$ that minimizes this sum relating to a locally dense scatter plot or could find $\phi_2 \in \Phi$ that maximizes the sum to find globally cluttered plots for another task $t_2 \in T$.

Mid-Level Perceptual Quality Metrics

The area of Mid-level Perceptual Quality Metrics leaves out all considerations about the user U; assessing his/her skill set or cognitive/physiological capabilities and does not (yet) deal with an explicit formulation of tasks T during the exploration process. The field of Mid-level Perceptual Quality Metrics is rather concerned with presenting heuristics and algorithms to statistically quantify the extent of an anti-pattern –e.g., how a cluttered view looks like– or which specific visual pattern is apparent –e.g., locally dense scatter plots can be used to reason about data similarity.

In the following, we will outline the components contributing the definition of a Mid-level Perceptual Quality Metric.

- A **Quality Metric (QM)** combines an optimization algorithm and quality criterion with the overarching goal to mimic parts of the human perception. QMs are developed with a specific goal in mind, such as finding clutter-free visualizations or visualizations with a specific interpretable visual pattern.
- **Visualization Definition** ϕ is an instantiation of the parameter space Φ defining the appearance of a specific visualization type. Following the information visualization of Card et al. [CMS99], as depicted in Figure 2.1, we will have to distinguish between *data*-dependent and *visualization*-dependent parameters. For a scatter plot, ϕ would define the necessary data transformations, such as which outliers will distract the view "too much" and the view-space parameters describing the visual appearance of data item (e.g., shape, color, texture, position) and the corresponding axis definition and appearance (e.g., offset, normalization type, aspect ratio).
- **Quality Criterion** $q(\dots)$ is an (heuristic) algorithm or function for quantifying the effectiveness of one visualization instantiation/view. In other words, a quality criterion evaluates heuristically whether or not a view follows established perceptual guidelines. In the most cases, the goal is to quantify the visual appearance of (anti-)patterns. We consider visual patterns as the target elements of the exploration process, while visual anti-patterns, such as noise, will distract the user without adding to his/her understanding about the dataset and task at hand.
- **Optimization Algorithm** makes use of a quality criterion and -concept to derive, e.g., a ranked or filtered list of visualization instantiations (or views). To achieve this goal an optimizer takes a quality criterion and improves the measure over the visualization method parameters ϕ. Most prominently, filtering

concepts are applied to discard cluttered views, while pattern-exploration systems categorize views in terms of the visual patterns they contain.

Note that *metric* has a precise meaning in mathematics, but is used more loosely in the present context. The characteristics of a metric, i.e., non-negativity, identity of indiscernibles, symmetry, and the triangle inequality, need not necessarily hold in all cases. As an example, many QM approaches are based on non-deterministic computations to retrieve (good) local optima in the visualization parameter space. Hence, the term quality metric should be rather understood as an artifact that developed over time from a mathematical understanding toward a more vague and indistinguishable field of more or less mathematically backed up research approaches.

2.2.3 Common Calculation Approaches

In our literature review, we identified three different concepts to compute quality metrics: a primarily *image space* dependent computation, a purely *data space* dependent computation, and *hybrid approaches* that efficiently combine both concepts. Moreover, we found that QMs are either used implicitly during the construction of visualizations or as a separate evaluation component complementing the construction and use of visualizations.

Image Space QMs assess the quality of a visualization solely based on the rendered image. Often, sophisticated feature descriptors are extracted from the image and used to measure clutter or perceivable patterns. For example, Tatu et al. [Tat+09] encode the visual quality of parallel coordinates by means of a Hough Space feature descriptor. With this approach, it is possible to distinguish visually noisy and strongly clustered axis combinations. In a quality metric driven analysis, we aim to mimic the perception of a human to identify patterns. The main advantage of an image-based quality assessment is therefore that we use the same visual information (i.e., image) that is also assessed by humans in an evaluation setting.

Data Space QMs measure the quality of a visualization before the rendering process starts. The approaches are based either on raw or transformed input data, or estimate how visual structures will most probably look like. As an example, Johansson and Johansson [JJ09] propose an interactive approach to weight multiple data spaces based quality metrics to reorder axes of parallel coordinates. Their metrics comprise a user-defined weighting of correlation dimensions (by a Pearson correlation coefficient), outlier analysis (by a grid and density based approach), and cluster detection (by applying a subspace clustering algorithm). The main advantage of data-based QMs is that many measures (such as cluster algorithms) exist and can be computed usually quite efficiently.

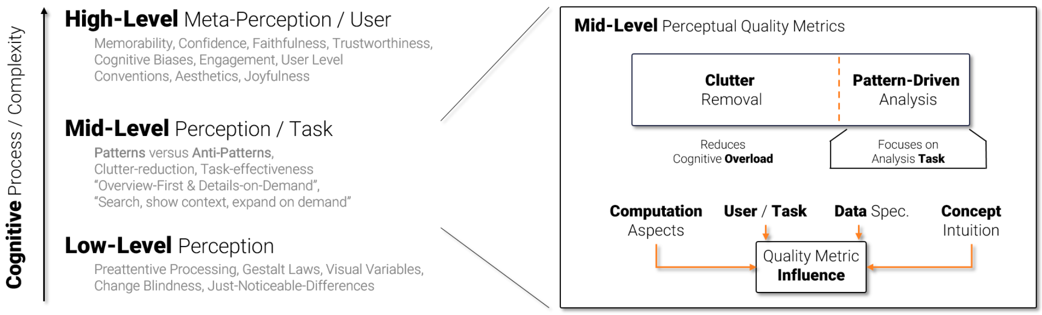


Fig. 2.3. A wide range of quality metric understandings exist in the literature. The left side shows a broad categorization of the field sorted according to the cognitive complexity these approaches try to reflect. The right side shows that the task focus of *Mid-level Perceptual Quality Metrics* comprises different granularity levels: (1) Overview: distinguish between noise/clutter and any kind of pattern, (2) quantify the quality of a visualization based on a specific pattern (depends on the task).

Hybrid QMs combine the advantages of image and data space approaches. For example, Bertini and Santucci [BS04] determine a good sampling rate in scatter plots by comparing the visible data density in image space with the relative data density in data space. The number of visible points at one specific location in the visualization is either 0 or 1 in the image space, while the data space can also count more than one points at one location. Combining these measures support most useful sampling strategies.

Implicit vs. Explicit Quality Metrics. Many approaches make use of implicit quality criteria *as part of* an optimization problem. Typically, these approaches do not explicitly externalize numeric scores for the quality of a visualization, but decide during the view construction which representations is more useful.

A practical example in the field of dimension reduction is presented by Wang et al. in [Wan+18a]. For labeled datasets, typically depicted by color-coded scatter plots, algorithms start with a (pseudo-)random placement of items in 2D. This placement is incrementally improved with respect to one or multiple visual class separation QMs by choosing the one perturbation of the current solution that improves the QMs. Integrated into a simulated annealing optimization, this approach helps to traverse the exploration space and find a locally optimal solution for the chosen class separation QMs.

An explicit quality criterion for parallel coordinates would quantify *to which extent* specific visual patterns (e.g., clusters) are present in the arrangement of axes, described by its applied reordering algorithm. But, explicit QMs can also be used to choose *between* various visualization types and configurations. For example, in "Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series" [Wan+18b] the authors quantify the visual consistency between the data set's trend curve and the trend described by a scatter plot or a line graph. Based on the numeric comparison of both QM scores, the better visual approximation is chosen.

2.2.4 Analysis Scenarios Supported by Quality Metrics

We can distinguish between QMs designed for *clutter reduction* and *pattern-driven analysis*, as depicted in Figure 2.3. Clutter reduction techniques reveal the contained set of visual structures by "only" filtering out noisy views. Therefore, they are most useful to obtain an overview of large and unknown datasets, as they keep all views with potentially interesting visual patterns. Hence, these QMs mitigate the cognitive overload problem. However, users typically have specific exploration or analysis foci in mind to understand the data structure and topology. Searching for visual patterns with particular properties is significantly more challenging and requires a quantification and distinguishing of visual structures. But, perceptually-inspired QMs have the benefit to support the user *directly* by contributing to their mental model and understanding of the data.

Overview of Analysis Tasks

Quality metrics identifying a particular pattern are typically related to one or more analysis tasks. We refer to these metrics as *task-specific quality metrics*. For all QM that we report in this chapter, we try to elaborate on the (potentially) underlying task(s). We do not stick to any of the established task taxonomies, since they are too specific compared to the analysis tasks supported by QMs. In contrast, we present a high-level overview of exploration tasks supported by the majority of metrics:

Clutter reduction. Users are interested in filtering out noisy views without a specific visual pattern in mind. This task is a typical used to get an overview of unknown datasets.

Preservation task. QMs for preservation tasks identify views that preserve the original data properties in the mapping process. The preserved aspects can be, for example, individual data points or topological structures.

Search for data groups and partitions (clusters). QMs aim to identify views in which a (useful) partition and/or dense groups of data records are visible.

Search for outliers. The goal is to identify views that highlight data points differing from the majority of other points.

Search for dimension relations. This task depict combinations of dimensions showing relationships between the data points (e.g., correlations).

Data and visualization specific tasks. For one data type, different visualization techniques exist; each with (dis-) advantages to reveal essential aspects. Some analysis tasks are specific to data or visualization types (e.g., readability of typographic visualizations) and cannot be generalized.

One example of a task-specific QM is shown in Figure 2.4. Imagine an analysis task in which users need to find data groupings (clusters) in scatter plots: While the first scatter plot contains only noise, the last plot reveals several clusters, detected by a quality metric. Although the second plot also shows an interpretable pattern

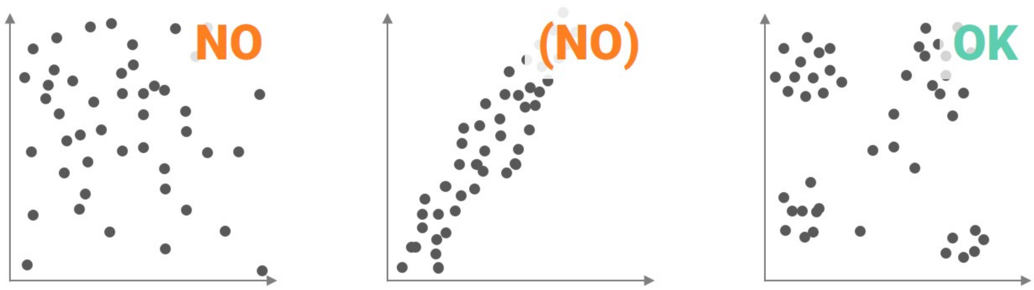


Fig. 2.4. Example of a *task-specific quality metric* for scatter plots. Task: finding data groupings or clusters.

(correlation of the data), it is not relevant to the current task. A task-specific quality metric needs to classify the plot as non-interesting due to the non-relevant visible pattern. One existing quality metric that can distinguish between a variety of patterns in scatter plots is Scagnostics [WAG05]. It captures the presence of the following nine visual features: outlying, skewed, clumpy, convex, skinny, striated, stringy, straight, and monotonic. In the example we would search for scatter plots with low monotonic and high clumpy features.

In second example, an analyst wants to measure how much information is preserved by projecting a high-dimensional dataset with class labels into a 2D representation. The analyst decides to use the RadViz technique and represent color with the class information. As shown by Figure 2.5, a task-dependent quality metric can help to optimize the ordering of dimensions such that the provided classes are well separated. A quality metric that facilitates this concept is presented by Albuquerque et al. [Alb+10]. Their approach is to measure the density of all classes in every 2D representation.

During our literature review, we recognized that a some tasks are well-supported by QMs, while others are not. We discuss well-adopted tasks in their respective visualization section and point to open research gaps for visualization technique.

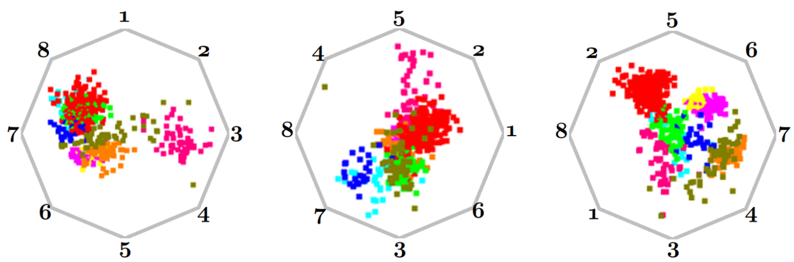


Fig. 2.5. Different dimension orderings in RadViz preserve and emphasize, respectively mask, given groupings in high-dimensional data. Groupings become increasingly visible from left to right. Figure adapted from [Alb+10].

2.3 Related Concepts

As mentioned earlier, this work surveys the recent advances and state-of-the-art for mid-level perceptual QMs. However, this subfield is embedded into an overall quality metric landscape, depicted in Figure 2.3. For the sake of completeness and delineation, we will enumerate the main concepts and relationships in this section.

The topic of quality metrics is not described in technical terms, but rather incorporates a wide range of understandings. Since one of the core contributions of this chapter is to establish a common vocabulary, we are categorizing QM related concepts along the axis of cognitive complexity.

2.3.1 Low-Level Perceptual Quality Metrics

Low-level perceptual quality metrics leverage the low-level processing of visual stimuli in human perception system such as preattentive processing [War20; HE12]. They are concerned with how basic visual encoding variables, such as position, length, area, shape, and color, and the interaction of the variables (e.g., integrable or separable) influence the efficiency of low-level perceptual tasks such as visual search, change detection, and magnitude estimation.

A great deal of prior visualization research has been devoted to understanding the effectiveness of different visual variables for encoding quantitative and qualitative data. For example, Cleveland & McGill [CM84] ran a series of graphical perception experiments to measure accuracy in comparing values and to derive the rankings of encoding variables. Similar experimental methods have been frequently employed to compare different chart types as well. The results of such experiments have also played a vital role in the automatic construction of visualizations [Mac86; MHS07].

2.3.2 High-Level Perceptual Quality Metrics

High-level perceptual quality metrics refer to cognitive measures such as memorability, aesthetics, and engagement [SES16]. While they are often considered as subjective dimensions of visualization design, recent studies attempt to quantify these measures based on experiments with human subjects. For example, Borkin et al. [Bor+16] showed that visualization memorability is consistent across people, suggesting that some visualizations are more memorable than others independent of subjects' context and biases.

Various factors can contribute to high-level perceptual quality metrics such as visual density and human recognizable objects for memorability [Bor+16], colorfulness and visual complexity for aesthetics [HRC15], and amount of interactions for engagement [SES16]. While mid-level and low-level perceptual quality metrics tend to focus on optimizing performance measures for data exploration and analysis tasks, high-level perceptual quality metrics put more emphasis on enhancing the communication aspect of visualization (e.g. whether a visualization can attract the attention of an audience and get the message across).

2.3.3 Design Recommendations

We consistently recognize two "end products" for quality metric design in all quality metric subfields:

Design Recommendations. In some visualization subfields, QM results are communicated via *Design Recommendations*; textual guidelines and arguments summarizing the findings about visualization design mostly derived from user-studies. Design recommendations have the great advantage that they represent reproducible evaluations of how a human perceives a view. They can summarize complex perceptual circumstances. Their biggest disadvantage is that these textual guidelines are often derived from simplified task- and context settings that often cannot be generalized to real-world environments and problem settings.

Heuristic Approaches. In other visualization subfields, purely *Heuristic Approaches* prevail. These algorithms model some form of understanding of how a visualization should look like in order to be effective/useful. The biggest advantage of heuristics is their reproducibility, thus allowing user- and context-independent, quantitative visualization comparisons. Their biggest negative point is that visualizations are often judged for their perceptual quality with quantitative scores that have never been proven to correspond to the humans' judgment.

We claim that both approaches are valid but should eventually be backed up with the other approach. Heuristics should be evaluated for their perceptual aspects and proven to be perceptual; design recommendations should be developed into quantifiable heuristics to allow for fair and quantitative evaluation schemes.

2.4 Related Work

Quality Metrics have been developed for different information visualization techniques. From a historical perspective, we are inspired by a range of survey works with a more or less specific notion of quality metrics. For example, Brath [Bra97] described several image space quality metrics, such as occlusion percentage or percentage of identifiable points, to assess the quality of business visualizations. Miller et al. [Mil+97] expressed the need for new metrics to compare visualizations. Similarly, Diaz et al. [DPS02] advocated the use of implicit and explicit quality metrics for assessing the quality of vertex ordering approaches. In this context, the term of aesthetics is used as same as it is traditionally used in the graph drawing community and refers to a set of measures to reduce the cognitive load for graph exploration tasks [Di +94; War+02].

A first survey focusing primarily on quality metrics for scatter plots and parallel coordinates was presented by Bertini et al. [BTK11]. Similar to our approach, their survey presents a systematic analysis focusing on the guiding questions: (1) What was measured? (2) Where was it measured (data/image space)? (3) What is the purpose of the QM? And, (4) does the QM allow to be interactively adapted? In total, 20 papers are surveyed in this work.

The evaluation of quality metrics has gained increasing importance in the recent years. For example, Lehmann et al.[LHT15] and Pandey et al. [Pan+16] study independently the questions about the connection of human perception and (heuristic) quality metrics and present both crowdsourcing studies to prove evidence that this connection exists. Sedlmair and Aupetit [SA15] even present a data-driven framework for quality measure evaluation. Their approach tries to mitigate the impact of (relative) human judgments by relying entirely on ground-truth data. However, this in turn also indirectly implies some sort of user involvement.

An information theoretic approach for assessing the effectiveness of information visualization has been mainly pursued by Chen et al [CJ10]. They built on the initial work by Yang-Peláez et al [YF00] and proposed a number of entropy-based measures, including visual-mapping ratio, information loss ratio, and display space utilization; these measures are akin to the data-ink ratio [TG83]. Chen et al also discussed visual multiplexing [Che+14] in relation to the information theoretic measures. They describe various mechanisms for overlaying multivariate data and discuss how to overcome perceptual difficulties such as occlusion and cluttering that arise from the interference among spatially overlapping visual channels. We consider that these measures concern low-level quality metrics and thus are not discussed in this chapter.

Saliency-based measures for evaluating the visualization quality have gained recent interest. They assess how well visually salient regions in a visualization can help users accomplish their goals and tasks. For instance, Chen and Jänicke [JC10] proposed a method for computing a saliency-based metric to measure the mismatches between visual salience and data characteristics (e.g., features detected by algorithms). Matzel at el [Mat+18] recently developed a saliency model to predict where people would look for a given visualization. Unlike models designed for images of natural scenes, their model attempts to incorporate top-down visual features (e.g., texts) that are crucial for visualization tasks. Tailoring the models for different visual analysis tasks is largely unexplored, however [Pol+18]. We believe that saliency-based measures touch on both low-level and high-level quality dimensions and thus not addressed in this chapter.

Although the field of quality metrics for color mapping can be safely categorized into low-level perceptual quality metrics research and is thus not in the focus of this survey, we decided to stress some shared argumentation paths by selectively summarizing some more recent works. Quality metrics for color mapping have been investigated amongst others in the work of Bernard et al. [Ber+15], Mittelstädt et al. [Mit+15; MK15], or recently by Gramazio [GLS17]. Szafir and Gleicher [SG16] argue for choosing colors based on a given context rather than in isolation. They identified three categories of design constraints and make design recommendations for effective color choices based on aesthetic constraints, perceptual constraints, and functional constraints. Eisemann et al. [EAM11] present an orthogonal approach. Based on a range of data analysis and transformation steps, a user-independent, data-driven color mapping approach is postulated.

While many approaches are targeted toward clutter removal [ED07], only very few are targeted toward describing the perceived appearance with respect to *visual patterns*. Our survey aims at describing quality metric approaches in a unified manner to better understand their differences and subtleties.

2.5 Methodology and Structure

We gathered an initial set of papers from an informal user study with domain experts (doctoral researchers and postdoctoral researchers with between 2–7 years of experience in respective visualization subdomains). Our paper selection was used to condense a set of high-level questions and evaluation criteria that guided in the following the expansion of the reference list by searching through the relevant visualization venues. Consequently, our survey should be seen as an educated selection of the concepts of quality metrics and does not claim comprehensiveness.

For each visualization technique we base our analysis and the organization of each content section on a structured questionnaire, which incorporates the following aspects:

- **Visualization Description** outlines the basic concept of a specific visualization type, its primary purpose, its inherent constraints and requirements.
- **Why do we need QMs?** motivates the use of QMs in this context, describing the perceptual/analytical benefits, sketches (computational) challenges, and refers back to the visualization definition part influenced most by the QMs.
- **Typical Analysis Tasks** outlines analysis scenarios for the respective visualization and mentions how QMs can improve efficiency and effectiveness.
- **Summary of Approaches** presents an overview of the influential QM work in the literature.
- **Evaluations Methods** shines a light on the evaluation approaches for QM-enhanced visualizations.
- **Open Research Questions** summarize the future challenges with respect to the visualization design and states how QMs could be applied to overcome these problems.

In order to come up with a structured and valid abstraction of the field, we reported all of our findings in a table format (encoding phase), which can also be found online at `http://visualquality.dbvis.de/summary`. We iterated on the table results for consistency and developed iteratively a more and more refined view of the landscape. The core findings of these iterations are reflected and abstracted in the background sections, while specifics are highlighted in the respective subsections.

In total, we collected for this survey 64 papers from the various information visualization subfields. While our coverage is not exhaustive and biased toward impactful publications illustrating the fundamental concepts of this field, our goal is to provide a central document where concepts for multiple visualization types are defined and related, algorithms grouped into broader categories and discussed in contrast to each other, and, finally, we give an overview, of how quality metrics are systematically evaluated.

2.6 Survey of Quality Metrics for Multi- and High-dimensional Data

Multi- and high-dimensional data is typically provided in a table-like format in which rows correspond to data records/objects, and columns to their dimensions, attributes, features, or descriptors. For example, consider a collection of cars (data objects) that are described by, e.g., their color, brand, and horsepower (dimensions). Often, these datasets comprise *combinations* of numerical, categorical, and complex types such as geo-locations, images, and texts. In the following, we restrict ourselves to quality measures for (combinations of) numerical and categorical dimensions.

Visualizations for multi- and high-dimensional data face two major challenges that also influence the computation of quality metrics: (1) datasets with a mix of numerical and categorical dimensions make it difficult to compute relations between objects (e.g., similarity) which is one of the fundamental concepts in many metrics. (2) The outstanding characteristic of datasets with a large number of dimensions is the *curse of dimensionality* [Bel61]. A huge number of dimensions increase the possible visual mappings and the arrangement of dimensions. Non-relevant, redundant, and conflicting dimensions may hide interesting patterns in a sea of noise. And, the number of dimensions highly influence the interpretability of similarity measures [Bey+99; HAK00].

In the remainder of this Section, we will use synonymously the term high-dimensional for multi-dimensional, and multivariate data. We will describe and categorize quality metrics for scatter plots (Section 2.6.1) and scatter plot matrices (Section 2.6.2), parallel coordinates (Section 2.6.3), pixel-based techniques (Section 2.6.4), radial visualizations (Section 2.6.5), and glyphs (Section 2.6.6). For each of the techniques, we describe the challenges and necessity of quality metrics, what they intend to measure, and outline the analysis tasks for the respective visualization. Afterward, we summarize the approaches and show their typical evaluation procedure, and outline open research questions.

2.6.1 Scatter Plots

A popular and intuitive way to visualize high-dimensional data is to use scatter plots or scatter plot matrices. A scatter plot represents each data point in a two-dimensional Cartesian coordinate view. The first dimension is typically denoted as x and the second dimension as y. Scatter plots support the analysis of single data instances, as well as data patterns and entire distributions. Figure 2.4 shows three different examples of distributions that are visualized by scatter plots. It is easy for analysts to see if the two variables correlate, contain groups or clusters, or more sophisticated patterns. For datasets with more than two dimensions, all pairwise combinations of dimensions are visualized using a scatter plot matrix (see: Section 2.6.2).

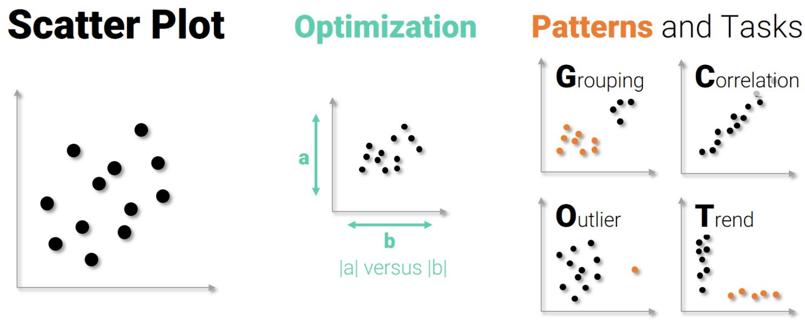


Fig. 2.6. Scatter plots – optimization goals, tasks & visual patterns

Why Do We Need Quality Metrics for Scatter Plots?

Using scatter plots to visualize large and high-dimensional datasets face several challenges. For example, dimensions may have different units that need to be scaled. However, an appropriate scaling of variables needs to be chosen carefully for reliable representations. Another well-known problem of scatter plots is visual clutter due to a high number of data instances that are plotted on top of each other. Such visual clutter may hide patterns in the data and make it hard for analysts to identify relationships among variables. Quality metrics are a means to reduce the number of displayed scatter plots to only the combinations of variables with interesting patterns. Recent quality metrics apply, for example, sampling, filtering, clustering, and distortion techniques, and then measure the remaining patterns in the data. Ultimately, quality metrics and taxonomy help determine the best reduction technique and settings for given data characteristics or analysis tasks.

Typical Analysis Tasks for Scatter Plots

A scatter plot is used to investigate the relation between two different variables. It is useful to get a quick overview and helps indicate problems, unique properties, or anything interesting about the data. Interesting insights are, for instance, correlating variables, outliers, or meaningful patterns (e.g., regression models, trends, well-separated clusters). Sarikaya and Gleicher [SG18] presented a taxonomy of twelve low-level analysis tasks that support the analysis in scatter plot views. The defined analysis tasks are: identify objects, locate objects, verify objects, search for known motifs, browse data, identify outliers, characterize distribution, identify a correlation, explore neighborhood, numerosity comparison, object comparison, and understand distances.

In more advanced analysis scenarios, dimension reduction techniques are often used to map high-dimensional features into 2D projection views [Wan+18a]. For instance, principal component analysis is a projection technique that uses traditional scatter plots to map high-dimensional data into a lower-dimensional space [WEG87].

Summary of Approaches

Our ability to perceive patterns and trends in Scatter Plots is highly influenced by the aspect ratio. Cleveland [Cle93] invented the principle called *banking to 45°*, which uses a midangle of 45°to enhance slope judgment for bivariate graphs. Applied to scatter plots, an improved aspect ratio selection, such as with the banking to 45°quality criterion, can be applied to emphasize trends in the dataset [Cle93]. This relationship between task (trend detection) and quality criterion (aspect ratio) was also examined and validated by Fink et al. in [Fin+13]. In addition, Fink et al. [Fin+13] uses Delaunay triangulation to generate scatter plot projections and measure the quality by calculating a small total edge length or large minimum angle of the triangles.

By assuming that the aspect ratio is chosen well, there still remains the question if the visual representation is appropriate for the data or not. A taxonomy of different visual factors to separate clusters in scatter plot well was given by Sedlmair et al. [Sed+12]. The presented taxonomy is based on classified data and considers *within-class* and *between-class* factors to guide design and evaluations of cluster separation measures. Furthermore, clutter must be considered to present point distributions clearly. An overview of different clutter reduction techniques including benefits and losses for scatter plot visualizations is given by Ellis and Dix [ED07]. Regarding quality metrics for clutter reduction, Bertini and Santucci [BS04; BS05] proposed a feature preservation approach to improving visual perception of 2D Scatter Plots. Their metric includes an automatic sampling strategy based on a perceptual user study to find an appropriate sampling ratio.

Evaluation Methods for Scatter Plot Quality Metrics

To assess improvements in visual perception, user studies were often conducted [LHT15; Pan+16]. Micallef et al. [Mic+17] implemented several models and metrics of human perception in a cost function to improve the visual low-level perception of a scatter plot. Based on input data and task, an optimizer automatically enhances design parameters such as marker size and opacity, aspect ratio. Various scatter plot design choices were investigated in [SG18] based on data characteristics and analysis tasks.

Open Research Questions

The traditional scatter plot is a well-known visualization technique and has been further developed over the last decades. Today, various kinds of visual optimizations and data preprocessing techniques exist to improve the final representation. For instance, there are density based modifications or combined representations to increase the information content on the visualization. However, what is missing in the literature are state-of-the-art reports about scatter plot related techniques and optimizations.

2.6.2 Scatter Plot Matrices

A scatter plot is only suited to visualize the relationship between two variables. To analyze the entire dataset with multiple dimensions, a scatter plot matrix (SPLOM) can be used. It shows all pair-wise combinations of variables in a matrix. For n variables, a SPLOM visualized n^2 cells. Each cell reflects a particular combination of two data dimensions. For example, analysts can inspect changes of an independent variable according to all dependent variables by scanning rows or columns of the matrix. Figure 2.7 shows an example SPLOM, together with different patterns.

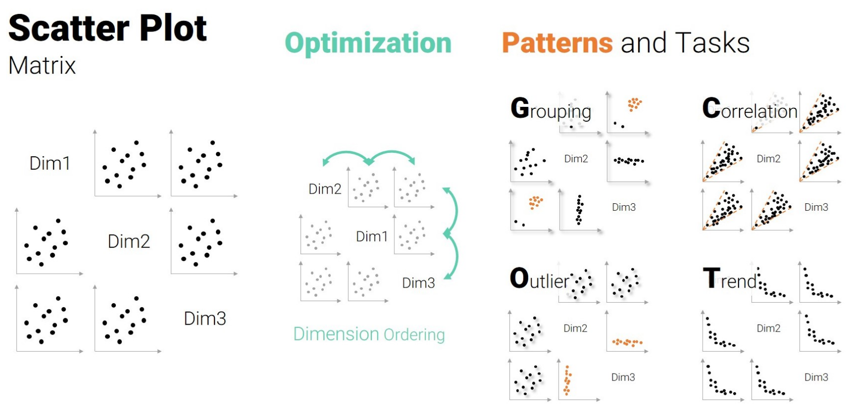


Fig. 2.7. Scatter plot matrix – optimization goals, tasks & visual patterns

Why Do We Need Quality Metrics for Scatter Plot Matrices?

Identifying interesting patterns in large scatter plot matrices is challenging. The number of views grows quadratically with the number of given variables. Furthermore, the goal of an exploratory data analysis is often based on a given analysis task. Therefore, not all scatter plot views are relevant. While a manual inspection of all combinations is often not feasible, quality metrics can help to show only the most interesting combinations. Popular quality metrics measure the quality of a visual pattern and then help to filter or sort SPLOMs accordingly.

Typical Analysis Tasks for Scatter Plot Matrices

A SPLOM is often used to get an overview of all bivariate correlations (via scatter plots) in a higher dimensional data space. This is particularly helpful to identify specific variables that might have similar patterns across various dimensions, e.g., correlation, classification, clusters, or trends.

Due to the orthogonal pairwise projections of dimensions in a SPLOM, a horizontal or vertical exploration enables the investigation of data transformations by exchanging one dimension. For example, a column-wise exploration allows the user to discover transformations by exchanging the independent variable and a row-wise exploration by exchanging the dependent variable. Shao et al. [Sha+16] used color coding in combination with a motif-based dictionary to highlight column-wise and row-wise coherence of segmented patterns in s SPLOM. This work also encourages to take the investigation of local patterns into the analysis process and focus on interest measures derived from local motifs in the data.

Furthermore, SPLOM-like representations are suitable for subspace analysis tasks, such as finding clusters or interesting subspaces. Yuan et al. [Yua+13] used a dimension projection matrix in which rows and columns represent multiple dimensions and the scatter plots are based on dimension projection.

Basically, all low-level perception task for single scatter plots (mentioned in Section 2.6.1) can be applied to a larger projection space. For the analysis in SPLOMs, these tasks are usually extended to a comparison task among multiple scatter plots (mid-level perception task). Sarikaya and Gleicher [SG18] derived twelve basic analysis tasks that are supported in scatter Plot and SPLOMs respectively (c.f., Typical Analysis Tasks for scatter plots in Section 2.6.1).

Summary of Approaches

In data analysis, methods for mapping multivariate data into lower dimensional space have been used for many decades[KW78; WEG87]. However, one of the major problems of these mappings is that the resulting outcome is often difficult to interpret. One influential approach by Friedman and Tukey [FT74] that tackles this issue is called Projection Pursuit. Projection Pursuit is a linear mapping algorithm that uses interpoint distances and the variance of point swarm to pursue optimum projections. Later, Tukey and Tukey [TT85] invented an exploratory visualization method for SPLOMs (Scagnostics). Wilkinson et al. [WAG05] followed up on their research and introduced graph-theoretic measures for computing scagnostic for large datasets. The method is based on proximity graphs and extracts nine characteristics that describe the point distributions of the scatter plot space. It has been shown that Scagnostics can serve for many applications and help to detect anomalies in time series, find specific patterns, or sort SPLOMs [WW08; NAW13; DW14; NW14].

Another common approach to index the interestingness of scatter plots is to consider the class consistency information of labeled points. For instance, Sips et al. [Sip+09] propose two quantitative measure of class consistency, one based on the distance to the class's center of gravity (distance consistency), and another based on the entropies of the spatial distributions of classes (*distribution consistency*). Tatu et al. [Tat+09] used similar ranking measures based on the image space of the scatter plot visualization to identify potentially relevant structures. For unclassified data, they used a *rotating variance measure* (RVM) to find linear and non-linear correlations in the data. For classified data, they measure the overlap between the different classes and rank scatter plots that show well-separated classes the best (*class density measure*). Later, Tatu et al. [Tat+11] extended their class density measure and introduced *class separating measure* to control the balance between the property of separation and dense clustering. A recent work of Matute et al. [MTL18] showed that a skeleton-based metric including shape and orientation information outperforms *RVM* and *Scagnostics* in perceptually-based similarity.

Albuquerque et al. [Alb+09] utilized the aforementioned quality measures for a quality-aware sorting of SPLOMs, the so-called class-based scatter plot matrix (C-SPLOM). Lehmann et al. [Leh+12] introduced another visualization scheme including detail-on-demand interactions that produces an abstract and interpretable SPLOM (A-SPLOM) by using known quality measures.

A current approach by Shao et al. [Sha+16] measures the interestingness by taking frequency properties of similar local patterns into account. The approach applies a Bag-of-Visual-Words concept that considers local motifs as visual word and ranks the interestingness based on the number of interesting motifs in a plot. Moreover, an extensive survey of quality metrics for scatter plot and other visualization techniques were carried out by Bertini et al. [BTK11]

Evaluation Methods for Scatter Plot Matrix Quality Metrics

The evaluation of quality metrics has gained increasing importance in the recent years. These works either focus on evaluating the connection between human perception and quality metrics (effectiveness) or demonstrating the usefulness of quality metrics base on various use case scenarios (efficiency).

For example, Projection Pursuit is demonstrated by various experiments on artificial and research data. Wilkinson et al. [WAG05; WW08; NAW13; DW14; NW14] evaluated the performance and usefulness of their Scagnostics tools by showing use cases and experimental results on different datasets. Actually, most approaches are evaluated by use cases and demonstrate the benefits by various scenarios [Tat+09; Alb+09; Tat+11; Sha+16]. For instance, the class consistency measures by Sips et al. [Sip+09] were applied to synthetic data and various well-known data sets from the UCI repository [DK17]. Classified data are ranked according to how consistently the high-dimensional classes are embedded in the 2D projection space. For unclassified data, a clustering algorithm is applied to generate high-dimensional class structures.

More recent work by Lehmann et al. [LHT15] and Pandey et al. [Pan+16] have investigated the human perception on scatter plot patterns and present both crowd-sourcing studies to prove evidence that this connection exists. Sedlmair and Aupetit [SA15] even present a data-driven framework for quality measure evaluation. Their approach tries to mitigate the impact of (relative) human judgments by relying entirely on ground-truth data. However, this in turn also indirectly implies some sort of user involvement. By using this framework, Aupetit and Sedlmair [AS16] evaluated a large number of visual separation measures for pre-classified data. They systematically generated 2002 visual separation measures by combining neighborhood graphs and class purity function with different parameterizations. As a result, they identified measures that outperforms the distance consistency measure. Sher et al. [She+17] conducted a study about the human perception of correlations in scatter plots. Their study reveals that humans perceive correlations differently compared to the statistical measure of Pearson's product-moment correlation coefficient.

Bertini et al. [BTK11] pointed out that all quality metrics that work in the image space try to simulate the human pattern recognition machinery and therefore, it is needed to validate and tune the metrics in a way that the parameters take models of human perception into account. Together with other colleagues [Tat+10], they presented a user study about human perception and quality metrics, where they compared the outcome of quality metrics with human rankings. The usefulness of Lehmann's A-SPLOM [Leh+12] was evaluated by a controlled experiment including 12 participants. The task of the study was to select relevant plots from different SPLOM configurations (A-SPLOM, unsorted A-SPLOM, sorted SPLOM, unsorted SPLOM). Finally, they compared mean and variance values of the number of selected plots to the values of the quality measures.

Open Research Questions

Even though a lot has been done in the field of quality metrics for scatter plot visualization, there are still some directions that can be further investigated. One possible direction could be the integration of human sensing technologies, e.g., eye tracking or motion tracking, to investigate the behavior of users during an analysis task. For instance, prior research of Shao et al. [Sha+17] has shown that eye tracking devices can be used to track already explored patterns, and thus support the exploration of varying patterns in large scatter plot spaces. Furthermore, eye tracking has also been used the evaluation of scatter plots and parallel coordinates. By using an area-of-interest (AOI) approach, Netzel et al. [Net+17] showed how participants act during analysis tasks and identified different reading strategies. Consequently, these sensing measurements could be integrated into the quality metrics-driven visual analytics pipeline and enrich the quality criterion inputs (user U, task T).

2.6.3 Parallel Coordinates

Parallel coordinates [Ins09b] are one of the most popular visualizations for multi- and high-dimensional data. Introduced to the information visualization community by Inselberg [Ins85], the technique gained popularity by enabling analysts to explore patterns across a large set of dimensions. Equally-spaced vertical axis represent the dimensions of the dataset; the top of the axis corresponds to the highest, the bottom to the lowest value in each dimension. Data points are mapped to polylines across the axis, such that the intersection between an axis and a polyline marks the data value. This visual mapping allows analysts to spot high-level patterns, as well as single data points of interest.

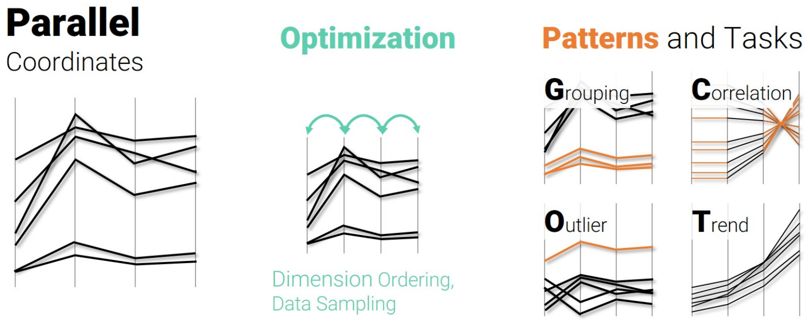


Fig. 2.8. Parallel coordinates – optimization goals, tasks & visual patterns

Why Do We Need Quality Metrics for Parallel Coordinates?

parallel coordinates face three major challenges: (1) With an increasing number of data records, the patterns start to disappear due to *overplotted* lines [ED07]. (2) A perceived pattern in parallel coordinates depends primarily on the ordering of the dimension axis [JJ09]. A proper ordering (for a specific task) can reveal unknown patterns while a non-useful ordering may hide them. Consider the example in Figure 2.9. Two different ordering strategies are applied to reveal clusters and correlations. (3) A large number of dimensions decreases the available screen space between two axes and results in cluttered plots; in particular when many data records are shown [DK10]. (4) The perception of positive and negative correlation is not symmetric: negative correlations are visible more clearly [Hei+12a; LMW10].

Quality metrics need to tackle these challenges by measuring the quality of a perceived pattern and the amount of clutter/overplotting in order to be able to guide ordering and sampling strategies. However, there are $n!$ possible dimension permutations (based on the assumption that we plot every dimension exactly once). Having a quality criterion that measures the quality of one particular permutation, Ankerst et al. [ABK98] prove that finding the optimal ordering can be reduced to the traveling salesman problem and is therefore $\mathcal{NP}$-complete. As a consequence, not only quality criteria but also efficient optimization algorithms are necessary.

Typical Analysis Tasks for Parallel Coordinates

Countless applications from various (research) domains have been tackled with parallel coordinates. In a recent state-of-the-art report by Heinrich and Weiskopf [HW13],

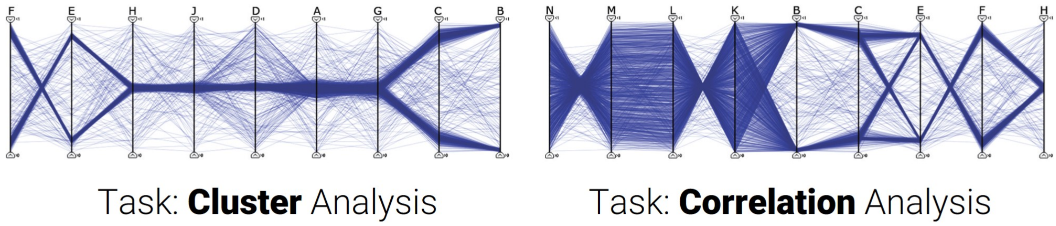


Fig. 2.9. Two partially different subsets of dimensions with axis ordering strategies optimized for cluster analysis (left) and correlation analysis (right). Figure adapted from [JJ09].

the tasks of these applications are categorized according to the established KDD taxonomy by Fayyad et al. [FPS96]: *classification*, *regression*, *clustering*, *summarization*, *dependency-modeling*, and *change and deviation detection*. In analogy, we show in Figure 2.8 four of the main visual patterns for that help to accomplish these tasks: *grouping*, *correlation*, *outlier*, and *trend*. Quality metrics should be able to re-order and de-clutter parallel coordinates such that these patterns are visible to the analyst (based on the current analysis task).

Summary of Approaches

A multitude of quality metrics has been presented for parallel coordinate plots. The approaches can be separated into *quality criteria* measuring the quality of one *visualization definition* and *optimization algorithms* that optimize the adjustable parameters. In the following, we will first describe and discuss the variety of quality criteria, followed by the applied optimization algorithms.

The first criterion described in the literature has been developed by Ankerst et al. [ABK98] with the argumentation that a similarity-based ordering will reduce visual clutter. In their data space approach, the authors propose finding a perceptually "good" ordering by measuring the Euclidean distance between two dimensions on a global level, or by partial similarity based on a defined threshold. The quality criterion measures the sum of distances between all neighboring dimensions, which needs to be minimized by the optimization algorithm. Yang et al. [Yan+03] extend the idea by applying a clustering on the dimensions first. Due to the resulting hierarchy, the search space of permutations can be reduced by considering only dimensions within one cluster.

Another similarity-based method is proposed by Peng et al. [PWR04]. The authors claim that the source of clutter can be caused by distortions of the data distribution, e.g. due to outliers. Peng et al. define an outlier based on the nearest neighbor algorithm and propose a quality criterion based on the proportion of outliers between two neighboring dimensions. Similar to the previous approaches, clutter is only measured between two neighboring dimensions in the visualization.

A quality criterion for supporting nearest neighbor searches is proposed by Peltonen and Lin [PL17b]. In their approach, the similarity between axes is computed using the Kullback-Leibler divergence of probabilistic neighborhood distributions.

Ellis and Dix [ED06a; ED06b] propose three methods to estimate the occlusion of lines in parallel coordinates: (1) overplotted% (percentage of pixels with more than one *plotted* point), (2) overcrowded% (percentage of pixels with more than one *existing* point), and (3) hidden% (percentage of plotted points *hidden* due to overplotting). Ellis and Dix propose several data space algorithms to count the number of pixels or points respectively. All criteria can be applied globally or in areas of interest, e.g., by a sampling lens [EBD05].

Several methods quantify the difference in the data distribution between the original space and a subset of data records or dimensions. Cui et al. [Cui+06] measure the difference of data density for all dimensions using a histogram approach. The quality criterion retrieves the difference between the histogram of the data sample and the histogram of the original data. In the same paper, the authors extend the idea by quantifying the similarity of each record in the original space with its nearest neighbor in the sample. An image space method by Johansson and Cooper [JC08] transforms the visualization into a so-called distance map [RP66] in which each pixel describes the distance to its closest object. The quality criterion measures the similarity between the distance maps of the original and the sampled data.

Several approaches argue that the first dimension attracts the most attention of the user. Therefore, it should be considered in the ordering. Lu et al. [LHZ16] use Singular Value Decomposition to measure the contribution of each dimension to the data space. Highly contributing dimensions are sorted up front. Yang et al. [Yan+03] consider the importance of a dimension (e.g., by variance) in their similarity-based ordering. A different method is proposed by Ferdosi and Roerdink [FR11] to promote grouping patterns. They do not only consider a pair-wise combination of dimensions, but rather search for high(er)-dimensional structures by using a subspace clustering algorithm. The quality of a subspace is measured by the density distribution [FR11] and an implicit algorithm sorts the dimensions based on the quality of each individual subspaces.

Tatu et al. [Tat+09] introduce three image-based quality criteria to measure the quality of perceived clusters. The assumption of all methods is that clusters are usually represented by clustered lines with a similar position and direction. The image of the visualization is transformed by a Hough transformation [VC62] into a new image, such that lines with a similar slope and interception are at a close location. The quality criterion measures the clusteredness between two dimensions within the Hough space. For datasets with given cluster labels, Tatu et al. adapt their measure and focus on (1) the intra-class similarity, and (2) cluster overlap by measuring the difference between the Hough space images per cluster.

One of the most central image-based QM approach is proposed by Dasgupta and Kosara [DK10]. Pargnostics, following idea of Scagnostics [WAG05] for Scatter Plots, are a set of seven quality criteria for parallel coordinates: *number of line crossings*, *angles of crossing*, *parallelism*, *mutual information* (dependency between variables), *convergence* and *divergence*, *overplotting*, and *pixel-based entropy* (randomness = uncertainty). The proposed measures are computed from 1D statistics and 1D/2D distance histograms, which allow for a rapid computation. The optimization algorithm can make use of a weighted combination of features.

Finally, Johansson and Johansson [JJ09] provide an interactive analysis of the whole high-dimensional dataset based on different quality metrics that can be selected and weighted by the user. The authors describe three criteria to measure the quality of a plot: (1) *Correlation analysis* by the Pearson correlation coefficient [LN88] between neighboring dimensions. The quality scores between neighboring dimensions are aggregated for the entire plot. (2) *Outlier detection* based on a grid-based density computation. The quality criterion combines the number of dimensions and the distance to the nearest neighbor across multiple dimensions. (3) *Cluster detection* by a subspace clustering approach (e.g., Mafia algorithm [NGC01]). For each subspace cluster, a quality score is computed representing density, dimensionality, and the fraction of the covered dataset.

So far, we have discussed quality criteria for combinations of two or more dimensions. In order to find an optimal ordering for the entire parallel coordinates plot, optimization algorithms are necessary. As shown by Ankerst et al. [ABK98], the reordering task in parallel coordinates is $\mathcal{NP}$-complete. The literature does not provide any novel algorithmic solutions, but rather applies existing approaches. To name a few: heuristic algorithms are used in [JJ09], a genetic approach is presented in [ABK98], and graph-based algorithm is applied in [Tat+09; DK10; HO10].

Evaluation Methods for Parallel Coordinates Quality Metrics

New quality metrics are mostly evaluated by showing examples based on synthetic or real-world datasets. Often, the performance of optimization algorithms is depicted in terms of efficiency. Only a few approaches compare multiple quality criteria: Ellis and Dix [ED06b] and Cui et al. [Cui+06] *empirically* compared their own approaches with each other. Ferdosi and Roerdink [FR11] systematically compared their subspace clustering approach with the similarity clustering method of Ankerst et al. [ABK98], the clutter-based method of Peng et al. [PWR04], the Hough space method by Tatu et al. [Tat+09], and the hierarchical dimension clustering method by Yang et al. [Yan+03].

The number of user-centered evaluations, investigating the perceptual aspects of QMs, is limited, as also discussed by a recent survey of Johansson and Forsell [JF16]. Few studies exist to measure the influence of clutter: Holten and van Wijk [HW10], Heinrich et al. [Hei+12a] and Palmas et al. [Pal+14] quantitatively analyze the reduction of clutter through edge bundling techniques or different variations of parallel coordinates extensions. Rosenbaum et al. [RZH12] evaluates the readability of parallel coordinates under different densities of data points, but focuses on progressive analytics argumentations. A qualitative and quantitative evaluation scheme considering the ordering of dimensions is discussed in Claessen and van Wijk [CV11] and Walker et al. [Wal+13]. Further studies exist to compare axis arrangements in 2D vs. 3D parallel coordinates: Forsell and Johansson [FJ07], Johansson et al. [JFC14; Joh+08], Lind et al. [LJC09]. A recent eye-tracking study of Netzel et al. [Net+17] revealed that participants pay more attention towards the center of parallel Coordinates Plots. This finding stands in contrast to the dimension contribution-based approaches, for example, by Lu et al. [LHZ16].

Open Research Questions

Promising future work is summarized by the survey of Johansson and Forsell [JF16]. The community has developed many quality metrics for parallel coordinates. The concepts of useful orderings and clutter reduction approaches for the underlying method differ significantly. However, there are no user studies that compare the different metrics for different tasks and different data characteristics. Based on such findings, the community could further develop task-dependent quality metrics that support the perception of humans.

2.6.4 Pixel-based Techniques

Pixel-based techniques create a separate view (called *subwindow*) for every dimension of a dataset. Within each subwindow, every data record is mapped to exactly one pixel, colored according to the value in the respective dimension [Kei00]. Pixel-oriented visualizations do not face overplotting issues, and they are designed to display large amounts of data without aggregation. The number of data points to be visualized is only limited by the available screen space.

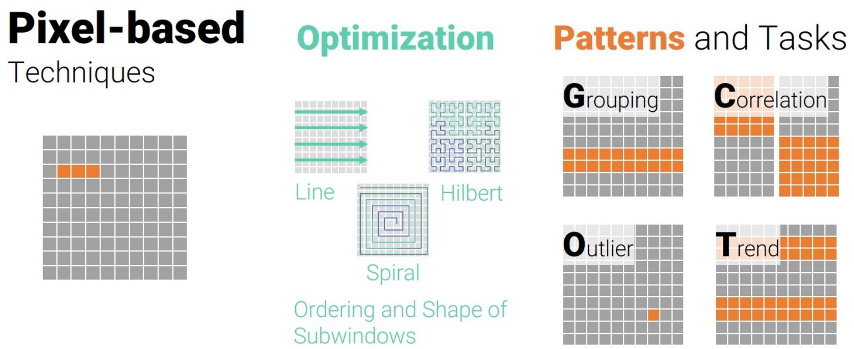


Fig. 2.10. Pixel-based visualizations – optimization goals, tasks & visual patterns

The most important aspect is the layout of pixels within each subwindow. For each window, the same layout is applied in order to make the dimensions comparable. Generally, the data points require an ordering, such as a natural order (e.g., by time or size), or the result of a function (e.g., order of nearest neighbors to a query object). Design recommendations by Keim [Kei00] and Wattenberg [Wat05] propose that data points need to be layouted such that the given ordering of the data is approximated in the subwindows. This means, data points that are nearby in the ordering, should end up nearby in the visualization. For rectangular-shaped subwindows, space-filling curves are proposed to optimize these recommendations, for example Hilbert Curves [Hil91] or H-Curves [NRS97].

Why Do We Need Quality Metrics for Pixel-based Techniques?

Pixel-based visualizations are designed to display large amounts of data, but only for individual attributes. With high-dimensional data with thousands of attributes being more and more common, it is practically impossible to manually inspect the visualization for each individual attribute for interesting patterns. Hence there is a need for quality metrics for pixel-based visualization techniques. These metrics help users to analyze high-dimensional data sets by calculating a quality for each attribute. These quality metrics can be used to identify interesting attributes or reordering attributes or data instances accordingly. According to Keim [Kei00], there are four properties that have to be considered when designing pixel-oriented visualizations. The color mapping, the arrangement of pixels, the shape of the subwindows, and the ordering of the dimensions. For each of these properties, Keim [Kei00] presents *design recommendations*. For instance, the usage of space-filling curves like the Morton curve [Mor66] for the arrangement of pixels. The problem is that the methods proposed by Keim such as the ordering of dimensions, the shape of the subimages, and arrangement of pixels require solving complex optimization problems. Some of which are proven to be $\mathcal{NP}$-hard [ABK98].

Typical Analysis Tasks for Pixel-based Techniques

Pixel-based visualization techniques are useful for solving four different tasks on large high-dimensional data, as depicted in Figure 2.10. When analyzing a single dimension, pixel-based visualizations can be used to identify clusters and outliers. Clusters, such as visible in Figure 2.11, can be identified by finding local regions of similar color. Outliers, in contrast, are depicted as points with outstanding colors in comparison to their surrounding region. Trends are depicted by consistently reoccurring occurrences of similar color spread out over the pixel plot. When considering multiple dimensions, pixel-oriented visualizations can be used to identify correlations between different dimensions. If a cluster occurs in multiple dimensions, this can be an indication for a positive correlation, if they share a color, or negative correlation, if they consistently depict a different color. However, finding these visual patterns is only possible if the ordering between and within dimensions is done appropriately.

Summary of Approaches

The existing approaches for pixel-based visualizations can be divided into data space, image space and hybrid approaches. Keim [Kei00], in addition to his general optimization algorithms for pixel-based visualizations, presents such data space quality criteria for geospatially-related data. The presented quality criteria focus on the layout and positioning of the pixels in the resulting visualization and measure,

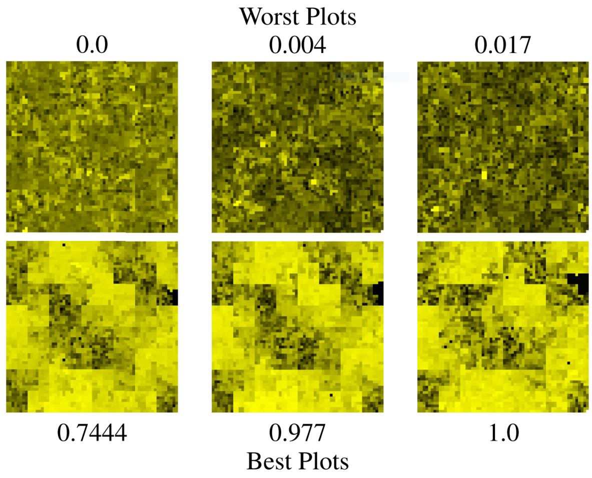


Fig. 2.11. Jigsaw maps and quality metric scores of the Ozone datasets created by Albuquerque et al. [Alb+10]. Depicted are six Jigsaw maps along with their calculated Noise Dissimilariy Measure. The top row shows the worst three plots and the bottom row the best three plots with clearly visible patterns.

for instance, the position-preservation of the layout algorithms, the relative position-preservation or the relative distance-preservation.

In addition to these data space approaches, also two image space approaches, Pixnostics [SSK06] and the Noise Dissimilarity Measure (NDM) [Alb+10] were presented for pixel-based displays. Pixnostics calculates the information content of a pixel-based visualization by calculating either the entropy or the standard deviation on the distribution of gray-level histograms in different grid cells. If the calculated score for a gray-level histogram of a cell is between two user-defined thresholds, it is considered to be interesting. However, this requires a manual setting of the interestingness thresholds. NDM uses the dissimilarity between a visualization and a noise image generated by a random permutation of the original visualization. Since the characteristic of the noise image is supposed to be the total absence of structure, visualizations with a large Noise Dissimilarity Measure are considered to have a higher potential relevance, as shown in Figure 2.11.

Evaluation Methods for Pixel-based Quality Metrics

In the presented works, there is no standard evaluation technique for pixel-based visualization techniques apparent. Keim [Kei00] provides a quantitative evaluation for geospatially-related data by measuring and comparing the position and distance preservation of different layout algorithms. Schneidewind et al. [SSK06] and Albuquerque et al. [Alb+10] both show the effectiveness of their quality metrics in a use case study, by showing the potential to find interesting visualization. Both start with a set of pixel-based visualization. Schneidewind et al. create their test set by randomly permuting the pixels of a Jigsaw map and Albuquerque et al. create multiple visualizations for the Ozone dataset, as shown in Figure 2.11. Both use their respective quality metric to calculate a score for each visualization in their test set and show that visualizations with a high QM score contain interesting visual patterns.

Open Research Questions

The quality metrics for geospatially-related data proposed by Keim [Kei00] focus on the , e.g. by calculating the position-preservation of the resulting visualization, while the quality metrics proposed by Schneidewind et al. [SSK06] and Albuquerque et al. [Alb+10] focus on the image space. Both of these image space approaches are particularly useful for identifying groupings. However, for other analysis tasks the pixel-visualization suitable quality metrics are missing. A beneficial line of research could seek to adapt e.g., TreeMap QM to this domain. Furthermore, a comparative and user-agnostic evaluation of the existing approaches could help to identify a baseline for further research directions.

2.6.5 Radial Visualizations

Radial visualizations for high-dimensional data arrange the data in a circular or elliptical fashion. Draper et al. [DLR09] presents a general survey on the topic referencing 88 works. Prominent techniques for high-dimensional data include, but are not limited to: the *MoireGraph* [JM03], the *TimeWheel* and *MultiComb* visualization both proposed by Tominski et al. in [TAS04] and the projection-based techniques, such as RadViz [Hof+97] and Star Coordinates [Kan00]. Note that visualizations, such as Pie Charts, Sunburst, or Radar Charts, albeit being radial visualization are explicitly excluded here, since their optimization focuses on storytelling and semantic aspects, c.f., high-level quality metrics in Section 2.3.2. The development of perceptual quality metrics was mainly driven by high-dimensional (projection-based) radial visualizations and thus will be the focus of this Section.

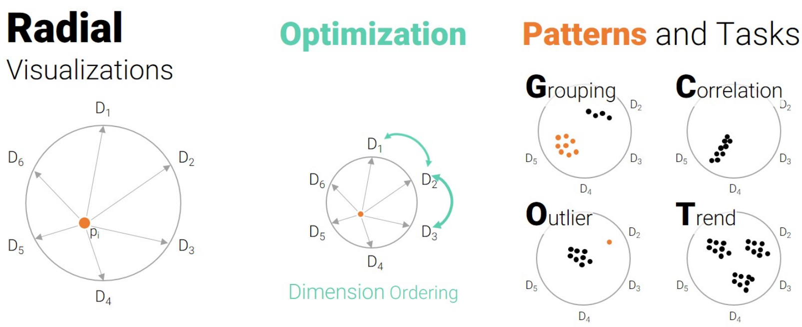


Fig. 2.12. Projection-based radial visualizations – optimization goals, analysis Tasks & visual Patterns

Projection-based radial visualizations are two-dimensional projections of high-dimensional data into a circle. For RadViz, the dimensions of a dataset X are represented as points that are evenly spread around the circumference of a circle. Each instance of the dataset $x_i \in X$ is also represented as a point inside the circle. The positioning of each instance x_i can be determined by connecting it with springs to each of the dimension representatives on the circumference of the circle. The final position of a point x_i is determined by the point p_i where the sum of all spring forces is zero and can be computed as $p_i = \frac{\sum_{j=1}^{n} d_i x_{i,j}}{\sum_{j=1}^{n} x_{i,j}}$, with d_j denoting the vector pointing from the center to the position of the respective dimension on the circumference [Alb+10]. Star coordinates apply a nearly identical mapping of points into a circle, but without the nonlinear normalization, for which the denominator of the previous equation is responsible. As a result of this, RadViz is especially advantageous for sparse data, but its nonlinearity may hamper several other exploratory analysis tasks [Rub+16]. Rubio et al. point out, that due to their similarity, many algorithms, such as the quality metrics developed by Albuquerque et al. for RadViz, can be directly applied to Star Coordinates and vice versa.

Why Do We Need Quality Metrics for Radial Visualizations?

Similar to Parallel Coordinates Plots (see: Section 2.6.3), radial visualizations are highly dependent on the ordering of dimensions, which is in turn dependent on the user's task. For example, if one data instance has high values in two neighboring dimensions, it is plotted more closely to the circumference, in addition, another data instance with high values in two opposite dimensions is plotted more closely to the center of the circle. Given an ordering of the dimensions, quality metrics can help to identify if the resulting visualization has interesting patterns for a specific user task. However, finding a suitable ordering is one of the key problems, which Ankerst et al. [ABK98] proved to be $\mathcal{NP}$-complete. Hence, QMs for radial visualizations are not only necessary but also efficient techniques to explore the search space of possible dimension orderings.

Typical Analysis Tasks for Radial Visualizations

RadViz was first proposed by Hoffman et al. [Hof+97] to help with the classification of DNA sequences. In their work they compare visualizations of multi-dimensional DNA sequence data. They compare, on the one hand, visualization techniques which are able to display all dimensions, i.e., RadViz and Parallel Coordinates, to techniques which use dimension reduction techniques to produce 2-dimensional visualization, such as Sammon Plots, on the other hand. They conclude that, although that some patterns can still be seen in the dimension-reduction techniques, the exact symmetry is lost, which is an inherent problem of such techniques due to the difficulty of choosing the important dimensions. RadViz can also be used to tackle various different tasks, as shown in Figure 2.12. Nováková and Štěpánková show how radial visualizations can be used to detect trends in time-series data [NŠ11]. Mrarmor et al. [Mra+07] use radial visualizations for outlier detection in lung cancer data based on gene expressions of six genes. Finally, Bertini et al. [BDS05] show how an extension of RadViz can help to detect correlations in data. Kandogan advertises star coordinates as a means for cluster, trend and outlier detection likewise [Kan00].

Summary of Approaches

Albuquerque et al. [Alb+10] show that due to the scatter properties of RadViz, most quality measures for Scatter Plots (see: Section 2.6.1), may be applied to RadViz as well, such as the *Class Density Measure* [Tat+09] for labeled datasets. They also introduce the *Cluster Density Measure (C_lDM)* as a new quality metric to rank visualizations based on how well-defined the clusters of the resulting projection are. This image-space based technique first applies an image clustering algorithm and then calculates the quality metric score based on the found cluster properties. They follow the following computational steps; calculate a density image based on the local neighborhood in the original visualization; smooth the density image by applying a Gaussian filter; identify clusters with the help of Laplace filters; and calculate the C_lDM measure, defined as:

$$C_lDM = \frac{1}{K}\sum_{k=1}^{K}\sum_{l=k+1}^{K}\frac{d_{k,l}^2}{r_k r_l}$$

where K is the number of detected clusters, $d_{k,l}$ the Euclidean distance between the cluster centers c_k and c_l and with r as the average radius of a cluster. Thus projection clusters with a small intra-cluster and large inter-cluster distance are assigned high values.

Another approach to calculating quality metrics for radial visualizations is presented by Di Caro et al. [CFF10]. They determine the visual usefulness of a projection by using the Davies-Bouldin (*DB*) index [DB79]. The *DB* index is known to be one of the best methods to measure the inter- and intra-cluster separation. A smaller *DB* index represents more compact and separated clusters. However, if a high-dimensional dataset d has a high DB index, it may become difficult for the projected data p to offer a high-quality visualization. Thus, the DB index is not directly used as a quality metric, but rather the ratio R between the index of the high-dimensional data DB_d and the projected data DB_p is taken, with a high R corresponding to a higher visualization quality.

Evaluation Methods for Radial Visualization Quality Metrics

Both quality metrics presented in the last section are used to evaluate new dimension-ordering techniques for RadViz. Di Caro et al. [CFF10] provide an independent and a RadViz-dependent formalization of the dimension arrangement problem, which was formalized by Ankerst et al. [ABK98] in a generic context. They provide an exhaustive evaluation of both of these dimension arrangement techniques, partly evaluating the visual quality of the resulting arrangements. Moreover, Albuquerque et al. [Alb+10] propose a greedy RadViz generation algorithm in which they start with a two-dimensional RadViz and iteratively add the remaining dimensions by checking which dimension they have to add for optimizing a quality metric. Additionally, they provide three comparisons of the resulting visualizations, using the original RadViz algorithm, the t-statistics algorithm of Sharko et al. [SGM08], and their algorithm, concluding that using their algorithm, the resulting projections show a better cluster separation.

Open Research Questions

So far two algorithms were proposed to measure the visual quality of visualizations generated by RadViz. One data space and one image space technique. Both approaches have shown, that their quality metric can be used to determine the visual quality of a resulting projection and they can even be applied during the construction of RadViz visualizations. However, the shortcomings are that both of these techniques focus on only one aspect, the intra- and inter-cluster separation. As previously shown, there are various possible applications of RadViz, with grouping only being one of these applications. In future work, quality metric for these different tasks, such as outlier or trend detection, or, if possible, a general quality metric usable for various tasks should be developed. Additionally, as Rubio et al. pointed in their comparative study of RadViz and Star Coordinates [Rub+16], algorithms designed for one technique, may be applied to the other. Therefore, when developing techniques for one technique, they recommend considering whether they would be appropriate for the other technique as well.

2.6.6 Glyphs

In the 1960s, the term glyph was used as a synonym for the metroglyph [And57]. However, over the years different glyph designs emerged and the initial definition was adapted to also describe new representations. As a result, the term glyph is used ambiguously in the visualization literature [Mun14]. Recent surveys tried to tackle this problem by identifying similarities across definitions and combining them in a more general statement [Fuc+17] or by categorizing already existing definitions into more specific or general interpretations [Bor+13].

In summary, all glyph designs are graphical representations of data points, which can be positioned independently from each other. This flexibility in the layout is the biggest advantage of glyphs. They can be easily combined with other established visualizations opening space for various application areas. Geo-spatial visualizations [AA04], graph visualizations [US09], tree-maps [FFM12], or scatter plots [WG11] are just a few examples where glyph designs can enrich other visualization techniques with additional information about the data.

Although the design space of data glyphs is nearly endless [Mun14], some designs have received more research attention than others. Chernoff faces [Che73], star glyphs [Sie+72], or profiles [DSS86] are prominent examples. However, in comparison to faces, star-like glyph designs and profiles are more often used in practice. Therefore, we want to focus on star-like glyphs and profiles to outline a very different approach to quality assessment and evaluation: *design recommendations*. For a better readability and didactic reasons, we enumerate recent and influential works on the evaluation aspect of design recommendations in the unified subsection *Summary of Evaluation approaches*.

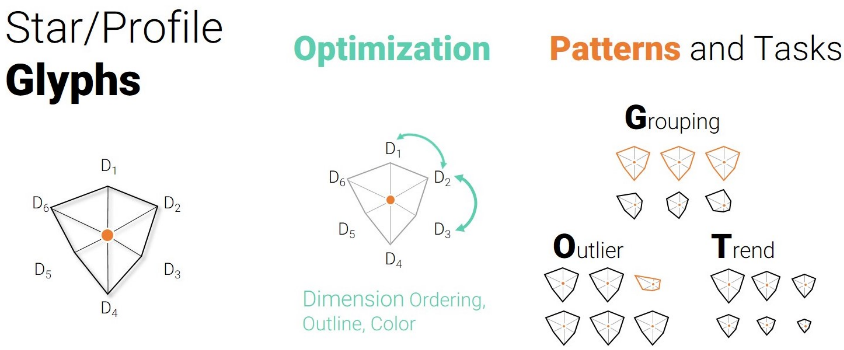


Fig. 2.13. Glyphs – optimization goals, tasks & visual patterns

Why Do We Need Quality Metrics for Glyphs?

In general, star or profile glyphs are closely related to parallel coordinate plots. They use a similar visual encoding to show dimensions and data values. Data lines are radiating from a center point to represent attribute dimensions. The length of each line is dependent on the underlying dimension value. The higher the data value, the longer the respective line. The endpoints of the data lines are connected to create a star-like shape. In comparison to parallel coordinates, the major differences are the reduced size, the circular layout of the axes and the presence of just a single data line in the plot.

As in parallel coordinates, the order of axis has a strong influence on the visual appearance of the individual stars and, therefore, need to be considered in the design process. Additionally, star glyphs can also be represented without the surrounding contour line [PG88], since this visual feature does not carry any information about the data. Adding color to the plot or highlighting certain visual features might also help to better solve the analysis task.

Profile glyphs are a more abstract term for small bar charts or line charts (i.e., sparklines [Tuf06]). They are easy to read and understand since they built upon a common mental model. Like in bar charts, the width of the bars, as well as their ordering, can be varied or single bars can be connected to show some trend information.

Although these two designs seem to be well-established, they still allow for some design variations. To come up with an optimal design is difficult, since the design of a glyph is a creative process with only limited guidance and nearly numberless design possibilities.

Typical Analysis Tasks for Glyphs

Data glyphs are used in various settings and for different analysis tasks. Based on Andrienko and Andrienko's task taxonomy [AA06], lookup tasks for single data values and similarity search are the most common analysis tasks followed by visual search and trend detection [Fuc+17]. Therefore, the optimal glyph design strongly depends on the task at hand. Is it important to perceive the entire shape as a whole (like in synoptic tasks) or is the focus on reading individual visual features (like in elementary tasks).

Summary of Evaluation approaches

Glyph designs are a good example of visualization techniques, that strongly profit from design considerations based on results from quantitative user evaluation. Star glyphs profit from the following recommendations that can be used to guide the design process. The surrounding contour line should be removed from the design. Studies have shown that participants are more accurate when comparing the similarity between data points using stars without a contour line [Fuc+14]. There are also guidelines for ordering the axes of stars. Results from experiments suggest avoiding salient shapes [KHW09]. This design consideration coincides with the clutter reduction quality metric proposed by Peng et al. [PWR04]. Additionally, the axes should be colored to reduce the negative influence from single spikes for visual classifications tasks [Kli+09].

To further improve the comparison between multiple stars, clustering results or statistical information should be added to the designs. Based on study results, researchers suggest adding the first and second principal component as additional axes to improve similarity comparisons [BS92]. Yang et al. [Yan+03] also proposed a quality metric to vary the angles between dimensions based on a hierarchical cluster analysis of the respective dimensions. Since no study has been conducted, this metric must be considered with caution.

However, the general public has to be careful about those recommendations, since all guidelines result from controlled experiments which are constructed to reflect specific conditions (e.g., analysis task, number of dimensions, layout). It is, therefore, difficult to generalize those findings [Fuc+17].

Open Research Questions

It would be interesting to transfer quality metrics from other visualization techniques to the data glyph domain. A good starting point can be parallel coordinates. Since star glyphs and parallel coordinates share many visual features, approaches for ordering dimensions could be adapted. Are star glyph specific orderings better compared to approaches used in parallel coordinate plots? Research has already made a first step in this direction by applying similar approaches to both visualization techniques [PWR04; Yan+03; HO10]. However, there is still much space for further research since the design space of data glyphs is huge.

2.7 Discussion

While this chart describes existing approaches for the quality assessment of visualizations, there are still many opportunities to improve and extend existing metrics. In this Section, we report general findings and highlight promising future research directions.

2.7.1 General Findings and Discussion

In the following, we discuss common aspects of quality metrics that span across most visualization techniques.

Which QM favors which visual pattern? One of the central questions for QM design is how an effective instance of a particular visualization type should look like. This *understanding* is implicitly modeled into a heuristic algorithm trying to capture if the *subjective* QM designer's *expectation* of the visual structure is met. However, in exploratory analysis settings, it is unclear which QM to apply. Some QMs favor one visual pattern, others another. But, it remains to the user to guess which data or visual pattern is in the dataset. What is even worse is that a majority of QMs is presented or published for the purpose of quantitative algorithm evaluations *without* describing which visual pattern they prefer.

What are the extreme cases that a QM can deal with? And what happens if the specifications are not met? Only a few of our surveyed approaches have been systematically investigated for their noise (in-) variances and robustness toward skewed data distributions. However, it is important that quality metrics can be applied independently of the quality of the data or the existence of patterns. A user should assume that no patterns exist in a dataset in case a quality metric does not provide a useful representation.

Is QM research transferable among visualization types? We found that some visualization subdomains share similar quality criterion. For example scatter plots and parallel coordinate Plots (see: Section 2.6.1 and Section 2.6.3) where the same clutter reduction techniques have been adapted for the respectively other visualization field (c.f., [ED06a; ED06b]). Another example are parallel coordinates and star glyphs where similar ordering strategies for the arrangement of axes are applied. Further work should explore if research efforts can be transferred between visualization techniques and subdomains.

Are QMs equally descriptive? In the case that QMs for different visualization techniques are able to assess equally well the same visual patterns, then QMs could be used as visualization type recommenders. This, in turn, presumes that a standard set of base patterns has been accepted and established in the respective visualization subfield. This itself is a challenging question, since not only the type of dataset (e.g., hierarchical, high-dimensional, relational) influences the to-be-expected patterns but also the domain.

Evaluation of Quality Metrics. Notably, many works cited in this survey acknowledge and explicitly mention the fact that the evaluation of QMs is not backed up with perception-focused user studies. This statement holds explicitly for quantitative quality metrics. As mentioned in Section 2.3.3, design recommendations are mostly derived from qualitative and quantitative user studies.

We claim that both approaches are valid but eventually should be backed up with the respectively other approach. Heuristics should be evaluated for their perceptual aspects and proven to correspond to the humans' perceptual properties. This can be only done in structured large-scale user studies. Especially, crowdsourcing studies, such as in [HB10], allow for more and more (statistically) sound statements to be made. Design recommendations, in contrast, should be translated eventually into algorithms for deriving quantifiable heuristics. This step allows one to make design recommendations generally usable, comparable, and unambiguous.

Another important aspect for the evaluation of quality metrics is the availability of *perceptually-inspired* benchmark datasets. To address the limitation for such datasets, Schulz et al. [Sch+16] propose generative data models for the validation, evaluation, and benchmark generation. In their paper, they survey various approaches that have been suggested to overcome the problem of the availability benchmark datasets for different types of data. They argue for the use of generative gold-standard data for a standardized evaluation of visualization approaches, in particular, with respect to perceptual quality.

2.7.2 Limitations of this Survey

This work surveys mid-level perceptual quality metrics for multi- and high-dimensional data by motivating the needs and benefits of quality metrics in the respective visualization subfield, summarizing the challenges and outlining analysis tasks supported by quality metrics in the literature. Our goal is to provide a central document where concepts from multiple visualization subdomains are enumerated and related, and their overarching concepts are discussed in contrast to each other.

While we discussed at length several alternatives to our present taxonomy, we finally opted to guide the reader through a structured questionnaire in each visualization section. We believe that the (missing) understanding of the visualization design challenges is a fundamental barrier to the effective use of visualizations in practice today. By providing a straightforward description of the problems and possible solutions in simple terms, we hope to help a wide audience better understand these algorithms and integrate them in future systems and libraries.

While we are trying to educate the user in the selection of QMs for a respective visualization type, a systematic answer to the question *“Which QM is the best one for my circumstances?”* remains extremely challenging. We decided against attempting to describe this matching formally. In particular, we do not think this is possible without considering domain-dependent, data-dependent, and user-dependent aspects.

2.8 Conclusion

This survey presents quality metric approaches for multi- and high-dimensional information visualizations. We summarize the efforts from six distinct visualization techniques.

We found that the major research developments in the field are increasingly abandoning the idea of pure clutter reduction approaches and focus on visual pattern retrieval. This in turn has significant implications for visualization techniques and visual analytics in the exploration process. Within an integrated QM-driven automation, as depicted in Figure 2.1), the user will be guided to the primary (visual) patterns within the data *and* will be presented with a birds-eye perspective allowing to assess the dataset-inherent importance of each pattern. Thus, not only clustering-, but also outlier-, correlation-, and trend analysis tasks can be accomplished more effective and more efficient.

One of our core contributions of this work is that we formalize, unify, and exemplify the major QM vocabulary. In future, we can expect that such a unified understanding will enable a more structured work on this problem.

By gathering the knowledge in a central document, we hope to inspire more research to develop novel quality metric measurement strategies, more externalized and quantifiable criteria proven to mimic the analysts perceptual system, as well as novel exploration approaches to harness the power of QMs.

3 Evaluating Ordering Strategies of Star Glyph Axes

Summary

Star glyphs are a well-researched visualization technique to represent multi-dimensional data. They are often used in small multiple settings for a visual comparison of many data points. However, their overall visual appearance is strongly influenced by the ordering of dimensions. To this end, two orthogonal categories of layout strategies are proposed in the literature: order dimensions by *similarity* to get homogeneously shaped glyphs vs. order by *dissimilarity* to emphasize spikes and salient shapes. While there is evidence that salient shapes support clustering tasks, evaluation, and direct comparison of data-driven ordering strategies has not received much research attention. This chapter contributes an empirical user study to evaluate the efficiency, effectiveness, and user confidence in visual clustering tasks using star glyphs. In comparison to similarity-based ordering, our results indicate that dissimilarity-based star glyph layouts support users better in clustering tasks, especially when clutter is present.

This chapter is *taken from* the following publication. Please refer to Sections 1.4 and 1.5 for the contribution clarification and general citation rules.

[Mil+19] Matthias Miller, Xuan Zhang, Johannes Fuchs, and **Michael Blumenschein**. *"Evaluating Ordering Strategies of Star Glyph Axes"*. In: IEEE Visualization Conference (VIS). 2019, pp. 91–95.

3.1 Introduction

Data glyphs are compact visual representations of multi-dimensional data points. Due to their small graphical appearance, they can be used in various settings like within node-link diagrams [Erb02], treemaps [FFM12], tables [KFM11], or geographic maps [Fuc+14]. For instance, star glyphs are employed in the medical domain [RP08] and can be used to show the spatial distribution of food production [Opa+18].

Due to their use of visual variables, star glyphs [Sie+72] are an adequate choice to encode single data points comprising numerical data. The glyph's axes represent the data dimensions, and their lengths encode numeric values. Since glyphs are versatile, different design variations of star glyphs emerged in literature. Many have already been extensively analyzed by the community (e.g., [Fuc+14], see [Fuc+17] for a full enumeration). However, there is not much empirical research about the effect of axes ordering strategy on visual comparison tasks.

The ordering influences the shape of a star glyph and affects its readability and similarity judgment [KHW09; Kli+09]. Therefore, we need (task-based) guidelines to arrange the dimensions in star glyphs [War08].

Numerous ordering strategies for star glyphs have been proposed [ABK98; AOL06; PWR04; War08; FK03; Yan+03; KHW09; Kli+09; KW09] which can be grouped into *similarity-based* (short: `SIM`), favoring homogeneous shapes, and *dissimilarity-based orderings* (short: `DIS`), emphasizing spikes and salient shapes. Some approaches also discuss symmetry, monotonicity, convexity and concavity, feature saliency, and user-driven relationships among neighboring dimensions. The ordering strategies typically analyze the relationship among all pair-wise dimensions and then adjust the axes of every star glyph simultaneously according to a metric (e.g., `SIM` or `DIS`). However, this also means that not all glyphs will result in the desired shape. In particular, outliers may be encoded by shapes which the reordering algorithm is trying to avoid.

We address the research question: "Which ordering strategy is most useful for similarity search and data grouping tasks (clustering) using star glyphs?". According to the task taxonomy by Andrienko and Andrienko [AA06], similarity search, and grouping are among the most common analysis tasks for glyphs [Fuc+14]. While different strategies have been proposed, they are not yet evaluated by empirical studies. Klippel et al. [KHW09; Kli+09] evaluated the influence of a star glyph's shape in grouping tasks. Although they found out that salient shapes, e.g., having spikes, can support grouping tasks, they did not apply a dimension ordering strategy that considers these salient properties.

Sorting the dimensions by dissimilarity favors the spikey-design, which Klippel et al. states to be promising for grouping. We compare this ordering strategy with the similarity-based design which is often proposed in the literature [War08; FK03; BS92; ABK98]. We conducted an empirical user study with 15 participants to evaluate the efficiency, data clustering quality, noise identification quality, and user confidence between the two different strategies (first independent variable). Our results show that star glyphs, ordered by a dissimilarity-based layout, support users better in a clustering task.

Real-world data often contains non-relevant dimensions with clutter and noise that may distort interesting patterns [GLH15]. Additionally, clusters do often not span across all dimensions but exist only in subspaces [KKZ09]. Therefore, we investigate impact of clutter on cluster identification and reordering strategies as a second independent variable. We use the term *clutter dimensions* to describe attributes that do not discriminate clusters but hinder the comprehension of feature relationships in the data [PWR04]. Therefore, we also investigate the influence of clutter separately, and in combination with the ordering strategies. For replicability and reproducibility, the material of the study (benchmark data, study results, analysis scripts, and code) is publicly available at https://osf.io/bje89.

3.2 Related Work

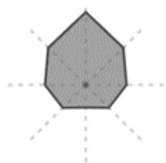
Finding an optimal star glyph axes ordering has proven to be NP-complete [ABK98] and more research is required [War08; Rze13]. It is related to the ordering of axes in parallel coordinates [ID90; ZMM12; ABK98; DK10; Tat+11], RadViz [Hof+97; CXM17; Alb+10; CFF10], ArcViz [Lon18], and other axes-based radial visualizations as summarized by Behrisch et al. [Beh+18]. Ordering algorithms typically define an objective function, modeling a good dimension order (according to their interpretation) and apply a heuristic to find an axes order which maximizes the objective function [War08].

3.2.1 Dimension Ordering Strategies

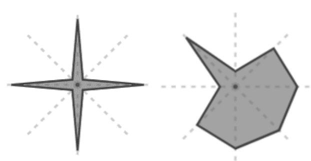
Different visual characteristics can be subject to shape optimization when applying specific ordering strategies of the star glyph axes. Ward [War08] summarizes four major strategies which have been extended by others: *user- and data-driven*, *correlation- and similarity-driven*, *spikes and salient shapes*, and *symmetry-driven*.

User-driven dimension orderings enable experts to adjust the shape of a star glyph based on their domain knowledge [Sac+14]. Users can select a data point to sort the data dimensions with ascending or descending order (*data-driven*) to reveal patterns between records [War08].

Correlation- and similarity-driven strategies improve star glyphs by adjacent placement of similar axes to support understanding of clusters, outliers, and relationships [Bor+13]. Ankerst et al. propose heuristic algorithms based on similarity for star glyphs to improve the overall perception [ABK98]. Similarly, Artero et al. use similarity heuristics of attributes to apply dimension-ordering and take perceptional aspects as Gestalt Laws into account by applying dimension reduction [AOL06]. Yang et al. combine similarity-based ordering with a hierarchical structure of the dimension to enable interactive exploration of high-dimensional subspaces [Yan+03]. Friendly and Kwan argue that using correlation-based ordering in star glyphs supports the identification of shape irregularities [FK03]. The authors did not conduct a survey to underpin their statement.

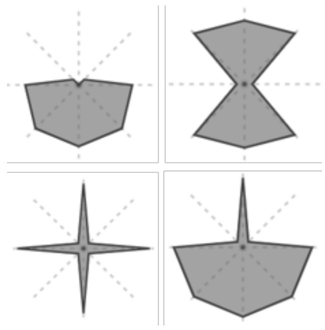


Spikes and salient shapes such as “*has-one-spike*” are helpful in visual grouping tasks of data points according to Klippel et al [Kli+09]. They argue, that concavity is more suitable for comparability than convexity, which is especially true for the “star" glyph, due to the large variations between adjacent dimensions. The salience of dissimilar neighboring axes shall enhance the comparison speed and help to detect



changes. Klippel et al. showed that the star glyph shape with eight dimensions influences classification tasks [KHW09]. Especially, in contrast to earlier work that state that similarity-driven orderings improve high-dimensional visualizations, dissimilarity between neighboring axes contribute *salient* properties that are perceptually more noticeable.

Symmetry-driven reordering methods help to reduce the visual complexity of star glyphs and, therefore, support comparison tasks by improving memorability [KW09]. By providing some examples, Peng et al. argue that orderings with simple and symmetric as well as monotonic shapes of the star glyphs facilitate the identification of value differences between multiple dimensions [PWR04]. They emphasize that symmetry and similarity are primary factors to identify patterns. For this, Gestalt Laws are a solid foundation for perception design [War20]. Peng et al. state that star glyphs can be optimized by aligning the symmetry on the vertical or horizontal axis [PWR04]. An additional rotation optimization step can be included in the pipeline to find the best global rotation for all star glyphs of a dataset. Rotation can be applied on top of other ordering approaches.



3.2.2 Empirical Studies and Research Gap

While many ordering strategies, algorithms, and heuristics have been proposed star glyph dimension ordering, empirical evaluation is missing. Previous approaches mainly argue by showing examples or providing arguments w.r.t. to e.g., Gestalt laws. While this is useful to find differences between strategies, we also need empirical evidence to directly compare strategies respecting scalability, performance, analysis tasks, data characteristics, and user perception [War08].

We are only aware of two studies conducted by Klippel et al. [KHW09; Kli+09]. Their results indicate that spikes and salient shapes have a positive effect on visual grouping tasks and colored axes positively affect the processing speed and reduce the negative influence of shape saliency on rotated data glyphs. However, Klippel et al. did not directly compare different ordering strategies or evaluated them against a benchmark. Instead, they designed different star glyph shapes and analyzed how participants grouped them by their understanding of similarity during an exploratory analysis task. In our study, we aim to close this research gap by comparing two proposed reordering strategies using a controlled, empirical user experiment.

3.3 Empirical User Study

We evaluate whether a similarity- (homogeneous shape, short: `SIM`) or dissimilarity-based layout (spike and salient shape, short: `DIS`) is more efficient and effective for a

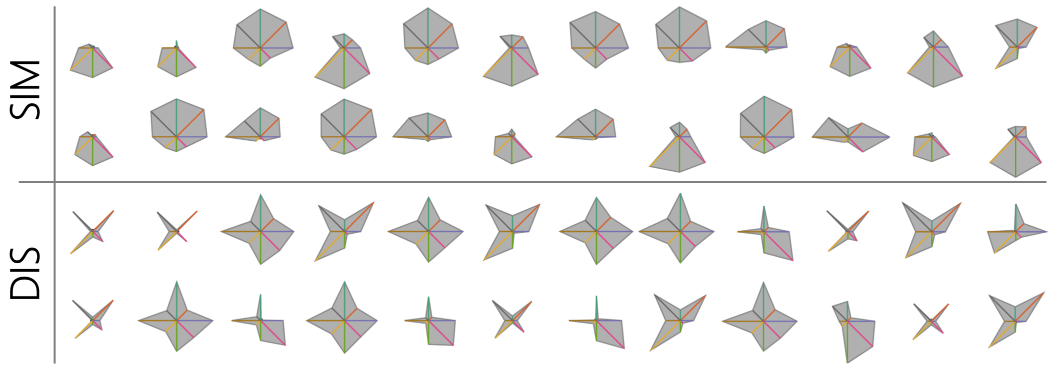


Fig. 3.1. Comparison of similarity (`SIM`) and dissimilarity (`DIS`) based ordering using the same data records.

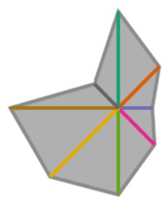
visual clustering task. We designed our study based on Klippel et al.'s work [KHW09; Kli+09]. We adopted the task, user interface, glyph design (including colored axes), and datasets' dimensionality (eight dimensions). However, in contrast, we applied two different reordering algorithms (`SIM` and `DIS`) and different clutter levels as independent factors, and evaluate the results using a benchmark dataset.

3.3.1 Experimental Design and Hypotheses

The participants had to manually assign star glyphs into reasonable clusters and identify noise, i.e., items not belonging to any cluster. In the study, we used the term *group* instead of cluster. To assess the performance, we use four dependent variables as quality measures: (i) *task completion time*, (ii) *quality of groups*, (iii) *quality of identified noise*, and (iv) the *confidence* of the participants.

Participants. We recruited 15 participants from the local student population (seven female, two bachelor, twelve master, one PhD student). The age ranged from 20 – 27 years with a median of 23. The participants had a different background in data analysis and visualization: ten had general knowledge in data analysis, four had data visualization experience, and one has used star glyphs before. All participants received a compensation of 10 EUR.

Glyph Design and Implementation. The glyphs are designed analog to Klippel et al.'s work [KHW09; Kli+09] using a contour, gray background, and colored axes. We used ColorBrewer [HB03] to select diverging colors and applied the ordering algorithm by Ankerst et al. [ABK98]. The Euclidean distance is used to measure the (dis)similarity between dimensions. We ran an exhaustive search to find the permutation with the highest (`SIM`) and lowest (`DIS`) similarity. An example of star glyphs with the two orderings is depicted in Figure 3.1. Orientation (rotation)



of the star glyphs is not considered and chosen randomly. All orderings are precomputed not to influence the run time during the study. We provide the study and the ordering strategy implementation on our websites[1].

Hypotheses

We address the following two hypotheses:

H1. Clutter negatively influences visual comparison. With increasing clutter, the performance of grouping tasks drops, independent of the axes ordering. In particular, we expect users to be **(a)** slower, **(b)** less accurate in grouping accuracy, **(c)** less accurate in noise identification, and **(d)** less confident of their grouping.

H2. Klippel et al. [KHW09; Kli+09] argue that spikes and salient shapes support users in similarity estimation and grouping tasks. Therefore, the performance of users should increase with a dissimilarity-based ordering. Furthermore, we hypothesize that the salient shapes should support users even more if the data contains clutter since sharp edges are more perceptually apparent. In particular, we expect users to be **(a)** faster, **(b)** more accurate in grouping accuracy, **(c)** more accurate in noise identification, and **(d)** more confident of their grouping when dimensions are ordered by dissimilarity.

3.3.2 Benchmark Datasets

We manually created 18 different datasets using the PCDC tool [Bre+12]. Every dataset contains 50 records of which 2–7 data points are selected as *noise* (randomly distributed across all dimensions). The remaining data points are grouped into 2, 3, or 4 clusters with similar cluster sizes. Besides, we introduced *clutter dimensions* which do not discriminate any cluster, since we uniformly distributed all data points across the clutter dimensions. 6 datasets contain no clutter (`0C`), 6 one- (`1C`), and 6 two clutter dimensions (`2C`). For instance, in condition `2C` a dataset consists of six dimensions discriminating the cluster, while the remaining two introduce clutter. Thus, we generated the datasets to keep the number of dimensions consistent. To verify the manually created clusters, we run a DBSCAN [Est+96] (parameters: $minPts = 3$, $\epsilon = 0.5$) on all datasets.

3.3.3 Tasks, Procedure, and Data Analysis

Tasks and Procedure

Each study took an hour on average. Participants filled out a consent form, demographics, and report on previous knowledge in data analysis, information visualization, and star glyphs. Afterward, we described how to read the visual encoding of a star glyph using an artificial car dataset as an example. Specifically, we clarified that

[1] `http://subspace.dbvis.de/sg-study` and `http://subspace.dbvis.de/sg-ordering`.

star glyphs with similar shapes on different axes are not similar (rotation invariance). The participants performed three training trials before the study was recorded.

To conduct the study, we used a 27-inch screen with 2560x1440 resolution and a mouse to execute given tasks. Every participant had to perform 18 trials, leading to $15 \texttt{ participants} \times 18 \texttt{ trials} = 270 \texttt{ trials}$ for the entire study. In between two trials, we showed a blank screen with the term 'break' to motivate the participants to have regular breaks. Each trial consisted of manually grouping all 50 star glyphs of one dataset into distinct groups and noise. Then, the participants stated the confidence level about their selection on a 7-point Likert-scale. We did not provide the number of clusters per dataset and explicitly told the participants that there might be glyphs which do not belong to any group (noise).

Figure 3.2 shows our interface. Participants were able to add new or delete groups. Glyphs can be interactively assigned to groups by drag&drop. If a glyph was considered to be noise, then it remained in the left panel. Participants were able to undo or change a grouping also using drag&drop. In the study, we did not constrain the task completion time. We ended the study with an interview about the participants' strategy and preferences regarding the `SIM` and `DIS` ordering by showing examples. Questions and answers were recorded.

Randomization

Each participant performed 18 trials, i.e., the grouping task on all benchmark datasets was equally distributed between `SIM` and `DIS`. We randomized the order of the trials as follows: First, we grouped the datasets into their level of difficulty based on the amount of clutter (`0C`, `1C`, `2C`). Then, participants performed the trials with increasing difficulty, i.e., 6 × `0C`, then 6 × `1C`, and finally 6 × `2C`. For every clutter condition, we randomized the dataset order and randomly assigned 3 × `SIM` and 3 × `DIS`. We attached the randomization algorithm and our configuration in the supplementary material. A summary of our trials:

3	levels of difficulty (clutter: `0C`, `1C`, `2C`)	×
2	ordering strategies (`SIM`, `DIS`)	×
3	trials (2, 3, 4 clusters)	×
15	participants	=
270	**trials in total**	

Data Collection and Analysis of Results

In each trial, we recorded the grouping task completion time, the selected groups and noise, and the participants' confidence. Some participants created groups with only one or two glyphs. Thus, in a post-processing step, we converted such small groups into noise to execute a more coherent analysis. We measured the quality of

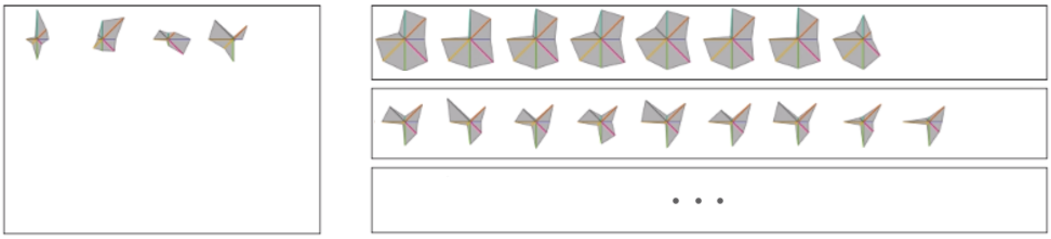


Fig. 3.2. Study prototype. Users can group visually similar star glyphs using drag&drop. Noise points remain in the left panel.

the identified noise by computing the Jaccard index between noise and ground truth noise.

The grouping quality is also based on the Jaccard index between the grouping and ground truth. However, since participants could have also selected too few or too many groups, we structured our quality computation as a two-step process: First, we computed the average Jaccard index of each group to its best match in the ground truth. Second, we computed the average Jaccard index of every ground truth cluster to its best match in the selection. Using this method, we considered too few, too many groups, as well as too few and too many records per group. The final clustering quality is the average score of both steps.

3.3.4 Results and Statistical Analysis

We executed a statistical analysis to summarize the study results. We report all statistically significant findings ($p < .05$) and some interesting trends visible in the data. We check for normal distribution using a *one-sample Kolmogorov-Smirnov test*. For a better comparison, we always report the median ($\bar{x}$) and, additionally, the mean (μ) for normally distributed samples. Analysis scripts and detailed results can be found in the supplementary material.

Statistical Tests used for Analysis

Confidence is measured as Likert-scale. Therefore, a *Pearson's Chi-squared test* is used for the analysis. Given the non-normal nature of the measures *time*, *cluster quality*, and *noise identification quality* w.r.t. `0C`, `1C`, and `2C`, we used a *non-parametric Friedman's test* and a *Wilcoxon signed rank test with Bonferroni correction* (Post-hoc). The same measures do also not follow a normal distribution w.r.t. the strategies `SIM` and `DIS`. Hence we used a *Wilcoxon signed rank test with continuity correction*. Considering `SIM` and `DIS` within `1C` and `2C` reveal normal distributed samples for the measures *time*, *quality of clustering*, and *quality of noise identification*. Hence, we use a *paired t-test* for the statistical analysis.

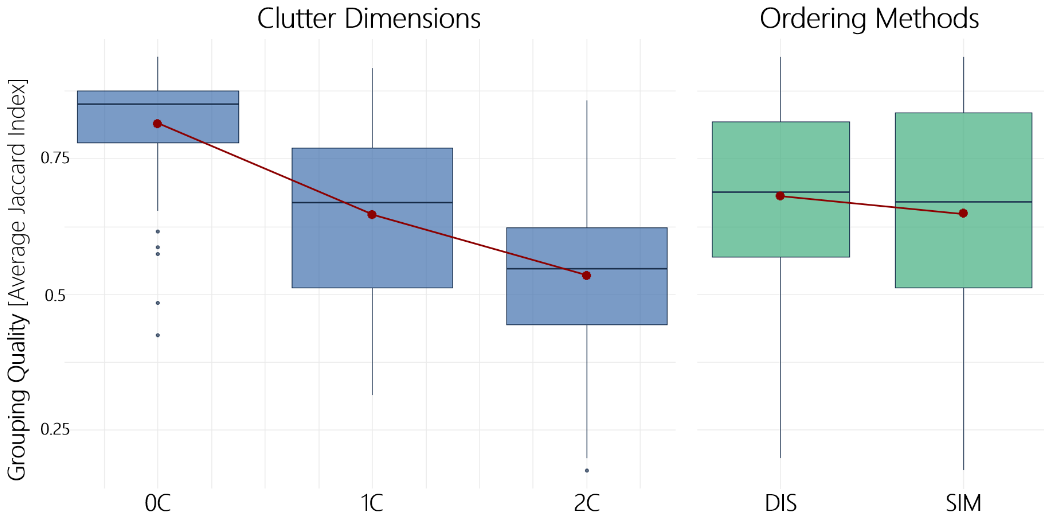


Fig. 3.3. Cluster quality analysis. Difference between clutter dimensions `0C`, `1C`, and `2C` as well as `DIS` and `SIM` ordering.

Task Completion Time

H1a. Task completion time increased with *clutter levels*, but not significantly: `0C` ($\bar{x} = 162.5s$), `1C` ($179.5s$), and `2C` ($184.0s$).

H2a. Using the *ordering* `DIS` ($176.0s$) users completed the grouping task slightly faster than `SIM` ($178.0s$), but only for datasets with clutter dimensions. `0C`: `DIS` ($168.0s$) vs. `SIM` ($158.0s$), `1C`: $180.0s$ / $179.0s$, and `2C`: $180.0s$ / $190.0s$. Differences are not significant.

Cluster Quality

An overview of the cluster quality is depicted in Figure 3.3.

H1b. There were significant effects of *clutter level* on *cluster quality* ($\chi^2(2, N = 270) = 109.92, p < .001$). Post-hoc tests showed a higher participants' accuracy with `0C` ($\bar{x} = 0.85$) compared to `1C` ($.67$, $p < .001$) and `2C` ($.55$, $p < .001$), and between `1C` and `2C` ($p < .001$).

H2b. When comparing *ordering strategies*, participants were more accurate with `DIS` ($\bar{x} = .69$) compared to `SIM` ($\bar{x} = .67$, $p < .05$), which is also true within clutter levels `1C` and `2C`, but not `0C`. `0C`: `DIS` ($\bar{x} = .85$, $\mu = .81$) vs. `SIM` ($\bar{x} = .85$, $\mu = .82$), `1C`: `DIS` ($\bar{x} = .68$, $\mu = .66$) vs. `SIM` ($\bar{x} = .66$, $\mu = .63$), `2C`: `DIS` ($\bar{x} = .58$, $\mu = .57$) vs. `SIM` ($\bar{x} = .52$, $\mu = .50$, $p < .001$).

Noise Identification Quality

H1c. There was a significant effect of *clutter level* on *noise identification* ($\chi^2(2, N = 270) = 80.02, p < .001$). Post-hoc tests revealed that participants were more accurate with 0C ($\bar{x} = .8$) compared to 1C ($\bar{x} = .5$, $p < .001$) and 2C ($\bar{x} = .33$, $p < .001$). In addition, there was a significant effect between clutter conditions 1C and 2C ($p < .001$).

H2c. In general, there is no difference between SIM and DIS w.r.t. noise identification quality (both $\bar{x} = .5$) There are no differences for 0C (both $\mu = .77$, DIS $\bar{x} = .88$, SIM $\bar{x} = .8$) and 1C (both $\bar{x} = .5$, $\mu = .52$). But for the 2C clutter condition, there was also a significant effect of *ordering strategy* on *noise identification* ($t(44) = 2.18$, $p = .05$). Participants working with DIS were more accurate ($\bar{x} = .4$, $\mu = .39$) in comparison to SIM ($\bar{x} = .33$, $\mu = .32$, $p < .05$).

Confidence

H1d. There was a significant effect of *clutter level* on *confidence* ($\chi^2(2, N = 270) = 28.816, p < .005$). Post-hoc tests revealed a higher confidence with 0C ($\bar{x} = 2$) compared to 2C (1, $p < .001$).

H2d. There is no significant effect between SIM (1) and DIS (1). While there is also no effect within the different clutter levels (0C: $2/2$, 1C: $1/1$, and 2C: $1/1$), there seems to be a tendency that participants are more confident with SIM without clutter dimensions and more confident with DIS with increasing clutter.

3.3.5 Quantitative User Feedback

Ordering Preferences. 11 out of 15 participants reported that they could see the clusters more clearly with dissimilarity reordering because they could use the orientation of the spikes as a determining factor. Some participants reported that they generally found the grouping tasks challenging, and they were not quite sure about the results. Interestingly, most of them said to have personal preferences towards the patterns with more smooth and convex shapes, namely the patterns produced by similarity reordering.

Similarity Estimation Strategies. The strategies reported by the participants can be grouped into three categories: (1) the majority of participants focused primarily on the spikes' orientation; (2) participants reported that they tried to find the center of a star glyph, and observe at which position of the glyph the center lies and how the gray area around the center is shaped; (3) a few participants searched for unique shape-parts and matched it with others.

3.4 Discussion

In summary, our study revealed two major findings.

Clutter Analysis. Clutter negatively influences the visual comparison of star glyphs. There is a significant drop in cluster quality, noise identification quality, and confidence with an increasing number of clutter dimensions. Also, task completion time changed considerably, although not statistically significant. Therefore, we can partially confirm our hypotheses **H1a – H1d**.

We expected these results as more clutter hampers similarity estimation in clustering tasks. As a result, cluster performance drops. While this is a general problem in information visualization [GLH15], it particularly affects star glyphs as clutter may change their shape considerably. Glyph designers should, therefore, think of using automatic algorithms to remove clutter dimensions, if possible.

Ordering Analysis. There are differences between the two evaluated ordering strategies. Generally, the quality of the clustering was significantly more accurate with `DIS`, in particular for datasets containing clutter (`1C`, `2C`). Participants also performed the task slightly faster using `DIS`. However, they were on average 10 seconds faster with `SIM` in non-cluttered datasets. We can see that `DIS` significantly supports noise identification for a cluttered dataset (`2C`), but we cannot see a difference for the other clutter conditions. While many participants reported that they prefer a dissimilarity-based layout, we cannot see a significant result from the study. However, analyzing the Likert-scale distributions reveal a tendency that participants are more confident with `SIM` for clutter-free datasets (`0C`) and with `DIS` for cluttered datasets (`2C`). Across all trials, we can confirm the hypotheses **H2b** and **H2c**, but completion time (**H2a**) and confidence (**H2d**) depend on the properties of the dataset.

These results are in line with Klippel et al. [KHW09; Kli+09]. We found it interesting that the difference between `SIM` and `DIS` is even more striking in cluttered datasets. The spikes seem to help users in identifying clutter dimension and improving the overall clusters. However, we could also see that, without clutter, participants were faster and more confident using a similarity-based ordering. The remaining question is whether it would be possible to combine `SIM` and `DIS` into a combined ordering strategy. Our study did not reveal whether participants need as many spikes as possible or whether a few important spikes are enough to improve the cluster quality. Further research needs to be done in this area. Another relevant question is also how the rotation of entire glyphs influences grouping quality in clustering tasks and further investigation in this direction is advisable as, for example, already started by Fuchs et al. [Fuc+14].

Design Considerations

Based on the study results, we derive the following design considerations:

(1) As the performance of users drop considerably when clutter dimensions are present, glyph designers should try to *avoid clutter by applying a feature selection* method first, if possible.

(2) Since, for *datasets with clutter*, salient shapes and spikes support grouping tasks, we recommend using `DIS` *strategies*.

(3) For *datasets without clutter*, we did not find a clear difference between `SIM` and `DIS`. As `SIM` seem to be slightly faster and less error prone to rotation [KHW09; Kli+09]. We recommend to use this strategy.

Limitations

We identified two main threats to our results' validity.

(1) The number of trials (270) is rather small, in particular, for the effectiveness and efficiency analysis of a specific clutter level. This affects not only the statistical analysis, but outliers may also distort the results. The number of trials per participant cannot be increased with the current study design; otherwise, the study would take much longer than one hour. Therefore, we suggest repeating the study with more participants to increase the number of trials. (2) While we designed our datasets with different cluster structures and distributions, we limited them by eight dimensions as Klippel et al. [KHW09]. There might be differences for datasets with less, more, or an odd number of dimensions.

3.5 Conclusion

We conducted an empirical user study to evaluate the impact of clutter and axes ordering to clustering performance with star glyphs. Our results show that users perform better when the glyphs represent salient shapes and spikes, which is achieved by a dissimilarity-based ordering of the dimensions. Furthermore, we elicited that there is a significant impact of clutter on the clustering performance in general.

As future work, we plan to extend and re-run the study based on our discussed limitations and include other reordering strategies, as well. Extending to that, we want to investigate whether there is an influence of the data characteristics and rotation (e.g., favor symmetrical glyph shapes) to the ordering strategy. If so, we are interested in developing techniques to select the most useful ordering strategy based on the given data and task. Finally, automatic ordering strategies should be compared to user-driven axes arrangements, which are determined by experts based on their domain knowledge.

4 Evaluating Ordering Strategies for Cluster Identification in Parallel Coordinates

Summary

The ability to perceive patterns in parallel coordinates plots (PCPs) is heavily influenced by the ordering of the dimensions. While the community has proposed over 30 automatic ordering strategies, we still lack empirical guidance for choosing an appropriate strategy for a given task. In this chapter, we first propose a classification of tasks and patterns and analyze which PCP reordering strategies help in detecting them. Based on our classification, we then conduct an empirical user study with 31 participants to evaluate reordering strategies for cluster identification tasks. We particularly measure time, identification quality, and the users' confidence for two different strategies using both synthetic and real-world datasets. Our results show that, somewhat unexpectedly, participants tend to focus on dissimilar rather than similar dimension pairs when detecting clusters, and are more confident in their answers. This is especially true when increasing the amount of clutter in the data. As a result of these findings, we propose a new reordering strategy based on the dissimilarity of neighboring dimension pairs.

This chapter is *taken from* the following publication. Please refer to Sections 1.4 and 1.5 for the contribution clarification and general citation rules.

[Blu+20b] **Michael Blumenschein**, Xuan Zhang, David Pomerenke, Daniel A. Keim, and Johannes Fuchs. *"Evaluating Reordering Strategies for Cluster Identification in Parallel Coordinates"*. In: Computer Graphics Forum 39.3 (2020), pp. 537–549.

4.1 Introduction

Parallel coordinates plots (PCPs) [Ins85; Ins09b] are a popular and well-researched technique to visualize multi-dimensional data. Dimensions are represented by vertical, equally spaced axes. Data records are encoded by polylines, connecting the respective values on each axis. PCPs have been applied to practical applications of various domains [JF16], such as finance [AZZ10], traffic safety [FWR99], and network analysis [Sto+05]. As discussed by Andrienko and Andrienko [AA01], PCPs are suited for a multitude of analysis tasks, such as cluster, correlation, and outlier analysis.

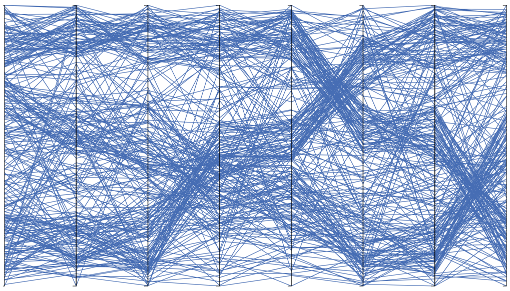
Compared to other visualizations for multi-dimensional data (e.g., RadVis [Hof+97], MDS and PCA projections, scatter plots, and scatter plot matrices), PCPs have the advantage to trace data records and patterns across a large set of dimensions. Empirical studies have shown that PCPs outperform scatter plots in clustering tasks, outlier, and change detection [KAC15], but are less suited for correlation analysis [LMW10; Har+14] and value retrieval [Kua+12].

A major challenge of visualizations is *visual clutter*, which influences the perception of visible patterns [Sun+15]. This problem is particularly given in PCPs, as line crossing and overplotting distort salient structures. Therefore, the community has proposed a multitude of enhancements, such as sampling [ED06a], edge bundling [MM08], interactive highlighting [MW95], and the usage of transparency [Joh+05] to reduce the impact of visual clutter.

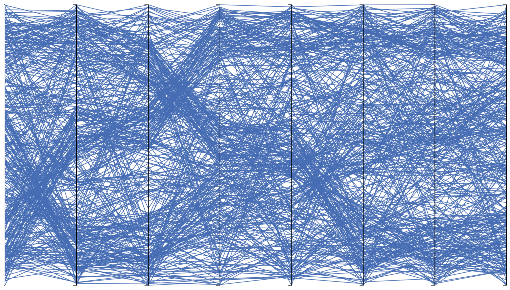
The ordering of axes plays a fundamental role in the design of a PCP and has a strong effect on the overall pattern structure [JJ09]. In contrast to data preprocessing, sampling, dimension filtering, and other enhancements, reordering does not remove data from the PCP [PWR04; PL17a], but changes the visual structure among neighboring axes. Depending on the user's analysis goal, some patterns are more interesting than others [DK10]. As a result, more than 30 different ordering strategies have been developed by the community to support a multitude of tasks. Some of these strategies group similar dimension pairs [ABK98], try to avoid line crossings [DK10], or put the most important dimensions first [Yan+03]. However, our community lacks empirical guidance and recommendations for choosing an appropriate strategy for a given task [Beh+18]. In this chapter, we address this limitation by summarizing the state-of-the-art in axes reordering strategies, as well as presenting a first empirical user study that measures the performance of two ordering methods for cluster identification in PCPs. Our study focuses on cluster analysis as the majority of ordering strategies are designed for this task.

We claim two main contributions. First, we provide guidance in selecting reordering strategies based on their intended patterns. Many existing algorithms follow similar concepts but differ in their implementation. To support users, we introduce a classification of the existing layout algorithms, group them according to their inner workings, and summarize their intended patterns and meta-characteristics. For more practical support, we implemented a set of 14 strategies in JavaScript and made them along with the source code available on our website for testing: http://subspace.dbvis.de/pcp.

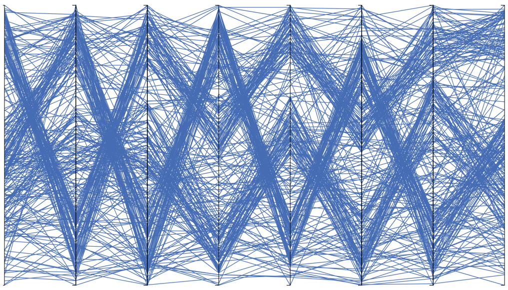
Second, we measure the performance of two reordering strategies for cluster identification tasks by an empirical user study with 31 participants. We realized that the often proposed similarity-based axes arrangement (e.g., [ABK98; Yan+03; AOL06]) is not always the most effective solution to identify clusters. As shown in Figure 4.1, arranging axes with a high dissimilarity next to each other produces more salient clusters, in particular in cluttered and noisy datasets. A reason for this effect is that lines with strong slopes are moving closer together, making clusters visually more prominent [Pom+19]. To find out whether this arrangement is more useful than a



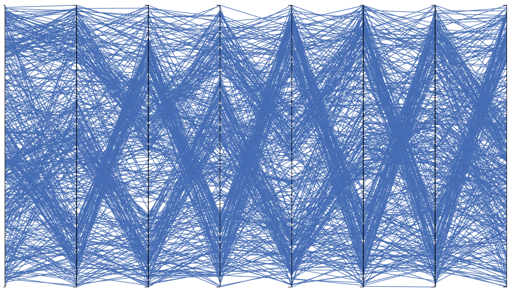
(a) Medium clutter, similarity-based layout



(b) High clutter, similarity-based layout



(c) Medium clutter, dissimilarity-based layout



(d) High clutter, dissimilarity-based layout

Fig. 4.1. A dataset with three clusters and two different clutter levels is sorted by *similarity* (a–b) and *dissimilarity* (c–d) of neighboring axes pairs. Clusters are more salient when arranging dissimilar dimensions next to each other. We show that in cluttered datasets, participants are more accurate and more confident when performing cluster identification tasks on such a layout. Figure adapted from [Blu+20b].

similarity-based layout, we conducted a user study and measured performance with respect to cluster quality, completion time, and users' confidence using synthetic and real-world datasets. Our results show that participants tend to focus on dissimilar axes pairs when selecting clusters and are more accurate and confident when doing so.

As a secondary contribution, we provide a benchmark dataset with 82 synthetic and real-world datasets for clustering analysis. For reproducibility, we make all our material and results, statistical analysis, and source code publicly available at https://osf.io/zwm69.

The remainder of this chapter is structured as follows: In the next section, we summarize important related work. Then, in Section 4.3, we survey existing reordering strategies for PCPs and classify them based on their intended patterns and inner workings (first contribution). Afterwards, in Section 4.4, we describe our user study design and report the statistical analysis results in Section 4.5 (second contribution). Finally, we discuss our findings and derive design considerations for axes orderings in cluster identification tasks.

4.2 Background and Related Work

In the following, we summarize the challenges of axes reordering, the results of existing user studies, and the relation of automatic ordering to interactive and semi-automatic analysis support.

4.2.1 Challenges of Axes Reordering of Parallel Coordinates

Linear ordering of an $n-$dimensional dataset in PCPs faces two main challenges. First, computing and evaluating all dimension permutations is computationally expensive. Ankerst et al. [ABK98] show that the ordering of axes according to some useful objective function is NP-complete. Therefore, the exhaustive search for a useful ordering is tedious, even for a modest number of dimensions [PWR04]. Second, the usefulness of a particular ordering highly depends on the analysis task of the user [DK10; PL17a], and is influenced by the complexity of the data [Tat+11]. More importantly, optimizing the axes ordering to highlight a particular pattern may even obstruct other patterns [JJ09], which are of relevance in a different scenario. Therefore, it is vital to carefully choose an appropriate strategy to arrange the axes in parallel coordinates.

More than 30 reordering strategies have been developed (see Section 4.3), many of which follow similar concepts but differ in their implementation affecting, for example, the runtime and quality of the results. Quality metrics and layout algorithms for PCPs have been summarized before: Heinrich and Weiskopf [HW13] give a comprehensive overview of the state-of-the-art for PCP research, including manual and automatic reordering approaches. Bertini et al. [BTK11] and Behrisch et al. [Beh+18] summarize quality metrics to optimize the visual representation. Ellis and Dix [ED07] discuss reordering from a clutter perspective. While Behrisch et al. [Beh+18] group the quality metrics by their analysis task, the literature still misses a summary of the different PCP patterns and a discussion on which reordering algorithms favor or avoid particular patterns [Tat+11]. We close this gap by introducing a classification along with a characterization of the reordering algorithms.

4.2.2 Evaluation of Axes Reorderings and Empirical Studies

There is a lack of empirical studies to measure the performance of specific axes orderings for different analysis tasks [JF16]. Most strategies are "evaluated" using examples of synthetic or real-world data (see Table 4.1) instead of comparing it to previous approaches. Exceptions are the works by Ferdosi & Roerdink [FR11] and Tatu et al. [Tat+11], which compare the resulting orders with competing approaches. However, no feedback from real users is provided.

Many reorderings claim to be useful for cluster analysis, but we do not know yet which patterns are most effective in identifying clusters. There is no user study that compares different reorderings for cluster identification in particular or different analysis tasks in general. Therefore, we want to push PCP reordering towards an empirically-driven research field by evaluating two axes reordering techniques for cluster identification. The works most closely related to ours are the empirical studies by Holten & van Wijk [HW10] (measuring response time and cluster identification correctness for nine PCP variations), Kanjanabose et al. [KAC15] (measuring response time and clustering accuracy in PCP, scatter plots, and classical tables), and the study by Johansson et al. [Joh+08] evaluating clutter threshold for the identification of patterns. However, none of these studies consider different axes orderings as an independent variable.

4.2.3 Relation to Interactive and Semi-Automatic Analysis

Besides axes reordering, countless enhancements have been developed to support the understanding of patterns in parallel coordinates. A comprehensive overview is out of the scope of this thesis but can be found in the surveys by Ellis & Dix [ED07], Bertini et al. [BTK11], Heinrich & Weiskopf [HW13], and Behrisch et al. [Beh+18]. Many techniques involve users within an interactive exploration workflow or combine the representation with automatic algorithms for pattern detection. Examples are the usage of clustering algorithms [FWR99; Joh+05; Mou11], automatic sampling techniques [ED06a], and interactive highlighting [MW95]. Inselberg [Ins09a], and Hurley & Oldford [HO10] propose to clone and arrange dimensions such that all pairwise permutations are visible, similar to a scatter plot matrix. Based on this initial view, the user can then start the exploration.

While the usefulness of such interactive and visual analytics approaches have been shown in many user studies, they are facing two challenges: First, most algorithms rely on sensitive parameters which influence the quality of the result. For example, k-means clustering [HKP11] requires the number of clusters as user input, which is typically unknown for a new dataset. Second, interactive exploration and highlighting are difficult if users do not know what they are searching for, and the initial configuration of a PCP does not show (parts of) interesting patterns. Often, this results in trial-and-error interactions, in which patterns are only detected 'by accident'. This is particularly true if the dataset contains a large number of dimensions, and relevant patterns only exist in smaller subspaces.

Therefore, we need methods to give analysts good starting conditions for their (interactive) analysis. Hereby, an important aspect is the arrangement of axes, as it significantly changes the visual patterns among neighboring axes [JJ09]. In this chapter, we provide a categorization of reordering algorithms and their intended patterns, which helps analysts to make an educated selection.

4.3 Classification of Ordering Strategies

This section addresses two questions: *Which patterns are emphasized by which reordering strategy?* And *which algorithms have been implemented to solve the reorder-*

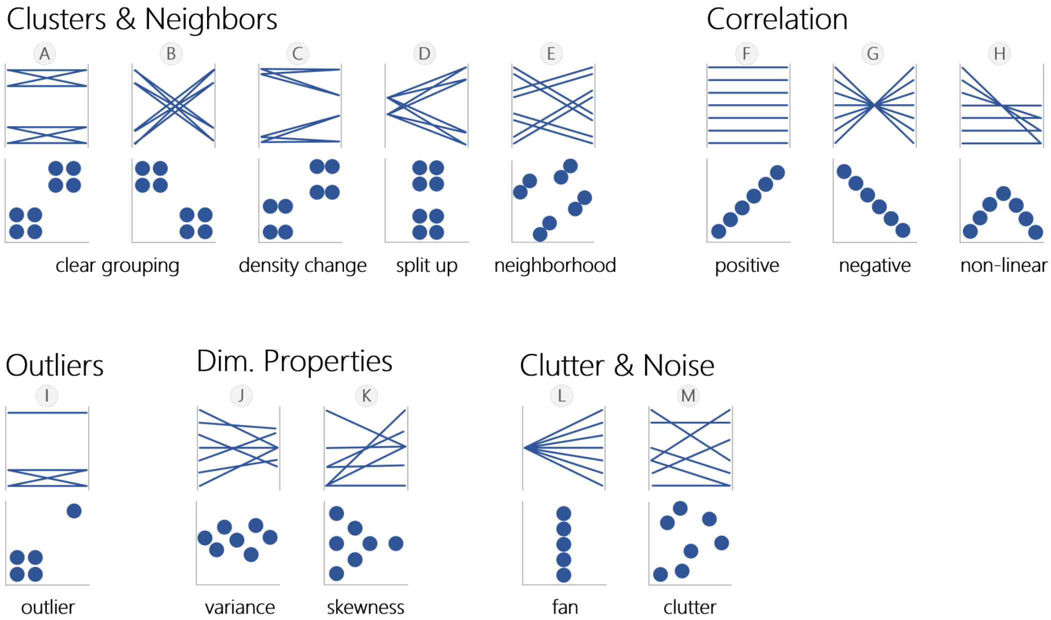


Fig. 4.2. Comparison of visual patterns in parallel coordinates and their scatter plot representation. Figure adapted from [Blu+20b].

ing problem? Before answering these questions, we provide an overview of important patterns.

4.3.1 Visual Patterns

Figure 4.2 shows five groups of the most common patterns in parallel coordinates and their representation in scatter plots:

Clusters & Neighbors Ⓐ – Ⓔ. Typical cluster structures show one or more groups of dense lines in a similar direction. While Ⓐ & Ⓑ seem similar in scatter plots, the visible structure in PCP differs significantly. Ⓒ shows clusters that change their density (cluster compactness) and Ⓓ, a cluster that splits up into sub-clusters.

Structures, preserving neighborhood information, are a special case of clusters. A (small) set of data records similar (close) to each other in one dimension are also similar in the neighboring dimension, which results in groups of parallel lines Ⓔ.

Correlations Ⓕ – Ⓗ. Positive and negative correlations look similar in a scatter plot. However, the PCP patterns differ: lines are parallel for positive Ⓕ, and in a star-like pattern for negative correlations Ⓖ. Variations of non-linear correlations may look different in both scatter plots and PCPs. Ⓗ shows only an example as the pattern depends on the type and degree of change in both dimensions.

Outliers Ⓘ. Outliers squeeze the majority of PCP lines together, resulting in a cluster-like pattern, hiding the underlying structure.

Dimension Properties Ⓙ – Ⓚ show patterns of dimensions, ordered by variance and skewness. The lines' slope indicates whether patterns stay consistent (parallel) or change across axes.

Clutter & Noise. Randomly distributed data without a clear pattern is considered noise or clutter *in the data*. The lines in the PCP cross without any particular structure Ⓜ. The fan pattern Ⓛ describes a cluster transitioning into clutter, a special case of density change Ⓒ.

Our selection of patterns is based on the work by Dasgupta & Kosara [DK10], Wegman [Weg90], Heinrich & Weiskopf [HW13; HW15], and Zhou et al. [Zho+08]. The cluster variations (i.e., patterns Ⓐ – Ⓓ) are inspired by Pattern Trails [Jäc+17], which introduces a taxonomy of pattern transitions between multi-dimensional subspaces. These patterns can be adapted to PCP, as two neighboring axes show a transition between one-dimensional subspaces. Finally, we add variance Ⓙ and skewness Ⓚ, which is produced by the algorithms described in [LHZ16; Sch+18; Yan+03]. We limit our patterns to 2D PCPs and ignore patterns in 3D PCPs (e.g., discussed in [PL17a; Ach+13]. We also consider only patterns among the two axes. Multi-dimensional patterns can be achieved by concatenating multiple two-dimensional patterns.

4.3.2 Ordering Strategies

To find ordering strategies, we took the 502 references of recent state-of-the-art reports [HW13; JF16; Beh+18] and combined it with 497 papers resulting from a search on the ACM, IEEE Xplore, EG, and DBLP digital library (see keywords and details in the supplementary material). We recursively scanned references and excluded papers that (1) did not propose an automatic axes ordering strategy (e.g., purely interactive approaches), (2) "just" apply a reordering method which has been proposed before, or (3) approaches which do not focus on "standard" parallel coordinates (e.g., 3D PCPs). Using this approach, we collected 18 papers with 32 different strategies.

Table 4.1 summarizes all reordering strategies, grouped by their ordering concept: strategies transforming the reordering into an *optimization problem* (Section 4.3.3), *implementing* efficient or sophisticated *algorithms* (Section 4.3.4), and approaches focusing on *properties* of *single dimensions* (Section 4.3.5). For each approach, we indicate the favored patterns, the involved axes, and the performed evaluation to show the usefulness (see caption of Table 4.1 for details).

Tab. 4.1. **Reordering Classification** summarizing the inner workings of reordering strategies for parallel coordinates. For each technique, we mark if it favors or avoids a particular pattern, if present in the data. Empty cells mean that the technique is not designed for this pattern and produces/avoids it primarily by change. We indicate the number of dimensions that are taken into account and mark the evaluation type used in the respective paper. Table adapted from [Blu+20b].

Patterns: favor ●, avoid ⊗ pattern, or depending on the algorithm's parameters ◓.
Axes considered for ordering: each dimension separately (∣), two neighboring dimensions (∥), or the majority of dimensions (∣∣∣∣).
Evaluation: case study or example Ⓔ, comparison of techniques ⬌, or empirical study 👤.

	Technique	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ	Ⓘ	Ⓙ	Ⓚ	Ⓛ	Ⓜ	Axes	Eval.
Objective Function & Optimization Algorithm	[Tat+11] no label	●	●		●									⊗	∥	⬌
	[Tat+11] label	●	●		◓									⊗	∥	⬌
	[Van15] consist.	●	●		⊗									⊗	∥	Ⓔ
	[Zho+18] trace	●	●	◓	⊗									⊗	∥	Ⓔ
	[PL17a] neighbor	●	●	◓	◓	●							⊗	⊗	∥	Ⓔ
	[DK10] overpl.	●	●	◓	◓		⊗								∥	Ⓔ
	[DK10] parall.		⊗				●	⊗						⊗	∥	Ⓔ
	[ABK98] similar	●	⊗				●	⊗						⊗	∥	Ⓔ
	[Yan+03] similar	●	⊗				●	⊗						⊗	∥	Ⓔ
	[Xia+12] interac.	●	⊗											⊗	∥	Ⓔ
	[ABK98] correl.	●	⊗				●	⊗						⊗	∥	Ⓔ
	[AOL06] clutter	◓	⊗				●	⊗					⊗	⊗	∣∣∣∣	Ⓔ
	★ **[Blu+20b] dis.**	⊗	●				⊗	●						⊗	∥	👤
	[DK10] cross.	⊗	◓				⊗	●						◓	∥	Ⓔ
	[DK10] angle	⊗	◓				⊗	●						◓	∥	Ⓔ
	[DK10] mutual i.						●	●	●					⊗	∥	Ⓔ
	[MTJ12] structure							●					●	●	∥	Ⓔ
	[PWR04] outlier									●					∥	Ⓔ
	[DK10] entropy													⊗	∥	Ⓔ
	[DK10] diverg.												●	⊗	∥	Ⓔ
Reordering Algorithm	[JJ09] cluster	●	●	●	●									⊗	∣∣∣∣	Ⓔ
	[FR11] subspace	●	●	●	●									⊗	∣∣∣∣	⬌
	[HHJ11] set theory	●	●	●										⊗	∥	Ⓔ
	[AOL06] similar	●	⊗				●	⊗						⊗	∥	Ⓔ
	[JJ09] correlation	●	⊗				●	⊗						⊗	∥	Ⓔ
	[LHH12] non-lin.						◓	◓	●					⊗	∥	Ⓔ
	[JJ09] outlier									●				⊗	∣∣∣∣	Ⓔ
Dimension QM	[Sch+18] class	●	●	●										⊗	∣	–
	[Sch+18] outlier									●				⊗	∣	–
	[LHZ16] svd										●	●	●	⊗	∣∣∣∣	Ⓔ
	[Yan+03] variance										●	●	●	⊗	∣	Ⓔ
	[Sch+18] skewn.											●		⊗	∣	–

4.3.3 Optimization Problem and Objective Functions

The largest group of reordering strategies transforms the axes arrangement into an optimization problem. These approaches measure the quality of a particular ordering by an objective function, which is then either minimized or maximized by an optimization algorithm.

Objective Functions Measuring Cluster Structures

Tatu et al. [Tat+09; Tat+11] argue that clusters consist of lines with a similar position and direction (patterns Ⓐ, Ⓑ, and Ⓓ). The authors take a rendered image of a PCP and apply a Hough space transformation [Hou62]. Each PCP line segment is mapped into one point within the Hough space. The point's location represents the position and slope of the line segment. The objective function measures dense areas (clusters) of points in the Hough space. Long [Van15] first computes a centroid for all given clusters. Then for each data record, the nearest centroid is identified (using the area between the lines as similarity function), and the objective function measures the ratio of correctly classified records. Cluster patterns Ⓐ and Ⓑ are highlighted, while a cluster split Ⓓ is avoided. Zhou et al. [Zho+18] aim at clusters that can be followed across neighboring axes (Ⓐ and Ⓑ). They compute a hierarchical clustering on every dimension and use the cluster similarity as quality. Dasgupta & Kosara [DK10] introduce seven different metrics, known as *pargnostics*. A metric aiming for clusters like Ⓐ and Ⓑ is *overplotting*. It measures the number of pixels that are not visible due to overlapping lines. When maximizing this measure, there is a high information loss, but a high-density of lines, i.e., clusters. Finally, Xiang et al. [Xia+12] try to avoid intersecting clusters Ⓑ by measuring the crossing of clusters among axes. This results in horizontal cluster structures Ⓐ.

Peltonen & Lin [PL17a] aim to preserve the neighborhood distribution of records (pattern Ⓔ). The objective function measures the similarity of nearest neighbors for all data records across two dimensions. Clusters like Ⓐ and Ⓑ are a special case of neighborhood relationships and are therefore considered as well.

Similarity-based Metrics for Clusters and Correlation

The main idea of the following approaches is to arrange similar dimensions next to each other. This results in cluster- and correlation patterns. The definition of similarity differs across the techniques. Ankerst et al. [ABK98] use a Euclidean distance and Pearson correlation for the similarity computation. The approach by Yang et al. [Yan+03] follows the same idea. However, they structure the dimensions into a hierarchy using a hierarchical clustering algorithm to highlight also clusters in subspaces of the dataset. The hierarchical structure also helps to speed up the computation time, as each subtree can be sorted independently. Depending on the similarity function, and whether the objective function is minimized or maximized, these approaches aim for the patterns Ⓐ, Ⓑ, Ⓕ, and Ⓖ.

Other metrics try to order axes such that lines are most parallel or diverging a lot. These patterns can help to identify correlations, but may also favor clusters to some extent. Artero et al. [AOL06] propose the total length of poly-lines as metric for pattern F. Similarly, there are four pargnostic [DK10] measures: (1) *Maximize the number of line crossings* to identify inverse relationships G in the data. (2) *Maximizing* G *the angle of crossings*. (3) *Maximizing parallelism*, resulting in less cluttered PCPs, which highlight positive correlations F. (4) *Maximizing the mutual information*, which measures the dependency between variables, i.e., optimizing for positive F, negative G, and non-linear correlations H.

Objective Functions for Clutter and Outliers

The pargnostic metric *maximizing divergence* results in fan pattern L, which helps to identify cluster-to-noise relationships. *Maximizing the entropy of neighboring axes* corresponds to a high information density, highlighting many line crossings and inverse relationships. The metric does not favor specific patterns, but results in busy, but very readable charts, according to the authors [DK10]. The metric by Makwana et al. [MTJ12] differs from previous metrics. The authors propose to order dimensions such that neighboring axes contain lines with different slopes, resulting in cluttered M PCPs.

Peng et al. [PWR04] interpret outliers as data points that do not belong to a cluster. They measure the ratio of outliers against the number of data points. When maximized, outliers are highlighted (pattern I), when minimized, outliers will not be highlighted.

Optimization Algorithms for Objective Functions

Except for [AOL06], all approaches measure the quality between neighboring axes and use the average as the objective function. To minimize or maximize this function, various heuristics are applied: Random swapping (particularly useful for very large datasets) [Yan+03; PWR04], Ant-optimization [ABK98], A*Search [Tat+09; Tat+11], Nearest-neighbor-based [PWR04], Branch and bound optimization [DK10; MTJ12], Non-linear optimization algorithm [PL17a], and Backtracking [Zho+18].

4.3.4 Reordering by Algorithms

The second class of strategies arranges dimensions based on layout algorithms. Compared to optimization procedures, this has two advantages: (1) Algorithms which approximate the understanding of an objective function, lead to more efficient but potentially less accurate results (e.g., [LHH12; AOL06; JJ09]). (2) Objective functions are typically defined only between neighboring axes. Using more advanced algorithms (e.g., based on subspace clustering) lead to PCP, which aims for higher-dimensional patterns [FR11].

Algorithms for Efficient Reordering

Artero et al. [AOL06] and Johansson & Johansson [JJ09] speed up the similarity and correlation-based axes arrangement, originally presented by Ankerst et al. [ABK98]. Both algorithms are identical, except for the similarity function. Artero et al. use a Euclidean distance, Johansson & Johansson, a Pearson correlation coefficient. The algorithm starts with the most similar dimension pair in the center of the PCP. Iteratively, the next most similar dimension is appended to the left or right side. While this approach is efficient, it also has the advantage that the most salient structure (the most similar dimensions) typically ends up close to the center of the PCP, which users are most attracted to [Net+17]. Lu et al.'s approach [LHH12] orders dimensions based on correlation. They use the nonlinear correlation coefficient (NCC), which is sensitive to any relationship Ⓕ–Ⓗ (not only linear ones) and can be used for partial similarity matching as well [ABK98]. The proposed algorithm combines the ordering by (non-linear) correlations together with an importance driven arrangement. The first (left) axis in the PCP is chosen based on the highest singular value after a singular value decomposition (SVD, highest contribution of the dataset). Afterwards, all dimensions are arranged from left to right according to their similarity of the NCC. The approach by Huang et al. [HHJ11] maximizes the uniform line crossings of clusters. Their approach is based on Rough Set Theory [Paw12], and the algorithm sorts the dimensions based on alternating sizes of high and low cardinality of the equivalence classes, leading to cluster patterns Ⓐ–Ⓒ.

Subspace Algorithms for Higher-dimensional Structures

Ferdosi & Roerdink [FR11] use a subspace search algorithm [Fer+10] to identify higher-dimensional clusters with patterns Ⓐ–Ⓓ. The quality of one subspace is based on a density distribution. Subspaces containing multiple clusters that are clearly separated are considered of high quality. First, the algorithm computes all one-dimensional subspaces and arranges the one with the highest quality on the very left of the PCP. Afterwards, all two-dimensional subspaces, which contain the first subspace, are computed, and the highest rank is attached as the second axis. The algorithm continues until all dimensions are part of the PCP, or no more subspace can be computed. Johansson & Johansson [JJ09] apply the MAFIA algorithm [NGC01], resulting in a set of subspaces, along with cluster structures and quality measures. The ordering algorithm then finds the longest sequence of connected variables shared by all detected subspace clusters. It starts with all dimensions of the first subspace (no specific ordering). Further subspaces are iteratively added based on their quality, but only if they share a substantial set of dimensions with the current PCP. The authors use the same algorithm to identify patterns with (multi-dimensional) outliers (pattern Ⓘ).

4.3.5 Reordering by Dimension-wise Quality Metrics

The third group of reordering techniques computes a quality for each dimension separately ( ) and sort the axes accordingly. Assuming the quality can be computed efficiently, reordering can be done in linear time. The techniques can also be extended by dimension filtering (e.g., considering only dimensions with a quality above a threshold). Relations *between dimensions* are not considered. Therefore, patterns may be scattered in different parts of the PCP [PL17a].

Lu et al. [LHZ16] sort the axes based on each of their contributions to the dataset. They compute an SVD and sort the dimensions according to their singular values. Yang et al. [Yan+03] propose a similar approach but sort the dimensions by variance. Both reorderings result in similar patterns (Ⓙ – Ⓛ). However, Lu et al.'s approach takes the distribution of the entire dataset into account.

Schloerke et al. [Sch+18] propose three different dimension metrics: (1) They use skewness for reordering, resulting in a Ⓚ pattern. (2) The dimensions are sorted by one of the Scagnostics [WAG05] measures. In particular, the *Outlying* measure is useful to highlight outliers in the data (pattern Ⓘ). (3) Finally, the authors order the dimensions such that existing clusters or classes are separated as well as possible. They compute an ANOVA on every dimension based on a given set of class labels and order the dimensions based on the F-statistic. Intuitively, the dimensions are ordered according to how well the given clusters are separated (patterns Ⓐ – Ⓒ).

4.3.6 Summary

In Table 4.1, we provide an overview of 32 different reordering approaches to arrange the axes of parallel coordinates. During our analysis, we made a few observations: (1) Many reordering algorithms follow similar concepts, but differ in their implementation and the applied metric. The main reason for this is that axes reordering is computationally complex, and more efficient approaches are necessary for interactive applications. (2) There seems to be a different understanding of the most important area in a PCP. While some reordering approaches try to put the most important dimensions upfront, others try to arrange them in the center. This is in line with the study by Netzel et al. [Net+17], who found out that people pay the most attention to the center part of a PCP. (3) The evaluation of novel reordering algorithms is primarily achieved by use cases and example applications. We are not aware of empirical user studies that compare different orderings for a particular analysis task. We want to close this gap and provide the first empirical study to evaluate ordering approaches for a particular analysis task.

4.4 User Study for Cluster Identification

Our reordering classification in Table 4.1 reveals that the majority of strategies are designed to support cluster analysis. Therefore, we select this task as the focus of our user study. In particular, we want to assess the performance of *cluster identification*, as this is the foundation for more sophisticated clustering analyses.

For cluster analysis, *similarity-based layouts* are proposed most often. Clusters, if present, can be followed across many axes, as algorithms try to minimize their variance. However, this strategy does not necessarily highlight clusters. This is especially true if the dataset contains noise or clutter, as shown in Figure 4.1a. While we can identify the clusters, they are visually less salient. We define the term *clutter* as data records that do not contribute to a particular pattern (e.g., randomly distributed), often also called *noise*. Cluttered datasets often end up in *visual cluttered* PCP due to many line crossings and overplotting.

Due to experiments with our implemented reordering algorithms, we realized that polylines and clusters with *strong slopes* are visually more prominent than horizontal ones. There are two reasons, as discussed by Pomerenke et al. [Pom+19]: (1) With an increasing slope, the distance between polylines decreases, and less whitespace (background) is visible. Hence, neighboring lines have higher contrast. (2) Compared to horizontal lines, diagonal lines need more pixels to encode a single data point, resulting in a low data-to-ink-ratio [Tuf01]. Both geometric effects make sure that neighboring lines are more easily perceived as a group or cluster. Interestingly, strong slopes are produced when dimensions are ordered by *dissimilarity*. An example can be found in Figure 4.1c. It shows the same data as in 4.1a, but with strong slopes due to reordering. Often this results in a zig-zag-like pattern, which makes the visual representation more complex but also ends up in more salient cluster structures.

4.4.1 Hypotheses

We address the question, *'whether there is a difference between a similarity-based* (`SIM`) *and dissimilarity-based* (`DIS`) *axes ordering for a cluster identification task'*. If yes, *'which ordering should be used, and why?'* As the majority of real-world datasets contain noise and clutter, we also want to investigate its influence generally, and in combination with the axes ordering. Hence, we analyze two independent variables: *ordering method* and *clutter level*.

To measure the performance, we use three dependent variables: (i) *time* to identify clusters, (ii) quality of manually *selected clusters* based on similarity to ground truth clusters, and (iii) the *confidence* of the users after the cluster identification. Additionally, we analyze axes-pairs which help to identify the clusters. In particular, we investigate whether users select clusters in similar or dissimilar axes pairs. For our study, we formulate the following three hypotheses:

H1. With an *increasing amount of clutter*, the cluster identification performance drops (independent of the ordering) as cluster structures are less salient in the PCP plot. We expect users to be **(a)** slower, **(b)** less accurate, and **(c)** less confident.

H2. *Without clutter*, users perform better in a cluster identification task when the axes are ordered by `SIM` instead of `DIS` as clusters can be followed more easily.

In particular, we expect users to be **(a)** faster, **(b)** more accurate, and **(c)** more confident with `SIM`.

H3. *With clutter*, users perform better in a cluster identification task when the axes are ordered by `DIS` instead of `SIM` as clusters are visually more prominent. In particular, we expect users to be **(a)** faster, **(b)** more accurate, and **(c)** more confident with `DIS`.

4.4.2 Benchmark Dataset and Ground Truth

To evaluate our hypotheses, benchmark datasets with ground truth information and increasing clutter levels are needed. We are not aware of such datasets for a cluster identification task. Therefore, we developed our own benchmark, consisting of ten popular real-world, and 72 synthetically created datasets along with the ground truth information. We make our dataset available in order to overcome the limitation of publicly available benchmark datasets [Sch+16] and to support the evaluation of PCP enhancements and reordering techniques in the future. For comparison, we show a PCP with each dataset and ordering strategy in the supplementary material.

Synthetic datasets

We limit the dimensionality of all synthetic datasets to eight. This allows us to create complex cluster structures while keeping the expected time for the study in a reasonable time frame. We alternated the number of clusters between one and four and varied the structures of the clusters – ranging from linear clusters towards a high variance on all scales of the different axes.

Using the PCDC tool [Bre+12], we created 24 *base datasets* $(\{1,2,3,4\}\ clusters \times 6\ variations)$ which fulfill the following properties: (i) clusters are clearly visible and separated from each other, (ii) there is only one clustering result per dataset, (iii) each cluster is present in all eight dimensions, and (iv) no outliers are added as they would distort the existing patterns [AA01]. In up to two dimensions, we merged two or more clusters such that participants need to investigate all dimensions to identify a cluster. To make the clusters comparable across datasets, we kept the cluster size constant with small randomization in the range of $45-50$ data points and vary the diameter of every cluster in each dimension randomly in the range $0.15-0.30$. All dimensions are normalized in $0.0-1.0$.

Next, we designed different clutter levels. Pomerenke et al. [Pom+19] show that random clutter (randomly and equally distributed records in all axes) produces visible patterns in PCPs which look similar to clusters (*Ghost clusters*). In order to not accidentally include 'fake patterns' in our dataset, but also be fair w.r.t. random clutter, we use a mixture of 30% random and 70% linear clutter (for every record: uniform and random distribution in one dimension $\pm\ 0-0.15$ in all other dimensions). Using several pilot experiments, this setting seemed to be complex enough, but also without any undesired patterns. For each base dataset, we created two copies with different clutter levels, one with 150 (`150N`), one with 300 data points (300N). We used the same clutter datasets for all base datasets to make them comparable and ensure we do not encode additional patterns in some of the datasets. After finalizing all 72 datasets (24 base datasets $\times\{0\text{N}, 150\text{N}, 300\text{N}\}$), we randomized the order of the records to remove potential effects in the drawing process.

Real-world datasets

We added ten frequently used datasets to see the performance in real settings. We selected the datasets based on common usage in PCP reordering (i.e., by choosing datasets used in the techniques described in Table 4.1). Hence, the number of dimensions and records differ compared to synthetic datasets. Dimensions range between $4 - 13$, and the number of records between $32 - 515$. Examples are the wine, mt-cars, and ecoli dataset. If present, we removed categorical dimensions and outliers. We used Ward's method [War63] to retrieve a hierarchical clustering and a visual inspection to determine the clusters in the data.

4.4.3 Implementation

To compare the 'optimal' SIM and DIS layout, we used Ankerst et al.'s reordering algorithm [ABK98] with an exhaustive search to find the axes ordering. For the SIM layout, we used the Euclidean distance and minimized the sum of distances. For DIS, we used the same algorithm but maximized the distances. We pre-computed the orderings for all datasets in advance. To run the study, we developed a web application that is available at http://subspace.dbvis.de/pcp-study. The parallel coordinates plots have a size of 960×500 pixels and use color for the polylines to separate them from the axes which are colored in black. We did not add any design variations to the chart (i.e., transparency, or edge bundling) to avoid confounding factors.

4.4.4 Tasks and Data Randomization

Our study consisted of 21 trials per participant, which are grouped into three tasks that build on top of each other. The tasks were executed in increasing difficulty: Tasks 1, 2, and 3. Between two trials of a task, we showed a white screen with the term '*break time*', and participants had to click a button to continue with the next trial.

Task 1 (Similarity of Axes-Pairs)

We wanted to find out which *visual structures* support users in a cluster identification task. In particular, we were interested whether users find neighboring axes with a high similarity or dissimilarity more useful. In each trial, we showed the participants a PCP in which the number of clusters had to be counted. Users selected the number which they identified using four radio buttons (i.e., 1, 2, 3, 4) and a *can't tell* option.

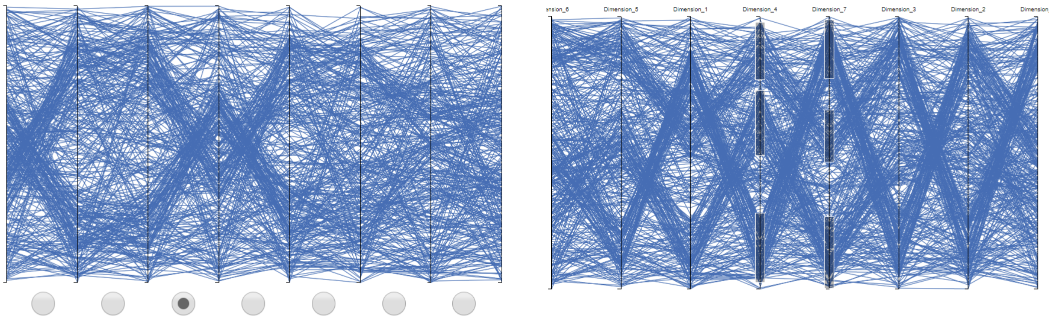


Fig. 4.3. User study interface for Task 1 (left) and Task 2 (right).

After the selection was confirmed, we showed a single radio button between each neighboring axes (see Figure 4.3 left) and asked the participants to select the pair which supported them best. Only one pair could be selected.

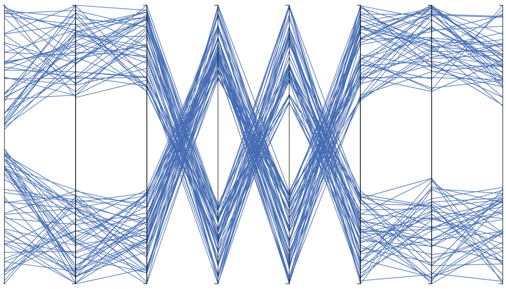
Randomization. We randomly picked three synthetic datasets with a different number of clusters. In the first trial, we showed `0N` clutter, in the second `150N`, and finally `300N` (increasing difficulty). As we are interested in whether participants prefer (dis-)similar neighboring axes, we arranged dimensions such that the PCP contains both similar and dissimilar axes pairs. To do so, we computed the similarity of neighboring axes using the Euclidean distance and used the maximal variance of similarities (`MaxVar`) as an objective function (see example in Figure 4.4). In summary:

3	levels of clutter (`0N`, `150N`, `300N`)	$\times$
31	participants	$=$
93	**trials in total**	

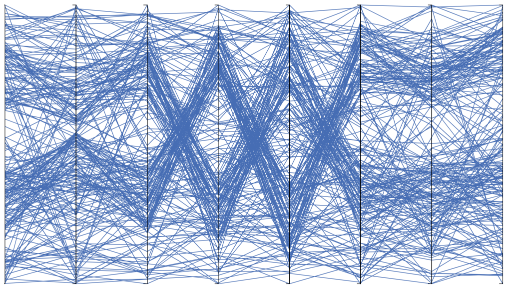
Post-processing. We collected the time to identify the number of clusters along with the similarity value of the selected axes pair. For comparison across datasets, we applied a linear min-max normalization to the similarity values of all neighboring axes pairs within each dataset. Pairs with the highest similarity are represented with 0.0, while high-dissimilarity pairs are represented by a value close to 1.0.

Task 2 (Cluster Identification and Selection)

We wanted to find out if participants are *better and more confident using a particular ordering strategy*. In each trial, we presented the participant one PCP, which was sorted by either `SIM` or `DIS`. The participant had to mark all clusters by choosing a pair of neighboring axes and marking every cluster in both axes using a brush feature (Figure 4.3 right). Brushing is applied by pressing the mouse button and marking the cluster along the axis. The selection can be moved, resized, or deleted. We do not highlight any data lines during or after brushing. After confirming their selections, participants rated their confidence in a correct clustering on a 5-point Likert scale.



(a) no clutter (0N)



(b) 150 clutter points (150N)

Fig. 4.4. Dimensions are ordered by MAXVAR (maximizing the variance of similarities among neighboring axes). The result combines similar and dissimilar dimension pairs in one PCP.

Randomization. We selected 12 synthetic base datasets, three for each number of clusters, and randomized the order. Then, we distributed the datasets into three equal-sized groups: `0N`, `150N`, and `300N`. Finally, we added four randomly selected real-world datasets in a new group `RW`. Within each group, we randomly applied twice a `SIM` and twice a `DIS` ordering. Participants worked on each group in order of increasing difficulty (i.e., clutter level). In summary:

4	levels of clutter (`0N`, `150N`, `300N`, `RW`)	×
2	repetitions	×
2	ordering strategies (`SIM`, `DIS`)	×
31	participants	=
496	**trials in total**	

Post-processing. We collected the time to mark the clusters and divided this by the number of clusters to be comparable across datasets. We also collected the selections and confidence levels.

We ignored clutter for the quality computation. We checked whether participants selected clusters in two neighboring dimensions and whether the number of clusters is therein consistent. In 132 trials, this was not the case, and we removed them from the data. The results are, however, still trustworthy as the removed trials are not skewed towards a particular reordering (66 trials each) or a clutter level (32, 28, 28, and 44 trials). For all correct trials, we then mapped the clusters between the selected axes together. First, we compute the Overlap coefficient [Szy34] between all cluster combinations and then merge the clusters with the highest overlap together. For each cluster combination, we keep the intersected set of data records as cluster members. The Overlap coefficient measures the overlap of members of the two clusters C_i and C_j: $overlap(C_i, C_j) = |C_i \cap C_j| / min(|C_i|, |C_j|)$.

The quality of an entire clustering is based on the Jaccard index [Jac01] between each selected cluster C_i and the corresponding ground truth cluster G_i. The Jaccard index measures the similarity between the clusters (record sets) C_i and G_i on a data record level: $jaccard(C_i, G_i) = |C_i \cap G_i| / |C_i \cup G_i|$. As participants can also select too few or too many clusters, our quality computation is a two-step process: First, we compute the average Jaccard index of each cluster to their best match in the ground truth. Second, we compute the average Jaccard index of every ground truth cluster to their best match of our selection. Our final clustering quality is then the average score of both steps.

Task 3 (Ordering Strategy Preferences)

We wanted to find out if participants have *preferences for a particular reordering* and why this is the case. We presented them two PCPs with the same dataset next to each other – one with `SIM`, one with `DIS` ordering. Using two radio buttons, participants had to select the preferred plot and then explain their choice in a free-text field.

Randomization. We randomly picked two synthetic datasets with a different number of clusters. In the first trial, we did not show any clutter (`0N`); in the second, we used either `150N` or `300N` (equally balanced across the participants). In the first trial, we used `SIM` ordering in the left, and `DIS` ordering in the right plot. In the second trial, we swapped the positions. In summary:

2	levels of clutter (`0N`, (`150N` $\vee$ `300N`))	$\times$
31	participants	=
62	**trials in total**	

Post-processing. We stored the preferred ordering and the text for each trial. Four participants reported not to see any preference between the options in one of the trials. We removed these participants from the statistical analysis, but report their choices in Section 4.5.3.

4.4.5 Participants and Procedure

Prior to the study, we conducted several pilot runs in order to determine appropriate clutter levels and the number of trials for each task.

Participants. To have participants with basic knowledge in information visualization and parallel coordinates, we conducted our user study during two lectures at the University of Konstanz, Germany. Both courses teach foundations in information visualization, one course for undergraduates, the other for graduates. The courses were taught by the same lecturer (not the authors), who also introduced and discussed the PCP technique two weeks prior to the study. We recruited 31 participants (17 male, 13 female, 1 NA). Their ages ranged from 19–31 years (median age 23). Each participant had finished high school, and 17 held a Bachelor's degree. The academic background was in the area of data analysis with 24 computer science, and 7 social and economic data analysis students. All participants reported having normal or corrected to normal vision.

Training and Procedure. All participants had to fill out a data privacy form in which we describe the data collected during the study. The participants sat scattered across the room and were not able to talk to each other. One of the authors started the study with a 30-minute recap on PCPs and comparing its visual patterns with scatter plots, discussing the advantages and disadvantages of the two techniques, and arguing about the effects of clutter, noise, and outliers. During the training, we did not provide strategies on how to identify clusters (or any other pattern) in PCP. Instead, we showed patterns in scatter plots and let all participants draw the respective patterns in a PCP (see training material). After this recap, we started the training. All participants opened their laptops and used a browser of their choice to access our online study. We provided a training platform, including all three tasks, but only two trials per task. We explained to the participants how to interact with the tool, and let the participants play around with the different trials. After answering the remaining questions, we made sure that all participants activated the full-screen mode within their browser and checked that the entire study could be conducted without scrolling for the different tasks. Participants needed between 20–30 minutes to complete the study.

4.5 User Study Results

We now report the summary statistics and highlight significant results ($p < .05$) in the data. For all tests, we checked the necessary preconditions, which can be found in the supplementary material along with the R scripts to reproduce the results. We used a one-sample Kolmogorov-Smirnov test to check if the data follows a normal distribution and Mauchly's test to check for sphericity.

4.5.1 Task 1 (Similarity of Axes-Pairs)

We used a repeated-measures ANOVA for the analysis of *completion time*. The post hoc analysis was done with a Bonferroni corrected t-test for dependent samples. As the *similarity of axes-pairs* was not normally distributed, a non-parametric Friedman's test was used.

Efficiency to Identify the Number of Clusters

There was a significant effect of *clutter* on *completion time* ($F(2, 60) = 6.07$, $p < .01$, $\eta^2 = .10$). Post hoc comparisons revealed that *completion time* was significantly lower for the *clutter* condition `0N` ($\mu = 9.44s$) compared to `150N` ($p < .01$, $\mu = 16.67s$), and `300N` ($p < .01$, $\mu = 17.47s$), but not between `150N` and `300N` ($p = 1.0$).

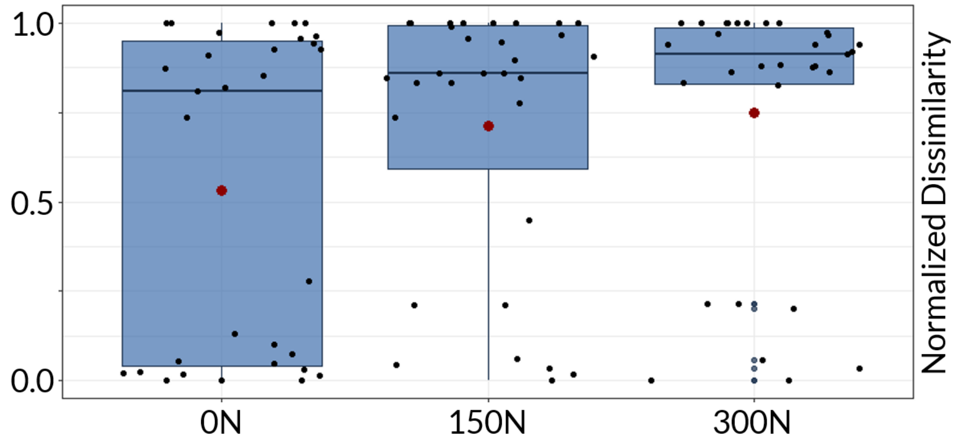


Fig. 4.5. Similarity of preferred neighboring dimensions (Task 1). With clutter, axes pairs with dissimilarity are preferred. Without clutter (0N), the preferences are almost equally balanced.

Similarity of Selected Axes-Pair

No significant results can be reported ($\chi^2(2) = 4.77$, $p = .09$). As shown in Figure 4.5, the mean of the distances for the different *clutter conditions* were 0N ($\mu = .53$), 150N ($.71$), and 300N ($.75$).

4.5.2 Task 2 (Cluster Identification and Selection)

For the comparison between *clutter* levels (independent of the ordering), we used a Kruskal-Wallis test, and a Bonferroni corrected Wilcoxon signed-rank test for post hoc analysis. For the *confidence*, we applied a Pearson's Chi-square test. To analyze the differences between the *ordering strategies* within each clutter level, we used a Wilcoxon signed-rank test for the analysis of *completion time*, *cluster quality*, and *confidence*. Data were split according to levels of *clutter* to compare the differences between SIM and DIS.

Efficiency to Identify and Mark Clusters

Between clutter levels, the medians of *completion time* for 0N, 150N, 300N, and RW were 10.79, 9.53, 10.26, and 15.02, respectively. A Kruskal-Wallis test showed a significant effect on *clutter* level ($\chi^2(3) = 23.31$, $p < .001$). A post hoc test using Wilcoxon signed-rank tests showed only significant differences between RW and 0N, 150N, and 300N (all $p < .01$).

As shown in Figure 4.6, the medians of the *completion time* for the 0N *clutter* condition, for SIM and DIS were $8.9s$ and $12.35s$, respectively. A Wilcoxon signed-rank test showed that there was a significant effect of *ordering* strategy ($W = 1$, $Z = -2.46$, $p < .05$, $r = .25$). For the other *clutter* conditions, no significant results can be reported. For the 150N *clutter* condition, the medians of *completion time*

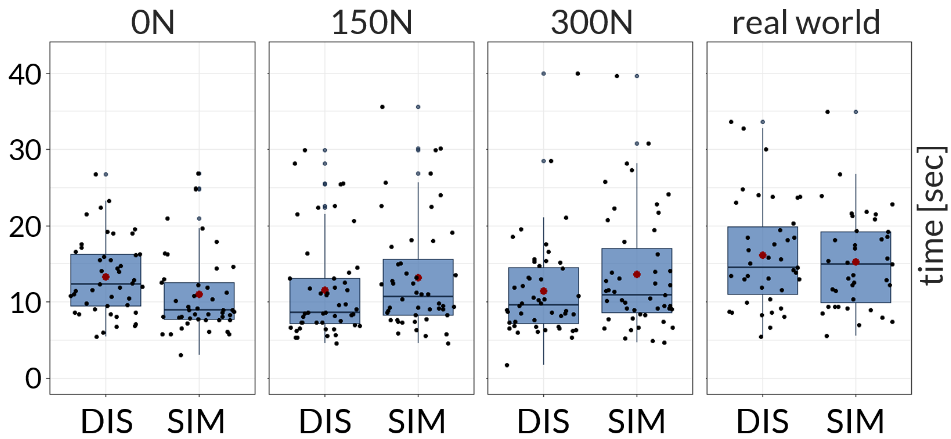


Fig. 4.6. Time to select clusters (Task 2).

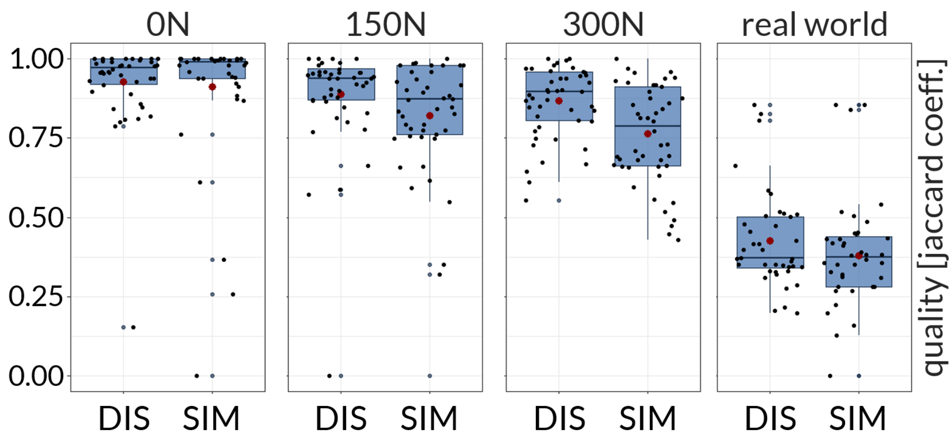


Fig. 4.7. Quality of selected clusters (Task 2).

for `SIM` and `DIS` were $10.67s$ and $8.58s$, respectively ($p = .05$), for the `300N` *clutter* condition, the medians of *completion time* for `SIM` and `DIS` were $11.12s$ and $9.62s$, respectively ($p = .05$), and for the `RW` condition, the medians of *completion time* for `SIM` and `DIS` were $15.07s$ and $14.79s$, respectively ($p = .82$).

Quality of Identified and Marked Clusters

Between *clutter* levels, the medians of *quality* for `0N`, `150N`, `300N`, and `RW` were $.98$, $.91$, $.85$ and $.37$, respectively. A Kruskal-Wallis test showed a significant effect on *clutter* level ($\chi^2(3) = 181.56$, $p < .001$). A post hoc test using Wilcoxon Sign-rank tests showed the significant differences between `0N` and `150N` ($p < .001$), `300N` ($p < .001$), and `RW` ($p < .001$). Also, there were significant effects between `150N` and `300N` ($p < .05$), and `RW` ($p < .001$). Finally, `300N` and `RW` were also significantly different ($p < .001$).

The results of the cluster quality are summarized in Figure 4.7. For the `150N` *clutter* condition, the medians of *quality* for `SIM` and `DIS` were $.87$ and $.94$, respectively. A

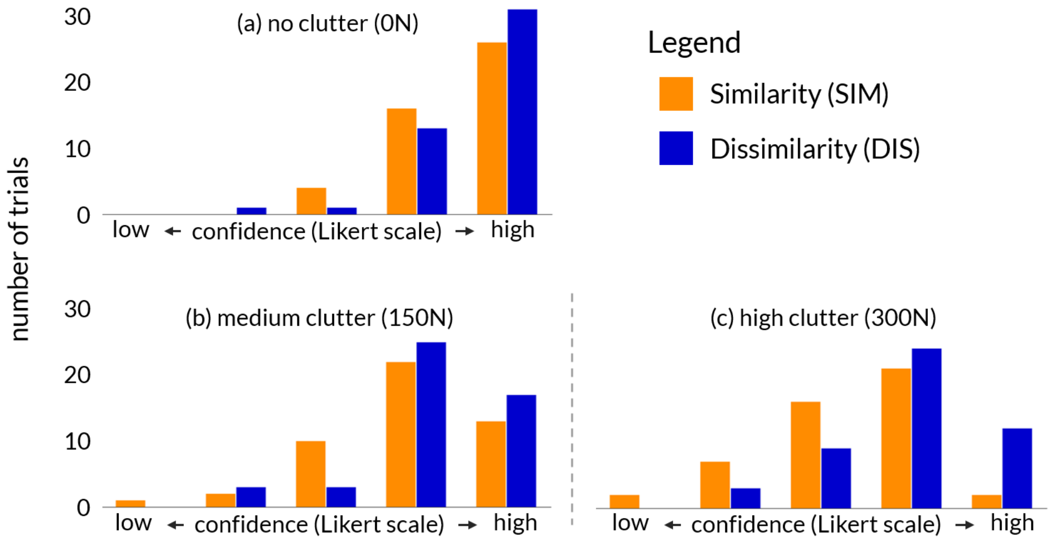


Fig. 4.8. Confidence of marked clusters (Task 2). Participants have higher confidence in their cluster selection using dissimilarity ordering. Figure adapted from [Blu+20b].

Wilcoxon signed-rank test showed a significant effect of *ordering* strategy (W $= 1$, Z $= -2.62$, $p < .001$, $r = .27$). For the `300N` *clutter* condition, the medians of *quality* for `SIM` and `DIS` were $.79$ and $.90$, respectively. A Wilcoxon signed-rank test showed a significant effect of *ordering strategy* (W $= 1$, Z $= -3.36$, $p < .001$, $r = .34$). The other levels of *clutter* did not show a significant difference. The medians of the quality score in the `0N` *clutter* condition were $.99$ for `SIM` and $.97$ for `DIS` ($p = .33$), and in the `RW` condition $.37$ for both `SIM` and `DIS` ($p = .25$).

Confidence of Marked Clusters

An overview of the participants' confidence is shown in Figure 4.8. Between *clutter levels*, the medians of *confidence* for `0N`, `150N`, `300N`, and `RW` were 2, 1, 1, and 0, respectively. A Pearson Chi-square test showed a significant effect of *clutter* level on *confidence* ($\chi^2(12) = 120.97$, $p < .001$). Post hoc analysis revealed significant differences between `0N` and `150N` ($p < .001$), `300N` ($p < .001$), and `RW` ($p < .001$); between `150N` and `300N` ($p < .05$), and `RW` ($p < .001$); and between `300N` and `RW` ($p < .005$).

For the `150N` *clutter* condition, the medians of *confidence* for `SIM` and `DIS` were both 1. A Wilcoxon signed-rank test showed a significant effect of *ordering* strategy ($W = 1$, $Z = -2.52$, $p < .05$, $r = .26$). For the `300N` *clutter* condition, the medians of *confidence* for `SIM` and `DIS` were 0 and 1, respectively. A Wilcoxon signed-rank test showed a significant effect of *ordering* strategy ($W = 1$, $Z = -3.75$, $p < .001$, $r = .38$). The remaining levels of *clutter* did not show a significant difference between *ordering* strategies with the same medians for `SIM` and `DIS` (`0N` = 2, $p = .25$; `RW` = 0, $p = .48$).

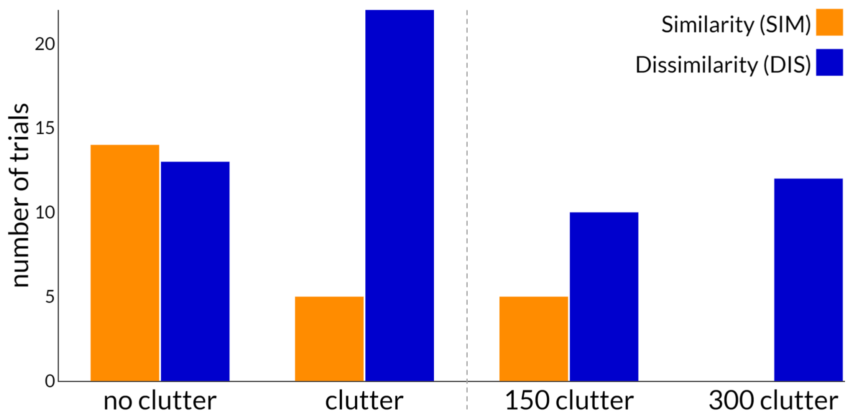


Fig. 4.9. Preference of reordering strategy (Task 3). No preference for datasets without clutter. Participants strongly preferred a dissimilarity-based layout with an increasing amount of clutter. The figure shows both clutter levels combined ('clutter') and separately (150 clutter vs. 300 clutter).

4.5.3 Task 3 (Understanding Preferences)

The distribution of preferences is shown in Figure 4.9. Two participants selected `SIM`, ten participants `DIS` for both *clutter* conditions. Twelve participants preferred `SIM` without *clutter* and changed their preference to `DIS` for the second trial, which included *clutter*. Vice versa, three participants changed from `DIS` to `SIM`. A binomial test showed a significant difference ($p < .05$) in the proportion of *preference* ($\chi^2(1, N = 27) = 12$). The probability of success was $.8$.

Three out of four of the removed participants (see Section 4.4.4) did not have a preference for `ON`, but preferred `DIS` for cluttered datasets. One participant preferred `SIM` for clutter-free datasets and had no preference for the dataset with clutter.

4.6 Discussion

We discuss the participants' performance according to our hypotheses and highlight findings from the qualitative feedback of task 3.

Influence of Clutter (H1)

Increasing the amount of clutter has a negative effect on the *quality* of the cluster identification and *confidence* of participants. These findings confirm **H1 (b)** and **(c)**. While we see an increasing *completion time* for higher clutter levels in task 1, we cannot verify this finding in the second task. Hence, we cannot make a final judgment on **H1 (a)**. As expected, clutter negatively influences the cluster identification. Patterns may vanish due to overlapping data lines, making the identification more difficult. Therefore, visualization experts need to carefully design PCPs and reduce the amount of clutter if possible (e.g., sampling).

Reordering for Clutter-free Datasets (H2)

Similarity-based ordering strategies are a good choice for datasets without clutter. Our results show that participants perform the identification of clusters more efficiently when working with a SIM layout, which confirms **H2 (a)**. It seems as if participants are faster in combining straight data lines into clusters in contrast to data lines with strong slopes. A possible explanation could be the Gestalt law [Wer23; War20] of *continuation*, which could help participants in tracking data lines across dimensions. Kellman and Shipley [KS91] support this argument: the angular parameters, determining the grouping of lines to clusters, may support the ability to find clusters across multiple sets of axes. The qualitative feedback also confirms our findings. Participants reported, for example, that *"The structure of clusters is clearer"*, *"[clusters] don't cross very often"*, or they prefer SIM *"[...] since they do not intersect with each other in the majority of areas between each two dimensions [...]"*. We cannot support hypotheses **H2 (b)** and **(c)**. There are no significant differences in the cluster identification *quality* or the *confidence* of the participants (see also Figure 4.7 and 4.8). Also, participants did not have a subjective preference for a particular reordering strategy, as shown in Figure 4.9.

Reordering for Cluttered Datasets (H3)

For datasets with clutter, there is strong evidence that DIS layout strategies are more suitable. The *quality* of marked clusters is significantly better when participants used a DIS ordering strategy, confirming **H3 (b)**. This finding coincides with the reported *confidence* of participants (see Figure 4.8), who are also significantly more confident when working with DIS in clutter conditions. Even if both options are available (SIM and DIS), there is statistical proof that participants will choose a DIS ordering in clutter conditions providing evidence for **H3 (c)**. The Gestalt law of *grouping by orientation similarly* [Wer23; War20] might be a reason for this preference. The orientation of the lines is more salient in the DIS ordering, which facilitates a stronger grouping compared to a SIM ordering.

We cannot see significant differences in the similarity values of the selected axes pairs in task 1. However, Figure 4.5 illustrates the distribution of similarity values, providing evidence that participants believe that DIS axes pairs support them better in a cluster identification task. These findings are also in line with the qualitative feedback. The majority stated a preference for DIS over a SIM layout (see Figure 4.9). The participants said, for example, *"the spikes make the clusters more obvious"*, *"clearer [in A] because the zig-zag makes it easier to see among the noise."*, or *"The lines are closer together"*.

As shown in Figure 4.6, participants performed the cluster identification task faster with a DIS layout compared to a SIM layout in cluttered datasets. However, the differences in the *completion time* are not significant ($p = 0.05$). Therefore, we cannot confirm **H3 (a)**.

4.6.1 Design Considerations

With the results gained from our study, we derive the following design considerations for using PCPs in a cluster identification task.

Whenever possible, clutter should be removed in a pre-processing step. Results from tasks 1 and 2 indicate that participants working with PCPs need more time, are less accurate, and are less confident in identifying clusters with an increasing amount of clutter.

For datasets without any clutter, a `SIM` layout should be preferred over a `DIS` layout. Participants working with a `SIM` ordering strategy were faster in identifying and marking clusters compared to a `DIS` layout, as results indicate in task 2. There is, however, no difference in the quality of the clustering or confidence.

When clutter is an issue, a `DIS` ordering strategy should be preferred over a `SIM` layout. As we can see in the results from tasks 2 and 3, participants performed more accurately and were more confident in their selection when working with a `DIS` layout.

Although we used two clutter conditions in our study, it is challenging to derive specific guidelines when a dataset is considered as *cluttered*. It depends on many properties, such as the number of records, the general density of data and patterns, and the size of the PCP. Therefore, this needs to be analyzed in follow-up studies.

4.6.2 Limitations and Future Work

In our study, we focused on cluster identification. Therefore, the proposed design considerations need to be considered with caution for other tasks like correlation analysis. There might be changes in performance due to different patterns of interest. The same is true for the cluster structures. Our synthetic benchmark consists of a strong cluster structure throughout all dimensions. The results of our study might not be representative if cluster structures are less compact or not present across all dimensions. The results of our study already show that real-world datasets perform significantly worse than the synthetic benchmark data, although we selected datasets, commonly used in PCP research. The reasons for this effect might be that (1) all records of the real-world data belong to a cluster (no clutter was present), (2) the clusters in the real-world dataset were less compact than in the synthetic datasets, and (3) the real-world trials were done right after the `300N` trials. Participants selected only a subset to be part of a cluster and interpreting the remaining points as clutter. To generalize the results, follow-up studies should be conducted.

Further limitations of the study are (1) **The number of dimensions and records:** while we believe that results are independent of the number of dimensions, we restricted ourselves to eight dimensions to keep the trials throughout the study comparable. With an increasing number of dimensions, the computation of ordering algorithms will take longer; however, this was out of the scope for our study to investigate. (2) **The population of participants:** The study was conducted during two InfoVis lectures at our university. Therefore, participants were recruited from an InfoVis-trained, local student population, limiting the generalizability. (3) **Study setup:** Participants used their own laptops with different screen space and resolution. Although we manually checked that the study was displayed correctly on each laptop, the experience might change due to different screen settings and browsers.

4.7 Conclusion

This chapter advances the field of axes reordering in parallel coordinates plots (PCPs). First, we classified existing reordering techniques based on their inner workings, preferred patterns, and meta characteristics. Using this classification, we provide guidance in selecting an appropriate approach for a given task. Second, we pushed the evaluation of axes reordering techniques towards empirical justification. We conducted the first controlled user study to assess the performance of PCPs with two different ordering strategies. Specifically, we investigated whether the often proposed similarity-based axes arrangement (`SIM`) is better to identify clusters than a dissimilarity-based layout (`DIS`), which produces more salient cluster patterns. Our results show that, depending on the clutter level, participants performed differently based on the used ordering strategy. When no clutter was present, a `SIM` layout was more efficient, whereas, for cluttered datasets, a `DIS` layout led to better results. The subjective preference of participants supported these findings. Thus, our study shows that the performance of participants can be increased by choosing the correct layout strategy based on the underlying task.

5 Reducing Density Distortion in Parallel Coordinates Plots

Summary

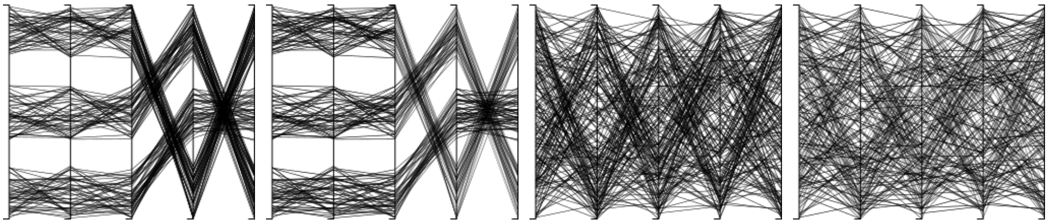
Parallel coordinates face a significant problem influencing the perception and interpretation of patterns. The distance between two parallel line segments differs based on their slope. Vertical lines are rendered longer and closer to each other than horizontal lines. This problem is inherent in the technique and has two main consequences: (1) clusters which have a steep slope between two axes are visually more prominent than horizontal clusters. (2) Noise and clutter can be perceived as clusters, as a few parallel vertical lines visually emerge as a ghost cluster. This chapter makes two contributions: First, we formalize the problem and show its impact. Second, we present a novel technique to reduce the effects by rendering the polylines of the parallel coordinates based on their slope: horizontal lines are rendered with the default width, lines with a steep slope with a thinner line. Our technique avoids density distortions of clusters, can be computed in linear time, and can be added on top of most parallel coordinate variations. To demonstrate the usefulness, we show examples and compare them to the classical rendering.

This chapter is *based on* the following publication. Please refer to Sections 1.4 and 1.5 for the contribution clarification and general citation rules.

[Pom+19] David Pomerenke, Frederik L. Dennig, Daniel A. Keim, Johannes Fuchs, and **Michael Blumenschein**. *"Slope-Dependent Rendering of Parallel Coordinates to Reduce Density Distortion and Ghost Clusters"*. In: IEEE Visualization Conference (VIS). 2019, pp. 86–90.

5.1 Introduction

Cluster identification is, among others, one of the most common tasks for parallel coordinates [AA01]. Every record of a dataset is represented by a single polyline, spanning across the different axes/dimensions of the dataset. Polylines running close together are considered a cluster as they have similar values across the dimensions. In Figure 5.1 (a), we can see three clusters spanning across the dataset. Between dimensions 1–3, the clusters are *horizontal*, meaning that the data values are approximately the same within all dimensions. Across dimensions 3–5, the clusters are *diagonal*, changing their values and cluster center, and have a steep *slope*. We can easily see a general problem of the PCP technique: diagonal changes of clusters are visually more prominent than horizontal trends.



(a) Regular rendering **(b)** Slope-dependent rendering **(c)** Regular rendering **(d)** Slope-dependent rendering

Fig. 5.1. Comparison of regular parallel coordinates with our slope-dependent polyline rendering. Parallel coordinates face two problems, which are inherent of technique: (a) depicts three clusters of the same diameter and size across all dimensions. *Diagonal changes of the clusters are visually more prominent*, as diagonal lines are rendered more closely. (c) shows 200 data points of uniform random noise in all dimensions. *Zig-zag clusters are visible* as diagonal lines are perceived as clusters, although there are no such clusters in the data (ghost clusters). We propose to render each line segment based on its slope between two axes. As a result, clusters are not distorted by their shape (b), and the ghost clusters effect is reduced (d).

Assuming all polylines have the same line thickness, there are two reasons for this effect: Diagonal lines need more area (=more pixels), and the background space between parallel lines is smaller for diagonal clusters compared to horizontal ones. As a consequence, there is a density distortion of clusters based on the slope or angle of the cluster. A second effect, also based on these rendering artifacts, are so-called *ghost clusters*. Figure 5.1 (c) depicts a dataset with 200 points, randomly and uniformly distributed across all dimensions. One can "see" two zig-zag patterns indicating two clusters. However, the data does not contain any specific structure – in particular, no clusters. This problem is not only relevant in pure clutter (or noise) datasets but also influences the perception of clusters in datasets that contain a limited amount of clutter and noise along with relevant patterns. *Ghost clusters* and distorted cluster density are related to human bias, but the core problem is based on the parallel coordinates technique itself. It can also occur in other variants of PCPs (e.g., different colors and transparency for lines, or edge-bundling).

This chapter makes two contributions: (1) we formalize the problem and show its impact. (2) we propose a novel approach that renders each line segment based on the slope between two dimensions. Horizontal lines are rendered with the default line thickness. Diagonal lines are rendered thinner. Two examples are depicted in Figure 5.1 (b) and (d). The technique can be computed in linear time and applied on top of most PCP variations. The approach by Zhou et al. [Zho+09] is closest to our work. It blends polylines based on their local neighborhood, which reduces the influence of noise but still suffers from the distortions caused by the over-emphasis on diagonal lines.

All material of this chapter is available at https://osf.io/sy3dv.

5.2 Related Work and Research Gap

A large number of approaches try to reduce clutter and highlight patterns in parallel coordinates. However, a formalization of distorted patterns, based on the polyline's slope, is missing, and none of the existing approaches specifically target this limitation. We describe the works most closely related to ours in the following, and highlight the relation to our slope-dependent rendering.

5.2.1 Sampling and Filtering Techniques

One of the most commonly used approaches to overcome clutter and overplotting issues is sampling and filtering techniques. With less data clutter decreases, while the general structures, typically represented by many data records, remain in the PCP [HW13]. The taxonomy by Ellis & Dix [ED07] provides a categorization of clutter reduction methods, including sampling, filtering, and clustering, as well as visual techniques such as adjusting the point size or opacity. Sampling often removes relevant data records or dimensions, reducing the truthfulness of the sampling concerning the dataset. Our technique reduces clutter by counterbalancing the distortion artifact inherent to PCPs. It can be applied on top of a sampled or filtered subset of the data. Dependent on the data characteristics, our technique increases the amount of data displayable in a given PCP by deemphasizing diagonal polyline segments.

5.2.2 Axes Reordering and Dimension Reduction

Another approach to minimize clutter in PCPs is to reorder the dimension axes or reduce the number of displayed dimensions. For example, Pargnostics by Dasgupta and Kosara [DK10] describes a set of quality metrics for PCPs which can be minimized or maximized (e.g., the number of line-crossings and parallelism). The authors also suggest the flipping of axes to reduce the number of line-crossings or diagonal clusters. The survey by Behrisch et al. [Beh+18] discusses a large number of quality metrics as objective functions for axes reordering. Axes reordering, dimension reduction, and axes flipping can reduce ghost clusters by favoring horizontal structures. Depending on the data, however, it cannot be avoided entirely. Axes reordering is highly dependent on the data and analysis task. It is an orthogonal concept to our approach and can be combined with it.

5.2.3 Density- and Cluster-based Rendering

Clusters and other patterns can also be highlighted by density-distributed rendering. The general idea is to render PCPs as density distributions rather than individual polylines. Johansson et al. [Joh+05] measure the density based on the number of overlapping polylines per pixel. This notion of density serves as input to a transfer function that allows highlighting areas according to their local density. Heinrich & Weiskopf [HW09] apply the concept of continuous scatterplots [BW08] to PCPs to derive a density model and thus interpolate the data. The resulting rendering is specifically useful for cluster identification. The work by Palmas et al. [Pal+14] provides a different approach, which bundles edges according to class membership. The resulting bundles are rendered as polygonal strips. Density- and cluster-based rendering may hide the underlying individual records and often require class labels to achieve a useful coloring or edge-bundling. While these approaches reduce clutter, they do not avoid the density distortion of clusters.

5.2.4 Polyline Modifications

A common technique is to modify the polylines of PCPs, specifically the overall line width, opacity, color, and shape. One example is the edge-bundling approach by Heinrich et al. [Hei+12b], which bundles polylines according to class membership and thus reshapes the line. The work by Zhou et al. [Zho+09] called line splatting is most closely to ours. Line splatting is iteratively adjusting the opacity of lines based on the local neighborhood. Users can interactively change the degree of polyline and segment splatting. In contrast to Zhou et al. [Zho+09], our work tries to mitigate the visual distortions intrinsic to PCPs, such as the perceived density of clusters and the effect of ghost clusters.

5.3 Problem Statement and Impact on Parallel Coordinates Patterns

We now formalize the line geometry of parallel coordinates and describe their effects on density distortions and ghost clusters.

5.3.1 Geometry of Classical Parallel Coordinates

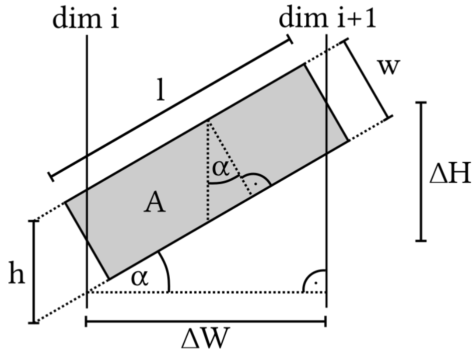
In classical parallel coordinates, polyline segments are rendered as *rectangles* (see Figure 5.2a). All lines have a *constant line width* (also called thickness or stroke width) $w \in \mathbb{R}^+$. ΔW denotes the space between the dimension axes and ΔH indicates the difference of data values.

Polyline Length

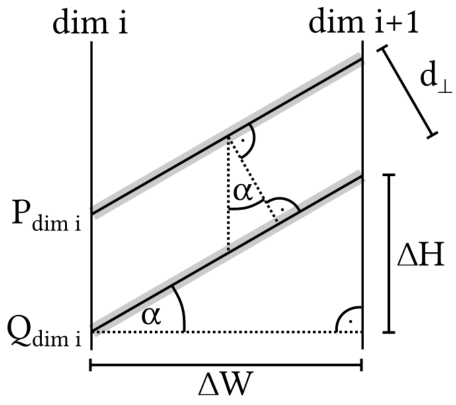
The polylines *differ in their length*, depending on their slope or the angle $\alpha \in [0, \frac{\pi}{2})$.

A higher α results in longer lines. The length $l(\alpha)$ is defined as:

$$l(\alpha) = \frac{l(\alpha)}{\Delta W} \cdot \Delta W = \sec(\alpha) \cdot \Delta W = \cos^{-1}(\alpha) \cdot \Delta W \tag{5.1}$$



(a) Rectangle-based rendering of one polyline segment.



(b) Distance between two parallel polyline segments P and Q.

Fig. 5.2. Geometry of a polyline segment. In regular PCPs, the stroke width w is constant and length l, height h, area A, and line distance $d_{\perp}$ are dependent on α.

Note that we define $\cos^{-1}(\alpha) = sec(\alpha) = \frac{1}{cos(\alpha)}$.

Keeping line length and surface area A consistent, polyline segments can also be rendered as *parallelograms* with sides h and l. For a constant line width w, the side length h is, similar to l, also dependent on the slope or the angle α and defined as:

$$h(\alpha) = \frac{h(\alpha)}{w} \cdot w = \sec(\alpha) \cdot w = \cos^{-1}(\alpha) \cdot w \tag{5.2}$$

A higher α results in a higher value of h.

Line Surface Area

The surface area A (= number of pixels) for each line segment depends on the line length as $A = w \cdot l$ (rectangle rendering) and $A = h \cdot \Delta W$ (polygon rendering). Combined with the slope-dependent definitions of l (Equation 5.1) and h (Equation 5.2), the line area is defined as:

$$A(\alpha) = \cos^{-1}(\alpha) \cdot \Delta W \cdot w \tag{5.3}$$

This formula is true for both rectangle- and polygon rendering. As a result, the line surface area increases when polyline segments change from *horizontal* to *diagonal*. In other words, the stronger the slope of a line, the higher the surface area of a line segment.

Distance between Polylines

The relation between the distance $|P_{\text{dim } i} - Q_{\text{dim } i}|$ of the data points P and Q, and the perceived orthogonal distance $d_{\perp}$ between the representing line segments, is shown in Figure 5.2b. $d_\perp(\alpha)$ is slope-dependent, and the lines are rendered closer for larger α leading to smaller free space between the lines:

$$d_\perp(\alpha) = cos(\alpha) \cdot |P_{\text{dim } i} - Q_{\text{dim } i}| \tag{5.4}$$

So far, the width w is neglected. If we add w to the formula, we get

$$d_\perp(\alpha) = cos(\alpha) \cdot |P_{\text{dim } i} - Q_{\text{dim } i}| - 2 \cdot \frac{w}{2} \tag{5.5}$$

as the distance between the two lines. For large α, the distance $d_\perp(\alpha)$ may disappear and two or more lines are perceived as a cluster.

Figure 5.3 summarizes the effects of the slope α on the rendered parallel coordinates. Based on the formalized geometric dependencies, we now derive the two main problems of perceived patterns in parallel coordinates.

5.3.2 Problem 1: Visual Distortion of Cluster Densities

The Gestalt law of proximity [Kof14; War20] indicates that the density of lines translates to a perception of cohesiveness and thereby enables users to recognize clusters in PCPs. Classical PCPs put undue emphasis on diagonal clusters, which is facilitated by the *increase of line lengths* and *decrease of line distances*. This contradicts the data-ink ratio coined by Tufte [Tuf01], which describes the proportion of ink devoted to the actual data relative to the total amount of ink. Thus, it adds unnecessary distortion: Diagonal clusters are emphasized more than horizontal clusters. Classical PCPs, therefore, induce a systematically inaccurate perception of clusters, when the observer would expect that the visualization is inherently neutral in this respect. We can see the effect in Figure 5.1 (a), where diagonal and horizontal clusters receive a significantly different emphasis.

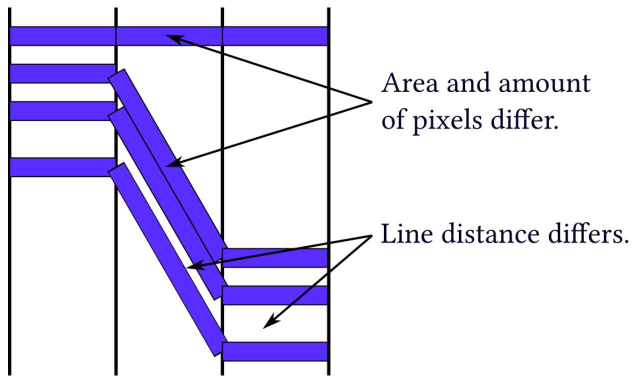
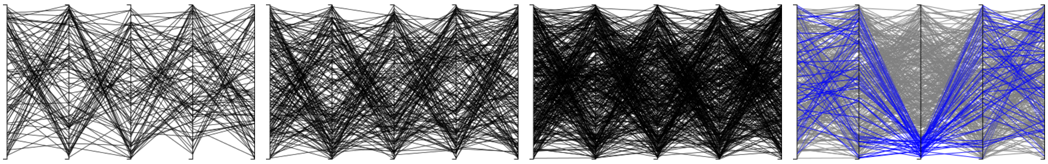


Fig. 5.3. Effect of angle α on PCP lines. (1) Diagonal lines have a *higher line surface area* (= more pixels) compared to horizontal lines. (2) Diagonal lines have a *smaller distance between lines*.



(a) $n = 100$ (b) $n = 200$ (c) $n = 400$ (d) $n = 400$

Fig. 5.4. Ghost clusters in uniformly distributed random data points. The number n of polylines is increased from (a) to (c). (d) = (c) but the data points of a ghost cluster are highlighted to demonstrate that they are indeed uniformly distributed even though (c) indicates otherwise.

5.3.3 Problem 2: Ghost Clusters

The rendering effects caused by the different slopes of the polyline segments can also produce artificial patterns in parallel coordinates plots. Figure 5.4 (a–c) show three PCPs with uniformly distributed random data points, i.e., there is no structure in the data. One can easily see that a *zig-zag pattern*, alternating between high and low values is visually present. The corresponding polylines seem to be parallel and close together, forming two clusters. With an increasing number of data points, the "clusters" are perceptually stronger. In Figure 5.4 (d), we mark one apparent cluster and highlight its polylines across the different dimensions. One can see that the data is indeed randomly distributed and not forming a cluster across the dimensions. We define these visible, but non-existing patterns as *ghost clusters*. Ghost clusters are not only a problem of datasets with clutter or noise. Also, in structured datasets, ghost clusters can be present and influence the interpretation of the data.

5.4 Slope-Dependent Rendering of Lines

To overcome the distortion of cluster densities and potential ghost clusters, we propose to render the polyline segments based on their angle α. The general idea is to render horizontal lines with the default width and diagonal lines with a thinner line. As a result, we increase the space between vertical lines and decrease the surface area, i.e., the number of pixels to draw a line. In the ideal case, all line segments should end up with the same area and the same distance between the segments. To achieve the same area for all line segments, the width w of the polyline segments needs to be scaled based on their length l.

5.4.1 Area Preserved Rendering

With our adjustment method, we aim to create equal areas for all polyline segments. We interpret all line segments as paralllelograms with an equal and constant area A. We cannot modify the line length (which is dependent on α according to Equation 5.1). Therefore, we conceptualize the rendering as change of the line height h. To obtain equal areas, we fix $h \in \mathbb{R}^+$ as a constant value for all line segments. Hence, $A = h \cdot \Delta W$ is now independent of α and therefore identical for all line segments.

Renderers require a specification of the line width w instead of the line height h. Therefore, we translate h to a new, slope-depndent line width ω. This is achieved by a simple permutation of Equation 5.2, and the replacement of w by ω:

$$\omega(\alpha) = \cos(\alpha) \cdot h \tag{5.6}$$

Rendering all lines with $\omega(\alpha)$ results in parallel coordinates with the same area for all polyline segments.

5.4.2 Overadjusted Rendering to Compensate Line Distance

While Equation 5.6 reduces the effect of a slope-dependent distance between the polylines, we cannot mitigate this effect entirely. Therefore, we compensate it by strengthening the adjustment and applying an exponent $P \in \mathbb{R}$ to the rendering of the line width:

$$\omega(\alpha) = \cos(\alpha)^P \cdot h \tag{5.7}$$

$P = 0$ results in classical rendering, i.e., surface area and distance between polylines depend on α. $P = 1$ corresponds to the area-preserving rendering (Equation 5.6). For $P > 1$, overadjusted rendering is applied.

5.4.3 Implementation and Computation of α

The angle α can be derived from $\tan(\alpha) = \frac{\Delta H}{\Delta W}$, given an axes distance ΔW and difference between values in neighboring dimensions $\Delta H = |P_{\text{dim } i} - P_{\text{dim } i+1}|$). Combined with Equation 5.7 we derive:

$$\omega(\alpha) = h \cdot \cos\left(\arctan\left(\frac{\Delta H}{\Delta W}\right)\right)^P \tag{5.8}$$

As $\cos(\arctan(x)) = (1 + x^2)^{-\frac{1}{2}}$, we can simplify Equation 5.8 to:

$$\omega(\alpha) = h \cdot \left(1 + \left(\frac{\Delta H}{\Delta W}\right)^2\right)^{-\frac{P}{2}} \tag{5.9}$$

5.4.4 Choosing the Adjustment Strength

$P = 0$ corresponds to classical PCP rendering, where all lines have the same width. $P = 1$ corresponds to rendering with equal line heights resulting in the same surface area A for all polylines. However, it does not fully correct the decreased line distances. Thus, we allow $P > 1$ as over-adjustment to further compensate

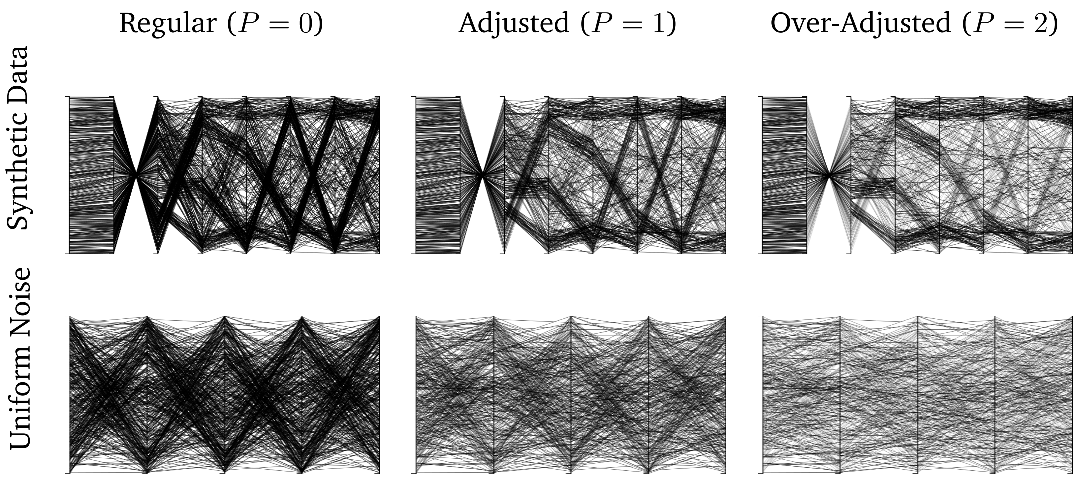


Fig. 5.5. Effect of parameter P on pattern visualization in synthetic data with uniformly distributed background noise, and in uniformly distributed random data only. Regular rendering ($P = 0$) significantly over-emphasizes diagonal clusters and causes the occurrence of ghost clusters. For $P = 1$, all clusters are equally emphasized, and the effect of ghost clusters is strongly mitigated. For $P = 2$ the distortion is reverted, and horizontal clusters are over-emphasized. Simultaneously, ghost clusters are further reduced.

overplotting of lines with strong slopes. In particular, the parameter P can be freely adapted to the degree of clutter, and the properties of the dataset. We want to highlight that our slope-dependent rendering can fully overcome the problem of different line surface area ($P = 1$), but the issue of varying distance between polylines can only be reduced with $P > 1$. Based on these geometric properties, we recommend $P = 1$ for truthful representation. However, many properties of a PCP and dataset influence the quality of the rendering (see Section 5.4.5), therefore an over-adjustment ($P > 1$) may be necessary. Our tests with various synthetic and real-world datasets showed that $P \approx 2$ is an upper bound for most applications.

In Figure 5.5, we apply our technique to a synthetic dataset and uniform random noise. We achieve a balanced emphasis of horizontal and diagonal clusters for $P = 1$ and an over-emphasis of horizontal lines for $P = 2$. Ghost clusters are also reduced for $P = 1$ because their density is corrected. However, the effect of smaller line distance cannot be avoided, and ghost clusters are still visible. We can compensate for the line distance effect by over-adjusting the line area effect (e.g., $P = 2$), nearly eliminating the ghost clusters, but introducing an over-emphasis of horizontal lines.

5.4.5 Influence of PCP Properties and Parameters

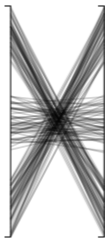
The following parallel coordinates parameters influence the impact of ghost clusters and the distortion of cluster densities and should be taken into account when applying the slope-dependent rendering.

PCP Size, Axis Height and Spacing. The overall size of a PCP has a direct impact on the axis height and spacing ΔW between the axes. Axis height and ΔW determine the range of α: Long axes and tight spacing, caused by high-dimensionality, increase the angles and distort cluster densities and increase the likelihood of ghost clusters.

Default Line Width. Manipulating the constant line-height h influences the detail and the clarity of the PCP. Thick lines increase the problem of overplotting, in particular for diagonal lines and clusters. Thin lines are more distinguishable and therefore produce more salient visualizations. The result of the slope-dependent rendering depends on the default line width, typically determined by the user. The default width directly influences the area covered by each line segment. It is advisable to consider a manual adaptation of the constant line-height h before applying a slope-dependent rendering.

Data Volume. The number of data records influences the visual representation a PCP and is strongly related to its size and the default line width. A high data volume visualized with a small PCP and/or a thick line width increases the problem of overplotting, but also the distortion of cluster densities and ghost clusters. For example, Figure 5.4 shows how the dataset size increases the perception of ghost clusters. Therefore, these properties should be optimized for a given dataset before applying the slope-dependent rendering.

Line Color and Transparency. When no transparency is used, then the color of the polylines does not affect PCPs and therefore also not our approach. Transparency can be used to avoid clutter and overplotting but introduces another artifact, which negatively influences the perception of patterns. Crossing lines introduce a darker color, which may be interpreted as a cluster. Combined with the slope-dependent rendering, new ghost clusters may occur, while other patterns may vanish: Adjusting the transparency of lines based on their slopes, as opposed to the line width, is not useful.



5.5 Discussion

To test the effectiveness of our slope-dependent rendering, we implemented a tool which is available on our website[1]. Users can upload their data, or try out various synthetic and real-world datasets, comparing the results of classical and slope-based rendering. During our testing with the implementation, we found out that our slope-dependent line adjustment technique performs well on various datasets, reduces ghost clusters, and counterbalances distortions. We also tested the impact of our approach with other patterns, such as positive and negative correlations (Figure 5.5). While positive correlations are not affected even with a large P value ($P = 2$), the slope-dependent rendering influences the diagonal lines of negative correlation. We

[1] Tool and source code available at `http://subspace.dbvis.de/pcp-adjustment`.

found that negative correlations also remain visible. However, the line representing data points at the ends of the dimension ranges are drawn with a small line width, making the visibility of this pattern susceptible to large P values ($P = 2$).

Our approach can be combined with other techniques, such as axes reordering and dimension reduction, as they do not manipulate the polylines of a PCP. It can also be combined with polyline modifications like edge-bundling. However, the line width should then be calculated relative to the line length rather than the slope. As described above, various PCP properties generally influence the visual distortion and ghost clusters in PCPs. To achieve optimal results, these parameters should be optimized before the slope-dependent rendering is applied, and focus on the reduction of overplotting and the average angles of polylines.

A careful selection of the parameter P is necessary. The usefulness of a particular P depends on many general PCP properties, as well as data characteristics such as the number of data records and dimensions. Therefore, P cannot be determined fully automatically based on a fixed parameter. However, we envision an algorithm which measures the density distribution, overlapping, and distortion and automatically selects an appropriate P to achieve a reliable representation of the data. We want to address this algorithm as part of future work. Furthermore, we want to evaluate the usefulness of our approach, in particular in comparison to other methods, by conducting a quantitative user study.

5.6 Conclusion

We formalize two general problems of parallel coordinates: The density of clusters are often distorted and non-existing ghost-clusters emerge. As a solution, we propose a novel rendering technique for the polyline segments: The line width is adjusted according to the angle of each line segment. Our method can be computed in linear time, depends on a single parameter, and can be combined with many existing parallel coordinates' variations.

Part II

User- and Task-Driven Design for Pattern Analysis

6 Designing Hybrid Charts for the Comparative Analysis of Data Distributions

Summary

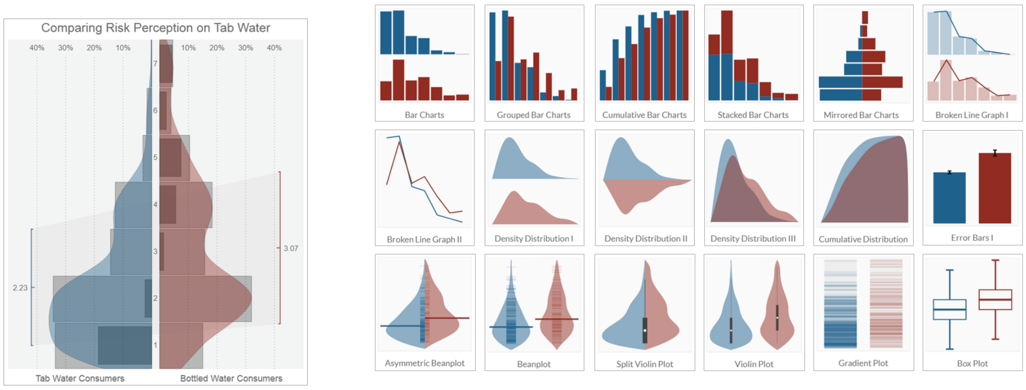
Comparing data distributions is a core focus in descriptive statistics, and part of most data analysis processes across disciplines. In particular, comparing distributions entails numerous tasks, ranging from identifying global distribution properties, comparing aggregated statistics (e.g., mean values), to the local inspection of single cases. While various specialized visualizations have been proposed (e.g., box plots, histograms, or violin plots), they are not usually designed to support more than a few tasks, unless they are combined. In this chapter, we present the v-plot designer; a technique for authoring custom hybrid charts, combining mirrored bar charts, difference encodings, and violin-style plots. v-plots are customizable and enable the simultaneous comparison of data distributions on global, local, and aggregation levels. Our system design is grounded in an expert survey that compares and evaluates 20 common visualization techniques to derive guidelines for the task-driven selection of appropriate visualizations. This knowledge externalization step allowed us to develop a guiding wizard that can tailor v-plots to individual tasks and particular distribution properties. Finally, we confirm the usefulness of our system design and the user-guiding process by measuring the fitness for purpose and applicability in a second study with four domain and statistic experts.

This chapter is *taken from* the following publication. Please refer to Sections 1.4 and 1.5 for the contribution clarification and general citation rules.

[Blu+20a] **Michael Blumenschein**, Luka J. Debbeler, Nadine C. Lages, Britta Renner, Daniel A. Keim, and Mennatallah El-Assady. *"v-plots: Designing Hybrid Charts for the Comparative Analysis of Data Distributions"*. In: Computer Graphics Forum 39.3 (2020), pp. 565–577.

6.1 Introduction

Analyzing and exploring empirical data and its distribution is a core task in descriptive statistics across various research disciplines, and often serves as a foundation



(a) v-plot

(b) Selection of common charts to visualize and compare data distributions.

Fig. 6.1. The *v-plot designer* enables the creation of custom hybrid charts *(v-plots)* for the comparative analysis of data distributions. Given a set of analysis tasks, v-plots can be tailored to highlight particular distribution properties (on a *local*, *global*, and *aggregated level*) using a *guiding wizard*. All the charts above represent the same data, showing the perceived risk of two groups for tap water consumption [Deb+18].

for in-depth analyses. Comparing data distributions is a multi-faceted process comprising a variety of different analysis tasks ranging from *global* aspects, such as identifying the type, shape, and skewness of a distribution, to *local* aspects, such as comparing value frequencies, or identifying differences on an instance level. Alongside global and local tasks, users also want to analyze *aggregated* statistical properties such as comparisons of mean and interquartile ranges. However, even with the abundance of statistical techniques, the visual inspection of distributions is essential to a successful analysis [Tuk77; FMF12] and can help to generate hypotheses, select appropriate statistical methods, and support the understanding and communication of analysis results.

Global, aggregated, and local analysis tasks focus on different properties of data distributions, which has led to the development of a broad range of *specialized* charts within the statistics and InfoVis communities. Some of the most prominent include box plots [Tuk77], violin plots [HN98], and bar charts. Each has its own strengths and weaknesses: for example, while box plots are useful for comparing medians and quartiles, they do not show whether a distribution is bi-modal, or in which value ranges two distributions differ most. Violin plots support these questions, but they are less useful for local tasks where users are interested in the frequencies of individual values (e.g., in discrete distributions). However, most analytical processes combine a number of different tasks. Analysts wishing to derive effective designs for a given dataset and a particular combination of tasks are faced with one of two choices: either using multiple charts to communicate their analysis on the different levels; or designing an expressive hybrid chart tailored to their analysis needs. Although hybrid designs have shown to be more effective for targeted analysis [Bor+13], most people do not have access to them and must use multiple charts instead. In addition, analysts have to rely on experience and knowledge about descriptive statistics when choosing appropriate charts, making this design space troublesome for non-experts to navigate.

To support the comparative analysis of data distributions for a wider audience, we need to make the design of hybrid charts more accessible and make task-dependent customization in the visualization of descriptive statistics less reliant to the user's expertise. Our chart authoring system, the *v-plot designer*, is developed to generate custom, widely applicable hybrid charts (*v-plots*) that concurrently support combinations of global, local, and aggregated tasks. v-plots consist of five ("**v**") layers, which inspired its name. We strove to ground our system design in an expert survey that aimed to externalize the implicit knowledge and experience of domain experts into *guidelines* used to inform a guiding wizard component. In this chapter, we reflect on the **multistage research process** of our system design that tackled the following research question: *How can we make hybrid charts for the visual comparison of data distributions (1)* ***simultaneously support local, aggregation-based, and global analysis tasks****; and (2)* ***accessible to analysts****?*

We provide a **grounding for our visualization design** by reviewing and categorizing existing tasks and charts for the comparative analysis of data distributions (Section 6.3). This lays the foundation for our expert survey (Section 6.4), which *induces guidelines* and an *automatic chart recommendation* for the selection of existing visualization techniques for specific task combinations.

In addition to this design study, we also provide a rationale on our **system design and guidance**. This is based on the review of the related work (Section 6.2), and describes the reasoning behind v-plots as custom hybrid charts that are designed to support a variety of concurrent tasks (Section 6.5). Based on the externalized guidelines, a *guiding wizard* in the *v-plot designer* can automatically tailor the v-plot through different visualization layers and transparency levels to highlight particular distribution properties. The guidelines are interchangeable and can be extended by new findings or the requirements of specific communities in the future. Single v-plots can also be combined to a *v-plot matrix* that is sorted according to the similarity of their visual structures and support the comparison of all pairwise distributions.

To summarize, the contribution of this chapter is two-fold. (1) We contribute **guidelines** based on a representative **expert survey** on the applicability and usefulness of the 20 most commonly-used statistical charts to 20 analysis tasks for descriptive statistics. For transparency and reproducibility, we make the survey and its results available at osf.io/jk8rp. (2) To make the acquired knowledge accessible while supporting the creation of custom hybrid charts, we contribute the *v-plot designer* (v-plot.dbvis.de). This **chart authoring approach** relies on a *guiding wizard* to enable users to adjust v-plots, which combine mirrored bar charts, direct difference encodings, a distribution shape, labels, and axes with statistic values.

6.2 Related Work

In addition to the landscape of comparative distribution analysis (described in Section 6.3), our work lies within the context of other chart authoring tools, as well as guiding and chart comparisons.

6.2.1 Chart Authoring Tools

Designing visualizations to communicate patterns is one of the key tasks in data science and descriptive statistics. However, implementing new visualizations for each analysis is not feasible for analysts. Therefore, in contrast to the vast amount of visualization coding libraries (e.g., Lyra [SH14] or Vega [Sat+16]), chart authoring systems have been developed to streamline design process and aid non-visualization-experts. Most of these systems are commercial applications [Beh+19]. Most prominently, Tableau [Tab18] has positioned itself as an easy-to-use, toolkit-based chart authoring system for the masses. More recently, Charticulator [RLB19] has been presented as a chart authoring tool for variable, user-defined layouts, enabling more flexibility with chart designs. RAWGraphs [Mau+17] is an example of a authoring system providing users with an open-access API. Lastly, approaches such as Voyager [Won+16; Won+17] combine the flexibility of declarative user-defined visualization designs with the ease of template-based authoring systems. While such tools target the broad spectrum of data-driven visualization designs, none are specifically focused on authoring visualizations for the comparative analysis of data distributions for descriptive statistics. Additionally, these approaches do not encode knowledge on choosing appropriate designs to provide useful constraints on the outputs of the systems, which has proven to be effective for non-visualization-experts [Mor+19].

6.2.2 Guiding and Chart Comparisons

Earlier research has addressed the weaknesses of particular visualizations for distribution analysis. For example, Silverman [Sil86], Tapia & Thompson [TT78], and Scott [Sco92] discuss the problem of histograms for continuous data. In the InfoVis community, many empirical user studies exist which compare charts or visual elements for distribution analysis. To name a few, Correll et al. [CG14] evaluate error bars, box plot variations, gradient plots, and violin charts for judging mean and standard error. Ondov et al. [Ond+19] evaluate the comparison of frequencies in distributions using different layouts such as mirrored and separate bar charts. Correll et al. [Cor+19] investigate the identification of outlier-distributions using bar charts, density plots, and density distributions. Gschwandtner et al. [Gsc+16] evaluate six different visual encodings, including gradient and violin plots, to visualize temporal uncertainty. Finally, Skau et al. [SHK15] evaluate the impact of embellishments in bar charts. While some of the links between analysis tasks and visualizations have been addressed, we still lack a set of guidelines for selecting appropriate charts for comparative distribution analyses, particularly if we are faced with a combination of tasks which should be combined in a single chart.

6.3 Landscape of Comparative Distribution Analysis

The goals of comparative distribution analysis are manifold. Exploratory data analysis [Tuk77] aims to identify interesting patterns across multiple dimensions, while other applications need sanity checks [Cor+19] to identify missing values, outliers, or skewed distributions that influence the analysis. Comparing distributions is another important aspect of descriptive statistics [FMF12] which can support data cleaning, choosing appropriate statistical models, and understanding potential reasons for significance.

Examples such as Anscombe's Quartet [Ans73] and the recent paper by Matejka & Fitzmaurice [MF17] illustrate that summary statistics are not enough to analyze distributions: visualizations are also needed to discover patterns, support analyses, and communicate results [FMF12]. Nevertheless, established charts often only support a limited number of tasks. Since a comprehensive statistical analysis typically comprises a combination of different aspects, it is often necessary to create and explore several different charts.

To guide analysts towards choosing an appropriate chart for a given application, we first provide a summary of the design space by categorizing existing analysis tasks (Section 6.3.1) and distribution charts (Section 6.3.2). Then, in Section 6.4, we conduct an expert survey to link analysis tasks with charts and derive guidelines.

6.3.1 Analysis Tasks

Many statistical books (e.g., [FMF12]) enumerate a subset of analysis tasks. However, we are unaware of an overview of all tasks for comparative distribution analysis. Often, tasks are not formally introduced, and different wordings are used for the description. We therefore establish a common vocabulary and introduce a classification of tasks, which is shown in Table 6.1. We group all tasks by their scope and type, classify them into four complexity levels, and discuss the involved distributions as elaborated below. In Table 6.1, we also provide examples from psychology applications for all tasks.

Scope of Analysis Task

We group the analysis tasks into three different scopes, which comprise the general analysis focus:

Local tasks concentrate on particular instances, for example reading the frequency of one value (L1) or comparing them across two or more distributions (L4). Local tasks are particularly interesting in discrete distributions (e.g., questionnaire results) in which the characteristic of specific values is of high interest.

Global tasks take the majority of data values into account and analyze the entire distribution. Typical questions are the identification of the distribution type (G1), or the comparison of skewness and kurtosis in different distributions (G4).

Aggregation tasks can be seen as the link between global and local tasks. They focus on aggregated statistical measures of a distribution, such as the identification of mean (A1) or median (A2), or the comparison of quartiles (A8) and standard errors (A10).

Analysis Type and Involved Distributions

We classify tasks into two types and investigate their involved distributions.

Comparison tasks ⬌ analyze the *relationships between distributions*. Some tasks compare the frequencies of one distribution (● vs ●), e.g., when identifying the most and least frequent values (L2). Others compare them across different distributions (● vs ●).

Identification tasks 🔍 focus on reading, measuring, and estimating the individual *properties of one distribution* (●). Examples include extracting the frequency of one value (L1), the identification of the distribution type (G1), and its skewness (G2).

Different analysis types are often interlinked and build on top of each other. For example, to identify the value(s) with the largest and smallest distribution difference (L5), one must first identify the frequency of each value (L1), then compare them across distributions (L4), and finally find the highest difference.

Complexity

We categorized the analysis tasks into four different complexity levels. We define complexity as the number of atomic, consecutive identification and comparison tasks needed to reach an analysis goal. It is not defined as the difficulty of extracting the relevant information from an (optimal) visual representation.

Complexity 1 ●○○○ comprises single identification tasks such as reading the frequency of a value (L1) within one distribution (●).

Complexity 2 ●●○○ are tasks which compare frequencies *within* a distribution (● vs ●). Examples are the identification of the least frequent value (L2), or the distribution type (G1).

Complexity 3 ●●●○ summarizes tasks which compare and relate frequencies *across* different distributions (● vs ●), e.g., the comparison of the distribution shapes (G3). Identifying aggregated statistics (A1–A5) is also considered as complexity 3.

Complexity 4 ●●●● quantifies the similarity and differences between two distributions (● vs ●) or aggregated statistical properties, e.g., identifying the value ranges with the largest and smallest distribution differences (G5), or comparing standard errors (A10).

6.3.2 Visualization Techniques

We summarize charts, particularly those designed for comparative distribution analysis. We first discuss charts for two distributions, and then for three or more distributions. We group charts into *histogram*, *shape*, and *statistical property based* approaches, as well as *hybrid methods*. We structure them perpendicularly by the taxonomy of Gleicher et al. [Gle+11]: **Juxtaposition** Ⓙ designs use a separate chart

Tab. 6.1. Classification of analysis tasks for comparative distribution analysis. Tasks are grouped by their scope (*local*, *aggregation*, and *global*) and ordered by complexity (●○○○ – ●●●●). We differentiate between identification (🔍) and comparison tasks (⬌), and highlight the involved distributions: lookup in one distribution (🔵), and comparison within (🔵 vs 🔵) and across distributions (🔵 vs 🔴).

		analysis task for descriptive statistics	example	type	complexity	distribution
local	**L1**	Identify the *frequency* of one value.	[Wri+12]	🔍	●○○○	🔵
	L2	Identify the *most & least frequent value(s) of* one distribution.	[Wri+12]	🔍	●●○○	🔵 vs 🔵
	L3	Compare *frequencies within* one distribution.	[Wri+12]	⬌	●●○○	🔵 vs 🔵
	L4	Compare *frequencies across* multiple distributions.	[Pin+16]	⬌	●●●○	🔵 vs 🔴
	L5	Identify the *value(s)* with the *largest and smallest difference*.	[Pin+16]	🔍	●●●●	🔵 vs 🔴
aggregation	**A1**	Identify the *mean* of one distribution.	[Lin15]	🔍	●●●○	🔵
	A2	Identify the *median* of one distribution.	[SPA11]	🔍	●●●○	🔵
	A3	Identify the *quartiles* of one distribution.	[MW07]	🔍	●●●○	🔵
	A4	Identify the *standard deviation* of one distribution.	[Lin15; SPA11]	🔍	●●●○	🔵
	A5	Identify the *standard error* of one distribution.	[Lin15]	🔍	●●●○	🔵
	A6	Compare the *means* of multiple distributions.	[Lin15; SPA11]	⬌	●●●●	🔵 vs 🔴
	A7	Compare the *medians* of multiple distributions.	[Gor+11]	⬌	●●●●	🔵 vs 🔴
	A8	Compare the *quartiles* of multiple distributions.	[Veg+98]	⬌	●●●●	🔵 vs 🔴
	A9	Compare the *standard deviations* of multiple distributions.	[FPH09]	⬌	●●●●	🔵 vs 🔴
	A10	Compare the *standard errors* of multiple distributions.	[Lin15; DRC13]	⬌	●●●●	🔵 vs 🔴
global	**G1**	Describe and identify the *shape* and *type* of one distribution.	[Lin15]	🔍	●●○○	🔵
	G2	Describe and identify the *skewness* and *kurtosis* of one distribution.	[Lin15]	🔍	●●○○	🔵
	G3	Compare the *similarity*, *shape*, and *type* of multiple distributions.	[Vil+11]	⬌	●●●○	🔵 vs 🔴
	G4	Compare the *skewness* and *kurtosis* of multiple distributions.	[Ric+15]	⬌	●●●○	🔵 vs 🔴
	G5	Identify the *value ranges* with the *largest and smallest difference*.	[Bir+05]	🔍	●●●●	🔵 vs 🔴

for each distribution. **Superposition** S designs represent multiple distributions within the same chart, supporting their comparison. **Explicit Encoding** E computes the similarity between distributions and directly encodes their difference.

Histogram-based Charts

Histogram-based charts represent the frequency as the height of bins. They are particularly useful for discrete distributions, as every value can be depicted by one bar. Binning is usually applied for continuous data. Figure 6.2 (top) depicts different arrangements of the bins. Separate (a) *bar charts* I place one chart per distribution next to each other, while (b) *grouped bar charts* S place bins of different distributions with the same value in one group. Similarly, (c) *stacked bar charts* S arrange these groups on top of each other, while (d) *mirrored bar charts* S place bins of the same value next to each other, mirrored by a horizontal axis. The charts can also be rotated (e) for a horizontal layout. The center of each bin can be connected with a (f) *broken line graph* I. Sometimes the bins are hidden. In other variations, the (g) lines can be arranged as a superposition S, similar to a line-chart, which allows a better comparison. Finally, (h) *cumulative bar charts* accumulate the frequencies of values and can be arranged as separate S or grouped bar charts I.

Statistic-property based Charts

Statistic-property based charts directly depict summary statistics. Popular examples are *box plots* [Tuk77] I and *error bars* I (Figure 6.1 (b), bottom right). Box plots indicate the median and interquartile range (IQR) as a box, representing 75% of the data. Additionally, the whiskers indicate $1.5\times$ IQR and outliers are marked with a dot. Variations have been proposed for the design, such as *notched* and *variable width box plots* [MTL78], along with numerous approaches for coping with skewed distributions [HV08]. Error bars typically represent the mean as a histogram and add the spread of standard error as an interval on top. Further designs comprise the mean and standard deviation, confidence interval, or any other uncertainty measure. A popular variation of the error bars shows only the uncertainty interval without the histogram underneath.

Shape-based Charts

Shape-based charts estimate the distribution using a probability density function. Most implementations use a kernel density estimation (KDE), which depends on the kernel type and bandwidth parameter. Both need to be carefully selected to provide reliable representations [Sil86]. As shown in Figure 6.2, simple *density distributions* compute a KDE and place the distributions next to each other I (i), on top of each other S (j), or in a mirrored fashion S (k). Violin-type charts are inspired by the idea of opening up the 'black box' of a box plot, illustrating the shape of the distribution. Early work comprised the *histplot* S and *vaseplot* S [Ben88], and the development of the *violin plot* (l) S [HN98] and *beanplot* (m) [Kam08] S was based on those

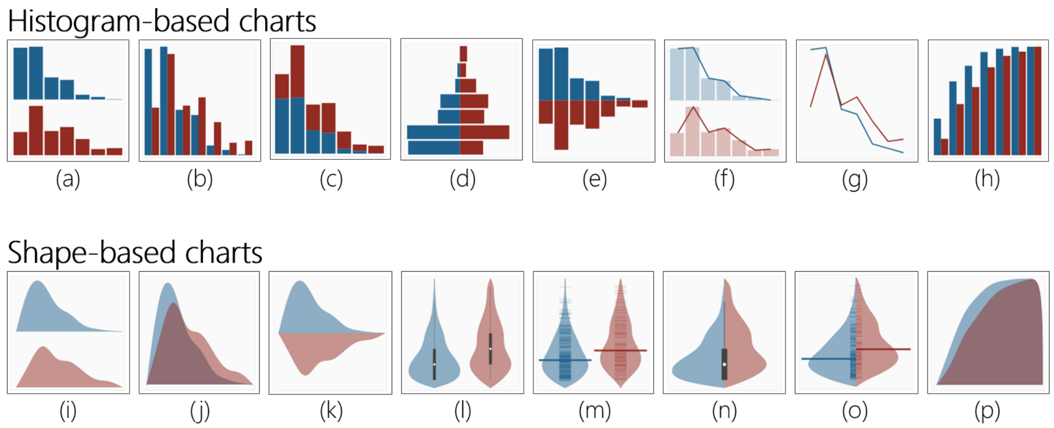


Fig. 6.2. Overview of charts for comparative distribution analysis.

ideas. Both encode mean or median and the distribution of the data in the form of a box plot or gradient plot. It can also be extended to an (n) *split violin plot* S or (o) *asymmetric beanplot* S, similar to the mirrored bar chart. The shape can also be used to encode the cumulative distribution (p) S.

Hybrid Charts

Hybrid charts are used to visualize different aspects of distributions in a single plot. Combinations of a histogram and a kernel density estimation, histogram and box plot, or boxplot with jitter are often used. Potter et al. [Pot+10] introduce a so-called Summary plot which combines a mirrored histogram with the majority of summary statistics, such as mean, skewness, and quartiles. These hybrid visualizations have proven to be effective in supporting multiple different analysis tasks at the same time. However, as discussed in Section 6.2, they are typically difficult to generate, particularly if the set of analysis tasks changes between applications.

Visualizations to Compare Multiple Distributions

Charts to compare multiple distributions are limited. Most approaches use Juxtaposition J or Superpositon S and 'just' concatenate single charts together, e.g., multiple box plots, error bars, stacked or grouped bar charts, line charts, and violin plots. We are not aware of direct encodings E of multiple distributions.

Related Visualization Techniques

Many other charts have been created which are often used to represent some aspects of distributions and their properties. A comprehensive summary can be found in the book by Wilkinson [Wil05]. Examples comprise *stem-leaf diagrams*, which follow the idea of histograms but use consecutive numbers to encode bins. This helps to understand the spread of aggregated values. Gradient plots use a transparent line (or point) for each data record. These plots explicitly use overplotting to help seeing all data records, as well as the overall shape. Other, partly related charts are stacked and jittered line plots [Cha+83], linear and nonlinear dotplots [SR96; RW18], and flow charts.

6.4 Connecting Analysis Tasks to Visual Representations

We now interlink the analysis tasks for descriptive statistics with the design space for visualizations that support distribution comparisons. Our aim is to find out which visualizations and visual elements support users for any given set of analysis task(s) to provide guidelines for application experts. To address these questions, our survey is based on the following methodology and survey design.

Methodology. Since we intend to investigate the usefulness of the most common visualizations for descriptive statistics for different analysis tasks, we selected a set of 20 representative charts covering histogram, shape, and statistical-property based visualizations. Part of our selection is shown in Figure 6.1b, which is extended by a rotated version of the mirrored bar chart and a commonly-used error bar representation without the median bar. As tasks, we use all local (L1 – L5) and all global tasks (G1 – G5). However, we selected only two representative examples from the aggregated group (A1 + A2). The reason is twofold: Firstly, the pilot study showed that with 20 tasks $\times$ 20 charts it would take participants almost two hours to complete the survey. And secondly, feedback from the pilot participants suggested that one would need the direct encoding of the statistical properties to successfully solve an aggregated task. This is true for all aggregated tasks and was also reflected in the results of the pilot study. We also limited the survey to the comparative analysis of two data distributions and included both InfoVis and statistic experts as participants to make sure we covered both statistical and information visualization expertise.

Survey design. For convenience, we designed the survey on paper. It comprises 16 pages and is divided into background material and five parts. Every analysis task was described, and a concrete example given. The participants first indicated the usefulness of every chart for a each analytical task using a Likert scale. Then, they selected up to three techniques which they felt were most useful for the respective task. At the end of the survey, we asked all the participants (a) whether any important and commonly used visualizations were missing; (b) which three plots they would be most likely to use in a paper when addressing most of the tasks, and why; and (c) which plots are most common in the literature, regardless of their usefulness. The survey and the results are available at osf.io/jk8rp.

Participants. 20 participants (6 female) participated in the survey. 17 were PhD students and 3 PostDocs. 13 participants reported their primary background in computer science (CS), 4 in psychology, and 3 in both CS and psychology. Their ages ranged from $25 - 52$ with a mean of 29.8 and a standard deviation (std) of 6.3. Their average experience (and std) was reported as 4.1 years (3.7) in statistics, and 4.0 years (3.0) in information visualization.

Survey procedure. We conducted several pilot runs and iteratively improved the descriptions and tasks in the survey. Since every participant was able to complete the survey at their own convenience we do not know whether the survey was completed in one go, or whether they had any additional help. However, as we are primarily interested in the participants' assessments rather than evaluating their knowledge about the different visualization techniques and supported analysis tasks, we do not see this as a limitation.

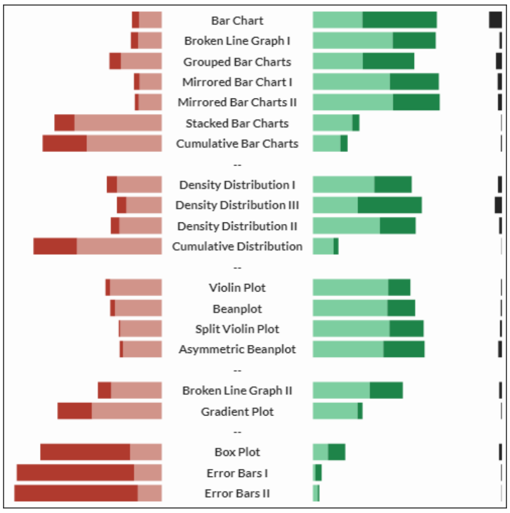
Data collection and cleaning. We manually extracted the answers and removed two participants from the results. Since the first had more than 50% missing values in the assessment of the different charts, we did not consider them as 'knowledgeable' in the topic. The second had a lot of outliers in the assessments of local analysis tasks, and most of the answers did not match the average of the other participants. For example, he/she said that bar charts are not useful for identifying the value with the highest frequency - which is obviously possible. As these outliers occurred across all local analysis tasks, we believed that the participant misunderstood local and global analysis tasks, and so he/she was removed. On average, all remaining participants have 3.3% missing values ($std = 4.9\%$) in the assessment of the different visualization techniques.

6.4.1 Usefulness of Visualizations Across all Analysis Tasks

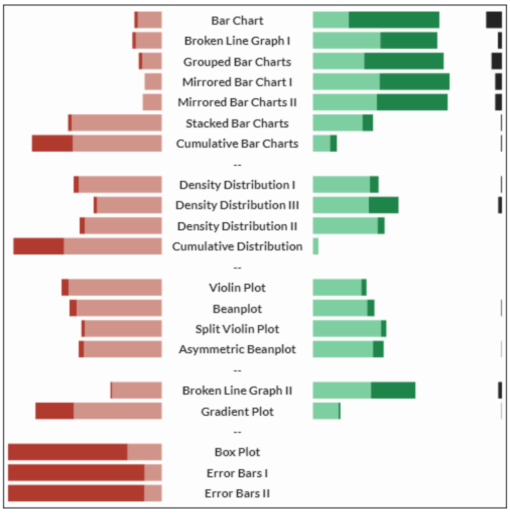
We report the results of the last tasks to provide general findings.

Most common charts. Based on the survey, bar charts (17 participants), box plots (15), error bars (4), broken line graphs (4), and grouped bar charts (3) are generally most common in the literature.

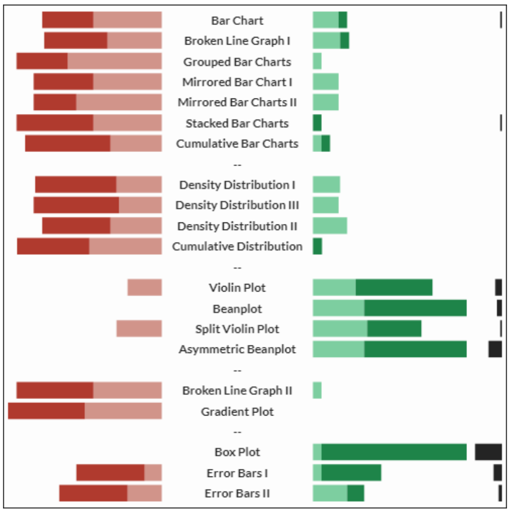
Visualizations most useful across all tasks. No single visualization technique was considered most useful for all analysis tasks. Many participants reported that the density distribution III is "*good to compare shapes*", "*great for global comparisons*", and enables the "*focus on difference analysis (difference is darker)*". Also, separate bar charts have been reported as useful because they are seen as "*simple*", "*a faithful representation of discrete data*", and "*easy to understand at first glance*". Box plots are considered useful as "*everybody knows them*", and "*you can see the median and quartiles*". Mirrored bar charts are "*easy to understand, have a clear baseline*



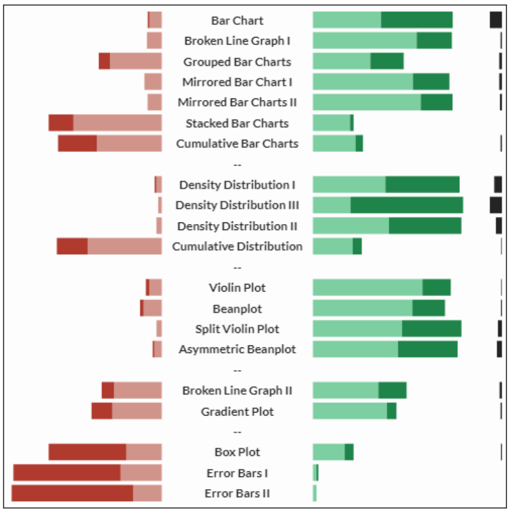
(a) All tasks



(b) Local tasks



(c) Aggregated tasks



(d) Global tasks

Fig. 6.3. Overview of expert survey results. Each chart represents a summary of a task scope according to Table 6.1. Every participant indicated the usefulness of each visualization for a particular task with **(++) very useful**, (+) useful, (-) not useful, and **(- -) not possible**. The vertical black bar charts on the right side represent the three visualizations, marked as most useful by the participants. We can see a clear difference for the usefulness of local, aggregated, and global analysis tasks. Generally, histogram representations are considered more useful for local tasks, while shape representations, such as a density distribution or violin plots are more useful for global tasks. Box plots and error bars, although often seen for comparing distributions, are only considered useful for aggregated tasks such as the identification and comparison of mean values and quartiles. The results of all individual tasks can be found in the supplementary material.

[compared to stacked bar charts], and good to compare distributions". Finally, the asymmetric bean plot comprises "*comprehensive information in a single graph*", "*looks pretty and captures a lot of information*", and makes it "*easy to compare distributions, means, and non-aggregated values*".

6.4.2 Usefulness of Visualizations for Different Analysis Tasks

In the following, we analyze the results of the survey on both a task level and across local, global, and aggregated analysis scopes.

Visual representation of (un)usefulness. We realized that the participants had different encoding strategies. While some marked almost all techniques that are useful with (++), others applied mainly (+), and (++) only for the most useful technique per analysis task. Therefore, we visualized the results of the survey (see Figure 6.3) for its exploration. While we cannot rely on the exact proportion between (+) and (++), and (-) and (- -), Figure 6.3 can provide us with a clear tendency towards (un)usefulness. In the following, we discuss the most important findings (**F**).

F1: Charts differ across local, aggregated, and global tasks. Considering the combined results in Figure 6.3, we can see that there are significant differences in the usefulness of the charts for local, aggregated, and global analysis tasks. In Figure 6.3b we can see that histogram-based charts are considered most useful for local analysis tasks. While they are also considered useful for global tasks (Figure 6.3d), the participants seemed to prefer shape-based visualizations such as density distributions and violin-type charts. For aggregated tasks, there was a clear preference for charts that directly encode statistical properties, such as box plots and error bars.

Another interesting observation is that six visualizations were not seen as being useful for most of the tasks. These were stacked and cumulative bar charts, cumulative distributions, box plots, and variations of error bars, as shown in Figure 6.3a. Surprisingly, while box plots and error bars are often used in the literature, they seem to be of limited usage for considering task variations.

F2: Local analysis tasks. On average, most histogram-based charts are considered useful for local analysis tasks, as shown by the green color in Figure 6.3b. As expected, stack and cumulative bar charts were an exception, being generally considered as not useful across all tasks, not only those with a local scope. We can see some differences of the other histogram charts within the scope of local tasks: For tasks considering a single distribution (L1 – L3), there is not much difference between single bar charts (with or without a broken line graph) and mirrored bar charts. They all support the identification of a value, as well as their comparison. Grouped bar charts are considered marginally less useful for these tasks. There was some disagreement on whether shape-based charts are useful for identifying the most (in)frequent values (L2).

Grouped bar charts, as well as density distribution III and broken line graph II, are considered most useful for comparative tasks across distributions, as they directly encode the similarity and difference between the distributions, making comparisons easier.

F3: Global analysis tasks. While many participants considered histograms to be useful for global tasks, the shape-based approaches outperformed them. In particular, when investigating the individual tasks, we can see that they are more useful, especially when comparing distribution shapes (G3), skewness (G4), and identifying differences (G5). Density distribution III, split violin, and asymmetric beanplots seem to be particularly useful for comparisons across distributions. Besides density distribution III, grouped bar charts are also considered very useful for identifying the value ranges with the largest and smallest differences (G5). Single and mirrored bar charts seem to be as useful as density distributions for the identification of the distribution type (G1), skewness, and kurtosis (G2), while grouped bar charts hinder the comparison of frequencies within a distribution and are therefore less useful.

Across tasks, we do not see much difference between the simple density distributions (I, II, and III), and the more complex violin and beanplot variants. Participants do not agree on the usefulness of the broken line graph II and gradient plots.

F4: Aggregated analysis tasks. Figure 6.3c shows that only charts which directly encode statistical measures are considered useful (i.e., all violin-typed charts, box plots, and error bars). The box plot is favored, followed by the asymmetric bean plot. While many participants rated error bars as one of their favorite charts for aggregated tasks, quite a few noted that error bars are not useful for identifying mean or median values. This is surprising, as the height of the error bars can directly encode this value.

F5: Impact of rotating charts. There seems to be no impact on the rotation of charts. Both versions of the mirrored bar charts are almost identical across all tasks. The same is true for the split violin plot and the density distribution II, which are the same except for the rotation (and an additional encoding of the mean value).

6.4.3 Summary

None of the presented charts support all tasks at the same time. Charts encoding statistical measures are either simplified to these values (e.g., error bars) or part of violin-typed charts, missing a histogram for local tasks. Vice versa, histograms support tasks on single values, but often miss comparisons on a global level or of statistical measures. Furthermore, charts that include a direct encoding of differences (e.g., grouped bar chart or density distribution III) support the comparison of multiple distributions, but they are more complex regarding the analysis of a single distribution as the frequencies of each distribution cannot be followed easily.

Based on our findings, we envision an ideal plot for comparative analysis of data distributions combining (1) a histogram representation for local analysis tasks, (2) a shape-representation for the global tasks, (3) a direct encoding of the differences between two or more distributions, and (4) a representation of statistical measures. Perpendicularly, such a chart should unify aspects of S and E in a single chart. While superposition S layouts help to identify the properties of single distributions, explicit encoding E helps to compare distributions by highlighting their differences.

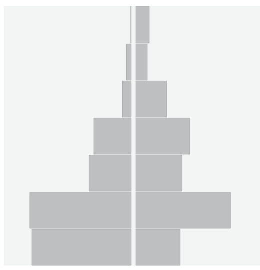
6.5 v-plots: Hybrid Distribution Chart Design

Based on the findings of our survey, we develop the *v-plot designer*; a chart authoring approach which facilitates the design of customizable hybrid charts, so-called *v-plots*. In particular, these charts combine the advantages of several established representations to support local, global, and aggregated tasks in a unified representation. One v-plot compares two distributions, while a combination of different distributions can be arranged in a *v-plot matrix* which can automatically be sorted by similarity using a matrix reordering algorithm. In the following, we discuss the design rationale of v-plots and introduce the guiding wizard of the v-plot designer, which automatically tailors the style of a v-plot to a given set of analysis tasks to highlight particular distribution properties.

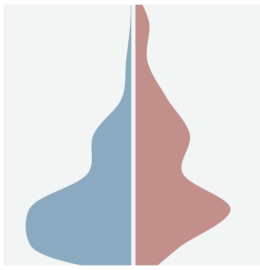
6.5.1 Design Rationale

The fundamental design rationale of v-plots is a layered representation, similar to the idea of the Summary plot [Pot+10]. Each layer supports different analysis tasks. The total number of layers is five ("**v**"), which gives this hybrid plot its name. The order of layers and their style can both be customized to focus on specific analysis scenarios. All layers are based on well-established visualizations, which makes them easy to use and interpret while still supporting a combination of complex analysis tasks. By default the v-plots are configured to enable all five layers to be visible but not highlighted. Users can adjust the layers and their highlighting by selecting certain tasks, either manually or through the guiding wizard, as described in Section 6.5.2.

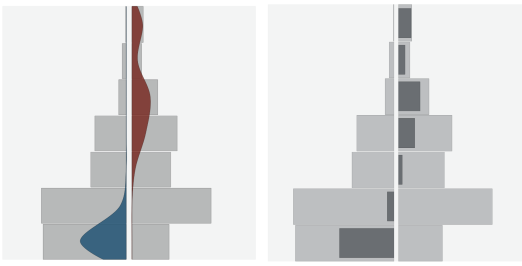
(i) Mirrored bar chart S. The first layer is a mirrored bar chart which supports local tasks on single distributions, such as the identification of frequencies (L1) or their comparison (L3). In discrete distributions, every bin corresponds to one particular value. For continuous distributions, an adjustable equal-width binning is applied. Small values are at the bottom; high values are at the top. The height of each bin corresponds to its relative frequency in the distribution, which also allows the comparison of distributions of different sizes.



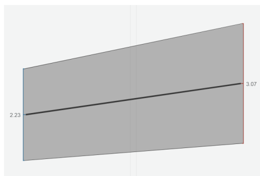
(ii) Density distribution Ⓢ. The global properties of distributions, such as type (G1) and skewness (G2), can be analyzed by a shape-based density distribution. This layer supports two implementations: (1) the center of each bin is used as a control point for a Catmull–Rom spline [CR74]; (2) a kernel-density estimation (KDE) with selectable parameters for the bandwidth and kernel type can be selected. The first option is the default, as it has three advantages compared to the KDE: (1) it is parameter-free, (2) it shows all peaks and valleys properly, and (3) it is directly linked to the underlying mirrored bar chart, linking global and local tasks together.



(iii) Direct difference encoding Ⓔ. We chose a vertical layout as the v-plots' default for supporting the comparisons between both distributions. The symmetrical arrangement allows one to easily see if the two distributions align or not. We also encode the differences of the distributions directly into the inner part of the mirrored bar chart as a *difference shape* or *difference histogram*. This encoding represents the absolute difference between the two relative frequencies and, for example, allows the highest difference between bins (L5), or the value ranges with low differences (G5), to be identified. While the difference histogram supports a direct comparison between bins, the difference shape supports the analysis of more general patterns and is often used in the v-plot matrix as shown below.



(iv) Statistic measures Ⓢ + Ⓔ. As a fourth layer, we support the encoding and comparison of statistical measures. For each distribution on the left and right side, the analyst can represent a value of central tendency (i.e., mean or median) and the spread of data (i.e., standard deviation, interquartile range, or standard error). The properties can be connected and highlighted with color for a comparative analysis.



(v) Labels. The final layer comprises various labels, such as the chart title, name of the distributions, a grid, and labels of the bins with the respective frequencies for a detail analysis. The size and position of the labels can be interactively adjusted.

6.5.2 Guiding Wizard for Task-Dependent v-plot Customization

Based on our survey results, we can guide analysts towards an optimized v-plot for specific analysis task combinations. After uploading their data, the user selects all tasks (c.f. Table 6.1) that are relevant for the current analysis question. Using different radio-buttons, the user can specify whether a task is *irrelevant*, *relevant*, or if the visualization should particularly *highlight* the corresponding property.

Automatic v-plot Optimization

The guiding wizard is implemented as a rule-based system. While the complete set of rules is documented in the supplementary material, we want to highlight the general concepts in the following: If all local tasks are marked as *irrelevant*, then the mirrored histogram layer is removed (the same holds for density distribution and statistical properties regarding global and aggregated tasks). Tasks marked as *relevant* will typically add elements to the v-plot. For example, if the user wants to identify the frequency of one value (L1), a grid and labels with the bin height are added. If the user wants to identify (A1) and compare the mean values (A6), the statistical layer is added and the mean values are connected, as shown in Figure 6.4. Changing a task from *relevant* to *highlight* usually results in a darker color and a higher level of the visualization layer. For example, if the user wants to highlight the differences (L5, G5), then the opacity of the difference histogram or difference shape is increased. We also change the difference histogram to the difference shape if only G5 is selected.

Automatic Recommendation for Basic Charts

The v-plot designer also provides an automatic chart recommendation which proposes basic charts which fit best for the given task combination. This allows for a comparison with the optimized v-plot. Based on the selected tasks, the system automatically provides a ranking of all visualizations (Figure 6.4 bottom) based on a score for each chart. This score is computed using a weighted linear combination of the Likert scale results in the survey, i.e. a chart considered *very useful* and *not possible* are weighted higher than charts considered *(not) useful*: $score_{vis_i} = w_1 \cdot ratio_{(++)} + w_2 \cdot ratio_{(+)} - w_3 \cdot ratio_{(-)} - w_4 \cdot ratio_{(--)}$. Here, $ratio_{(++)}$ corresponds to the ratio of participants rating the visualization as very useful **(++)**. $w_1 \dots w_4$ are the weights and by default are set to $w_1 = w_4 = 1.5$ and $w_2 = w_3 = 1.0$.

The ranked visualizations all show the distributions of the uploaded data so the user can compare them easily. The system also provides two perpendicular views of the ranked charts. First, we customize a usefulness chart based on the selected tasks and show the usefulness of every visualization to this task combination (Figure 6.4 top right). Second, we create a similar representation illustrating the usefulness of every selected task to each chart (Figure 6.6 bottom). Since not all task combinations can be covered with the existing charts, the user can then select one or multiple charts

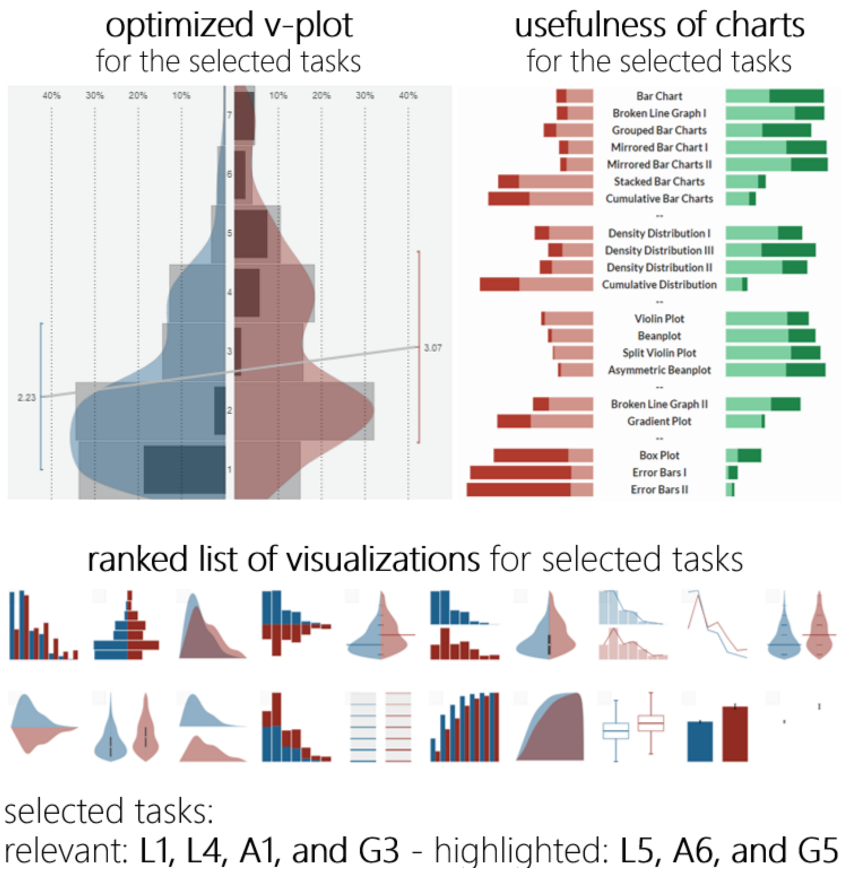


Fig. 6.4. Based on a set of analysis tasks, the v-plot designer automatically adjusts the v-plot to highlight distribution properties. We show 20 alternative visualizations using the same data, ranked by the perceived usefulness of the task combinations. For each chart type, we also select the top-ranked visualization (see Figure 6.6).

that cover the underlying analysis question, based on the automatic recommendation and the supportive charts.

While the findings in our survey are the grounding for the guiding wizard and chart recommendation, we want to highlight that this basis is interchangeable. New findings based on other quantitative user studies or the recommendation and guidelines of specific communities can be exchanged by replacing CSV files in our publicly available source code of the v-plot tool (see Section 6.5.4).

6.5.3 v-plot Matrix

v-plots are particularly designed to compare *two distributions*. For the comparative analysis of *many distribution* pairs, we extend the v-plot designer to generate a *v-plot matrix*, which arranges all pair-wise distributions in a matrix (see Figure 6.5). This layout allows analysts to compare one distribution against all others (one row or column), but also helps to find similarities and differences across all distribution pairs. To improve the perception of similarities, we allow users to apply a matrix

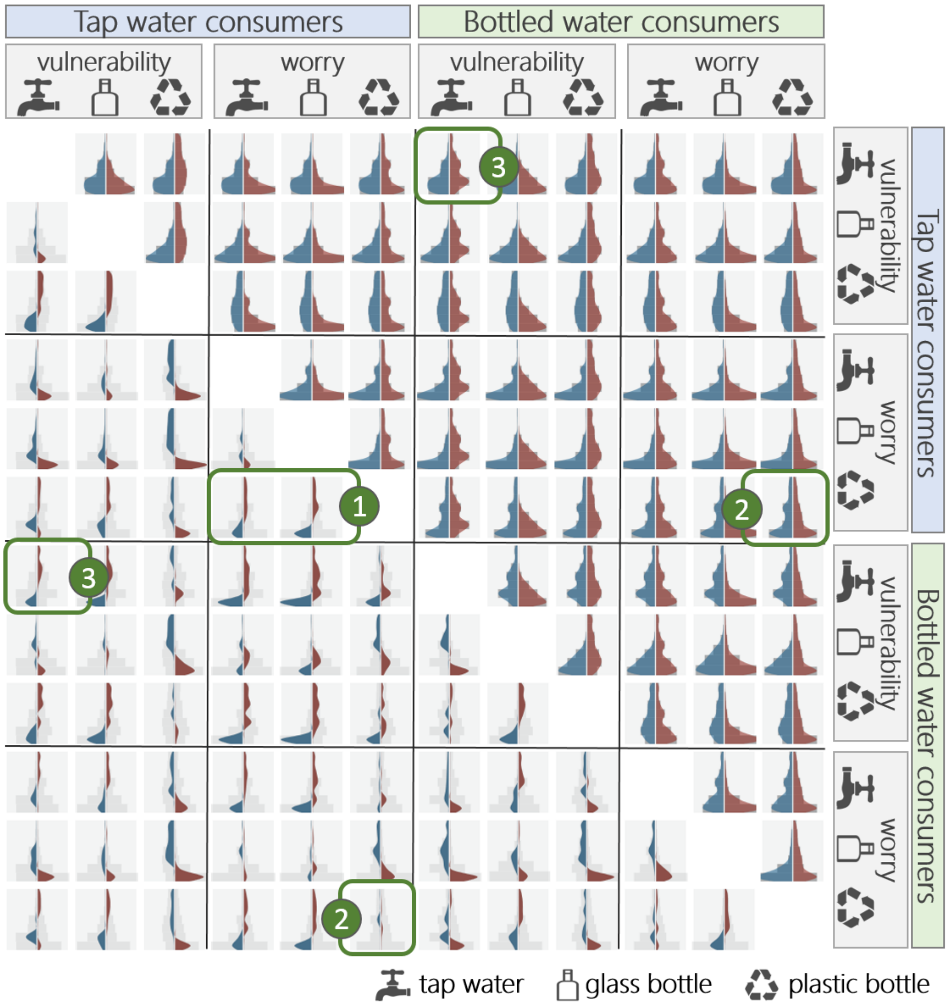


Fig. 6.5. A v-plot matrix comparing the risk perception of tap and bottled water consumers w.r.t. their *self-vulnerability* and *worry* when consuming water from a tap, a glass, or a plastic bottle. The upper triangle of the matrix shows a density distribution for global analysis, the lower triangle a difference encoding for comparison.

reordering algorithm to sort rows and columns such that similar distributions or similar difference patterns appear close together. However, rows and columns can also be arranged by semantics, as we show in Figure 6.5.

Matrices are symmetric, i.e., the upper and lower triangles depict the same distribution pairs. Hence, we support different v-plot styles for both triangles. Each style can be designed independently, either manually or with the help of the guiding wizard. In the example of Figure 6.5, we are interested in the general shape of the distributions (G1) and the frequency of each value (L1) in the upper triangle, and we want to know in which value ranges the distributions differ most (G5) in the lower triangle. Our guiding wizard therefore proposes starting with a *histogram + shape representation*, and a *difference encoding*. This layout has also generally proven to be useful for providing an overview and initiating a detailed analysis.

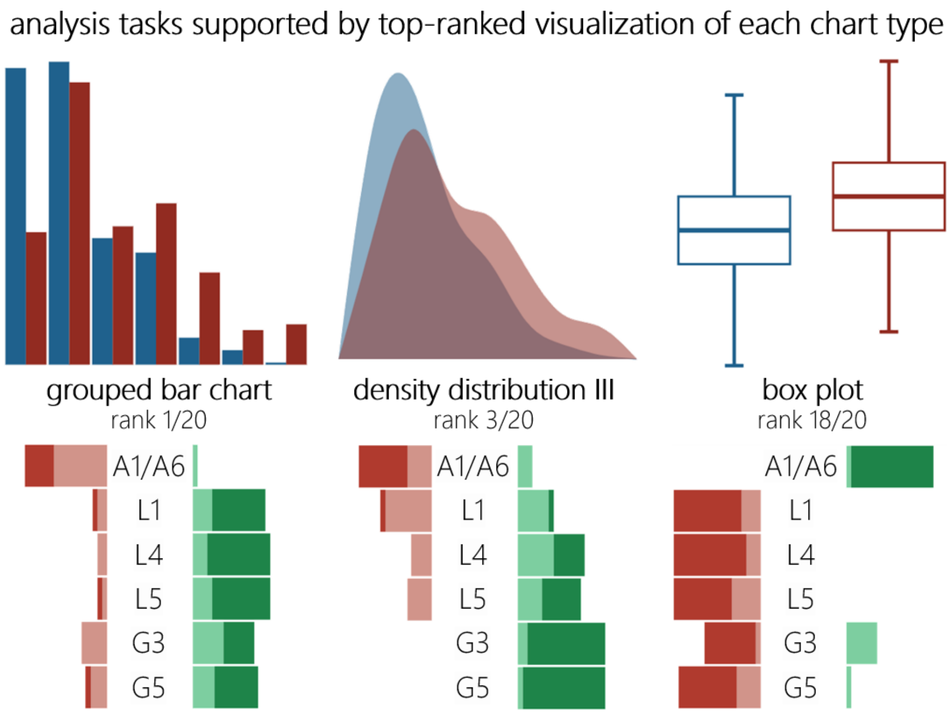


Fig. 6.6. Based on a set of analysis tasks, we select the top-ranked visualizations for each chart type and show their fitness to all tasks. For a comparison, we select the same tasks as in Figure 6.4.

6.5.4 Usability and Implementation Details

We implemented the v-plot designer using d3[1] and angular.js[2]. The tool is available at v-plot.dbvis.de, along with the source code, which will allow researchers to adjust the guiding wizard and the chart recommendation engine. Users can upload CSV files and directly compare the corresponding v-plot with 20 alternative charts, all using the same data for a useful comparison. All properties can be changed in an interactive menu, which immediately updates them in the v-plot. As shown in the supplementary video, users can, for example, change the size and aspect ratio, add a tile, modify the grid granularity, and adjust the color and transparency of the histogram, distribution shapes, and the statistical measures. Together with reordering the layers, users can thereby tailor the v-plots to highlight specific distribution properties. The resulting v-plot, as well as the customized style, can be downloaded.

6.6 Evaluation

To show the effectiveness and usability of the v-plot designer and guiding wizard, we conducted a qualitative expert user study and show a use case with survey results

[1] `https://d3js.org/`
[2] `https://angularjs.org/`

from a psychology application. All designs and interactions in this use case are inspired by the participants of our expert user study.

6.6.1 Use Case: Risk Perception in Drinking Water

Consider health psychologists investigating why many people buy water in (plastic) bottles, despite the numerous advantages of tap water. The psychologists conducted an experiment [Deb+18] in which tap and bottled water consumers reported their perceived risk with respect to *self-vulnerability* and *worry* when consuming water from the tap, glass, or plastic bottles. To investigate differences and similarities between the two consumer groups on a general level, we need to compare the distributions of 2 consumer groups (tap water and bottled water consumers) $\times$ 2 attributes (self-vulnerability and worry) $\times$ 3 drinking water categories (water from the tap, a glass bottle, or a plastic bottle) $= 12 \times 12 = 144$ combinations. To do so, we create a v-plot matrix (see Figure 6.5) and manually sort rows and columns by consumer groups and water types. By simultaneously analyzing the general shape and differences of the groups we make three interesting observations, as highlighted in Figure 6.5.

1 When analyzing *tap water consumers* and comparing *worry* of consuming bottled water in *glass vs. plastic*, we can see that tap water consumers are more worried about plastic packaging than glass. We can also see the same distribution difference in *worry* of consuming *water from plastic bottles* (more worried) vs. *water from the tap*. In summary, it seems that tap water consumers are generally more worried about drinking water from plastic bottles, compared to glass bottles or water from the tap.

2 When comparing *worry* for tap and bottled water consumers, we see that both consumer groups have very similar data distributions. This is also reflected in the difference encoding (lower-left triangle). This means that both tap and bottled water consumers have a similar risk perception about water from plastic bottles. Both distributions are also skewed towards smaller values, indicating a general tendency towards a low risk perception for both groups.

3 When comparing both consumer groups w.r.t. *self-vulnerability* when *drinking water from the tap*, we can see that the distribution of tap water consumers (blue distribution) is visually more skewed towards smaller values while the distribution of bottled water consumers is skewed towards higher risk perception values. This may indicate that bottled water consumers see water from the tap as a higher risk than consumers that generally drink tap water.

We investigate the last observation in more detail, focusing on the identification and comparison of mean values (A1 + A6), the frequencies across the distributions (L1 + L4), comparisons of distribution shapes (G3), and, in particular, highlighting distribution differences (L5 + G5). These tasks can be seen as an often-used combination for a comprehensive comparison of two data distributions. We start the guiding wizard, select the tasks above, and receive a v-plot tailored towards the underlying tasks. Figure 6.4 shows the v-plot, along with a ranked list of basic visualizations. For each chart type, we also select the top-ranked visualization (see Figure 6.6). Below each chart, we show its usefulness for each of the selected tasks. We can clearly see that these basic charts only support a subset of the tasks. For example, the density distribution support the global tasks to identify and compare the shapes of the distribution, while box plots support the identification and comparison of median values. The optimized v-plot from the same data, also shown in Figure 6.1a, supports all analysis question in a single chart.

6.6.2 Qualitative Expert User Study

We conducted a pair analytics study [KF14] with four domain experts, **DE1** – **DE4**, to evaluate and get feedback on our v-plot designer. All participants were PhD students in the field of psychology. Two were female, the age range was 25 to 33 years, and the reported experience in statistics varied from 4 to 10 years. We explained our aim of evaluating a chart authoring visualization to compare data distributions and asked all participants to bring a dataset which they are currently exploring. Each session took one hour and was structured as follows: The **DE** first explained the dataset and the visualizations that s/he commonly uses. Then, the visualization expert (**VE**) introduced v-plots with the different layers and parameter settings, and the **DE** analyzed this own dataset. Thereafter, the **VE** introduced the guiding wizard and the **DE** started tailoring the analysis toward specific task combinations. In a second step, we provided a new dataset [Kön+17], not known to the **DE** before, and let the **DE** create hypotheses and explore the data freely, allowing us to observe the participant's action and approach. As a last step, we introduced the v-plot matrix and let the participants explore the pair-wise relationships. We ended the study with questions about the general assessment of the usability of the v-plot designer with the guiding wizard. Occasionally, the **VE** asked for feedback during the study and guided the **DE** towards new tasks. The **DE** was encouraged to think aloud during the study so that we were able to capture their thought process.

Findings

Due to privacy constraints of the participants' own datasets, we will only summarize general findings. All participants reported that they normally use histograms as a first approach to get an overview of the data. **DE2** said that his/her biggest challenge is "*to keep an overview over the data and not get lost in the jungle of variables*". **DE2** also mentioned that s/he is satisfied with the visualizations s/he uses, but sometimes they are difficult to generate. **DE4** uses Tableau [Tab18] on a regular basis to create more advanced visualizations to compare global vs. local aspects.

After introducing the v-plots, **DE2** mentioned that having the distributions and aggregated statistics combined in one graph is useful. **DE2+3** liked that changes in the menu are translated directly to the chart, making step-by-step adjustments easily possible. **DE2** said "*the adjustment of the v-plot is very easy. Particularly compared to R and SPSS, where you need 100 clicks to do 10 changes. Here you need 10 clicks to do 10 changes.*" **DE3+4** raised some skepticism as the v-plots may show more than they actually need, which can make the analysis more complex. But they liked that "*there is lots of information that you usually only get by combining information from two different graphs and one table*". All participants particularly endorsed the difference encoding. Spotting differences was particularly easy with this feature, and the participants agreed, that this is a core task in their common analyses that is not supported well in the charts they typically use. All participants were able to find new insights into their data. **DE3** and **DE4** felt that even if the mean values did not differ between groups, there were interesting differences between the distributions that they spotted with the help of the v-plots and did not know beforehand.

We then introduced the guiding wizard, along with our classification of analysis tasks. By exploring different tasks and task combinations, the participants automatically optimized the v-plots and compared them to the alternative visualizations. We regularly asked for the visualization they liked most. **DE1** said, "*if I wanted to display one attribute (i.e. standard error), I would choose the box. For displaying a single task, I do not think the v-plots are necessarily better compared to other visualizations. But if I want to display multiple tasks, I would choose the v-plots. They are very good at displaying the combination of tasks.*" **DE4** agreed and added "*[for multiple tasks] I would definitely choose the v-plot, because the other charts only display single attributes. The more complex the attributes get [...] the more I would tend towards the v-plot.*" **DE2** preferred the bar charts for the global tasks, because they entailed more information in the view. Adding an aggregated task to a global task, s/he would manually add the mean level to the graph.

The wizard was generally liked by the participants. **DE3+4** stated that the wizard was particularly useful when getting started, because it suggests a quick and good starting point based on what is important for the current analysis. The alternative visualizations were positively emphasized by all participants. **DE4** said that "*the wizard shows me which visualization I can normally use and how attributes are displayed by the v-plot*". **DE3** said that the recommendations of alternatives are reasonable.

After introducing the v-plot matrix, **DE3** and **DE4** said that they initially found the v-plot matrix overwhelming. But all participants agreed that if one worked through the matrix and explored the different patterns, it's a good way to get an overview over the dataset and extract interesting attribute combinations. **DE2** particularly liked the matrix: "*I do not have to create v-plots one by one, I get the combination of all plots right away. I like it!*"

We asked the participants for general feedback at the end of the study. All mentioned that the v-plot designer was very intuitive, but one needs a few minutes to understand all the v-plots' layers. All participants agreed that they would use the tool for more than just getting an overview over new datasets; they would also try to incorporate the v-plots in a paper, poster or presentation as an eye-catcher and a dense source of information, if enough time or space for explanations was given. **DE1** said "*I would include a v-plot in my paper if displaying the combination of several attributes was important to me*".

6.7 Discussion

This section summarizes the main findings, lessons learned, and limitations by reflecting on our multistage research process, in particular summarizing the results of the qualitative expert study.

6.7.1 Summary and Lessons Learned

Our evaluation shows the advantages and usefulness of the v-plots and v-plot matrix when *simultaneously analyzing different analysis tasks*. Based on feedback from the participants and our own observations, we can summarize the following lessons learned.

Understanding the visual elements. Even though v-plots are designed in an easy way by combining existing charts into a unified representation, one still needs some training to understand and interpret all visual elements. **DE1** said that while s/he would also like to use the v-plots in papers and conference talks they are not very well-known, meaning extra time would be needed for explanations.

Usability and direct feedback. All participants in our study liked that changes in the menu are directly reflected in the v-plots. This helps to understand the impact of specific parameters and to adjust and tailor the v-plot design incrementally.

Difference encoding. All participants repeatedly highlighted that the difference encoding of the v-plots (i.e., the difference histogram for local analysis and the difference shape for global analysis tasks) is one of the most important visual elements. Compared to other representations, this facilitates concentrating on the differences (only) and so helps significantly when comparing distributions.

Single vs. combination of analysis tasks. Some participants reported that v-plots might be too complex for single analysis tasks. For example, they would prefer a box plot if the main goal is to compare the median and quartiles of different distributions. However, our user case and the feedback of the participants also show that existing visualizations often fall short when a combination of local, global, and aggregated tasks is required. In this case, the layered concept of the v-plots supports a comprehensive analysis.

Guided analysis and automatic chart recommendation. Comparing the v-plots with alternative charts was well received by the participants. In particular, our system automatically proposes the most useful charts, after which the participants could make an educated decision on whether v-plots are the appropriate technique for a specific task. Furthermore, they liked that the v-plots can be tailored to the analysis by highlighting specific distribution properties.

6.7.2 Limitations and Future Work

Our main focus is to design a guided authoring approach for hybrid charts, supporting comparative distribution analysis. To make such visualizations accessible and useful for non-experts, a central part of the v-plot designer is the guidance component that is grounded in a design study of analysis tasks and visualization techniques. In this section, we discuss four limitations of the current state of our approach and highlight potential for future research.

Coverage of visualization techniques. As presented in Section 6.3.2, there is a wide range of visualizations available for the analysis of data distributions. In this chapter, we deliberately focus on analyzing elements of some of the most-commonly-used, basic charts. However, we plan to extend this work to include more charts and visual elements. Most notably, we plan to include *other visual representations* such as dotplots as a potential additional layer to the customizable hybrid plots. We also intend to further investigate approaches for communicating *uncertainty* in the v-plot design.

Guidelines limited to scope of expert survey. To establish a foundation for user guidance, we relied on surveying the usage and analysis patterns of 20 practitioners. In addition to the theoretical foundation provided by the related work, the design guidelines were grounded in our survey. Further research and replication studies are needed to avoid potential sample biases based on the number of survey participants and their background. We provide all data from our survey at osf.io/jk8rp for transparency and have implemented the recommendations of our guiding wizard to be modular, i.e., subject to *adaptation and renewal through the availability of new findings*.

Lack of quantitative evidence. Going beyond the qualitative analysis and evaluation of our approach, there is a research opportunity to examine the *cognitive effects* of combining chart elements. Our evaluation suggests that the correct interpretation of the v-plot layers might be explained by the familiarity of the chart elements, as well as learning effects through usage. However, more studies are needed to determine the usefulness of individual components and their combinations. In particular, we plan a quantitative user study to evaluate the performance of the v-plots and to identify when it is beneficial to switch from a simple representation to the v-plots. This can further improve our guiding wizard.

Complexity of the v-plot matrix. We also plan to further improve our v-plot matrix. Some participants mentioned that presenting so many charts at the beginning of the analysis might be overwhelming. We therefore plan to add interaction concepts such as linking and brushing, highlighting, and attribute filtering directly to the matrix. We further want to automatically highlight interesting v-plots, for example with pattern matching and similarity search, as well as an automatically applied statistical analysis which only extracts significantly different distribution pairs.

6.8 Conclusion

*How can we make hybrid charts for the visual comparison of data distributions which (1) **simultaneously support local, aggregation-based, and global analysis tasks**; and are (2) **accessible to analysts**?* The current chapter addresses this research question by first classifying existing tasks for comparative distribution analysis and exploring the design space of appropriate visualizations. Based on a representative expert survey with 20 participants, we develop an automatic chart recommendation which proposes appropriate charts for a given combination of analysis tasks. As a second main contribution, we develop the v-plot designer as a chart authoring tool for hybrid v-plots, allowing data distributions to be compared simultaneously on global, local, and aggregated levels. Furthermore, we introduce a guiding wizard which tailors the style of the v-plots towards given analysis tasks. Our evaluation shows that this wizard helps to design effective v-plots through highlighting specific distribution properties. Once a combination of analysis tasks is relevant, v-plots outperform other techniques.

7 Simplifying High-Dimensional Data Analysis through a Table-based Visual Analytics Approach

Summary

We present SMARTEXPLORE, a novel visual analytics technique that simplifies the identification and understanding of clusters, correlations, and complex patterns in high-dimensional data. The analysis is integrated into an interactive table-based visualization that maintains a consistent and familiar representation throughout the analysis. The visualization is tightly coupled with pattern matching, subspace analysis, reordering, and layout algorithms. To increase the analyst's trust in the revealed patterns, SMARTEXPLORE automatically selects and computes statistical measures based on dimension and data properties. While existing approaches to analyzing high-dimensional data (e.g., planar projections and Parallel coordinates) have proven effective, they typically have steep learning curves for non-visualization experts. Our evaluation, based on three expert case studies, confirms that non-visualization experts successfully reveal patterns in high-dimensional data when using SMARTEXPLORE.

This chapter is *taken from* the following publication. Please refer to Sections 1.4 and 1.5 for the contribution clarification and general citation rules.

[Blu+18] **Michael Blumenschein**, Michael Behrisch, Stefanie Schmid, Simon Butscher, Deborah R. Wahl, Karoline Villinger, Britta Renner, Harald Reiterer, and Daniel A. Keim. *"SMARTexplore: Simplifying High-Dimensional Data Analysis through a Table-Based Visual Analytics Approach"*. In: IEEE Conference on Visual Analytics Science and Technology. 2018, pp. 36–47.

7.1 Introduction

Users need to *find* and *understand* clusters, correlations, and complex patterns in high-dimensional (HD) data for many applications. Consider, for example, diabetes experts, seeking to understand the eating habits of individuals or groups of patients. Factors to explore could include similarities in meal ingredients between patients from different cultural backgrounds, whether location and environment influence

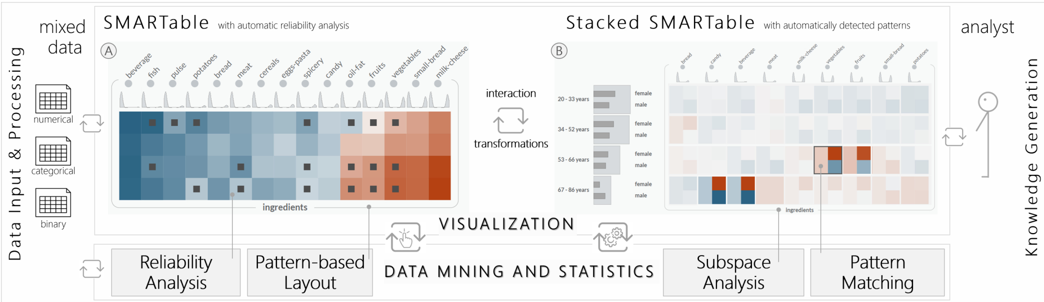


Fig. 7.1. The visual representation of SMARTEXPLORE is a so-called SMARTABLE. Descriptors like mean, variance, or deviation are computed, normalized per dimension or subspace, and mapped to a bi-polar or linear colormap. Manual and (semi-)automatic algorithms are executed through the visualization and support analysts in identifying and understanding clusters, correlations, outliers, and application-specific patterns in subspaces of the data. To increase trust in the patterns, statistical measures are computed on-the-fly and visualized along with missing values as overlays. Details on demand and a *stacked* SMARTABLE support detail analysis.

the subjective enjoyment of a meal, or which combination of influences do (not) correlate with age. Often, the datasets are not only high-dimensional but contain a mixture of different data types, such as, numerical, categorical, and binary.

To analyze such patterns, the InfoVis community has acknowledged the need for visualizations and interactive tools to deal with the overwhelming complexity and the large amount of data. A broad number of approaches have been developed. However, they usually transform the data into abstract representations. Popular examples are Scatter plots, Parallel coordinates [Ins85], and linear and non-linear projections, such as PCA [Jol86] and MDS [Tor52]. While these and other approaches have proven to be effective for the analysis of HD data, they often require long training for non-visualization experts and influence the analyst's trust in the revealed patterns [Jen+18]. Even after applying the concept of an abstract visualization, interacting with records and dimensions is seldom intuitive. Instead, it requires mental effort to interrelate records, dimensions, and values in the original format with the representation in the visualization and vice versa.

We present SMARTEXPLORE, an intuitive approach which injects visual analytics (VA) concepts into a table-based visualization. Rows represent records or record groups and columns, dimensions. A broad number of (statistical) measures, such as mean or deviation, can be computed, normalized, and represented with different colormaps, as shown in Figure 7.1. Pattern analysis algorithms, reordering techniques, and interaction concepts support visualization experts and novice users, alike, to reveal patterns in large HD data. Whenever possible, algorithms are automatically applied to reduce the number of tedious or complex tasks. Our decision to develop an enhanced analysis system around a table representation is backed by the fact that HD data is usually given in a table format and that the majority of analysts are familiar with spreadsheet tools, like Excel. Over a long period, they have been trained to read such tables, modify, filter or reorder rows and columns; or compute new derivative measures, such as mean or variance. While table representations naturally have the disadvantage of an inflexible layout, recent tabular-based visualizations [Gra+13; Gra+14; CW16; Fur+19] have shown to be intuitive for a variety of user groups, even for complex analysis tasks. However, none of the existing approaches is designed to identify and understand patterns, such as clusters or correlations in HD (sub-)spaces.

The primary contribution of this chapter is to *simplify the identification and understanding of HD patterns* through a table-based VA approach. First, we describe a set of 13 requirements for table-based visualizations supporting the identification and understanding of clusters, correlations, outliers, and complex patterns. Second, we introduce SMARTEXPLORE with the following four contributions:

Automatic handling and aggregation of mixed data types. SMARTEXPLORE supports datasets with a combination of numerical, categorical, and binary dimensions which are displayed in a consistent, unified representation. Hence, patterns across mixed types can be analyzed easily. Appropriate similarity functions, statistical tests, and algorithms are automatically selected and applied based on the dimension type and its properties such as the distribution.

Simplification of complex data transformations. SMARTEXPLORE implements complex data transformations such as (recursive) record grouping, pattern analysis, and subspace detection with a similar interaction design as known from classical table manipulations such as filtering and sorting.

Automation of pattern identification and highlighting. Based on visual template matching and (semi-)automatic table reordering, SMARTEXPLORE supports analysts to identify and understand patterns across a large set of dimensions and record groups.

Trust-building through automatic reliability analysis. To increase trust, SMARTEXPLORE automatically computes and visualizes uncertainty and statistical significance. An appropriate test is selected based on the dimension type, sample size, and distribution.

To guide the reader through the different visual mappings and various interaction techniques, we introduce a guiding dataset called `food`. The dataset contains 2,571 meals consumed by 99 participants over a period of eight days [Vil+17; Wah+17]. Each meal (data record) contains a combination of numerical, categorical, and binary dimensions: For example, the amount of kcal, sugar, vitamins (numerical), where and with whom the meal was consumed (categorical), and a binary representation of ingredients such as meat, fish, potato, and milk. Each participant occurs multiple times in the data with all of his/her consumed meals. Potential analysis questions for research include *"How age and gender affect the eating behavior of people?"* Due to data privacy restrictions, we removed dimensions with sensitive information for the examples in this chapter of the thesis. Although we use this dataset as a running example, SMARTEXPLORE can be applied to any HD dataset with homogeneous and mixed data types.

To evaluate the usefulness of our proposed technique, we implemented a prototype. The source code and a running version, which allows data uploads, is available on our website: `http://smartexplore.dbvis.de`. As a secondary system design contribution, SMARTEXPLORE stores the visualization properties *and* the applied interactions in the URL parameters of the web application. This URL allows easy sharing of findings and intermediate analysis results among researchers and fosters academic discussions.

Next, we collect requirements for table-based visualizations and discuss them in relation to related work in Section 7.3. In Section 7.4, we introduce the visual design of SMARTEXPLORE and the interpretability of visual patterns, user-guided analysis concepts (Section 7.5), and the fully automatic pattern matching and verification provided by SMARTEXPLORE (Section 7.6). Afterward, we present the expert case study evaluation and conclude the chapter with a discussion.

7.2 Requirement Analysis

SMARTEXPLORE has been developed in close collaboration with domain experts from the psychology domain. Although this is not the only analysis domain with HD datasets, psychologists are especially often confronted with large tabular datasets from user studies. Based on their common analysis tasks, we collected an initial list of requirements for tabular visualizations. To be of practical use to a broader number of domains, we generalized the requirements by our own experience and requirements by related table-based VA tools. We see our requirement analysis tailored towards the vision of a **pattern-driven analysis** of HD data, in which *finding* and *understanding* of clusters, correlations, and other patterns is of imminent importance. In contrast, Gratzl et al. [Gra+13] propose a set of ten requirements to compare rankings of data records, Perin et al. [PDF14] specify eight requirements to encode, modify, and reorder raw data within a table, and the twelve requirements by Furmanova et al. [Fur+19] support the dynamic and hierarchical aggregation of rows.

Among all of these requirement lists, there is some overlap. All approaches call for a visual encoding of data values, manual or automatic rearranging and sorting of rows and columns, an interactive and responsive analysis refinement strategy, and data manipulation possibility. Most approaches require details-on-demand, applicability to datasets with missing values, and applying operations only on subsets of the data and/or dimensions.

However, while these and other requirements sound similar, their underlying purpose and implementation differs significantly (e.g., reordering to compare rankings vs. reordering to identify patterns like clusters). Therefore, we derived and generalized our requirements specifically for a pattern-driven analysis in HD data.

General- and System Requirements

R0: Support for Data- and Dimension Analysis. A system should support *finding* and *understanding* the following basic patterns in the data- and the feature space: ***(a) clusters of data records*** according to the given feature space and a chosen similarity notion; ***(b) clusters of dimensions*** for a given grouping/clustering of data records; ***(c)*** linear and non-linear ***correlations*** among two or more dimensions; and ***(d) outliers*** in records, groups of records/clusters, and dimensions.

R1: Persistent Representation. To reflect the analysts' mental model and to mitigate potential misinterpretations, the visual representation of the data *and* the analysis results should be kept consistent.

R2: Capabilities for Mixed Dimension Type Analysis. To find *patterns across multiple mixed dimensions*, a system should be able to analyze *numerical*, *categorical*, and *binary dimensions* in a single view. Separate views for different data types avoid revealing relationships among those dimensions.

R3: Interactive Response. Interaction with a visualization should run smoothly. Whenever possible, results of user interactions should be shown directly and without large delays. For operations on records and dimensions (e.g., **R9**, **R10**) the user should be able to interact directly with the visible data (elements) instead of becoming lost in abstract or non-related handles.

Scalability on Data Record- and Dimension Level

R4: Grouping Data Records. Users should be able to group a set of records into a group/cluster to reflect its similarity, and reduce the complexity of data. Besides manual grouping, established procedures, such as grouping by a given category, binning, and clustering of (a subset of) dimensions should be supported.

R5: Value Aggregation. All data records within groups/clusters should be meaningfully aggregated to support group comparisons. For every combination of group and dimension, multiple aggregated measures should be available. Users require standard statistical aggregations, such as *mean*, *median*, *min*, *max*, *variance*, and *standard deviation* for numerical dimensions. For all dimensions, users are typically interested in *distributions* plots.

R6: (Visual) Encoding of Aggregated Values. Aggregated values and distributions should be visually encoded, such that users can reliably assess their similarity and quickly retrieve relationships. The encoding should not only support a two-way comparison but also alleviate the challenging task of manifold comparisons (e.g., across multiple dimensions or clusters; see **R8**).

R7: Grouping Dimensions into Subspaces. A system should support users in finding subspaces that are semantically meaningful or revealing dimension and pattern relationships (**R0b**). Naturally, a dimension may be part of multiple subspaces. Thus, users should be able to interactively adjust subspace memberships (see **R10**) to reflect their personal understandings of the data.

Comparative Analysis of Record- and Dimension Level

R8: Comparison of Records and Dimensions. The visual arrangement, as well as the encoding (**R6**), should support users to compare records or record groups across large sets of dimensions. Simultaneously, a dimension or a subspace should be compared among multiple record (groups). The concurrent comparison of records and dimensions supports the user in comprehending the visible patterns.

Data Handling- and Transformation

R9: Operations on Record Groups. Users should be able to operate intuitively on record groups to find and understand patterns (**R0**), thereby facilitating the comparison of records and dimensions (**R8**). The following group operations are required: ***(a) select*** and ***highlight***, ***(b) filter*** and ***remove***, ***(c) change ordering*** (manual) and ***automatic sorting*** based on similarity or by dimension/subspace, ***(d) merge*** one or more groups, and ***(e)*** extend grouping by ***recursively grouping*** records within a cluster.

R10: Operations on Subspaces. Users should be able to interact with dimensions and subspaces to facilitate records and dimensions comparisons (**R8**). All operations should be provided for each individual dimension and subspace: ***(a) select*** and ***highlight***, ***(b) remove***, ***(c)*** change ***ordering*** of subspaces and dimensions within a subspace and ***automatic sorting*** based on similarity, ***(d) add new*** subspaces, and ***(e) copy and move***, dimensions across subspaces.

Reliability and Trust

R11: Reliability of Perceived Patterns. Users require support to assess the reliability of findings. In particular, users need visual/algorithmic support for assessing: ***(a) missing data***, ***(b) too small groups***, or ***(c) statistically (non-)significant*** patterns. A system should be able to remove record groups or dimensions, classified as unreliable by the user (see **R9** and **R10**).

R12: Provenance of Visualizations and Interactions. A system should support storing intermediate analysis results and their associated visualizations, including all applied operations. Hence, results can be shared among researchers and analysts can reiterate previous results, or follow promising new analysis paths (see also **R11**).

7.3 Related Work

In the following, we delineate SMARTEXPLORE from other tabular-based and general HD visualization approaches, and show similarities to existing works for mixed datasets and trust-building in VA.

7.3.1 Table-based Visualizations

The most commonly used representation for HD data is a spreadsheet, such as Microsoft Excel [Mic18b] or Google Sheets [Goo18]. These tools typically allow a wide range of row and column interactions, and let the user augment current analysis results with basic visualizations, e.g., bar charts and scatter plots. More interactive approaches support a larger set of visualizations, e.g., Tableau [Tab18], Spotfire [Spo18], Power BI [Mic18a], and JMP [SAS18]. All these tools use tables for their data representation and use more or less intuitive mappings into different visual representations. Tableau and Spotfire focus on visual analysis, while JMP represents the model-building and statistical analysis spectrum. However, the approaches miss a tight integration between algorithmic support and visualization. Although the sophisticated visualizations for parts of the analysis process are usually linked to the table, they still require frequent mental model adoptions and changes. Table lens [RC94] is one of the first approaches to overcome this problem. The data stays in a table format, but the values in the rows and columns are approximated by sparklines [Tuf06]. An interactive focus+context approach enlarges rows and columns of interest. FOCUS [SBB96] extends the idea through interactive queries that focus on data areas of interest.

The three approaches most related to ours are Bertifier [PDF14], Taggle [Fur+19; Fur+17], and I^F, F^I-Tables [CW16]. Bertifier implements the idea of Bertin's reorderable (glyph) matrix [Ber75] into an interactive tool. As in the original work, row and column ordering is the primary interaction concept for identifying patterns. Yet, it does not allow aggregating records or dimensions. Taggle features hierarchical aggregation of records, but compared to SMARTEXPLORE, the analysis goal differs. Taggle is used to compare aggregations on different granularity levels, rather than finding patterns across a large set of dimensions. I^F, F^I-Table uses two interlinked tables to compare records across a large set of dimensions and vice versa.

The visual representation of SMARTEXPLORE is also inspired by recent work in matrix visualization [Elm+08; AH04; Twe+15; KZG10]. Most matrix visualizations are static and cannot be interactively adjusted. In particularly, matrices mostly feature algorithmic approaches that optimize the layout for one particular visual pattern. However, as also stated in a recent survey [Beh+16b], these visual patterns do not necessarily align with the user's analysis question. As envisioned in [Beh+16b], SMARTEXPLORE implements a more adaptive, user-guided process, which goes further than just drag&drop approaches, such as presented in [SM05]. Many other table-based visualizations exist. However, their core analysis tasks and contributions differ from SMARTEXPLORE. LineUp [Gra+13] identifies multi-attribute rankings in a table-like representation containing stacked bar-chart-like visualizations. Similarly, Podium [Wal+18] lets the user adjust rankings update the weights of the underlying ranking function. Taco [Nie+17] visualizes change over time within an aggregated table. A popular technique to visualize the relation between sets is UpSet [Lex+14]. Domino [Gra+14] lets users interactively combine, arrange, and extract subsets of data from different sources within a combined table-based view.

The last category of related table-based approaches are tools that let the user find patterns using sophisticated sorting algorithms. Examples are SimulSort [HY09], Matchmaker [Lex+10], and StratomeX [Lex+12].

In recent years, the InfoVis community presented a multitude of novel table-based visualization and VA systems. Most of these systems show that a representation in a table supports the users in the analysis process. However, the focus of the presented techniques is different from SMARTEXPLORE, as it combines sophisticated aggregation and grouping features, with pattern matching and an automatic reliability analysis.

7.3.2 Visualizations for High-dimensional Data

The community has presented many approaches for the analysis and visualization of HD data. Each approach has its advantages and disadvantages, and their discussions fill entire surveys [Liu+17].

With respect to our work, most approaches present specific solutions and trade-offs by focusing either on data vs. visual scalability, and complexity vs. understandability. For example, the seminal work of Inselberg on Parallel coordinate (PC) [Ins85] advanced the field by focusing likewise on dimension and record scalability. Many improvements for PC have been proposed. For example, highlighting density [Zho+08; MM08] and quality metrics [Beh+18] which reduce visual clutter [ED06b; PWR04], or reveal specific patterns [DK10] by reordering the axis. Analogue to the idea of Ankerst et al. [ABK98], the dimensions of SMARTEXPLORE can be reordered by visual similarity or particular visual patterns across multiple dimensions.

Orthogonal projections, such as in its bivariate form in Scatter plots, are also used for HD analysis. Here, the dimension interpretability and scalability is sacrificed for a better understandability of data record relations. Yet, a large set of possible dimension combinations has to be assessed for its usefulness. Tatu et al. and Albuquerque et al. present image and data-space quality metrics to quantify patterns in large sets of Scatter plots [Tat+09; Alb+10]. Non-linear [MH08; Tor52] and linear projections [Jol86] are classic approaches for HD analysis and visualization. In the context of dynamic graph analysis, van den Elzen et al. use a 2D projection (t-SNE) of topologically similar graph snapshots for their argumentations [Elz+16]. Visual complex glyph designs layouted with 2D projections are presented e.g., in [Keo+06; Cao+11].

To improve the understandability of HD datasets, navigation and user-guided exploration techniques have been presented recently with LDSScanner [Xia+18], in Subspace Voyager [WM18], and in Dimension Projection Matrix/Tree [Yua+13]. Fernstad et al. [FSJ13] propose a quality-metric guided framework for exploratory dimensionality reduction. Based on a large set of quality metrics, users can interactively rank and weight variables to reveal HD patterns. SMARTEXPLORE also supports metrics to identify visual patterns in the aggregated table. However, the metrics are computed in the image space which mimic the perception of analysts [Beh+18].

Only few approaches tackle HD dataset with mixed data types. The reason for this is the incompatibility of types w.r.t. distance functions and visual mappings. Often, different representations for different types (e.g., [Fur+19]) are used. Approaches focusing on the relation between data records typically apply the Gower distance [Gow71] and project the data using MDS [CC00] into a Scatter plot. To identify relations between dimensions, categories are transformed into comparable numbers based on an application-dependent ordering or distribution [Gre17]. The transformation can be done automatically [Ros+04] or with the help of analysts [JJJ08]. SMARTEXPLORE also transforms the distribution of categories into a numerical representation and visualizes it with the same encoding as numerical and binary dimensions within the table. This allows an easy identification of outliers and patterns in record groups across large sets of mixed dimensions.

Building trust in analysis results requires showing potential uncertainty along the analysis process. SMARTEXPLORE presents an automatic reliability analysis, which *automatically* determines and executes the correct statistical test from a set of 15 mathematical models. SMARTEXPLORE is influenced, by Correa et al.'s work on reflecting uncertainty aspects with visual mappings [CCM09]. Similarly to Buchmüller et al. [Buc+15], we use a semi-transparent random noise and colored overlays to represent the uncertainty of the computed descriptors.

7.4 Visual Design in SMARTEXPLORE

We introduce the SMARTEXPLORE technique and show how it addresses the specified requirements. In Section 7.5 and 7.6 we show how interaction and automatic algorithms can support users when finding and exploring high-dimensional patterns.

7.4.1 Visual Design for Aggregated Features

We define the basic visual representation of SMARTEXPLORE as SMARTABLE: data records can be grouped into clusters and dimensions to meaningful subspaces. The values in every record group are aggregated to its distribution or (statistical) measures such as mean or variance. We show an example of a SMARTABLE in Figure 7.2 based on the `food` dataset. The analyst has grouped the meals by type. The first row contains breakfast meals, then lunch, supper, meals consumed during coffee breaks, and snacks. Only dimensions in the ingredient and nutrition subspace are visible. The color gradient (white $\rightarrow$ red) describes the average number of meals containing a particular ingredient. Analysts can clearly see that ingredients towards the right dominate breakfast meals (except for the dimension *bread*), and ingredients on the left are mostly consumed during lunch and supper. On the left side of the table, users can compare the size of the record groups with the help of a histogram. The distribution of values in every dimension is visualized as distribution plot on top of each dimension. During the entire analysis, SMARTEXPLORE remains in this representation (**R1**) to reflect the mental model of users. Different visual overlays

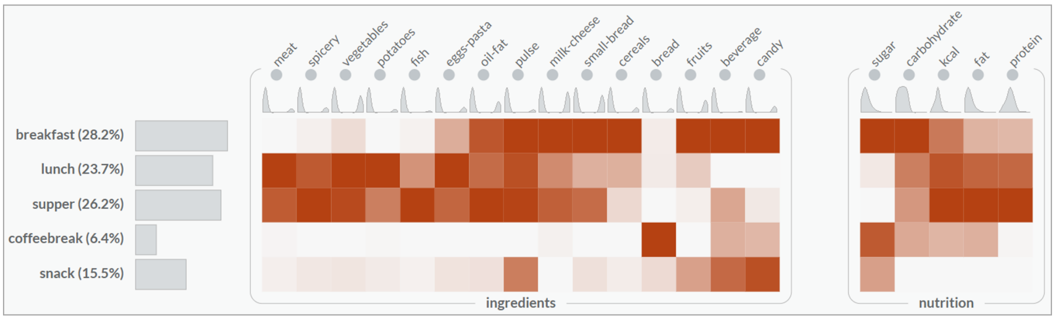


Fig. 7.2. Example of a SMARTABLE: meals are grouped by type (rows). A color gradient (white $\rightarrow$ red) is used to describe how often a specific ingredient and nutrient (dimension, column) is part of a meal.

help to visualize mixed dimensions in a homogeneous view (**R2**) and highlight the results of automatic algorithms for pattern reliability analysis (**R11**).

We do not claim any superiority of our approach compared to established HD visualizations. However, in this chapter, we show that a table contrasts well with abstract visualizations when equipped with VA tools. The well-known structure of rows and columns corresponding to records and dimensions respectively, appears advantageous for visualization experts and non-experts alike. It supports analysts in easily understanding the visual structures, and intuitively operate on record groups (**R4**, **R9**) and dimensions (**R7**, **R10**). Although the layout of tables is restricted to a grid, table cells can be arbitrarily complex. We show this aspect with our automatic reliability glyph (**R11**), which descriptively summarizes statistical reliability tests.

Data Record Grouping

Record grouping and clustering (**R4**) are useful means for spotting global patterns in the data and reducing the complexity. Additionally, analysts are often interested in understanding the properties of a given natural grouping in the data, e.g., compare different meal types as shown in Figure 7.2. SMARTEXPLORE supports different record grouping strategies, useful for different applications and data types.

Existing groups. Categories naturally provide semantic groupings. All records with the same category can be combined into one group.

Binning. To find groups in numerical dimensions, *equal-width* or *equal-height binning* can be applied. In our implementation, we show an interactive preview in which the user can freely experiment with the bin size and instantly see the binning result in a histogram.

Clustering. An algorithmic solution for finding groups of data records is to apply clustering. However, not all dimensions in a dataset may be relevant to determining application dependent clusters. Therefore, a user can select a subspace of dimensions to be considered. Finding a good parameter setting and a good number of clusters, in particular, is challenging. In SMARTEXPLORE, we compute a hierarchical clustering [HKP11] and let the user interactively adapt cluster numbers using a threshold in a visualized dendrogram. For numerical subspaces, a Euclidean distance, for subspaces with mixed data types (**R2**), the Gower metric [Gow71] is used.

Descriptors: Aggregated Values of Record Groups

Every record or record group is represented by one row in the SMARTABLE. Comparing record groups across dimensions, and dimensions across record groups (**R8**) is a central analysis task for SMARTEXPLORE. Consequently, all values within a group need to be *aggregated* (**R5**) and visually *encoded* (**R6**) to foster comparability. We define *aggregated values* synonymously as *descriptors*.

Descriptors for consistent dimension views. In our prototype, we decided to implement the following descriptors: For **numerical dimensions**, we support the mean, median, min, max, variance, and standard deviation. The values in **binary dimensions** are $true = 1$ and $false = 0$. As a descriptor, we compute the mean, which corresponds to the percentage of records with the value $true$. This descriptor is also used in the example in Figure 7.2 to show the frequency of ingredients for meal types. In **categorical dimensions**, a user is oftentimes interested in the distribution of categories. Here SMARTEXPLORE supports the visualization of the distribution as an overlay.

Descriptors for mixed dimension views. The aim of SMARTEXPLORE is to visualize all dimensions, independent of its type in a consistent representation **(R2)**. Therefore, the aggregated values of mixed dimensions need to be represented by a descriptor that can be compared across data types. SMARTEXPLORE proclaims a so-called *deviation descriptor*. It measures the deviation between the descriptor of a record group and the same descriptor for the entire dimension. For example, for numerical and binary dimensions, we compute the difference between the mean of a group and the mean of the dimension. The deviation descriptor of categorical dimension is defined as the Euclidean distance between the frequency histogram of a dimension and within one group. While the deviation is computed differently for all data types, the intuitive understanding is the same: the value or distribution differs (much) from the overall dimensions. Hence, users can quickly spot dimensions or groups with less/more deviation and then continue the analysis with different descriptors.

Visual Representation of Descriptors

To allow a fast comparison between record groups and dimensions, SMARTEXPLORE encodes (**R6**) the computed descriptors by color (**R8**). Similar measures are represented by similar colors, thus helping analysts to spot patterns. In the literature, there is a myriad of guidelines which help users to select an appropriate colormap for a specific task (e.g., [Mit+15; ZH16; Buj+18]). During the development of

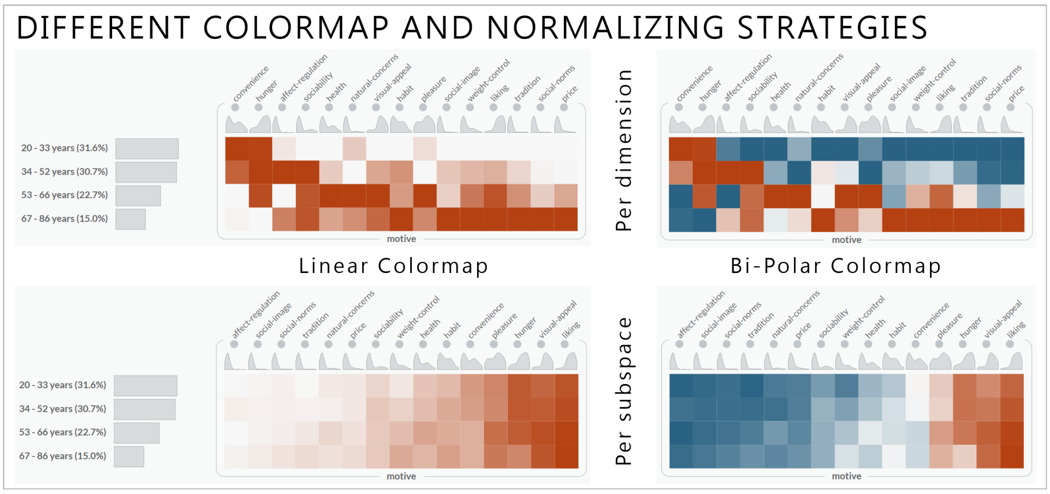


Fig. 7.3. Comparison of *linear* (left) and *bi-polar* colormap (right) with *normalizing per dimension* (top) and *per subspace* (bottom). Visualized *descriptor*: *mean* per record group. All four examples represent the same data. The ordering is based on visual similarity.

SMARTEXPLORE, as well as many discussions with potential and active users, we found that there are two classes of colormaps that appear to be useful for analysis: *linear* and *bi-polar* as depicted in Figure 7.3. We implement a linear colormap white (low) $\rightarrow$ red (high), and a bi-polar colormap blue (low) $\rightarrow$ white $\rightarrow$ red (high value). While linear colormaps enable users to directly compare two descriptors, bi-polar colormap are a great tool for identifying descriptors with high and low values.

Normalizing Strategies for Descriptors

Normalizing is essential for promoting the visual prominence of patterns in SMARTEXPLORE. Since we apply the concept of aggregations with respect to records and dimensions, we need a flexible mechanism to normalize distributions. The intuition of the two implemented strategies is given in Figure 7.3. Per default, descriptors are **normalized per dimension**, considering the descriptors of all record groups *within one dimension*. This strategy supports users to easily spot high, middle, or low values, but sacrifice the descriptors' comparability across dimensions. As a result, users can find patterns across multiple dimensions - even of dimensions with a different scale. Users can directly compare descriptors by **normalizing across dimensions** of a subspace (Figure 7.3 bottom). In this case, the min and max within an entire subspace are used. This strategy only makes sense if all dimensions have semantic connections and have the same dimension scale. However, if this strong requirement holds, we allow users to derive conclusions from this fact, e.g., quantify descriptors across multiple dimensions. While scaling can be checked automatically, semantic interpretability needs to be determined by the user. Descriptors can be normalized *linearly* or *logarithmically*. Additionally, we allow users to inject domain knowledge by manually setting min and max; e.g., for a manual outlier correction or scale capping.

Subspaces and Dimension Grouping

SMARTEXPLORE allows to group a subset of dimensions into a so-called *subspace* (**R7**). Every subspace contains at least one dimension and has a label which can be set by the user. A particular dimension can be part of more than one subspace. The reason for grouping dimensions into subspaces is twofold: First, it reduces the complexity of the visualization by introducing *visual gaps* between groups of dimensions that are semantically meaningful. Second, all visualization properties, such as normalizing strategy, colormap, sorting of dimensions etc., can be adjusted per subspace. This means, a user can group dimensions that should be treated similarly into a subspace, and select different properties for different subspaces.

7.4.2 Visual Design for Stacked Record Grouping

So far we have considered the *elementary aggregations* of the data records: A single dimension or a clustering algorithm determines the grouping of the data records. Every aggregated group is visualized as one row in the SMARTABLE. In many applications, users are interested in details of the aggregated rows. Consider for example Figure 7.1 (B). Records are grouped by *age* into four groups. Users may now be interested in *similarities/differences* between male/female *within each aggregation*. To support this analysis task, SMARTEXPLORE implements *stacked aggregations* (**R4b**). Each age group is further aggregated into a second level by the dimension *sex*. The distribution of both aggregation levels is shown by the histograms on the left side. The descriptor (here: mean) of the first aggregation level is represented by the upright rectangular and the descriptors for the stacked aggregation by the smaller squares on the right side. The stacked aggregation help users to analyze whether there is a difference in the descriptor when considering a more fine-grained aggregation. In Figure 7.1 (B), we can see that the mean value of the dimension *vegetables* for the age group 53-66 years (marked) is light red. Male participants within this group have a much smaller mean value (dark blue) compared to female participants (dark red).

Stacked aggregations can also be created for more than two groups in the second level. For example, we can aggregate the records first by the attribute *age* into four groups and then by the *meal type* into five groups in the second level.

7.4.3 Interpretation of Patterns in SMARTexplore

Every system with an elaborated visual design allows identifying and describing *how* the occurring visual patterns need to be interpreted and how they support the analysis process. An overview of SMARTEXPLORE's most common visual patterns can be found in Figure 7.4 and 7.5, along with a mapping of the analysis task **(R0)**.

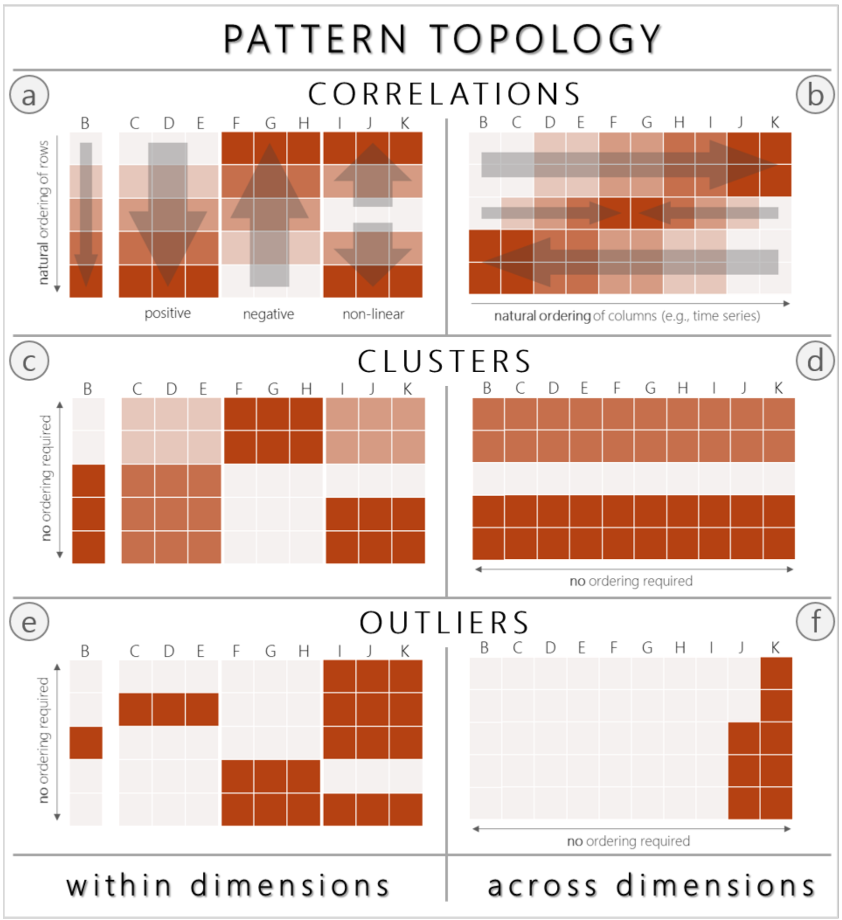


Fig. 7.4. Most important patterns in SMARTABLE. Correlations, clusters, and outliers can occur *within* and *across* dimensions.

Patterns Within and Across Dimensions

We have to distinguish between patterns existing *within a single dimension* and patterns *across multiple dimensions*. *Within dimensions* refers to patterns within a single dimension based on the current record grouping. For example, we see correlations, clusters, and outliers for dimension B in Figure 7.4 (left). Patterns *across dimensions* allow relating and comparing descriptors across multiple dimensions - typically all dimensions of a subspace. For example, we can see correlations, clusters, and outliers for different record groups across the dimensions $B, ..., K$ in Figure 7.4 (right).

Understanding Correlations

Analysts have to distinguish two notions of correlations (**R0c**):

Correlations between dimensions and record groups stand out as color gradients within one dimension (e.g., dimension B). Assuming that aggregated rows are in ascending order, Figure 7.4 (a) shows dimensions with *positive*, *negative*, and *non-linear correlations*. Dimension groups can be clustered into a subspace to foster interpretability (e.g., dimensions C, D, and E).

Correlations across dimensions are independent of an ordering of the aggregated rows; however, they require an ordering among the dimensions (e.g., dimensions representing the values of a time series). Figure 7.4 (b) shows an example of correlations across the dimensions $B, ..., K$.

Understanding Clusters

Three cluster types can be analyzed with SMARTEXPLORE:

Clusters of similar dimensions. Visually similar dimensions can be clustered into a subspace **(R0b)**. Hereby, the structure of the pattern does not matter. In Figure 7.4 (a) and (e), dimensions with the same correlation and outlier pattern are clustered (e.g., F, G, and H).

Clusters within a dimension. Data records or record groups with similar descriptors can be perceived as a cluster **(R0a)**. For example, the same value distribution in dimension B is shared among the first two and last three record groups in Figure 7.4 (c).

Clusters across dimensions. Figure 7.4 (d) depicts three clusters of record groups which are described by all dimensions of the subspace.

Understanding Outliers

An outlier is defined as a computed descriptor which differs substantially from all other descriptors. Based on the normalizing strategy, all descriptors of a dimension, or the descriptors of all dimensions of a subspaces, need to be taken into consideration when determining an outlier. Two types of outliers can be analyzed:

Outliers within a dimension. Figure 7.4 (e) depicts an outlier in the dimension B in the third record group.

Outliers across multiple dimension can be found in Figure 7.4 (f). To find this pattern, dimensions need to be normalized per subspace. All record groups of a dimension can be considered outliers (dim. K), but also only as a subset of the groups, as shown in dimension J.

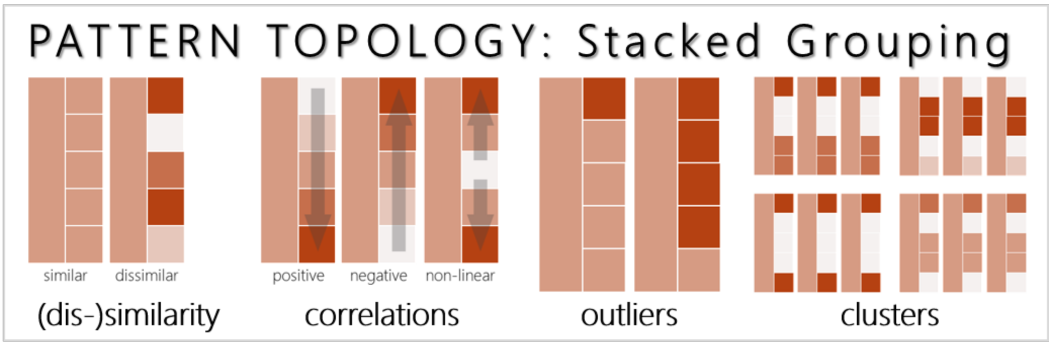


Fig. 7.5. Common patterns in stacked SMARTABLE. Most importantly is whether the base and stacked descriptor are similar or not. If they are dissimilar, correlations, outliers, and clusters can be present.

Understanding Patterns in Stacked Aggregations

Stacked aggregations help users retrieving commonalities and differences across subcategories. Generally, there are two possibilities: stacked and base descriptors have the *same color*, or have a *different color*, implying descriptor similarity or dissimilarity, respectively. As shown in Figure 7.5, correlation, outlier, and cluster patterns exist in stacked groupings. Of course, the pattern depends on the ordering of the records in the stacked aggregation.

Application Specific Patterns

In our experiments with psychologists, we came across interesting patterns that cannot universally be described by a topology, instead, they depend on the analysis. For example, multi-modal distributions within dimensions, or 'expected' outliers in clusters of records or dimensions. SMARTEXPLORE can also be used to find and understand such application dependent patterns. However, analysts are necessary for the pattern interpretation.

7.5 User-Guided Analysis in SMARTEXPLORE

SMARTEXPLORE provides easy to use interaction concepts to find interesting patterns, analyze records across dimensions, and dimensions across records, which are introduced in the following.

7.5.1 Interaction on Record Groups (R9)

As known from spreadsheet applications, users can *select and highlight* one or more record groups to compare the descriptors across all dimensions. By means of drag&drop, the *order of rows* can manually be *changed*, e.g., to compare record groups temporarily. Users can change the ordering of clusters to reflect semantic relationships, such as the temporal order of meals (Figure 7.2). After the grouping, analysts can also select a dimension and reorder the groups based on the dimension's descriptors. This 'sorting operation' helps identify potential correlations, clusters, and outliers of the selected dimension.

Users can *delete* entire record *groups* to remove outliers, and to tailor the analysis to a specific task. Record groups can be *merged*, the grouping granularity can be *customized*, and non-uniform binning of records is supported. One application scenario is shown in Figures 7.1 and 7.3, where participants were grouped into ten bins, and then manually merged into four application-specific groups. During the analysis, users can freely adapt the grouping and *apply stacking*.

7.5.2 Interaction on Dimensions and Subspaces (R10)

Grouping dimensions to subspaces provides a useful mechanism to reduce the complexity of a HD dataset. While we still keep all dimensions for our analysis, we introduce a *visual* gap that separates subspaces. Thus, we are subdividing the SMARTABLE into small, semantically meaningful, and cognitively graspable subsets.

Analog to record groups, both dimensions and subspaces can be *selected*, *highlighted*, and *rearranged* for a better comparison. Assuming the user has built a mental model of the underlying relationship, SMARTEXPLORE allows *dragging&dropping* dimensions into interpretable subspaces, as shown in Figure 7.2.

The *ordering* of subspaces along the x-axis can be changed by drag&drop. This enables the user to arrange subspaces close together, fosters comparability and understandability, or to drag subspaces to prominent positions at the start or end of the table. Subspace can be *deleted* (irrelevant for analysis) or *cloned* (show in other context), and *new* subspaces with a customized name can be created on-the-fly (further semantic relationship). Users can *copy* or *move* dimensions from one subspace to another in order to reflect their interrelation in the current analysis task. The properties of a dimension (e.g., colormap, computed descriptor, normalizing strategy) can either be changed globally for all dimensions, or per subspace. For example, a user can clone a subspace, and visualize its different statistical facets, e.g., its mean and variance. Within a subspace users reorder dimensions by dragging&dropping or *removing* them entirely.

A semi-automatic grouping of dimensions based on their similarity helps with deriving non-obvious subspaces. We apply a *hierarchical clustering* on all dimensions and map similar dimensions to a subspace. As in record grouping, a slider allows interactively changing the *granularity* of clusters. SMARTEXPLORE applies a Euclidean distance between all descriptors of two dimensions. Intuitively, this means visually similar dimensions will end up in a cluster. The Euclidean distance can be weighted by statistical significance.

As one algorithmic contribution, SMARTEXPLORE supports *semi-automatic pattern highlighting* and *sorting*. Users can select a dimension of interest and sort the remaining dimensions based on similarity. Our pattern matching algorithm can also highlights all dimensions similar to this selection. Finding similar dimensions can help users to identify patterns across multiple dimensions. The similarity search can be restricted to a subspace or can be applied on all dimensions of the dataset. Similar to the hierarchical clustering of dimensions, a Euclidean distance, optionally weighted by significance, is used to determine the similarity between two dimensions. SMARTEXPLORE proposes the number of dimensions to be highlighted based on the calculated distance distribution. The user can modify the expected highlighting accuracy (precision vs. recall) with the help of a slider. Highlighted dimensions can be *copied* or *moved* to a new or different subspace. This analytic guidance feature allows users to define subspaces with specific visual patterns.

7.5.3 Interaction on Aggregated Descriptors

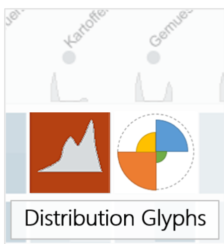
Semi-automatic pattern highlighting is also implemented for descriptors. In SMARTABLE, users can select a *cell of interest* and highlight the k-nearest neighbors. Searching for similar descriptors is most useful for stacked aggregations, as shown in Figure 7.1 (B). For the pattern highlighting of stacked descriptors, users can either consider only the patterns in the stacking, and ignore the value of the base descriptor (as applied in Figure 7.1 (B)), or choose a $50 : 50$ weighting to incorporate the base and the stacked descriptors. Both options have valid argumentations based on their use case. Here also, a slider allows defining the degree of (dis-)similarity, which should be considered in the analysis.

7.5.4 Details on Demand for Record-Level Analysis

A computed descriptor represents a data distribution in one aggregated value. However, the entire distribution should often be taken into account to obtain a valid pattern interpretation.

Distribution Overlay

The user can add a distribution overlay on top of each visualized descriptor. A kernel-density estimation is used for numerical dimensions, a histogram for categorical and binary dimensions. The kernel-density curve depends on the parameter *bandwidth*. We estimate a good selection of the method proposed by Silverman [Sil86]. Additionally, the user can change the kernel-density curve with a histogram, and, for the categorical dimension, change the histogram into a glyph representation, which is inspired by Star Glyphs [Sie+72]. The overall distribution



of a dimension gives users a first impression of the data and can help interpreting measures and removing outliers.

Table Lens and Tooltip

Often, a user is interested in seeing all distribution details for one record group and/or one dimension. For this purpose, SMARTEXPLORE implements a tooltip for a single cell and a table lens [RC94] for entire rows/columns. Hovering over a cell depicts the data distribution for the overall dimension and the record group, along with information about missing values, and results of statistical tests, such as the $p - value$ and the applied test (see Section 7.6.3 for details). Hovering over a record group or dimension enlarges the visualized descriptors and add data distributions, and values for descriptors, and/or statistical significance as shown in Figure 7.7.

7.6 Automatic Pattern Detection and Verification

SMARTEXPLORE has fully automatic exploration support, such as reliability analysis or table ordering, to increase trust in the findings.

7.6.1 Pattern-based Layout

The perception of patterns in the SMARTABLE depends on the ordering of rows and columns. Therefore, SMARTEXPLORE implements automatic sorting strategies to reveal these patterns. Since SMARTEXPLORE allows visualizing numerical, categorical, and binary dimensions, our internal heuristics can automatically select the correct distance functions for the involved data type (-combination). For all sorting strategies, similar descriptors, record groups, and dimensions should be placed close to each other. However, finding a good table reordering can be seen as an optimization problem in which row- and column positions can be freely changed without affecting underlying data interpretation [Beh+16b]. Yet, finding an 'optimal' solution is often computationally impossible or reveals the problem that reordering algorithms are inherently designed to foster the visual appearance of *one* visual pattern [Beh+16b].

Automatic Sorting of Groups and Dimensions

In SMARTEXPLORE, we can luckily restrict our search for an appropriate reordering algorithm to those approaches that are known to promote the visual patterns presented in Section 7.4.3. Hence, at least three options are possible: (a) the Barycenter reordering [MS05], the Bond-Energy algorithm (BEA) [McC+69], or Correspondence Analysis (CA) algorithms [Hil74]. Barycenter and CA are both fast algorithms but are designed to *only* retrieve groupings around the diagonal. CA implements a Singular Value Decomposition, which is not applicable if distances are less discriminative (i.e., binary or categorical dimensions). We decided to implement the BEA algorithm as it is a more conservative approach. This algorithm internally optimizes the 'measure of effectiveness', which fosters the visual appearance of groups independent of their relative location to the main diagonal. Moreover, these groups do not necessarily have to have a quadratic shape, but can also be rectangular.

Automatic Sorting of Dimensions

SMARTEXPLORE implements two strategies to automatically sort dimensions within a subspace based on the given ordering of dimensions. Same as before, this ordering is automatically applied until the user changed the ordering manually.

Sorting by average descriptor. The first approach sorts all dimensions of a subspace ascendingly by the average descriptor per dimension. An example can be found in Figure 7.1. While this sorting can be applied to both normalizing strategies, it is most useful when the dimensions are normalized across the subspace. As a result, users can quickly see which dimensions generally have higher/lower measures.

Sorting by visual similarity. The second strategy sorts the dimension by visual similarity. First, SMARTEXPLORE computes a distance matrix by all pairs of dimensions within a subspace. To do so, the previously introduced distance measures based on the (weighted) Euclidean distance are used. Afterward, we compute a one-dimensional multi-dimensional scaling projection of the distance matrix, similar to proposed in [Jäc+16]. We ignore the actual position in the one-dimensional layout but use the ordering of the projected dimensions. For stacked grouping, users can select which parts to consider for the layout: the base measure, the stacked measures, or a combination of both.

7.6.2 Automatic Pattern Detection

In Section 7.5.2, we describe how users can select a dimension of interest and highlight all dimensions that are visually similar. This user-guided analysis is particularly interesting for the application of specific patterns of interest. However, in most applications, users are primarily interested in linear correlations, clusters, and outliers as introduced by our pattern topology. SMARTEXPLORE supports users in automatically identifying these patterns: For each pattern-type, we defined a *template* describing the 'optimal' pattern for a single dimension. These templates correspond to the examples of the pattern topology, as shown in Figures 7.4 and 7.5. We adapt the size of the pattern to the number of rows in the (stacked) SMARTABLE. For patterns like the outliers in Figure 7.5 (e), we iterate the position of the pattern (here: outlier) through all rows of the dimension. Finally, the different templates for each pattern are matched against each dimension in the dataset - analog to the manual similarity search.

7.6.3 Reliability of Visual Patterns

While the visual design supports finding different data patterns, SMARTEXPLORE automatically and transparently supports the analyst in the question *"How reliable are these findings?"* (**R11**).

Statistical Significance and Visual Representation

Different colors for visualized descriptors naturally indicate that the underlying values are different. However, based on the normalizing strategy and the chosen colormap (e.g., bi-polar), the minimum descriptor (min) is mapped to blue and the maximum (max) to red. In the visualization, users cannot quantify the difference between min and max without using the tooltip or table lens. The same is true for all descriptors in-between. Therefore, SMARTEXPLORE *automatically* computes various statistical tests to assess whether differences are statistically significant or not. The following two levels-of-detail are considered:

S1: Significance of a descriptor. For every computed descriptor, a statistical test is used to decide whether it is significantly different from the overall dimension. To measure this difference, classical tests are t-tests to compare the mean (descriptor) with the mean of a dimension, Chi^2 tests for categorical dimensions, and a binomial test for binary dimensions.

S2: Significance of a dimension. To measure the significance of multiple descriptors at the same time, classical tests are an ANOVA for numerical, and a Chi^2-test for categorical and binary dimensions. These tests generalize the understanding of **S1** to an entire dimension, but do not indicate the significance of each descriptor.

Assumption-based selection of statistical test. Each statistical test relies on different assumptions that need to be fulfilled in order to achieve reliable results. In numerical dimensions, for example, analysts have to check whether the data follows a normal distribution (e.g., using the Kolmogorov-Smirnov test), for variance homogeneity for independent samples (using Levence's test), and sphericity for dependent samples (for ANOVA with rep. measures, using Mauchly test). The same applies to categorical and binary dimensions in which, for example, the sample size has to be taken into account. Following Andy Field [Fie13], there are 11 tests for numerical, three for categorical, and one for binary dimensions that apply to our application. SMARTEXPLORE supports the user by automatically selecting the appropriate test for each dimension. Based on the data type, the (in)dependence of samples, and the significance type **S1** or **S2**, SMARTEXPLORE computes all statistical tests and their assumptions. Appropriate test are selected as proposed by Andy Field [Fie13]. Then the test's $p - value$ is compared to a user-defined α to determine the significance of a dimension or a computed descriptor. The tooltip shows the $p - values$ of all tests and assumptions such that the user can compare their difference and reproduce the

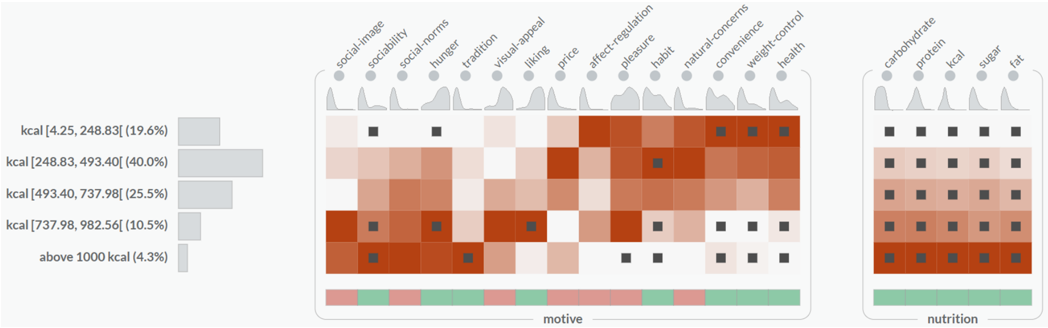


Fig. 7.6. Comparison of nutrition, eating motives, and calories per meal (rows). Meals, rich in calories, are merged into one record group (bottom). Significant dimensions and descriptors (mean) are marked.

system's selection. Users can also manually determine the applied test for a single dimension, a subspace, or globally for the entire dataset.

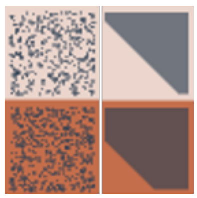
A $p - value$ only informs whether a statistical effect exists; it does not show its magnitude. The appropriate effect size (e.g., Cohen's d, Cramer's V) for the selected tests is also automatically computed.

Visual representation of statistics. The statistic results can be added to SMARTABLE. The significance of a descriptor **(S1)** is visualized by an overlay. Users can choose between a dot for significant descriptors (Figure 7.6) and a glyph that uses full size for significant and a smaller size for non-significant descriptors. Applying the first option, users can concentrate on the patterns and use the statistical information as added value. The second option modifies the visual representation such that significant results jump out and users can concentrate on areas with mainly significant descriptors.

To show the significance of a dimension **(S2)**, users can enable a red or green icon below each dimension (Figure 7.6). Also, an adaptive colormap can be used. Significant dimensions use the full range of colors, non-significant, only the inner part. As a result, users can still perceive differences in the descriptors, but they are visually less dominant as significant ones.

Missing Values

Missing values are common in many applications and influence the reliability of descriptors. Therefore, the visualization should highlight the areas in the data space which contain missing values and show their proportions. Otherwise, the uncertainty of calculated descriptors is not shown, and the visualization pretends a reliable pattern which does not exist in the underlying data. SMARTEXPLORE supports different visual overlays to show the amount of missing values. For example, the *glyph covering* adds a gray layer on top of the visualized descriptor



in order to reduce its expressiveness. The *texture overlay* covers the visualized descriptors with random noise, as used by Buchmüller [Buc+15]. Estimating the exact proportion of missing values is not possible. However, it is more intuitive as it seems there are 'holes' in the data, analog to missing values.

7.7 Effectiveness and Generalizability Evaluation

We evaluate SMARTEXPLORE for two general criteria: First, its usability and understandability for pattern analysis tasks, and second its generalizability to different datasets and domains.

Evaluating effectiveness

To assess the effectiveness and usability of SMARTEXPLORE, we conducted a qualitative expert user study with six participants. Our evaluation process is structured in a multi-stage evaluation process:

(1) We generate a set of 'ground truth findings' from the `food` dataset derived by two participants, who are familiar with the data due to earlier analysis using established statistics. Both subjects have no far-reaching VA experience, but continuously provided feedback during the development and use SMARTEXPLORE on a regularly basis. We refer to these participants as **E1** and **E2** as they are **e**xperts in both, the data and SMARTEXPLORE.

(2) We target the usability across *different expertise levels*, by conducting four pair analytics [KF14] studies with two different user groups. In the first group, two psychologists without VA experience, but good knowledge of the `food` dataset participated. We refer to these participants as **d**ata **e**xperts (**DE1** and **DE2**). In the second group, two **v**isual **a**nalytics experts (PhD students with one to three years of experience), **VE1** and **VE2**, participated. **VE1** and **VE2** did not have knowledge about the `food` dataset. We planned about one hour per pair analytics session per participant and conducted a semi-structured interview for gathering feedback, feature requests, and potential improvements. None of these participants (**DE1+2** and **VE1+2**) has been using SMARTEXPLORE before.

Evaluating generalizability

In order to showcase SMARTEXPLORE's applicability on datasets of various domains, we also let our VA experts **VE1** and **VE2** analyze the `university- ranking` dataset[1]. This dataset contains the top-1000 universities for the years 2014-2017, ranked according to nine different metrics, such as the quality of education, number of publications and patents. The metrics result in a numerical score, used to derive an

[1]Source: http://cwur.org; last accessed: 2018-06-26.

Tab. 7.1. Overview of expert users and their role during the evaluation.

	VA	Dataset(s)	Role	Method
E 1+2 Psych.	novice	`food`	ground truth generation	indep. analysis & feedback
DE 1+2 Psych.	novice	`food`	compare across expertise	pair analytics & interview
VE 1+2 CS	expert	`food` `university`	compare across expertise & data	pair analytics & interview

overall ranking. As before, we conducted a pair analytics study combined with a semi-structured interview which took 30 minutes in total.

An overview of all user groups, expertise levels, roles, and evaluation methods is shown in Table 7.1. In the following we will describe the results of each experiment in detail.

7.7.1 Insight Generation

During the last year **E1** and **E2** have been using SMARTEXPLORE in different stages of the implementation. Both experts primarily analyzed the `food` dataset. We are not able to report all findings in this chapter, but we will describe interesting usage scenarios and depict the general analysis process of the experts. According to **E1** and **E2**, finding *statistically significant* commonalities among a large set of semantically grouped dimensions, (e.g. eating motives or ingredients), is the most convincing argument for using SMARTEXPLORE.

The experts analyzed how age influences the preference towards certain ingredients (Figure 7.1 (A)). The dimensions are, hereby, normalized within a subspace to find coinciding products that are generally consumed a lot. The experts found (statistically) obvious insights easily, such that *milk*, *small bread*, and *vegetables* are generally consumed more often (dark red colors) than *fish*, *potatoes*, and *pulse* (dark blue colors). Older people (last row) seem to use more milk than younger people, a finding which could be later rejected due to its unreliability ($p - value$ of 0.06). **E1** and **E2** found that there is variance based on the gender, so they created a stacked SMARTABLE (Figure 7.1 (B)). In the group 53-66 years, the amount of vegetables is slightly above average (light red color), but differs strongly for male (less vegetables) and female (more vegetables). The experts made use of our automatic pattern retrieval functionality by selecting this pattern and searched for similar findings. The experts extended the analysis by comparing the age also to different motives (reasons why people consumed a specific meal; Figure 7.3). Different normalizing strategies and colormaps were applied. The SMARTABLE illustrates that motives like *convenience*, *hunger*, *affect-regulation*, and *sociability* might be more important for younger people, while older people are more motived by *price*, *tradition*, and *social* *norms* (top row). The experts also found that, generally, the motives *liking*, *visual-appealing*, and *hunger* are the most common motives. Further analysis results can be found in Figure 7.2 and 7.6 in which **E1** and **E2** analyzed the relation between ingredients and nutritions, respectively motives to consume meals with high/low calories.

7.7.2 Comparability across Expertise-levels

To analyze the influence of expertise-levels on the usefulness of SMARTEXPLORE we asked **DE1+2** and **VE1+2** a non-trivial analysis question: *"Which meal type is generally most unhealthy?"* Based on this controversial question, we gave the participants a ten minute introduction and showed them the most important features. All experts showed active interest in our available normalizing strategies, how to interpret particular visual patterns, and asked for the internals of our automatic computation of statistical tests; a circumstance of significant importance, especially in the psychology domain. After understanding that SMARTEXPLORE automatically selects the test based on all assumptions, **DE1** stated that SMARTEXPLORE "[. . .] not only lets us validate [hypothesis] significantly faster, but also mitigates the problem of choosing accidentally a wrong test".

After all open questions were answered, we asked the participants a second, more open analysis question: *"Which motives and ingredients relate to meals with high, middle, and low calories?"* All participants started analyzing the dataset by grouping the record over the dimension *kcal*. The grouping granularity, however, changed between the different user groups. While **VE1+2** used a grouping with more bins, **DE1+2** created only five bins based on a similar grouping in the literature; Figure 7.6. Independent of the grouping granularity, both participant groups where able to identify a (linear) correlation between *kcal* and all dimensions within the *nutrition* subspace (all statistically significant). **VE1** then merged all record groups with $kcal > 900$ into a single group to remove the distorted distribution of group size. The resulted groups are similar to manual groups of **DE1+2**. Based on this grouping, **VE1** could identify that meals with higher number of calories might be associated with the motives *social-norms*, *hunger*, *tradition*, and *visual-appealing*, while a lower number of calories corresponds to motives like *natural-concerns*, *weight-control*, and *health* (see Figure 7.6). Changing the granularity levels of the grouping by *kcal*, the computed descriptors alternated between significant and non-significant. These findings are in line with the 'ground truth' identified by **E1** and **E2**.

Afterward, all participants were motivated to continue analyzing the dataset based on their own interest. **DE1** expanded the search to other dimensions and continued with stacked aggregations separating male vs. female for different meal types. **DE2** started a completely new analysis and looked for patterns w.r.t. stress and mood before and after meals. **VE2** analyzed which ingredients and motives are related to a high body-mass-index. Surprisingly, a small number of participants tried to avoid food and ingredients with a high amount of sugar and calories. As *weight control* is one of the outstanding motives of this record group, **VE2** hypothesized that these participants may be planning or conducting a diet.

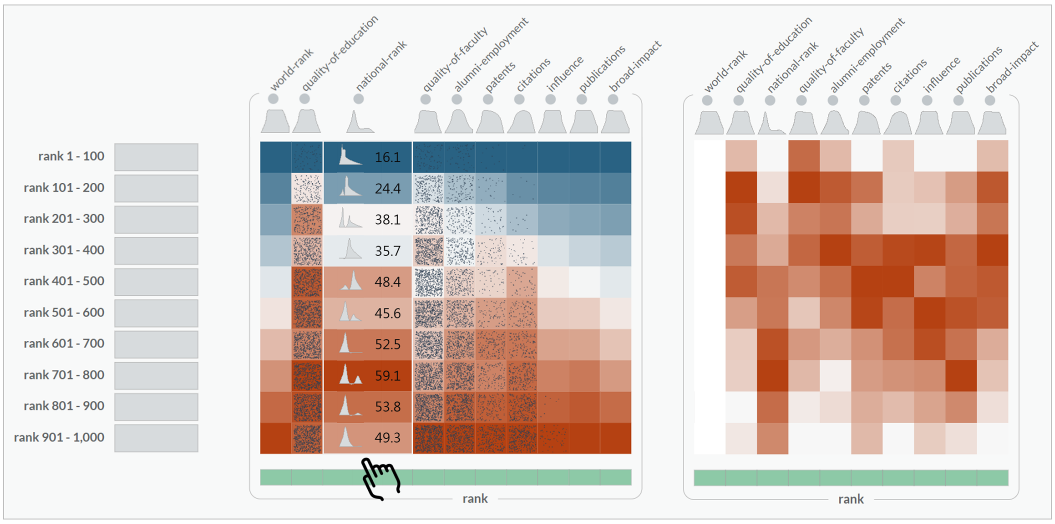


Fig. 7.7. Dataset `university`, grouped by world-rank, and visualized by mean (left) and variance (right subspace). The ten dimensions represent different ranking measures. Left: (blue → good rank and red → low rank); right (white → low and red → high variance). Missing values are shown by noise overlay. Table lens is used to investigate the data distribution and the descriptor of dimension *national rank*.

7.7.3 Comparability across Datasets

In a separate session, **VE1+2** started analyzing the `university` dataset. Both VA experts directly applied their experience from the first study and wanted to find out, which aspect correlates mostly with the overall ranking of universities. Therefore, the universities were grouped and binned by their world rank. Figure 7.7 shows the mean and variance descriptors for all dimensions. Missing values (universities with a $rank > 1000$ within one dimension) are visually highlighted with our random noise overlay. Both participants found effortlessly that all of the attributes correlate to the world rank (first dimension). However, there were two observations: (1) the ranking is not linear, and (2) there is a strong variance in all dimension. The dimension *national rank* is visually outstanding as the variance seems to be linearly correlated with the world rank. **VE1** continuously used the tooltip to get the data distribution while **VE2** used the stacked-aggregations to analyze differences in the different years. He found, for example, that the influence of the rank by *patents* changed significantly between 2014 and 2017. Both experts made use of the statistical tests for verification, but relied mainly on the pattern taxonomy, and the distribution overlay to generate findings.

The reported findings of the `university` dataset are rather an illustrative example than a comprehensive user study. However, we could show that SMARTEXPLORE can be used for other datasets as well and the usefulness is acknowledged by VA experts (see below).

7.8 Discussion

In our expert case studies we have shown that SMARTEXPLORE can be applied to various applications. Users with different data and VA expertise are able to identify and understand interesting patterns in HD data. Based on their feedback and our observations during the study, we summarize the following lessons learned:

Lessons Learned

Instant applicability through familiar representation. Both, the **E1+2** and **D1+2** participants have not been using sophisticated VA tools before to find patterns across a large set of dimensions. When we asked them to apply SMARTEXPLORE to their data and give feedback (**E1+2**), and to participate in our study (**D1+2**) the experts showed some skepticism on the usefulness. However, after only a few minutes the familiar representation of the SMARTABLE convinced them instantly to see its usefulness for their own data. Of course, applying SMARTEXPLORE to their own data helped them building a mental relationship between previous findings and the visual patterns. We were able to see that the participants fully understood SMARTEXPLORE by the following observations: (1) During the feedback sessions **E1+2** proposed useful extensions based on the concept of SMARTEXPLORE. For example, they suggested to clone entire subspaces for a comparative analysis using different statistic tests, and initiated the discussion for the automatic reliability analysis. (2) After a short training, **D1+2** directly applied the concepts of SMARTEXPLORE to their own analysis questions. They did not question our design choices but were immediately able to make sense of the visible patterns and explain interesting relationships. Therefore, we conclude that they were able to effectively use SMARTEXPLORE after only a short training phase.

Findings by automatic support. We realized that most participants acknowledged the automatic support of SMARTEXPLORE. For example, they liked that similar dimensions are arranged next to each other by default and appropriate statistic test are proposed. As a consequence, the participants were able to spot interesting patterns without any pre-configuration and parameter choices. Once an interesting pattern has been identified, the participants investigated the automatic selections and adjusted the settings.

Linking to classical approaches. Even though the layout of the SMARTABLE is quite fixed, sophisticated patterns could be detected by **VE1+2**. Both argued that our design choices along with the automatic support is helpful to identify and explain various patterns. Especially, they liked the possibility to analyze datasets with mixed data types. However, to confirm some of the hypothesis they proposed to transform a subset of the data to other visualization approaches. For example, to see the actual values of records across all dimension (and not just its descriptors), Parallel coordinates are useful.

Future Work

Although SMARTEXPLORE presents a sophisticated table-based VA system, we identified five areas for future improvements:

Data types. We have limited ourselves to datasets with numerical, categorical, and binary dimensions. While the analysis of these mixture datasets is itself challenging, e.g., due to the problematic definition of similarity and aggregations, a broad range of further data types exist. Text-, geo-spatial-, time series-, or relational datasets impose further challenges to both visualization and analytics.

Layout flexibility. SMARTEXPLORE's main visualization is a table which borrows the static layout of rows and columns. While it has significant advantages for a broad range of users, we envision a system that lets the user freely change back and forth between known layouts and, e.g., projection-based layouts to facilitate more intuitively high-dimensional similarity assessments.

Data and analysis provenance. In SMARTEXPLORE, we present an implicit data provenance approach: All analysis stages are encoded in the URL. However, we found that an explicit gallery or journal view would be highly appreciated by our user group.

Supporting hypothesis generation. Within the user study, **VE1** argued that even for unknown datasets users will need an initial hypothesis. **VE1** suggested to show small previous of different aggregations and orderings. Ideally these previous should be sorted and incorporate the user's interaction provenance. **VE2** had a similar idea by proposing to generally highlight relations in the data (e.g., correlation matrix) in order to guide the analysis.

Trust-building. One of SMARTEXPLORE's primary contributions is its automatic reliability analysis, which builds trust in the tool and its findings. Further, an algorithmic 'helper', such as subspace clusterings [PHL04] or subspace nearest neighbor search [Hun+15a], could be explored into (semi-) automated exploration processes.

7.9 Conclusion

Finding and understanding clusters, correlations, and complex patterns in high-dimensional data is a challenging task, especially if the underlying dataset contains a mixture of different data types. With SMARTEXPLORE, we present a fully functional table-based visual analytics technique that combines automatic analysis with user-guided- and purely interactive exploration. In an easy to use interface, our system automatically guides users to interpretable patterns and supports the exploration through semi-automated pattern matching and user invoked reordering. Our interaction concept, based on drag&drop, context-dependent menus, and on-the-fly sliders, allows the user to effectively explore datasets along the record and dimension axis. While some of our approaches are inherent to SMARTEXPLORE's design, we claim that, e.g., our automatic reliability analysis is generalizable to other systems. By means of an expert case studies with users of different expertise, we show that SMARTEXPLORE is effective for a broad audience and application domains.

8 Conclusion and Future Work

This chapter concludes the doctoral thesis. We will take a retrospective view of the user studies and implemented tools, summarize the contributions in a bigger context, and point to promising research directions for the next years for a pattern-driven design of visualizations.

Recap and Summary of Contributions

Knowledge needs to be extracted from high-dimensional and complex datasets to support data analysis processes across various research and economic disciplines. Visualizations play hereby an essential role in the identification and understanding of patterns within such datasets. They can be used as a primary analysis method, or support the understanding of automatic analysis methods. However, the choice of visual mappings heavily influences the effectiveness of the visualization. While one design choice is useful for a particular task, the very same design can make another analysis task more difficult, or even impossible. The design space of visualizations is huge, making it hard to ultimately develop a design that fits a particular analysis task and corresponds to the characteristics of the data. Data analysts are overwhelmed with numerous choices during the design process, such as selecting data attributes, visualization types, and the particular properties of each visualization. Hence, this thesis supports the effective design of visualizations by tackling the following research question: *"How can we effectively design visualizations to highlight patterns – using automatic and user-driven approaches"*. The thesis thereby advanced the quality and pattern-driven design and optimization of visualizations in two core areas, namely automatic optimization through quality metrics, and user-centered approaches. These two core areas also structured the two parts of the thesis.

Part I: Quality Metric-Driven Design for Pattern Analysis

The first part focused on *quality metrics* and addressed the research question *"how can we automatically measure the quality of a particular design to optimize the layout?"*

In Chapter 2 we **contributed a survey of quality metric research for visualizations for high-dimensional data**. We unified the vocabulary, enumerated the different existing metrics, and highlighted research gaps. In the survey, we focused on high-dimensional visualizations in which quality metrics play a key role, namely, scatter plots, scatter plot matrices, parallel coordinates, pixel-based techniques, radial visualizations, and glyph representations. Our results showed that a large body of research exists, and various quality metrics have been developed in the past. However, many metrics follow similar concepts (also across visualization types) but differ in their vocabulary or understanding of what quality means. Therefore, our presented survey provides a broad overview of the state-of-the-art and helps analysts choose metrics appropriately.

In Chapter 4, we extended the overview of quality metrics for parallel coordinates. We concentrated on metrics measuring the quality of a particular axes arrangement. We **introduced a classification of the existing metrics, grouped them according to their inner workings, and summarized their intended patterns and meta-characteristics**. For more practical support, we implemented a set of 14 strategies in JavaScript and made them available, along with the source code.

One of the main findings of the survey in Chapter 2 is that many metrics are proposed in the literature, but not evaluated in empirical settings. Often authors, who are proposing new quality metrics, evaluate them only with examples and case studies – instead of conducting user studies with synthetic or real datasets. In particular, we identified that for two of the most common visualizations for high-dimensional data, parallel coordinates, and star glyphs, necessary user studies are missing. The order of the axes plays a fundamental role in the design of both visualization techniques. Different arrangements of axes show different visual patterns, which can either support or prevent a successful analysis. In this thesis, we pushed axes reordering for both visualization techniques towards empirical guidance by **conducting a user study for parallel coordinates (Chapter 4) and star glyphs (Chapter 3)**. For both studies, we selected cluster identification as analysis task due to its primary focus in the literature. Our main finding for both studies was that ordering dimensions based on dissimilarity (place dimensions with a high dissimilarity next to each other) outperform the classical similarity-based arrangement. Most likely, the salient shapes, produced by dissimilar-based arrangements, help to identify groups of similar data records (i.e., clusters). Based on these findings, we **proposed a new reordering method for parallel coordinates and star glyphs**.

While experimenting with different axes orderings in parallel coordinates, we found out that standard parallel coordinates systematically distort the perception of patterns, in particular clusters. In Chapter 5, we proved that this problem is inherent to the technique itself: diagonal line segments are rendered longer (=need more pixels) and closer to each other (=less background color), compared to horizontal lines. Consequently, clusters are distorted, and ghost clusters (fake clusters, not existing in the data) can emerge. Based on these findings, **we contributed a formalization of the problem and provided an automatic method to adjust the rendering of the polylines based on their slope**. Our proposed technique can be computed in linear time and added on top of most parallel coordinate variations.

For reproducibility, and to help researchers applying our findings to new applications, **we made all materials of the papers** (data, analysis scripts, analysis results, and source codes) **publicly available on different repositories at the Open Science Framework (OSF)**. An overview can be found at `https://osf.io/yjxw2`.

Part II: User- and Task-Driven Design for Pattern Analysis

In many applications, the design of visualizations and the selection of visual elements depends on the underlying analysis tasks and may even need a highly iterative approach to describe and identify the patterns of interest. The second part of this thesis, therefore, provided *user- and task-driven* approaches to (semi-)automatically optimize visualizations. We addressed the question *"how can analysts support the design of visualization to highlight particular patterns?"* As a result, we contributed two techniques that advance the automatic design of visualizations from a user-centered research perspective.

The first technique, called *v-plot designer* (Chapter 6) is build for the comparative analysis of data distributions. Based on the selection of one or multiple analysis tasks, the v-plot designer **proposes an automatic recommendation of basic charts (e.g., box plots, violin-typed visualizations, and bar charts), along with a customized hybrid chart which is called a *v-plot*.** v-plots are automatically optimized to support all selected analysis tasks, and highlight required distribution properties. The automatic recommendations and the system design were developed based on the findings of a user study of 20 InfoVis and statistic practitioners. This study provided a solid foundation for the automation of the v-plot designer. We evaluated the designer and the v-plot itself, by measuring the fitness for purpose and applicability in a second study with four domain and statistic experts. The v-plot designer's focus was to select and adapt the visual properties of a particular visualization (v-plot) based on a given set of analysis tasks.

For the second tool, *SMARTexplore* (Chapter 7), our motivation was different. We used a table-based representation to simplify the interactive analysis of high-dimensional datasets. SMARTexplore is intended for both novice and expert users alike. Rows of a table can be aggregated manually, or with the help of clustering algorithms. Dimensions can be grouped into semantically meaningful subspaces, or automatically into groups of similar dimension patterns. **SMARTexplore combines easy-to-apply interaction concepts with the automatic and pattern-driven layout of rows and columns of the table.** The reliability of the perceived patterns can be verified by an automatic performed statistical analysis, which is encoded as possible overly in the visualization. The focus of SMARTexplore is to provide analysts with a tool to explore a dataset and identify their desired customized patterns. Based on these patterns, SMARTexplore allows to optimize the layout and to search for similar patterns throughout the entire datasets. Furthermore, an automatic reliability analysis helps analysts to see whether an identified pattern is statistically relevant or not.

Similarly to the first part of the thesis, we provide the material, source code, and a runnable version of the tools on our websites (`https://smartexplore.dbvis.de` and `https://v-plot.dbvis.de`) and on OSF. The available tools also allow to upload new data, making it possible to compare them to other approaches.

The different techniques presented in this thesis have a strong focus on classical information visualization. However, most of the techniques build the foundation of more advanced visual analytics frameworks. For example, a framework that uses parallel coordinates as a (meta-)visualization, the reordering classification in

Chapter 4 can help to find a useful initial ordering, based on the analysis task at hand. Afterwards, analysts can continue with a more interactive analysis.

To give an overview, the core contributions of this thesis have been summarized in Table 1.1 of Chapter 1.

Open Research Questions and Promising Directions

All chapters discuss future work with respect to their particular contribution. In the following, we want to highlight promising research directions for the next years from a more general perspective. We see great potential for future work in the following foru research areas:[1]

Multi-Criterion Quality Metrics for Customized Patterns

The current design of quality metrics mostly follows *one* straight path. However, the underlying data and the resulting visualizations rarely expose just one clear visual pattern but present rather a mixture of several patterns. Accordingly, it is challenging to say under which circumstances particular quality metrics and automatic chart designs work and fail. What is needed are "flexible" quality metrics that adapt to the underlying dataset at hand and promote the main visual pattern. Then, what is even more important, these *multi-criterion quality-metric* should notify the user upon usage that their dataset contains more than just the main pattern and offer a faceted visual pattern space view. To give an example for parallel coordinates: A multi-criterion QM would first let the user see the expected primary pattern (e.g., data clusters) and then promote –optimally related– other aspects (e.g., correction) by suggesting a different reordering of the axes. Advanced techniques should go even one step further, and show only relevant attribute combinations (i.e., subspaces) to the analysts. Of course, the relevance of a particular attribute or subspace may also change based on the pattern of interest. Here, researchers may apply or adapt concepts and algorithms from subspace analysis techniques, such as subspace clustering [PHL04; KKZ09] and subspace outlier detection [Kri+09; ZSK12].

Task-Adapted Optimization of Visualizations

In interactive and exploratory systems, the notion of quality, especially concerning the current analysis task, may change over time. However, the existing approaches are often not integrated into an exploration workflow and cannot change their quality notion by adapting to the currently conducted task or analysis direction. Research towards *task-adapted optimization of visualizations* should, therefore, be focusing in two different areas.

[1] The following arguments are *based on* Section 11 of my publication: [Beh+18]. Please refer to Sections 1.4 and 1.5 of this thesis for the contribution clarification and general citation rules.

First, there are analysts or particular applications in which the analysis tasks can clearly be defined. Of course, in such applications, analysts also know when a particular analysis task is changing. Hence, we need to provide helpful user interfaces in which analysts can select or change particular tasks. Here, the necessary research is not about finding a particular task but about changing the properties of visualizations such that it is accessible to the analyst. Analysts need to understand the design properties (e.g., the order of dimensions or the applied colormap) before and after a change. Otherwise, they cannot connect the patterns from one visualization design to the next. Here, research has to investigate different methods in terms of animation, visualizing provenance, and connected small-multiple designs.

Second, there are applications in which analysts can neither describe the task at hand nor directly indicate when a particular exploration path changes over time. Here, the research needs to be conducted to identify the analyst's task and identify once the analysis task is changing. There is a body of research on extracting the intention of users in visualizations and visual analytics systems. For example, Xu et al. [Xu+20] provides a survey of provenance and user interaction, Brown et al. [Bro+14] is learning user intention from low-level interactions, Endert et al. [EFN12] provides an example for semantic interactions, and Blascheck et al. [Bla+17] surveys how eye tracking can be used to implicitly learn from users.

A promising research direction is to combine these existing approaches with quality metrics and develop analysis systems that capture the users' intentions (both explicit and implicit) and adapt visualizations in a traceable way. Such changes in the visual design should also include possibilities to (automatically) switch the representation between different visualization types.

Interactive and Human-Supported Quality Steering

Related to the two aspects above is interactive and human-supported quality assessment. Interactive and reactive systems should facilitate the same exploration flexibility as the user in a manual process. Several approaches have already been presented in this growing research field with different foci on how quality metrics can be integrated into the exploration workflow:

Behrisch et al. [Beh+14] and Dennig el al. [Den+19] present a *relevance feedback* approach for a user-defined notion of interestingness in scatter plots. Users iteratively rank presented candidate views for their perceived interestingness. A gradually adapted classification model and a similarity advisor try to mimic the current understanding of interestingness in a given feature space. At the same time, a so-called "*decision support system*" constantly monitors the user and assesses the relevance-driven search process for convergence and stability.

Another interesting approach is presented by Wongsuphasawat et al. [Won+16] in their Voyager system. Voyager is designed as a mixed-initiative system, in which intelligent services and the user collaborate to achieve the analysis goals [Hor99] – an idea also inherently incorporated in the Visual Analytics mantra [Kei+08]. Upon startup, the user is provided with a gallery of automatically-generated visualizations for each (statistically) interesting data variable. The user navigates in the data space by a drill-down on one meaningful/expressive data variable and the underlying *Compass* recommendation engine enumerates, clusters, and ranks related visualizations according to both their underlying data properties and the resulting perceptual principles.

The interactive navigation and query definition for (complex) visual patterns are in the focus of the work of Shao et al. [Sha+14]. Confronted with a scatter plot pattern retrieval task, the user draws a vague idea of an expected visual pattern into a canvas. Upon each stroke, the system retrieves the most similar, respectively, most dissimilar plots; an idea referred to as *guided-sketching*. Visually similar results are clustered and can be taken over to the canvas to adapt the search in this specific direction.

Following these lines of how analysts, quality metrics, and classifications work together, there is a great potential to extend these ideas in several directions. One of the most interesting and promising directions is to understand the user's input and feedback better and encode this knowledge into adaptive quality metrics.

Machine Learning

Deep-learning based approaches have proven to be good visual pattern detectors. This could make *deep learning-based quality metrics* a viable research direction. Two preconditions must hold [Var+16]: (1) a sufficiently large training dataset must be provided or generated, (2) an appropriate network structure has to be found that is able to deal not only with one expected visual pattern but rather a mixture-model of the pattern space.

To date, only a few approaches exist which apply deep learning to measure the quality of a visualization or propose potentially interesting views [Sak+18]. VizDeck [Key+12] is one of the first attempts for a machine-learning based approach. Using users' up and down votes on a large set of visualization, VizDeck learns to recommend charts that the user will most likely vote up. Data2Vis [DD19] is a more recent model that generates visualization specifications in the Vega-Lite grammar [Sat+17] from descriptions of a dataset. The model was trained on thousands of example pairs of data and visualizations recommended by the rule-based system CompassQL [Won+17]. Similar approaches are VizML [Hu+19] and DeepEye [Luo+18] which have been published recently. Sabour et al. [SFH17] have shown an interesting approach in which the activity vector of *groups of neurons*, so-called capsules, represents a specific type of entity, such as an object or an object part. This approach could be used for learning capsules, one for each visual pattern. The network routing scheme decides which of the visual patterns are visually outstanding (have the most information content).

While these approaches could lead to satisfactory results, proving their perceptual correspondence will be even harder since these approaches suffer inherently from the interpretability gap.

Concluding Remarks

In summary, this thesis set out from the research question *"How can we effectively design visualizations to highlight patterns – using automatic and user-driven approaches?"* In the previous chapters, we have contributed various techniques and user studies to push the quality-driven automation of visualization forward. Based on our findings, we now have more evidence on how to reorder the dimension of parallel coordinates and star glyphs, know how to reduce distortion in parallel coordinates, and have an overview of the state-of-the-art in quality-metric research. We proposed an automatic chart recommendation engine for the comparative analysis of data distributions, along with task-specific customization of a hybrid chart, grounded in an expert survey. Finally, we presented SMARTexplore, a table-based visual analytics system allowing analysts to identify complex patterns in high-dimensional data.

However, we can also see that there are numerous ways to build on top of this thesis' findings. We are looking forward to the upcoming research in the next years, which will make the (semi-) automatic design of visualizations even more effective.

List of Figures

List of Tables

Bibliography

[AH04] James Abello and Frank van Ham. “Matrix Zoom: A Visual Interface to Semi-External Graphs”. In: *10th IEEE Symposium on Information Visualization (InfoVis 2004)*. 2004, pp. 183–190. DOI: 10.1109/INFVIS.2004.46 (cit. on p. 139).

[Ach+13] Elke Achtert, Hans-Peter Kriegel, Erich Schubert, and Arthur Zimek. “Interactive data mining with 3D-parallel-coordinate-trees”. In: *Proceedings of the International Conference on Management of Data*. 2013, pp. 1009–1012. DOI: 10.1145/2463676.2463696 (cit. on p. 73).

[Alb+10] Georgia Albuquerque, Martin Eisemann, Dirk J. Lehmann, Holger Theisel, and Marcus A. Magnor. “Improving the Visual Analysis of High-dimensional Datasets Using Quality Measures”. In: *Proc. of the IEEE Conf. on Visual Analytics Science and Technology*. 2010, pp. 19–26. DOI: 10.1109/VAST.2010.5652433 (cit. on pp. 26, 44–47, 57, 140).

[Alb+09] Georgia Albuquerque, Martin Eisemann, Dirk J. Lehmann, Holger Theisel, and Marcus A. Magnor. “Quality-Based Visualization Matrices.” In: *International Symposium on Vision, Modeling, and Visualization (VMV)*. 2009, pp. 341–350 (cit. on p. 36).

[AZZ10] Jamal Alsakran, Ye Zhao, and Xinlei Zhao. “Tile-based parallel coordinates and its application in financial visualization”. In: *Visualization and Data Analysis*. 2010, p. 753003. DOI: 10.1117/12.838819 (cit. on p. 67).

[And57] Edgar Anderson. “A Semigraphical Method for the Analysis of Complex Problems”. In: *Proceedings of the National Academy of Sciences of the United States of America* 43.10 (1957), pp. 923–927 (cit. on p. 48).

[AA01] Gennady Andrienko and Natalia Andrienko. “Constructing parallel coordinates plot for problem solving”. In: *International Symposium on Smart Graphics*. 2001, pp. 9–14 (cit. on pp. 67, 80, 93).

[AA04] Natalia Andrienko and Gennady Andrienko. “Interactive Visual Tools to Explore Spatio-temporal Variation”. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI ’04. New York, NY, USA: ACM, 2004, pp. 417–420. DOI: 10.1145/989863.989940 (cit. on p. 48).

[AA06] Natalia V. Andrienko and Gennady L. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer, 2006. DOI: 10.1007/3-540-31190-4 (cit. on pp. 49, 56).

[ABK98] Mihael Ankerst, Stefan Berchtold, and Daniel A. Keim. “Similarity Clustering of Dimensions for an Enhanced Visualization of Multidimensional Data”. In: *IEEE Symposium on Information Visualization*. 1998, pp. 52–60. DOI: 10.1109/INFVIS.1998.729559 (cit. on pp. 38, 39, 41, 43, 46, 47, 56, 57, 59, 68, 70, 74–77, 81, 140).

[Ans73] Francis J Anscombe. “Graphs in statistical analysis”. In: *The American Statistician* 27.1 (1973), pp. 17–21. DOI: 10.2307/2682899 (cit. on pp. 2, 111).

[AOL06] Almir Olivette Artero, Maria Cristina Ferreira de Oliveira, and Haim Levkowitz. “Enhanced High Dimensional Data Visualization through Dimension Reduction and Attribute Arrangement”. In: *Intl. Conf. on Information Visualisation*. 2006, pp. 707–712. DOI: 10.1109/IV.2006.49 (cit. on pp. 56, 57, 68, 74, 76, 77).

[AS16] Michaël Aupetit and Michael Sedlmair. “SepMe: 2002 New visual separation measures”. In: *2016 IEEE Pacific Visualization Symposium (PacificVis)*. Apr. 2016, pp. 1–8. DOI: 10.1109/PACIFICVIS.2016.7465244 (cit. on p. 37).

[BW08] Sven Bachthaler and Daniel Weiskopf. “Continuous Scatterplots”. In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008), pp. 1428–1435. DOI: 10.1109/TVCG.2008.119 (cit. on p. 96).

[Beh+16a] Michael Behrisch, Benjamin Bach, Michael Hund, Laura Von Rüden, Michael Delz, Jean-Daniel Fekete, and Tobias Schreck. “Magnostics: Image-based Search of Interesting Matrix Views for Guided Network Exploration”. In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2016), pp. 31–40. DOI: 10.1109/TVCG.2016.2598467 (cit. on pp. 13, 20).

[Beh+16b] Michael Behrisch, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck, and Jean-Daniel Fekete. “Matrix Reordering Methods for Table and Network Visualization”. In: *Comput. Graph. Forum* 35.3 (2016), pp. 693–716. DOI: 10.1111/cgf.12935 (cit. on pp. 139, 151).

[Beh+18] Michael Behrisch, Michael Blumenschein, Nam Wook Kim, Lin Shao, Mennatallah El-Assady, Johannes Fuchs, Daniel Seebacher, Alexandra Diehl, Ulrik Brandes, Hanspeter Pfister, Tobias Schreck, Daniel Weiskopf, and Daniel A. Keim. “Quality Metrics for Information Visualization”. In: *Comput. Graph. Forum* 37.3 (2018), pp. 625–662. DOI: 10.1111/cgf.13446 (cit. on pp. 9, 17, 57, 68, 70, 71, 73, 95, 140, 166).

[Beh+14] Michael Behrisch, Fatih Korkmaz, Lin Shao, and Tobias Schreck. “Feedback-Driven Interactive Exploration of Large Multidimensional Data Supported by Visual Classifier”. In: *Proc. IEEE Conference on Visual Analytics Science and Technology*. Peer-reviewed full paper. 2014, pp. 43–52. DOI: 10.1109/VAST.2014.7042480 (cit. on p. 167).

[Beh+19] Michael Behrisch, Dirk Streeb, Florian Stoffel, Daniel Seebacher, Brian Matejek, Stefan Hagen Weber, Sebastian Mittelstädt, Hanspeter Pfister, and Daniel A. Keim. “Commercial Visual Analytics Systems – Advances in the Big Data Analytics Field”. In: *IEEE Trans. Vis. Comput. Graph.* 25.10 (2019), pp. 3011–3031. DOI: 10.1109/TVCG.2018.2859973 (cit. on p. 110).

[Bel61] Richard E Bellman. *Adaptive control processes: a guided tour*. Princeton university press, 1961 (cit. on p. 31).

[Ben+18] Houssem Ben Lahmar, Melanie Herschel, Michael Blumenschein, and Daniel A. Keim. "Provenance-Based Visual Data Exploration with EVLIN". In: *Proceedings of the 21th International Conference on Extending Database Technology (EDBT)*. 2018, pp. 686–689. DOI: 10.5441/002/edbt.2018.85 (cit. on p. 12).

[Ben88] Yoav Benjamini. "Opening the box of a boxplot". In: *The American Statistician* 42.4 (1988), pp. 257–262. DOI: 10.2307/2685133 (cit. on p. 114).

[Ber+15] Jürgen Bernard, Martin Steiger, Sebastian Mittelstädt, Simon Thum, Daniel A. Keim, and Jörn Kohlhammer. "A survey and task-based quality assessment of static 2D colormaps". In: *Visualization and Data Analysis 2015*. 2015, p. 93970M. DOI: 10.1117/12.2079841 (cit. on p. 29).

[Ber75] Jacques Bertin. "La graphique et le traitement graphique de l'information". In: *Nouvelle bibliothèque scientifique, Flammarion* (1975) (cit. on p. 139).

[Ber83] Jacques Bertin. *Semiology of Graphics*. University of Wisconsin Press, 1983 (cit. on p. 5).

[Ber81] Jacques Bertin. "Théorie matricielle de la graphique". fre. In: *Communication et langages* 48.1 (1981), pp. 62–74. DOI: 10.3406/colan.1981.1409 (cit. on p. 17).

[BDS05] Enrico Bertini, Luigi Dell'Aquila, and Giuseppe Santucci. "Springview: Cooperation of radviz and parallel coordinates for view optimization and clutter reduction". In: *Third International Conference on Coordinated and Multiple Views in Exploratory Visualization*. IEEE. 2005, pp. 22–29 (cit. on p. 46).

[BS05] Enrico Bertini and Giuseppe Santucci. "Improving 2D scatterplots effectiveness through sampling, displacement, and user perception". In: *Ninth International Conference on Information Visualisation (IV'05)*. 2005, pp. 826–834. DOI: 10.1109/IV.2005.62 (cit. on p. 33).

[BS04] Enrico Bertini and Giuseppe Santucci. "Quality Metrics for 2D Scatterplot Graphics: Automatically Reducing Visual Clutter". In: *Smart Graphics*. 2004 (cit. on pp. 20, 24, 33).

[BTK11] Enrico Bertini, Andrada Tatu, and Daniel Keim. "Quality Metrics in High-Dimensional Data Visualization: An Overview and Systematization". In: *IEEE Trans. Vis. Comput. Graph.* 17.12 (2011), pp. 2203–2212. DOI: 10.1109/TVCG.2011.229 (cit. on pp. 18, 19, 29, 36, 37, 70, 71).

[Bey+99] Kevin S. Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. "When Is "Nearest Neighbor" Meaningful?" In: *Proceedings of the 7^{th} International Conference on Database Theory*. ICDT '99. Springer-Verlag, 1999, pp. 217–235 (cit. on p. 31).

[Bir+05] Marius-Victor Birsan, Peter Molnar, Paolo Burlando, and Martin Pfaundler. "Streamflow trends in Switzerland". In: *Journal of Hydrology* 314.1 (2005), pp. 312–329. DOI: 10.1016/j.jhydrol.2005.06.008 (cit. on p. 113).

[Bla+17] Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. “Visualization of Eye Tracking Data: A Taxonomy and Survey”. In: *Comput. Graph. Forum* 36.8 (2017), pp. 260–284 (cit. on p. 167).

[Blu+18] Michael Blumenschein, Michael Behrisch, Stefanie Schmid, Simon Butscher, Deborah R. Wahl, Karoline Villinger, Britta Renner, Harald Reiterer, and Daniel A. Keim. “SMARTexplore: Simplifying High-Dimensional Data Analysis through a Table-Based Visual Analytics Approach”. In: *IEEE Conference on Visual Analytics Science and Technology*. 2018, pp. 36–47. DOI: 10.1109/VAST.2018.8802486 (cit. on pp. 11, 133).

[Blu+20a] Michael Blumenschein, Luka J. Debbeler, Nadine C. Lages, Britta Renner, Daniel A. Keim, and Mennatallah El-Assady. “v-plots: Designing Hybrid Charts for the Comparative Analysis of Data Distributions”. In: *Computer Graphics Forum* 39.3 (2020), pp. 565–577. DOI: 10.1111/cgf.14002 (cit. on pp. 11, 107).

[Blu+20b] Michael Blumenschein, Xuan Zhang, David Pomerenke, Daniel A. Keim, and Johannes Fuchs. “Evaluating Reordering Strategies for Cluster Identification in Parallel Coordinates”. In: *Computer Graphics Forum* 39.3 (2020), pp. 537–549. DOI: 10.1111/cgf.14000 (cit. on pp. 10, 67, 69, 72, 74, 88).

[BS92] Ingwer Borg and Thomas Staufenbiel. “Performance of Snow Flakes, Suns, and Factorial Suns in the Graphical Representation of Multivariate Data”. In: *Multivariate Behavioral Research* 27.1 (1992), pp. 43–55. DOI: 10.1207/s15327906mbr2701_4 (cit. on pp. 50, 56).

[Bor+13] Rita Borgo, Johannes Kehrer, David H. S. Chung, Eamonn Maguire, Robert S. Laramee, Helwig Hauser, Matthew Ward, and Min Chen. “Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications”. In: *Eurographics - State of the Art Reports*. 2013, pp. 39–63. DOI: 10.2312/conf/EG2013/stars/039-063 (cit. on pp. 48, 57, 108).

[Bor+16] Michelle A. Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. “Beyond Memorability: Visualization Recognition and Recall”. In: *IEEE Transactions on Visualization and Computer Graphics* 22.1 (Jan. 2016), pp. 519–528. DOI: 10.1109/TVCG.2015.2467732 (cit. on p. 27).

[Bra97] Richard Brath. “Metrics for Effective Information Visualization”. In: *IEEE Symposium on Information Visualization*. IEEE Computer Society, 1997, pp. 108–111 (cit. on p. 28).

[BM13] Matthew Brehmer and Tamara Munzner. “A Multi-Level Typology of Abstract Visualization Tasks”. In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2376–2385. DOI: 10.1109/TVCG.2013.124 (cit. on p. 20).

[Bre+12] Sebastian Bremm, Martin Hess, Tatiana von Landesberger, and Dieter W. Fellner. “PCDC - On the Highway to Data - A Tool for the Fast Generation of Large Synthetic Data Sets”. In: *EuroVis Workshop on Visual Analytics*. Eurographics Association, 2012. DOI: 10.2312/PE/EuroVAST/EuroVA12/007-011 (cit. on pp. 60, 80).

[Bro+14] Eli T. Brown, Alvitta Ottley, Helen Zhao, Quan Lin, Richard Souvenir, Alex Endert, and Remco Chang. "Finding Waldo: Learning about Users from their Interactions". In: *IEEE Trans. Vis. Comput. Graph.* 20.12 (2014), pp. 1663–1672 (cit. on p. 167).

[Buc+15] Juri Buchmüller, Halldór Janetzko, Gennady L. Andrienko, Natalia V. Andrienko, Georg Fuchs, and Daniel A. Keim. "Visual Analytics for Exploring Local Impact of Air Traffic". In: *Computer Graphics Forum* 34.3 (2015), pp. 181–190. DOI: 10.1111/cgf.12630 (cit. on pp. 141, 155).

[Buj+18] Roxana Bujack, Terece L Turton, Francesca Samsel, Colin Ware, David H Rogers, and James Ahrens. "The Good, the Bad, and the Ugly: A Theoretical Framework for the Assessment of Continuous Colormaps". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), pp. 923–933 (cit. on p. 143).

[Cao+11] Nan Cao, David Gotz, Jimeng Sun, and Huamin Qu. "Dicon: Interactive visual analysis of multidimensional clusters". In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2581–2590 (cit. on p. 140).

[CMS99] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, eds. *Readings in Information Visualization: Using Vision to Think*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999 (cit. on pp. 1–3, 18, 22).

[CFF10] Luigi Di Caro, Vanessa Frías-Martínez, and Enrique Frías-Martínez. "Analyzing the Role of Dimension Arrangement for Data Visualization in Radviz". In: *Advances in Knowledge Discovery and Data Mining*. Springer, 2010, pp. 125–132. DOI: 10.1007/978-3-642-13672-6_13 (cit. on pp. 47, 57).

[CR74] Edwin Catmull and Raphael Rom. "A class of local interpolating splines". In: *Computer aided geometric design*. Academic Press, 1974, pp. 317–326. DOI: 10.1016/B978-0-12-079050-0.50020-5 (cit. on p. 122).

[CGM19] Davide Ceneda, Theresia Gschwandtner, and Silvia Miksch. "A Review of Guidance Approaches in Visual Data Analysis: A Multifocal Perspective". In: *Comput. Graph. Forum* 38.3 (2019), pp. 861–879 (cit. on p. 5).

[Cha+83] John M. Chambers, William S. Cleveland, Beat Kleiner, and Paul A. Tukey. *Graphical Methods for Data Analysis*. Wadsworth, 1983 (cit. on p. 116).

[CJ10] Min Chen and Heike Jaenicke. "An information-theoretic framework for visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), pp. 1206–1215 (cit. on p. 29).

[Che+14] Min Chen, Simon J. Walton, Kai Berger, Jeyarajan Thiyagalingam, Brian Duffy, Hui Fang, Cameron Holloway, and Anne E. Trefethen. "Visual Multiplexing". In: *Computer Graphics Forum* 33.3 (2014), pp. 241–250. DOI: 10.1111/cgf.12380 (cit. on p. 29).

[CXM17] Shenghui Cheng, Wei Xu, and Klaus Mueller. "RadViz Deluxe: An Attribute-Aware Display for Multivariate Data". en. In: *Processes* 5.4 (2017), p. 75. DOI: 10.3390/pr5040075 (cit. on p. 57).

[Che73] Herman Chernoff. “The Use of Faces to Represent Points in K-Dimensional Space Graphically”. In: *Journal of the American Statistical Association* (1973), pp. 361–368 (cit. on p. 48).

[CV11] Jarry HT Claessen and Jarke J. Van Wijk. “Flexible linked axes for multivariate data visualization”. In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2310–2316 (cit. on p. 41).

[Cle93] William S. Cleveland. “A Model for Studying Display Methods of Statistical Graphics”. In: *Journal of Computational and Graphical Statistics* 2.4 (1993), pp. 323–343 (cit. on p. 33).

[CM84] William S. Cleveland and Robert McGill. “Graphical perception: Theory, experimentation, and application to the development of graphical methods”. In: *Journal of the American statistical association* 79.387 (1984), pp. 531–554 (cit. on pp. 5, 27).

[CW16] Paul van der Corput and Jarke J. van Wijk. “Exploring Items and Features with I^F, F^I-Tables”. In: *Computer Graphics Forum* 35.3 (2016), pp. 31–40. DOI: 10.1111/cgf.12879 (cit. on pp. 135, 139).

[CCM09] Carlos D. Correa, Yu-Hsuan Chan, and Kwan-Liu Ma. “A framework for uncertainty-aware visual analytics”. In: *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*. 2009, pp. 51–58. DOI: 10.1109/VAST.2009.5332611 (cit. on p. 141).

[CG14] Michael Correll and Michael Gleicher. “Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error”. In: *IEEE Trans. Vis. Comput. Graph.* 20.12 (2014), pp. 2142–2151 (cit. on pp. 3, 110).

[Cor+19] Michael Correll, Mingwei Li, Gordon L. Kindlmann, and Carlos Scheidegger. “Looks Good To Me: Visualizations As Sanity Checks”. In: *IEEE Trans. Vis. Comput. Graph.* 25.1 (2019), pp. 830–839. DOI: 10.1109/TVCG.2018.2864907 (cit. on pp. 110, 111).

[CC00] T.F. Cox and A.A. Cox. *Multidimensional Scaling, Second Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2000 (cit. on p. 141).

[Cui+06] Qingguang Cui, Matthew O. Ward, Elke A. Rundensteiner, and Jing Yang. “Measuring Data Abstraction Quality in Multiresolution Visualizations”. In: *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006), pp. 709–716 (cit. on pp. 40, 41).

[DW14] Tuan Nhon Dang and L. Wilkinson. “ScagExplorer: Exploring Scatterplots by Their Scagnostics”. In: *Pacific Visualization Symposium (PacificVis), 2014 IEEE*. 2014, pp. 73–80. DOI: 10.1109/PacificVis.2014.42 (cit. on pp. 35, 36).

[DK10] Aritra Dasgupta and Robert Kosara. “Pargnostics: Screen-Space Metrics for Parallel Coordinates”. In: *Visualization and Computer Graphics, IEEE Transactions on* 16.6 (2010), pp. 1017–1026. DOI: 10.1109/TVCG.2010.184 (cit. on pp. 20, 38, 40, 41, 57, 68, 70, 73–76, 95, 140).

[DB79] David L. Davies and Donald W. Bouldin. “A cluster separation measure”. In: *IEEE transactions on pattern analysis and machine intelligence* 2 (1979), pp. 224–227 (cit. on p. 47).

[Deb+18] Luka J. Debbeler, Martina Gamp, Michael Blumenschein, Daniel A. Keim, and Britta Renner. “Polarized but illusory beliefs about tap and bottled water: A product- and consumer-oriented survey and blind tasting experiment”. In: *Science of The Total Environment* 643 (2018), pp. 1400–1410. DOI: 10.1016/j.scitotenv.2018.06.190 (cit. on pp. 11, 12, 108, 127).

[Den+19] Frederik L. Dennig, Tom Polk, Zudi Lin, Tobias Schreck, Hanspeter Pfister, and Michael Behrisch. “FDive: Learning Relevance Models Using Pattern-based Similarity Measures”. In: *IEEE Conference on Visual Analytics Science and Technology*. IEEE, 2019, pp. 69–80 (cit. on p. 167).

[DK17] Dua Dheeru and Efi Karra Taniskidou. *UCI Machine Learning Repository*. 2017 (cit. on p. 36).

[Di +94] Giuseppe Di Battista, Peter Eades, Roberto Tamassia, and Ioannis G Tollis. “Algorithms for drawing graphs: an annotated bibliography”. In: *Computational Geometry* 4.5 (1994), pp. 235–282 (cit. on p. 29).

[DPS02] Josep Diaz, Jordi Petit, and Maria Serna. “A Survey of Graph Layout Problems”. In: *ACM Computing Surveys* 34.3 (2002), pp. 313–356. DOI: 10.1145/568522.568523 (cit. on p. 28).

[DD19] Victor Dibia and Çagatay Demiralp. “Data2Vis: Automatic Generation of Data Visualizations Using Sequence-to-Sequence Recurrent Neural Networks”. In: *IEEE Computer Graphics and Applications* 39.5 (2019), pp. 33–46 (cit. on p. 168).

[DLR09] Geoffrey M Draper, Yarden Livnat, and Richard F Riesenfeld. “A survey of radial methods for information visualization”. In: *IEEE Transactions on Visualization and Computer Graphics* 15.5 (2009), pp. 759–776 (cit. on p. 45).

[DRC13] Adam Drewnowski, Colin D. Rehm, and Florence Constant. “Water and beverage consumption among children age 4–13y in the United States: analyses of 2005–2010 NHANES data”. In: *Nutrition Journal* 12.1 (2013), p. 85. DOI: 10.1186/1475-2891-12-85 (cit. on p. 113).

[DSS86] Stephen HC Du Toit, A Gert W Steyn, and Rolf H Stumpf. *Graphical Exploratory Data Analysis*. Springer, 1986 (cit. on p. 48).

[EAM11] Martin Eisemann, Georgia Albuquerque, and Marcus Magnor. “Data driven color mapping”. In: *Proceedings of EuroVA: International Workshop on Visual Analytics*. 2011 (cit. on p. 30).

[ElA+13] Mennatallah ElAssady, Daniel Hafner, Michael Hund, Alexander Jäger, Wolfgang Jentner, Christian Rohrdantz, Fabian Fischer, Svenja Simon, Tobias Schreck, and Daniel A. Keim. “Visual Analytics for the Prediction of Movie Rating and Box Office Performance”. In: *VAST Challenge 2013 - Award for Effective Analytics*. 2013 (cit. on p. 14).

[EBD05] Geoffrey Ellis, Enrico Bertini, and Alan Dix. “The sampling lens: making sense of saturated visualisations”. In: *CHI’05 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2005, pp. 1351–1354 (cit. on p. 40).

[ED07] Geoffrey P. Ellis and Alan J. Dix. “A Taxonomy of Clutter Reduction for Information Visualisation”. In: *IEEE Trans. Vis. Comput. Graph.* 13.6 (2007), pp. 1216–1223. DOI: 10.1109/TVCG.2007.70535 (cit. on pp. 19, 30, 33, 38, 70, 71, 95).

[ED06a] Geoffrey P. Ellis and Alan J. Dix. “Enabling Automatic Clutter Reduction in Parallel Coordinate Plots”. In: *IEEE Trans. on Vis. and Comp. Graph.* 12.5 (2006), pp. 717–724. DOI: 10.1109/TVCG.2006.138 (cit. on pp. 40, 51, 68, 71).

[ED06b] Geoffrey P. Ellis and Alan J. Dix. “The plot, the clutter, the sampling and its lens: occlusion measures for automatic clutter reduction”. In: *Proceedings of the working conference on Advanced visual interfaces, (AVI)*. 2006, pp. 266–269. DOI: 10.1145/1133265.1133318 (cit. on pp. 40, 41, 51, 140).

[Elm+08] Niklas Elmqvist, Thanh-Nghi Do, Howard Goodell, Nathalie Henry, and Jean-Daniel Fekete. “ZAME: Interactive Large-Scale Graph Visualization”. In: *Proceedings of IEEE Pacific Visualization Symposium*. 2008, pp. 215–222. DOI: 10.1109/PACIFICVIS.2008.4475479 (cit. on p. 139).

[Elz+16] Stef van den Elzen, Danny Holten, Jorik Blaas, and Jarke J. van Wijk. “Reducing Snapshots to Points: A Visual Analytics Approach to Dynamic Network Exploration”. In: *IEEE Transactions on Visualization and Computer Graphics* 22.1 (2016), pp. 1–10. DOI: 10.1109/TVCG.2015.2468078 (cit. on p. 140).

[EFN12] Alex Endert, Patrick Fiaux, and Chris North. “Semantic Interaction for Sensemaking: Inferring Analytical Reasoning for Model Steering”. In: *IEEE Trans. Vis. Comput. Graph.* 18.12 (2012), pp. 2879–2888 (cit. on p. 167).

[Erb02] Robert F. Erbacher. “Glyph-based generic network visualization”. In: *Visualization and Data Analysis*. 2002, pp. 228–237. DOI: 10.1117/12.458790 (cit. on p. 55).

[Est+96] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In: *Proc. of the Intl. Conf. on Knowledge Discovery and Data Mining*. 1996, pp. 226–231 (cit. on p. 60).

[FPS96] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. “From data mining to knowledge discovery in databases”. In: *AI magazine* 17.3 (1996), p. 37 (cit. on p. 39).

[Fer+10] Bilkis J. Ferdosi, Hugo Buddelmeijer, Scott C. Trager, Michael H. F. Wilkinson, and Jos B. T. M. Roerdink. “Finding and visualizing relevant subspaces for clustering high-dimensional astronomical data using connected morphological operators”. In: *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. 2010, pp. 35–42. DOI: 10.1109/VAST.2010.5652450 (cit. on p. 77).

[FR11] Bilkis J. Ferdosi and Jos B. T. M. Roerdink. “Visualizing High-Dimensional Structures by Dimension Ordering and Filtering using Subspace Analysis”. In: *Comput. Graph. Forum* 30.3 (2011), pp. 1121–1130. DOI: 10.1111/j.1467-8659.2011.01961.x (cit. on pp. 40, 41, 70, 74, 76, 77).

[FSJ13] Sara Johansson Fernstad, Jane Shaw, and Jimmy Johansson. “Quality-based guidance for exploratory dimensionality reduction”. In: *Information Visualization* 12.1 (2013), pp. 44–64. DOI: 10.1177/1473871612460526 (cit. on pp. 6, 140).

[Fie13] Andy Field. *Discovering statistics using IBM SPSS statistics*. sage, 2013 (cit. on p. 153).

[FMF12] Andy Field, Jeremy Miles, and Zoë Field. *Discovering statistics using R*. Sage publications, 2012 (cit. on pp. 108, 111).

[Fin+13] Martin Fink, Jan-Henrik Haunert, Joachim Spoerhase, and Alexander Wolff. “Selecting the Aspect Ratio of a Scatter Plot Based on Its Delaunay Triangulation”. In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2326–2335. DOI: 10.1109/TVCG.2013.187 (cit. on p. 33).

[FFM12] Fabian Fischer, Johannes Fuchs, and Florian Mansmann. “ClockMap: Enhancing Circular Treemaps with Temporal Glyphs for Time-Series Data”. In: *Eurographics Conference on Visualization, EuroVis*. 2012. DOI: 10.2312/PE/EuroVisShort/EuroVisShort2012/097-101 (cit. on pp. 48, 55).

[FJ07] Camilla Forsell and Jimmy Johansson. “Task-based evaluation of multirelational 3D and standard 2D parallel coordinates”. In: *Visualization and Data Analysis 2007*. Vol. 6495. International Society for Optics and Photonics, 2007, p. 64950C (cit. on p. 41).

[FPH09] Miguel de França Doria, Nick Pidgeon, and Paul R. Hunter. “Perceptions of drinking water quality and risk and its effect on behaviour: A cross-national study”. In: *Science of The Total Environment* 407.21 (2009), pp. 5455–5464. DOI: 10.1016/j.scitotenv.2009.06.031 (cit. on p. 113).

[FT74] Jerome H. Friedman and John W. Tukey. “A Projection Pursuit Algorithm for Exploratory Data Analysis”. In: *IEEE Transactions on Computers* C-23.9 (1974), pp. 881–890. DOI: 10.1109/T-C.1974.224051 (cit. on p. 35).

[FK03] Michael Friendly and Ernest Kwan. “Effect ordering for data displays”. In: *Computational Statistics & Data Analysis* 43.4 (2003), pp. 509–539. DOI: 10.1016/S0167-9473(02)00290-6 (cit. on pp. 56, 57).

[FWR99] Ying-Huey Fua, Matthew O. Ward, and Elke A. Rundensteiner. “Hierarchical Parallel Coordinates for Exploration of Large Datasets”. In: *Proceedings of the IEEE Conference on Visualization*. 1999, pp. 43–50. DOI: 10.1109/VISUAL.1999.809866 (cit. on pp. 67, 71).

[Fuc+14] Johannes Fuchs, Petra Isenberg, Anastasia Bezerianos, Fabian Fischer, and Enrico Bertini. “The Influence of Contour on Similarity Perception of Star Glyphs”. In: *IEEE Trans. on Vis. and Comp. Graph.* 20.12 (2014), pp. 2251–2260. DOI: 10.1109/TVCG.2014.2346426 (cit. on pp. 50, 55, 56, 65).

[Fuc+17] Johannes Fuchs, Petra Isenberg, Anastasia Bezerianos, and Daniel A. Keim. "A Systematic Review of Experimental Studies on Data Glyphs". In: *IEEE Trans. on Vis. and Comp. Graph.* 23.7 (2017), pp. 1863–1879. DOI: 10.1109/TVCG.2016.2549018 (cit. on pp. 48–50, 55).

[Fur+19] Katarina Furmanova, Samuel Gratzl, Holger Stitz, Thomas Zichner, Miroslava Jaresova, Martin Ennemoser, Alexander Lex, and Marc Streit. "Taggle: Combining Overview and Details in Tabular Data Visualizations". In: *Information Visualization* 19.2 (2019), pp. 114–136. DOI: 10.1177/1473871619878085 (cit. on pp. 135, 136, 139, 141).

[Fur+17] Katarina Furmanova, Miroslava Jaresova, Bikram Kawan, Holger Stitz, Martin Ennemoser, Samuel Gratzl, Alexander Lex, and Marc Streit. "Taggle: Scaling Table Visualization through Aggregation". In: *Poster @ IEEE Conference on Information Visualization (InfoVis '17)* (2017) (cit. on p. 139).

[GLH15] Salvador García, Julián Luengo, and Francisco Herrera. *Data Preprocessing in Data Mining*. Vol. 72. Intelligent Systems Reference Library. Springer, 2015. DOI: 10.1007/978-3-319-10247-4 (cit. on pp. 56, 65).

[Gle+11] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. "Visual comparison for information visualization". In: *Information Visualization* 10.4 (2011), pp. 289–309 (cit. on p. 112).

[Goo18] Google. *GoogleSheets*. https://www.google.com/sheets/about/, Last accessed on 2018-03-30. 2018 (cit. on p. 139).

[Gor+11] Marc H. Gorelick, Lindsay Gould, Mark Nimmer, Duke Wagner, Mary Heath, Hiba Bashir, and David C. Brousseau. "Perceptions About Water and Increased Use of Bottled Water in Minority Children". In: *Archives of Pediatrics & Adolescent Medicine* 165.10 (2011), pp. 928–932. DOI: 10.1001/archpediatrics.2011.83 (cit. on p. 113).

[Gow71] John C. Gower. "A general coefficient of similarity and some of its properties". In: *Biometrics* (1971), pp. 857–871 (cit. on pp. 141, 143).

[GLS17] Connor C. Gramazio, David H. Laidlaw, and Karen B. Schloss. "Colorgorical: creating discriminable and preferable color palettes for information visualization". In: *IEEE Transactions on Visualization and Computer Graphics* (2017) (cit. on p. 29).

[GTS10] Lars Grammel, Melanie Tory, and Margaret-Anne D. Storey. "How Information Visualization Novices Construct Visualizations". In: *IEEE Trans. Vis. Comput. Graph.* 16.6 (2010), pp. 943–952 (cit. on p. 3).

[Gra+14] Samuel Gratzl, Nils Gehlenborg, Alexander Lex, Hanspeter Pfister, and Marc Streit. "Domino: Extracting, Comparing, and Manipulating Subsets Across Multiple Tabular Datasets". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 2023–2032 (cit. on pp. 135, 139).

[Gra+13] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. “LineUp: Visual Analysis of Multi-Attribute Rankings”. In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2277–2286. DOI: 10.1109/TVCG.2013.173 (cit. on pp. 135, 136, 139).

[Gre17] Michael Greenacre. *Correspondence analysis in practice*. CRC press, 2017 (cit. on p. 141).

[Gsc+16] Theresia Gschwandtner, Markus Bögl, Paolo Federico, and Silvia Miksch. “Visual Encodings of Temporal Uncertainty: A Comparative User Study”. In: *IEEE Trans. Vis. Comput. Graph.* 22.1 (2016), pp. 539–548. DOI: 10.1109/TVCG.2015.2467752 (cit. on p. 110).

[HKP11] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques, 3rd edition*. Morgan Kaufmann, 2011 (cit. on pp. 71, 143).

[Hao+13] Ming C. Hao, Manish Marwah, Sebastian Mittelstädt, Halldor Janetzko, Michael Hund, Daniel A. Keim, Umeshwar Dayal, Collin Bash, Carlos Felix, Chandrakant Patel, and Meichun Hsu. “Visual analytics of cyber physical data streams using spatio-temporal radial pixel visualization”. In: *In Proceedings of Visualization and Data Analysis* (2013), pp. 865404–865412. DOI: 10.1117/12.2002948 (cit. on p. 14).

[HRC15] Lane Harrison, Katharina Reinecke, and Remco Chang. “Infographic Aesthetics: Designing for the First Impression”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. New York, NY, USA: ACM, 2015, pp. 1187–1190. DOI: 10.1145/2702123.2702545 (cit. on p. 27).

[Har+14] Lane Harrison, Fumeng Yang, Steven Franconeri, and Remco Chang. “Ranking Visualizations of Correlation Using Weber’s Law”. In: *IEEE Trans. Vis. Comput. Graph.* 20.12 (2014), pp. 1943–1952 (cit. on pp. 3, 68).

[HB03] Mark Harrower and Cynthia A Brewer. “ColorBrewer. org: an online tool for selecting colour schemes for maps”. In: *The Cartographic Journal* 40.1 (2003), pp. 27–37 (cit. on pp. 3, 59).

[HE12] Christopher Healey and James Enns. “Attention and visual memory in visualization and computer graphics”. In: *IEEE Transactions on Visualization and Computer Graphics* 18.7 (2012), pp. 1170–1188 (cit. on p. 27).

[HB10] Jeffrey Heer and Michael Bostock. “Crowdsourcing graphical perception: using mechanical turk to assess visualization design”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2010, pp. 203–212 (cit. on p. 52).

[Hee+08] Jeffrey Heer, Frank van Ham, Sheelagh Carpendale, Chris Weaver, and Petra Isenberg. “Creation and Collaboration: Engaging New Audiences for Information Visualization”. In: *Information Visualization*. Vol. 4950. Lecture Notes in Computer Science. Springer, 2008, pp. 92–133 (cit. on p. 3).

[Hei+12a] Julian Heinrich, Yuan Luo, Arthur E. Kirkpatrick, and Daniel Weiskopf. "Evaluation of a Bundling Technique for Parallel Coordinates". In: *GRAPP & IVAPP 2012: Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications*. 2012, pp. 594–602 (cit. on pp. 38, 41).

[Hei+12b] Julian Heinrich, Yuan Luo, Arthur E. Kirkpatrick, and Daniel Weiskopf. "Evaluation of a Bundling Technique for Parallel Coordinates". In: *Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications*. Feb. 2012, pp. 594–602 (cit. on p. 96).

[HW09] Julian Heinrich and Daniel Weiskopf. "Continuous Parallel Coordinates". In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (2009), pp. 1531–1538. DOI: 10.1109/TVCG.2009.131 (cit. on p. 96).

[HW15] Julian Heinrich and Daniel Weiskopf. "Parallel Coordinates for Multidimensional Data Visualization: Basic Concepts". In: *Computing in Science and Engineering* 17.3 (2015), pp. 70–76. DOI: 10.1109/MCSE.2015.55 (cit. on p. 73).

[HW13] Julian Heinrich and Daniel Weiskopf. "State of the Art of Parallel Coordinates". In: *Eurographics (State of the Art Reports)*. 2013, pp. 95–116. DOI: 10.2312/conf/EG2013/stars/095-116 (cit. on pp. 38, 70, 71, 73, 95).

[Hil91] David Hilbert. "Ueber die stetige Abbildung einer Line auf ein Flächenstück". In: *Mathematische Annalen* 38.3 (1891), pp. 459–460 (cit. on p. 43).

[Hil74] M. O. Hill. "Correspondence Analysis: A Neglected Multivariate Method". English. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 23.3 (1974), pp. 340–354 (cit. on p. 151).

[HAK00] Alexander Hinneburg, Charu C. Aggarwal, and Daniel A. Keim. "What Is the Nearest Neighbor in High Dimensional Spaces?" In: *Proceedings of the 26th International Conference on Very Large Data Bases*. VLDB '00. Morgan Kaufmann Publishers Inc., 2000, pp. 506–515 (cit. on p. 31).

[HN98] Jerry L. Hintze and Ray D. Nelson. "Violin Plots: A Box Plot-Density Trace Synergism". In: *The American Statistician* 52.2 (1998), pp. 181–184. DOI: 10.1080/00031305.1998.10480559 (cit. on pp. 108, 114).

[Hof+97] Patrick Hoffman, Georges G. Grinstein, Kenneth A. Marx, Ivo Grosse, and Eugene Stanley. "DNA visual and analytic data mining". In: *IEEE Visualization Proc.* 1997, pp. 437–442. DOI: 10.1109/VISUAL.1997.663916 (cit. on pp. 45, 46, 57, 68).

[HW10] Danny Holten and Jarke J. van Wijk. "Evaluation of Cluster Identification Performance for Different PCP Variants". In: *Comput. Graph. Forum* 29.3 (2010), pp. 793–802. DOI: 10.1111/j.1467-8659.2009.01666.x (cit. on pp. 41, 71).

[Hor99] Eric Horvitz. "Principles of Mixed-Initiative User Interfaces". In: *Proceeding of the CHI '99 Conference on Human Factors in Computing Systems: The CHI is the Limit*. 1999, pp. 159–166. DOI: 10.1145/302979.303030 (cit. on p. 167).

[Hou62] Paul VC Hough. *Method and means for recognizing complex patterns*. US Patent 3,069,654. 1962 (cit. on p. 75).

[Hu+19] Kevin Zeng Hu, Michiel A. Bakker, Stephen Li, Tim Kraska, and César A. Hidalgo. "VizML: A Machine Learning Approach to Visualization Recommendation". In: *CHI*. ACM, 2019, p. 128 (cit. on p. 168).

[HHJ11] Tze-Haw Huang, Mao Lin Huang, and Jesse S. Jin. "Parallel Rough Set: Dimensionality Reduction and Feature Discovery of Multi-Dimensional Data in Visualization". In: *Neural Information Processing - 18th International Conference (ICONIP)*. 2011, pp. 99–108. DOI: 10.1007/978-3-642-24958-7_12 (cit. on pp. 74, 77).

[HV08] M. Hubert and E. Vandervieren. "An adjusted boxplot for skewed distributions". In: *Computational Statistics & Data Analysis* 52.12 (2008), pp. 5186–5201. DOI: 10.1016/j.csda.2007.11.008 (cit. on p. 114).

[Hun+15a] Michael Hund, Michael Behrisch, Ines Färber, Michael Sedlmair, Tobias Schreck, Thomas Seidl, and Daniel A. Keim. "Subspace Nearest Neighbor Search - Problem Statement, Approaches, and Discussion". In: *Similarity Search and Applications*. Vol. 9371. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 307–313. DOI: 10.1007/978-3-319-25087-8_29 (cit. on pp. 13, 160).

[Hun+16a] Michael Hund, Dominic Böhm, Werner Sturm, Michael Sedlmair, Tobias Schreck, Torsten Ullrich, Daniel A. Keim, Ljiljana Majnaric, and Andreas Holzinger. "Visual Analytics for Concept Exploration in Subspaces of Patient Groups". In: *Brain Informatics* 3.4 (2016), pp. 233–247. DOI: 10.1007/s40708-016-0043-5 (cit. on p. 13).

[Hun+16b] Michael Hund, Ines Färber, Michael Behrisch, Andrada Tatu, Tobias Schreck, Daniel A. Keim, and Thomas Seidl. "Visual Quality Assessment of Subspace Clusterings". In: *KDD 2016 Workshop on Interactive Data Exploration and Analytics (IDEA'16)*. 2016 (cit. on p. 13).

[Hun+15b] Michael Hund, Werner Sturm, Tobias Schreck, Torsten Ullrich, Daniel A. Keim, Ljiljana Majnaric, and Andreas Holzinger. "Analysis of Patient Groups and Immunization Results Based on Subspace Clustering". In: *Brain Informatics and Health*. Vol. 9250. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 358–368. DOI: 10.1007/978-3-319-23344-4_35 (cit. on p. 13).

[HY09] Inkyoung Hur and Ji Soo Yi. "SimulSort: Multivariate Data Exploration through an Enhanced Sorting Technique". In: *Proceedings of Human-Computer Interaction. Novel Interaction Methods and Techniques*. 2009, pp. 684–693. DOI: 10.1007/978-3-642-02577-8_75 (cit. on p. 140).

[HO10] Catherine B Hurley and RW Oldford. "Pairwise display of high-dimensional information via eulerian tours and hamiltonian decompositions". In: *Journal of Computational and Graphical Statistics* 19.4 (2010), pp. 861–886 (cit. on pp. 41, 50, 71).

[Ins09a] Alfred Inselberg. "Data Mining and Other Applications". In: *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer, 2009, pp. 379–427. DOI: 10.1007/978-0-387-68628-8_10 (cit. on p. 71).

[Ins09b] Alfred Inselberg. *Parallel Coordinates*. Boston, MA: Springer US, 2009, pp. 2018–2024. DOI: 10.1007/978-0-387-39940-9_262 (cit. on pp. 37, 67).

[Ins85] Alfred Inselberg. "The plane with parallel coordinates". In: *The Visual Computer* 1.2 (1985), pp. 69–91. DOI: 10.1007/BF01898350 (cit. on pp. 38, 67, 134, 140).

[ID90] Alfred Inselberg and Bernard Dimsdale. "Parallel Coordinates: A Tool for Visualizing Multi-dimensional Geometry". In: *Proceedings IEEE Visualization '90, San Francisco, California, USA, October 23-26, 1990.* 1990, pp. 361–378. DOI: 10.1109/VISUAL.1990.146402 (cit. on p. 57).

[Jac01] Paul Jaccard. "Étude comparative de la distribution florale dans une portion des Alpes et des Jura". In: *Bulletin de la Société vaudoise des sciences naturelles* 37 (1901), pp. 547–579 (cit. on p. 83).

[Jäc+16] Dominik Jäckle, Fabian Fischer, Tobias Schreck, and Daniel A. Keim. "Temporal MDS Plots for Analysis of Multivariate Data". In: *IEEE Transactions on Visualization and Computer Graphics* 22.01 (2016). DOI: 10.1109/TVCG.2015.2467553 (cit. on p. 152).

[Jäc+17] Dominik Jäckle, Michael Hund, Michael Behrisch, Daniel A. Keim, and Tobias Schreck. "Pattern Trails: Visual Analysis of Pattern Transitions in Subspaces". In: *IEEE Conference on Visual Analytics Science and Technology*. 2017, pp. 1–12. DOI: 10.1109/VAST.2017.8585613 (cit. on pp. 12, 73).

[JC10] Heike Jänicke and Min Chen. "A Salience-based Quality Metric for Visualization". In: *Computer graphics forum*. Vol. 29 - 3. Wiley Online Library. 2010, pp. 1183–1192. DOI: 10.1111/j.1467-8659.2009.01667.x (cit. on p. 29).

[JM03] T. J. Jankun-Kelly and Kwan-Liu Ma. "MoireGraphs: Radial Focus+Context Visualization and Interaction for Graphs with Visual Nodes". In: *9th IEEE Symposium on Information Visualization (InfoVis 2003), 20-21 October 2003, Seattle, WA, USA*. 2003, pp. 59–66. DOI: 10.1109/INFVIS.2003.1249009 (cit. on p. 45).

[Jen+18] Wolfgang Jentner, Dominik Sacha, Florian Stoffel, Geoffrey Ellis, Leishi Zhang, and Daniel A. Keim. "Making machine intelligence less scary for criminal analysts: reflections on designing a visual comparative case analysis tool". In: *The Visual Computer* (2018), pp. 1–17. DOI: 10.1007/s00371-018-1483-0 (cit. on p. 134).

[JC08] Jimmy Johansson and Matthew Cooper. "A screen space quality method for data abstraction". In: *Computer Graphics Forum*. Vol. 27. Wiley Online Library, 2008, pp. 1039–1046 (cit. on p. 40).

[JF16] Jimmy Johansson and Camilla Forsell. “Evaluation of Parallel Coordinates: Overview, Categorization and Guidelines for Future Research”. In: *IEEE Trans. on Vis. and Comp. Graph.* 22.1 (2016), pp. 579–588. DOI: 10.1109/TVCG.2015.2466992 (cit. on pp. 41, 42, 67, 70, 73).

[JFC14] Jimmy Johansson, Camilla Forsell, and Matthew Cooper. “On the usability of three-dimensional display in parallel coordinates: Evaluating the efficiency of identifying two-dimensional relationships”. In: *Information Visualization* 13.1 (2014), pp. 29–41 (cit. on p. 42).

[Joh+08] Jimmy Johansson, Camilla Forsell, Mats Lind, and Matthew D. Cooper. “Perceiving patterns in parallel coordinates: determining thresholds for identification of relationships”. In: *Information Visualization* 7.2 (2008), pp. 152–162. DOI: 10.1057/palgrave.ivs.9500166 (cit. on pp. 42, 71).

[Joh+05] Jimmy Johansson, Patric Ljung, Mikael Jern, and Matthew D. Cooper. “Revealing Structure within Clustered Parallel Coordinates Displays”. In: *IEEE Symposium on Information Visualization*. Oct. 2005, pp. 125–132. DOI: 10.1109/INFVIS.2005.1532138 (cit. on pp. 68, 71, 95).

[JJJ08] Sara Johansson, Mikael Jern, and Jimmy Johansson. “Interactive Quantification of Categorical Variables in Mixed Data Sets”. In: *Proceedings of the International Conference on Information Visualisation (IV)*. 2008, pp. 3–10. DOI: 10.1109/IV.2008.33 (cit. on p. 141).

[JJ09] Sara Johansson and Jimmy Johansson. “Interactive Dimensionality Reduction Through User-defined Combinations of Quality Metrics”. In: *IEEE Trans. Vis. Comput. Graph.* 15.6 (2009), pp. 993–1000. DOI: 10.1109/TVCG.2009.153 (cit. on pp. 3, 23, 38, 39, 41, 68, 70, 71, 74, 76, 77).

[Jol86] Ian T. Jolliffe. *Principal Component Analysis*. Springer, New York, 1986. DOI: 10.1007/978-1-4757-1904-8 (cit. on pp. 134, 140).

[KF14] Linda T. Kaastra and Brian D. Fisher. “Field experiment methodology for pair analytics”. In: *Proceedings of the Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*. 2014, pp. 152–159. DOI: 10.1145/2669557.2669572 (cit. on pp. 128, 155).

[Kam08] Peter Kampstra. “Beanplot: A Boxplot Alternative for Visual Comparison of Distributions”. In: *Journal of Statistical Software, Code Snippets* 28.1 (2008), pp. 1–9. DOI: 10.18637/jss.v028.c01 (cit. on p. 114).

[Kan00] Eser Kandogan. “Star Coordinates: A Multi-dimensional Visualization Technique with Uniform Treatment of Dimensions”. In: *Proceedings of the IEEE Information Visualization Symposium*. Vol. 650. 2000, p. 22 (cit. on pp. 45, 46).

[KAC15] Rassadarie Kanjanabose, Alfie Abdul-Rahman, and Min Chen. “A Multi-task Comparative Study on Scatter Plots and Parallel Coordinates Plots”. In: *Comput. Graph. Forum* 34.3 (2015), pp. 261–270. DOI: 10.1111/cgf.12638 (cit. on pp. 68, 71).

[KW09] Greet Kayaert and Johan Wagemans. “Delayed shape matching benefits from simplicity and symmetry”. In: *Vision Research* 49.7 (2009), pp. 708–717. DOI: `https://doi.org/10.1016/j.visres.2009.01.002` (cit. on pp. 56, 58).

[Kei00] Daniel A Keim. “Designing pixel-oriented visualization techniques: Theory and applications”. In: *IEEE Transactions on Visualization and Computer Graphics* 6.1 (2000), pp. 59–78 (cit. on pp. 42–45).

[Kei+08] Daniel A. Keim, Gennady L. Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. “Visual Analytics: Definition, Process, and Challenges”. In: *Information Visualization - Human-Centered Issues and Perspectives*. Vol. 4950. Lecture Notes in Computer Science. Springer, 2008, pp. 154–175. DOI: `10.1007/978-3-540-70956-5_7` (cit. on pp. 20, 167).

[KS91] Philip J. Kellman and Thomas F. Shipley. “A theory of visual interpolation in object perception”. In: *Cognitive Psychology* 23.2 (1991), pp. 141–221. DOI: `10.1016/0010-0285(91)90009-D` (cit. on p. 90).

[Keo+06] Eamonn J. Keogh, Li Wei, Xiaopeng Xi, Stefano Lonardi, Jin Shieh, and Scott Sirowy. “Intelligent Icons: Integrating Lite-Weight Data Mining and Visualization into GUI Operating Systems”. In: *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM)*. 2006, pp. 912–916. DOI: `10.1109/ICDM.2006.90` (cit. on p. 140).

[Key+12] Alicia Key, Bill Howe, Daniel Perry, and Cecilia R. Aragon. “VizDeck: self-organizing dashboards for visual analytics”. In: *SIGMOD Conference*. ACM, 2012, pp. 681–684 (cit. on p. 168).

[KFM11] Christopher Kintzel, Johannes Fuchs, and Florian Mansmann. “Monitoring large IP spaces with ClockView”. In: *Intl. Symp. on Visualization for Cyber Security*. 2011, p. 2. DOI: `10.1145/2016904.2016906` (cit. on p. 55).

[Kli+09] Alexander Klippel, Frank Hardisty, Rui Li, and Chris Weaver. “Colour-Enhanced Star Plot Glyphs: Can Salient Shape Characteristics Be Overcome?” In: *Cartographica* 44.3 (2009), pp. 217–231. DOI: `10.3138/carto.44.3.217` (cit. on pp. 50, 56–60, 65, 66).

[KHW09] Alexander Klippel, Frank Hardisty, and Chris Weaver. “Star Plots: How Shape Characteristics Influence Classification Tasks”. en. In: *Cartography and Geographic Information Science* 36.2 (2009), pp. 149–163 (cit. on pp. 50, 56, 58–60, 65, 66).

[Kof14] Kurt Koffka. *Principles of Gestalt psychology*. Mimesis International, Sept. 2014 (cit. on p. 98).

[KZG10] Shawn Konecni, Jianping Zhou, and Georges Grinstein. “Advanced Interactions with Heatmaps for Analyzing High-Dimensional Datasets”. In: (2010) (cit. on p. 139).

[Kön+17] Laura M. König, Helge Giese, F. Marijn Stok, and Britta Renner. “The social image of food: Associations between popularity and eating behavior”. In: *Appetite* 114 (2017), pp. 248–258. DOI: `10.1016/j.appet.2017.03.039` (cit. on p. 128).

[Kos19] Robert Kosara. “The Impact of Distribution and Chart Type on Part-to-Whole Comparisons”. In: *EuroVis (Short Papers)*. Eurographics Association, 2019, pp. 7–11 (cit. on p. 3).

[KS16] Robert Kosara and Drew Skau. “Judgment Error in Pie Chart Variations”. In: *EuroVis (Short Papers)*. Eurographics Association, 2016, pp. 91–95 (cit. on p. 3).

[Kri+09] Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. “Outlier Detection in Axis-Parallel Subspaces of High Dimensional Data”. In: *PAKDD*. Vol. 5476. Lecture Notes in Computer Science. Springer, 2009, pp. 831–838 (cit. on p. 166).

[KKZ09] Hans-Peter Kriegel, Peer Kröger, and Arthur Zimek. “Clustering High-Dimensional Data: A Survey on Subspace Clustering, Pattern-Based Clustering, and Correlation Clustering”. In: *ACM Transactions on Knowledge Discovery from Data* 3.1 (2009), 1:1–1:58. DOI: `10.1145/1497577.1497578` (cit. on pp. 56, 166).

[Krs+12] Milos Krstajic, Christian Rohrdantz, Michael Hund, and Andreas Weiler. “Getting There First: Real-Time Detection of Real-World Incidents on Twitter”. In: *Published at the 2nd IEEE Workshop on Interactive Visual Text Analytics ”Task-Driven Analysis of Social Media” as part of the IEEE VisWeek 2012*. 2012 (cit. on p. 14).

[KW78] Joseph B Kruskal and Myron Wish. *Multidimensional scaling*. Vol. 11. Sage, 1978 (cit. on p. 35).

[Kua+12] Xiaole Kuang, Haimo Zhang, Shengdong Zhao, and Michael J. McGuffin. “Tracing Tuples Across Dimensions: A Comparison of Scatterplots and Parallel Coordinate Plots”. In: *Comput. Graph. Forum* 31.3 (2012), pp. 1365–1374. DOI: `10.1111/j.1467-8659.2012.03129.x` (cit. on p. 68).

[Lee+06] Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. “Task taxonomy for graph visualization”. In: *Proc. of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*. BELIV ’06. ACM. New York, NY, USA: ACM, 2006, pp. 1–5. DOI: `10.1145/1168149.1168168` (cit. on p. 20).

[LN88] Joseph Lee Rodgers and W Alan Nicewander. “Thirteen ways to look at the correlation coefficient”. In: *The American Statistician* 42.1 (1988), pp. 59–66 (cit. on p. 41).

[Leh+12] Dirk J. Lehmann, Georgia Albuquerque, Martin Eisemann, Marcus Magnor, and Holger Theisel. “Selecting Coherent and Relevant Plots in Large Scatterplot Matrices”. In: *Computer Graphics Forum* 31.6 (2012), pp. 1895–1908. DOI: `10.1111/j.1467-8659.2012.03069.x` (cit. on pp. 36, 37).

[LHT15] Dirk J. Lehmann, Sebastian Hundt, and Holger Theisel. “A study on quality metrics vs. human perception: Can visual measures help us to filter visualizations of interest?” In: *it - Information Technology* 57.1 (2015), pp. 11–21. DOI: `10.1515/itit-2014-1070` (cit. on pp. 29, 33, 37).

[Leh+15] Dirk J. Lehmann, Fritz Kemmler, Tatsiana Zhyhalava, Marco Kirschke, and Holger Theisel. "Visualnostics: Visual Guidance Pictograms for Analyzing Projections of High-dimensional Data". In: *Computer Graphics Forum* 34.3 (2015), pp. 291–300. DOI: `10.1111/cgf.12641` (cit. on p. 20).

[Lex+14] Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. "UpSet: Visualization of Intersecting Sets". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 1983–1992. DOI: `10.1109/TVCG.2014.2346248` (cit. on p. 139).

[Lex+10] Alexander Lex, Marc Streit, Christian Partl, Karl Kashofer, and Dieter Schmalstieg. "Comparative Analysis of Multidimensional, Quantitative Data". In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), pp. 1027–1035. DOI: `10.1109/TVCG.2010.138` (cit. on p. 140).

[Lex+12] Alexander Lex, Marc Streit, Hans-Jörg Schulz, Christian Partl, Dieter Schmalstieg, Peter J. Park, and Nils Gehlenborg. "StratomeX: Visual Analysis of Large-Scale Heterogeneous Genomics Data for Cancer Subtype Characterization". In: *Computer Graphics Forum* 31.3 (2012), pp. 1175–1184. DOI: `10.1111/j.1467-8659.2012.03110.x` (cit. on p. 140).

[LMW10] Jing Li, Jean-Bernard Martens, and Jarke J. van Wijk. "Judging correlation from scatterplots and parallel coordinate plots". In: *Information Visualization* 9.1 (2010), pp. 13–30. DOI: `10.1057/ivs.2008.13` (cit. on pp. 38, 68).

[LJC09] Mats Lind, Jimmy Johansson, and Matthew Cooper. "Many-to-many relational parallel coordinates displays". In: *Information Visualisation, 2009 13th International Conference*. IEEE, 2009, pp. 25–31 (cit. on p. 42).

[Lin15] Sander van der Linden. "Exploring Beliefs About Bottled Water and Intentions to Reduce Consumption: The Dual-Effect of Social Norm Activation and Persuasive Information". In: *Environment and Behavior* 47.5 (2015), pp. 526–550. DOI: `10.1177/0013916513515239` (cit. on p. 113).

[Liu+17] Shusen Liu, Dan Maljovec, Bei Wang, Peer-Timo Bremer, and Valerio Pascucci. "Visualizing High-Dimensional Data: Advances in the Past Decade". In: *IEEE Transactions on Visualization and Computer Graphics* 23.3 (2017), pp. 1249–1268. DOI: `10.1109/TVCG.2016.2640960` (cit. on p. 140).

[Lon18] Tran Van Long. "ArcViz: An Extended Radial Visualization for Classes Separation of High Dimensional Data". In: *Intl. Conf. on Knowledge and Systems Engineering*. 2018, pp. 158–162. DOI: `10.1109/KSE.2018.8573428` (cit. on p. 57).

[LHH12] Liang Fu Lu, Mao Lin Huang, and Tze-Haw Huang. "A New Axes Re-ordering Method in Parallel Coordinates Visualization". In: *Intl. Conf. on Machine Learning and Applications (ICMLA)*. 2012, pp. 252–257. DOI: `10.1109/ICMLA.2012.148` (cit. on pp. 74, 76, 77).

[LHZ16] Liang Fu Lu, Mao Lin Huang, and Jinson Zhang. "Two axes re-ordering methods in parallel coordinates plots". In: *J. Vis. Lang. Comput.* 33 (2016), pp. 3–12. DOI: `10.1016/j.jvlc.2015.12.001` (cit. on pp. 40, 42, 73, 74, 78).

[Luo+18] Yuyu Luo, Xuedi Qin, Nan Tang, and Guoliang Li. “DeepEye: Towards Automatic Data Visualization”. In: *International Conference on Data Engineering (ICDE)*. IEEE Computer Society, 2018, pp. 101–112 (cit. on p. 168).

[MH08] Laurens van der Maaten and Geoffrey Hinton. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research* 9 (2008), pp. 2579–2605 (cit. on p. 140).

[Mac86] Jock Mackinlay. “Automating the Design of Graphical Presentations of Relational Information”. In: *ACM Trans. Graph.* 5.2 (Apr. 1986), pp. 110–141. DOI: `10.1145/22949.22950` (cit. on p. 27).

[MHS07] Jock D. Mackinlay, Pat Hanrahan, and Chris Stolte. “Show Me: Automatic Presentation for Visual Analysis”. In: *IEEE Trans. Vis. Comput. Graph.* 13.6 (2007), pp. 1137–1144 (cit. on pp. 20, 27).

[MS05] Erkki Mäkinen and Harri Siirtola. “The Barycenter Heuristic and the Reorderable Matrix”. In: *Informatica (Slovenia)* 29.3 (2005), pp. 357–364 (cit. on p. 151).

[MTJ12] Hemant Makwana, Sanjay Tanwani, and Suresh Jain. “Axes Re-Ordering in Parallel Coordinate for Pattern Optimization”. In: *Intl. Journal of Computer Applications* 40.13 (2012), pp. 43–48 (cit. on pp. 74, 76).

[MW95] Allen R. Martin and Matthew O. Ward. “High Dimensional Brushing for Interactive Exploration of Multivariate Data”. In: *Proceedings of the Conference on Visualization*. 1995, pp. 271–278. DOI: `10.1109/VISUAL.1995.485139` (cit. on pp. 68, 71).

[MF17] Justin Matejka and George W. Fitzmaurice. “Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing”. In: *Proceedings of the Conference on Human Factors in Computing Systems*. 2017, pp. 1290–1294. DOI: `10.1145/3025453.3025912` (cit. on p. 111).

[MTL18] Jose Matute, Alexandru C. Telea, and Lars Linsen. “Skeleton-Based Scagnostics”. In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), pp. 542–552. DOI: `10.1109/TVCG.2017.2744339` (cit. on p. 36).

[Mat+18] Laura E. Matzen, Michael J. Haass, Kristin M. Divis, Zhiyuan Wang, and Andrew T. Wilson. “Data Visualization Saliency Model: A Tool for Evaluating Abstract Data Visualizations”. In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), pp. 563–573. DOI: `10.1109/TVCG.2017.2743939` (cit. on p. 29).

[Mau+17] Michele Mauri, Tommaso Elli, Giorgio Caviglia, Giorgio Uboldi, and Matteo Azzi. “RAWGraphs: A Visualisation Platform to Create Open Outputs”. In: *Proceedings of the Conference on Italian SIGCHI*. ACM, 2017, pp. 1–5. DOI: `10.1145/3125571.3125585` (cit. on p. 110).

[May+14] Thomas Mayer, Bernhard Wälchli, Michael Hund, and Christian Rohrdantz. "From the extraction of continuous features in parallel texts to visual analytics of heterogeneous areal-typological datasets". In: *Language Processing and Grammars. The role of functionally oriented computational models* (2014), pp. 13–38. DOI: 10.1075/slcs.150.02may (cit. on p. 13).

[MW07] Gregory J. McCabe and David M. Wolock. "Warming may create substantial water supply shortages in the Colorado River basin". In: *Geophysical Research Letters* 34.22 (2007). DOI: 10.1029/2007GL031764 (cit. on p. 113).

[McC+69] William T McCormick, Stephen B Deutsch, John J Martin, and Paul J Schweitzer. *Identification of Data Structures and Relationships by Matrix Reordering Techniques*. Tech. rep. DTIC Document, 1969 (cit. on p. 151).

[MM08] Kevin T. McDonnell and Klaus Mueller. "Illustrative Parallel Coordinates". In: *Computer Graphics Forum* 27.3 (2008), pp. 1031–1038. DOI: 10.1111/j.1467-8659.2008.01239.x (cit. on pp. 68, 140).

[MTL78] Robert McGill, John W Tukey, and Wayne A Larsen. "Variations of box plots". In: *The American Statistician* 32.1 (1978), pp. 12–16. DOI: 10.2307/2683468 (cit. on p. 114).

[Mer+17] Leonel Merino, Johannes Fuchs, Michael Blumenschein, Craig Anslow, Mohammad Ghafari, Oscar Nierstrasz, Michael Behrisch, and Daniel A. Keim. "On the Impact of the Medium in the Effectiveness of 3D Software Visualizations". In: *VISSOFT'17: Proceedings of the 5th IEEE Working Conference on Software Visualization*. 2017, pp. 11–21. DOI: 10.1109/VISSOFT.2017.17 (cit. on p. 13).

[Mic+17] Luana Micallef, Gregorio Palmas, Antti Oulasvirta, and Tino Weinkauf. "Towards Perceptual Optimization of the Visual Design of Scatterplots". In: *IEEE Transactions on Visualization and Computer Graphics* 23.6 (2017), pp. 1588–1599. DOI: 10.1109/TVCG.2017.2674978 (cit. on p. 33).

[Mic18a] Microsoft. *Power BI Software, Interactive Data Visualization BI Tools*. https://powerbi.microsoft.com/en-us/, Last accessed on 2018-03-30. 2018 (cit. on p. 139).

[Mic18b] Microsoft Excel. *Spreadsheet Software*. https://products.office.com/en/excel, Last accessed on 2018-03-30. 2018 (cit. on p. 139).

[Mil+19] Matthias Miller, Xuan Zhang, Johannes Fuchs, and Michael Blumenschein. "Evaluating Ordering Strategies of Star Glyph Axes". In: *IEEE Visualization Conference (VIS)*. 2019, pp. 91–95. DOI: 10.1109/VISUAL.2019.8933656 (cit. on pp. 9, 55).

[Mil+97] Nancy Miller, Beth Hetzler, Grant Nakamura, and Paul Whitney. "The Need for Metrics in Visual Information Analysis". In: *ACM Workshop on New Paradigms in Information Visualization and Manipulation*. ACM, 1997, pp. 24–28 (cit. on p. 28).

[Mit+15] Sebastian Mittelstädt, Dominik Jäckle, Florian Stoffel, and Daniel A Keim. "ColorCAT: Guided design of colormaps for combined analysis tasks". In: *Proc. of the Eurographics Conference on Visualization (EuroVis 2015: Short Papers)*. Vol. 2. 2015 (cit. on pp. 29, 143).

[MK15] Sebastian Mittelstädt and Daniel A. Keim. "Efficient Contrast Effect Compensation with Personalized Perception Models". In: *Computer Graphics Forum* 34.3 (2015), pp. 211–220. DOI: `10.1111/cgf.12633` (cit. on p. 29).

[Mor+19] Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith, Bill Howe, and Jeffrey Heer. "Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco". In: *IEEE Trans. Vis. Comput. Graph.* 25.1 (2019), pp. 438–448 (cit. on p. 110).

[Mor66] Guy M Morton. *A computer oriented geodetic data base and a new technique in file sequencing*. International Business Machines Company New York, 1966 (cit. on p. 43).

[Mou11] Rida E. Moustafa. "Parallel coordinate and parallel coordinate density plots". In: *WIREs Computational Statistics* 3.2 (2011), pp. 134–148. DOI: `10.1002/wics.145` (cit. on p. 71).

[Mra+07] Minca Mramor, Gregor Leban, Janez Demšar, and Blaž Zupan. "Visualization-based cancer microarray data classification analysis". In: *Bioinformatics* 23.16 (2007), pp. 2147–2154 (cit. on p. 46).

[Mun14] Tamara Munzner. *Visualization Analysis and Design*. A.K. Peters visualization series. A K Peters, 2014 (cit. on pp. 1, 2, 5, 48).

[NGC01] Harsha S. Nagesh, Sanjay Goil, and Alok N. Choudhary. "Adaptive Grids for Clustering Massive Data Sets". In: *Proceedings of the International Conference on Data Mining*. 2001, pp. 1–17. DOI: `10.1137/1.9781611972719.7` (cit. on pp. 41, 77).

[Net+17] Rudolf Netzel, Jenny Vuong, Ulrich Engelke, Seán I. O'Donoghue, Daniel Weiskopf, and Julian Heinrich. "Comparative eye-tracking evaluation of scatterplots and parallel coordinates". In: *Visual Informatics* 1.2 (2017), pp. 118–131. DOI: `10.1016/j.visinf.2017.11.001` (cit. on pp. 37, 42, 77, 78).

[NAW13] Dang Tuan Nhon, Anushka Anand, and Leland Wilkinson. "TimeSeer: Scagnostics for High-Dimensional Time Series". In: *IEEE Transactions on Visualization and Computer Graphics* 19.3 (2013), pp. 470–483. DOI: `10.1109/TVCG.2012.128` (cit. on pp. 35, 36).

[NW14] Dang Tuan Nhon and Leland Wilkinson. "Transforming Scagnostics to Reveal Hidden Features". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 1624–1632 (cit. on pp. 35, 36).

[Nie+17] Christina Niederer, Holger Stitz, Reem Hourieh, Florian Grassinger, Wolfgang Aigner, and Marc Streit. "TACO: Visualizing Changes in Tables Over Time". In: *IEEE Transactions on Visualization and Computer Graphics* (2017) (cit. on p. 139).

[NRS97] Rolf Niedermeier, Klaus Reinhardt, and Peter Sanders. "Towards optimal locality in mesh-indexings". In: *International Symposium on Fundamentals of Computation Theory*. Springer. 1997, pp. 364–375 (cit. on p. 43).

[Nor06] Chris North. "Toward measuring visualization insight". In: *IEEE Computer Graphics and Applications* 26.3 (2006), pp. 6–9 (cit. on p. 18).

[NŠ11] Lenka Nováková and Olga Štěpánková. "Visualization of trends using RadViz". In: *Journal of Intelligent Information Systems* 37.3 (2011), p. 355 (cit. on p. 46).

[Ond+19] Brian D. Ondov, Nicole Jardine, Niklas Elmqvist, and Steven Franconeri. "Face to Face: Evaluating Visual Comparison". In: *IEEE Trans. Vis. Comput. Graph.* 25.1 (2019), pp. 861–871. DOI: `10.1109/TVCG.2018.2864884` (cit. on p. 110).

[Opa+18] Tomasz Opach, Stanislav Popelka, Jitka Dolezalova, and Jan Ketil Rød. "Star and polyline glyphs in a grid plot and on a map display: which perform better?" In: *Cartography and Geographic Information Science* 45.5 (2018), pp. 400–419. DOI: `10.1080/15230406.2017.1364169` (cit. on p. 55).

[Pal+14] Gregorio Palmas, Myroslav Bachynskyi, Antti Oulasvirta, Hans-Peter Seidel, and Tino Weinkauf. "An Edge-Bundling Layout for Interactive Parallel Coordinates". In: *IEEE Pacific Visualization Symposium*. Mar. 2014, pp. 57–64. DOI: `10.1109/PacificVis.2014.40` (cit. on pp. 41, 96).

[Pan+16] Anshul Vikram Pandey, Josua Krause, Cristian Felix, Jeremy Boy, and Enrico Bertini. "Towards Understanding Human Similarity Perception in the Analysis of Large Sets of Scatter Plots". In: *Conference on Human Factors in Computing Systems CHI*. 2016, pp. 3659–3669. DOI: `10.1145/2858036.2858155` (cit. on pp. 29, 33, 37).

[PHL04] Lance Parsons, Ehtesham Haque, and Huan Liu. "Subspace Clustering for High Dimensional Data: A Review". In: *ACM SIGKDD Explorations Newsletter* 6.1 (2004), pp. 90–105. DOI: `10.1145/1007730.1007731` (cit. on pp. 160, 166).

[Paw12] Zdzisław Pawlak. *Rough sets: Theoretical aspects of reasoning about data*. Vol. 9. Springer Science & Business Media, 2012 (cit. on p. 77).

[PL17a] Jaakko Peltonen and Ziyuan Lin. "Parallel Coordinate Plots for Neighbor Retrieval". In: *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (IVAPP)*. 2017, pp. 40–51. DOI: `10.5220/0006097400400051` (cit. on pp. 68, 70, 73–76, 78).

[PL17b] Jaakko Peltonen and Ziyuan Lin. "Parallel Coordinate Plots for Neighbor Retrieval." In: *Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (IVAPP)*. 2017, pp. 40–51 (cit. on p. 40).

[PWR04] Wei Peng, Matthew O. Ward, and Elke A. Rundensteiner. "Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering". In: *Symposium on Information Visualization*. 2004, pp. 89–96. DOI: `10.1109/INFVIS.2004.15` (cit. on pp. 39, 41, 50, 56, 58, 68, 70, 74, 76, 140).

[PDF14] Charles Perin, Pierre Dragicevic, and Jean-Daniel Fekete. "Revisiting bertin matrices: New interactions for crafting tabular visualizations". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 2082–2091 (cit. on pp. 136, 139).

[PG88] Ronald M Pickett and Georges G Grinstein. "Iconographic displays for visualizing multidimensional data". In: *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*. Vol. 514. 1988, p. 519 (cit. on p. 49).

[Pin+16] An-Sofie Pinket, Marieke De Craemer, Lea Maes, Ilse De Bourdeaudhuij, Greet Cardon, Odysseas Androutsos, Berthold Koletzko, Luis Moreno, Piotr Socha, Violeta Iotova, and et al. "Water intake and beverage consumption of preschoolers from six European countries and associations with socio-economic status: the ToyBox-study". In: *Public Health Nutrition* 19.13 (2016), pp. 2315–2325. DOI: 10.1017/S1368980015003559 (cit. on p. 113).

[Pol+18] Patrik Polatsek, Manuela Waldner, Ivan Viola, Peter Kapec, and Wanda Benesova. "Exploring visual attention and saliency modeling for task-based visual analysis". In: *Computers & Graphics* 72 (2018), pp. 26–38. DOI: 10.1016/j.cag.2018.01.010 (cit. on p. 29).

[Pom+19] David Pomerenke, Frederik L. Dennig, Daniel A. Keim, Johannes Fuchs, and Michael Blumenschein. "Slope-Dependent Rendering of Parallel Coordinates to Reduce Density Distortion and Ghost Clusters". In: *IEEE Visualization Conference (VIS)*. 2019, pp. 86–90. DOI: 10.1109/VISUAL.2019.8933706 (cit. on pp. 10, 11, 68, 79, 80, 93).

[Pot+10] Kristin Potter, Joe Kniss, Richard F. Riesenfeld, and Chris R. Johnson. "Visualizing Summary Statistics and Uncertainty". In: *Comput. Graph. Forum* 29.3 (2010), pp. 823–832. DOI: 10.1111/j.1467-8659.2009.01677.x (cit. on pp. 115, 121).

[RC94] Ramana Rao and Stuart K. Card. "The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information". In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. 1994, pp. 318–322. DOI: 10.1145/191666.191776 (cit. on pp. 139, 151).

[RLB19] Donghao Ren, Bongshin Lee, and Matthew Brehmer. "Charticulator: Interactive Construction of Bespoke Chart Layouts". In: *IEEE Trans. Vis. Comput. Graph.* 25.1 (2019), pp. 789–799. DOI: 10.1109/TVCG.2018.2865158 (cit. on p. 110).

[Ric+15] Joshua S. Rice, Ryan E. Emanuel, James M. Vose, and Stacy A. C. Nelson. "Continental U.S. streamflow trends from 1940 to 2009 and their relationships with watershed spatial characteristics". In: *Water Resources Research* 51.8 (2015), pp. 6262–6275. DOI: 10.1002/2014WR016367 (cit. on p. 113).

[RW18] Nils Rodriguez and Daniel Weiskopf. "Nonlinear Dot Plots". In: *IEEE Trans. Vis. Comput. Graph.* 24.1 (2018), pp. 616–625. DOI: 10.1109/TVCG.2017.2744018 (cit. on p. 116).

[Roh+12] Christian Rohrdantz, Michael Hund, Thomas Mayer, Bernhard Wälchli, and Daniel A. Keim. "The World's Languages Explorer: Visual Analysis of Language Features in Genealogical and Areal Contexts". In: *Computer Graphics Forum* 31.3 (2012), pp. 935–944. DOI: 10.1111/j.1467-8659.2012.03086.x (cit. on p. 14).

[RP08] Timo Ropinski and Bernhard Preim. "Taxonomy and Usage Guidelines for Glyph-based Medical Visualization". In: *Simulation and Visualization*. 2008, pp. 121–138 (cit. on p. 55).

[Ros+04] Geraldine E. Rosario, Elke A. Rundensteiner, David C. Brown, Matthew O. Ward, and Shiping Huang. "Mapping nominal values to numbers for effective visualization". In: *Information Visualization* 3.2 (2004), pp. 80–95. DOI: 10.1057/palgrave.ivs.9500072 (cit. on p. 141).

[RZH12] René Rosenbaum, Jian Zhi, and Bernd Hamann. "Progressive parallel coordinates". In: *Visualization Symposium (PacificVis), 2012 IEEE Pacific*. IEEE, 2012, pp. 25–32 (cit. on p. 41).

[RP66] Azriel Rosenfeld and John L Pfaltz. "Sequential operations in digital picture processing". In: *Journal of the ACM* 13.4 (1966), pp. 471–494 (cit. on p. 40).

[Rub+16] Manuel Rubio-Sánchez, Laura Raya, Francisco Diaz, and Alberto Sanchez. "A comparative study between RadViz and Star Coordinates". In: *IEEE Transactions on Visualization and Computer Graphics* 22.1 (2016), pp. 619–628 (cit. on pp. 46, 48).

[Rze13] Tomasz Rzeźniczak. "Evaluation of multidimensional visualization techniques for medical patterns representation". In: *Journal of Theoretical and Applied Computer Science* 7.4 (2013), pp. 70–85 (cit. on p. 57).

[SFH17] Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic Routing Between Capsules". In: *CoRR* abs/1710.09829 (2017). arXiv: 1710.09829 (cit. on p. 168).

[Sac+17] Dominik Sacha, Michael Sedlmair, Leishi Zhang, John Aldo Lee, Jaakko Peltonen, Daniel Weiskopf, Stephen C. North, and Daniel A. Keim. "What you see is what you can change: Human-centered machine learning by interactive visualization". In: *Neurocomputing* 268 (2017), pp. 164–175. DOI: 10.1016/j.neucom.2017.01.105 (cit. on p. 20).

[Sac+14] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey P. Ellis, and Daniel A. Keim. "Knowledge Generation Model for Visual Analytics". In: *IEEE Trans. on Vis. and Comp. Graph.* 20.12 (2014), pp. 1604–1613. DOI: 10.1109/TVCG.2014.2346481 (cit. on p. 57).

[SED19] Bahador Saket, Alex Endert, and Çagatay Demiralp. "Task-Based Effectiveness of Basic Visualizations". In: *IEEE Trans. Vis. Comput. Graph.* 25.7 (2019), pp. 2505–2512 (cit. on p. 3).

[SES16] Bahador Saket, Alex Endert, and John Stasko. "Beyond Usability and Performance: A Review of User Experience-focused Evaluations in Visualization". In: *Proc. of 6th Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. ACM, 2016, pp. 133–142. DOI: `10.1145/2993901.2993903` (cit. on p. 27).

[Sak+18] Bahador Saket, Dominik Moritz, Halden Lin, Victor Dibia, Çagatay Demiralp, and Jeffrey Heer. "Beyond Heuristics: Learning Visualization Design". In: *CoRR* abs/1807.06641 (2018) (cit. on p. 168).

[SG18] Alper Sarikaya and Michael Gleicher. "Scatterplots: Tasks, Data, and Designs". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2018). DOI: `10.1109/TVCG.2017.2744184` (cit. on pp. 33–35).

[SAS18] SAS. *JMP Software, JMP Software from SAS*. `https://www.jmp.com`, Last accessed on 2018-03-30. 2018 (cit. on p. 139).

[SR96] Peter D. Sasieni and Patrick Royston. "Dotplots". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 45.2 (1996), pp. 219–234. DOI: `10.2307/2986156` (cit. on p. 116).

[SH14] Arvind Satyanarayan and Jeffrey Heer. "Lyra: An Interactive Visualization Design Environment". In: *Comput. Graph. Forum* 33.3 (2014), pp. 351–360. DOI: `10.1111/cgf.12391` (cit. on p. 110).

[Sat+17] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. "Vega-Lite: A Grammar of Interactive Graphics". In: *IEEE Trans. Vis. Comput. Graph.* 23.1 (2017), pp. 341–350 (cit. on p. 168).

[Sat+16] Arvind Satyanarayan, Ryan Russell, Jane Hoffswell, and Jeffrey Heer. "Reactive Vega: A Streaming Dataflow Architecture for Declarative Interactive Visualization". In: *IEEE Trans. Vis. Comput. Graph.* 22.1 (2016), pp. 659–668 (cit. on p. 110).

[SPA11] Amber Saylor, Linda Stalker Prokopy, and Shannon Amberg. "What's Wrong with the Tap? Examining Perceptions of Tap Water and Bottled Water at Purdue University". In: *Environmental Management* 48.3 (2011), pp. 588–601. DOI: `10.1007/s00267-011-9692-6` (cit. on p. 113).

[Sch+19] Christin Schätzle, Frederik L. Dennig, Michael Blumenschein, Daniel A. Keim, and Miriam Butt. "Visualizing Linguistic Change as Dimension Interactions". In: *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*. Association for Computational Linguistics, 2019, pp. 272–278 (cit. on p. 12).

[Sch+17] Christin Schätzle, Michael Hund, Frederik L. Dennig, Miriam Butt, and Daniel A. Keim. "HistoBankVis: Detecting Language Change via Data Visualization". In: *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*. Linköping University Electronic Press, 2017, pp. 32–39 (cit. on p. 12).

[Sch+18] Barret Schloerke, Jason Crowley, Di Cook, Heike Hofmann, Hadley Wickham, Francois Briatte, Moritz Marbach, Edwin Thoen, Amos Elberg, and Joseph Larmarange. *Ggally: Extension to ggplot2*. R package version 1.4.0. 2018 (cit. on pp. 73, 74, 78).

[SSK06] Jorn Schneidewind, Mike Sips, and Daniel A Keim. "Pixnostics: Towards measuring the value of visualization". In: *2006 IEEE Symposium On Visual Analytics Science And Technology*. IEEE. 2006, pp. 199–206 (cit. on pp. 20, 44, 45).

[Sch+16] Christoph Schulz, Arlind Nocaj, Mennatallah El-Assady, Steffen Frey, Marcel Hlawatsch, Michael Hund, Grzegorz Karol Karch, Rudolf Netzel, Christin Schätzle, Miriam Butt, Daniel A. Keim, Thomas Ertl, Ulrik Brandes, and Daniel Weiskopf. "Generative Data Models for Validation and Evaluation of Visualization Techniques". In: *Proceedings of the Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. 2016, pp. 112–124. DOI: 10.1145/2993901.2993907 (cit. on pp. 13, 52, 80).

[Sco92] David W Scott. *Multivariate density estimation: theory, practice, and visualization*. Wiley, 1992. DOI: 10.1002/9780470316849 (cit. on p. 110).

[SA15] Michael Sedlmair and Michaël Aupetit. "Data-driven Evaluation of Visual Quality Measures". In: *Computer Graphics Forum* 34.3 (2015), pp. 201–210. DOI: 10.1111/cgf.12632 (cit. on pp. 29, 37).

[Sed+12] Michael Sedlmair, Andrada Tatu, Tamara Munzner, and Melanie Tory. "A Taxonomy of Visual Cluster Separation Factors". In: *Computer Graphics Forum* 31.3pt4 (June 2012), pp. 1335–1344. DOI: 10.1111/j.1467-8659.2012.03125.x (cit. on p. 33).

[Sha+14] Lin Shao, Michael Behrisch, Tobias Schreck, Tatiana von Landesberger, Maximilian Scherer, Sebastian Bremm, and Daniel A. Keim. "Guided Sketching for Visual Search and Exploration in Large Scatter Plot Spaces". In: *Proc. EuroVA International Workshop on Visual Analytics*. Peer-reviewed short paper. 2014, pp. 19–23. DOI: 10.2312/eurova.20141140 (cit. on p. 168).

[Sha+16] Lin Shao, Timo Schleicher, Michael Behrisch, Tobias Schreck, Ivan Sipiran, and Daniel A. Keim. "Guiding the exploration of scatter plot data using motif-based interest measures". In: *Journal of Visual Languages and Computing* 36 (2016), pp. 1–12. DOI: 10.1016/j.jvlc.2016.07.003 (cit. on pp. 35, 36).

[Sha+17] Lin Shao, Nelson Silva, Eva Eggeling, and Tobias Schreck. "Visual Exploration of Large Scatter Plot Matrices by Pattern Recommendation Based on Eye Tracking". In: *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*. ESIDA '17. New York, NY, USA: ACM, 2017, pp. 9–16. DOI: 10.1145/3038462.3038463 (cit. on p. 37).

[SGM08] John Sharko, Georges Grinstein, and Kenneth A Marx. "Vectorized radviz and its application to multiple cluster datasets". In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008) (cit. on p. 47).

[She+17] Varshita Sher, Karen G. Bemis, Ilaria Liccardi, and Min Chen. "An Empirical Study on the Reliability of Perceiving Correlation Indices using Scatterplots". In: *Computer Graphics Forum* 36.3 (2017), pp. 61–72. DOI: 10.1111/cgf.13168 (cit. on p. 37).

[Shn96] Ben Shneiderman. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations". In: *IEEE Symposium on Visual Languages*. VL '96. IEEE Computer Society, 1996, pp. 336–343 (cit. on p. 20).

[Sie+72] John H Siegel, Edward J Farrell, Roger M Goldwyn, and Herman P Friedman. "The surgical implications of physiologic patterns in myocardial infarction shock". In: *Surgery* 72.1 (1972), pp. 126–141 (cit. on pp. 48, 55, 150).

[SM05] Harri Siirtola and Erkki Mäkinen. "Constructing and Reconstructing the Reorderable Matrix". In: *Information Visualization* 4.1 (Mar. 2005), pp. 32–48. DOI: 10.1057/palgrave.ivs.9500086 (cit. on p. 139).

[Sil86] Bernard W Silverman. *Density estimation for statistics and data analysis*. Vol. 26. CRC press, 1986. DOI: 10.1201/9781315140919 (cit. on pp. 110, 114, 150).

[Sip+09] Mike Sips, Boris Neubert, John P. Lewis, and Pat Hanrahan. "Selecting good views of high-dimensional data using class consistency". In: *Computer Graphics Forum* 28.3 (2009), pp. 831–838 (cit. on p. 36).

[SHK15] Drew Skau, Lane Harrison, and Robert Kosara. "An Evaluation of the Impact of Visual Embellishments in Bar Charts". In: *Comput. Graph. Forum* 34.3 (2015), pp. 221–230 (cit. on pp. 3, 110).

[Sno55] John Snow. *On the mode of communication of cholera*. John Churchill, 1855 (cit. on pp. 1, 2).

[SL91] Ian Spence and Stephan Lewandowsky. "Displaying proportions and percentages". In: *Applied Cognitive Psychology* 5.1 (1991), pp. 61–77. DOI: 10.1002/acp.2350050106 (cit. on p. 3).

[Spe14] Robert Spence. *Information Visualization - An Introduction*. Springer, 2014 (cit. on p. 5).

[SBB96] Michael Spenke, Christian Beilken, and Thomas Berlage. "FOCUS: The Interactive Table for Product Comparison and Selection". In: *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology (UIST)*. 1996, pp. 41–50. DOI: 10.1145/237091.237097 (cit. on p. 139).

[Spo18] Spotfire. *Spotfire Software, Data Visualization & Analytics Software*. https://spotfire.tibco.com/, Last accessed on 2018-03-30. 2018 (cit. on p. 139).

[Sto+05] Kurt Stockinger, Kesheng Wu, Scott Campbell, Stephen Lau, Mike Fisk, Eugene M. Gavrilov, Alex Kent, Christopher E. Davis, Richard D. Olinger, Robert J. Young, Jim Prewett, Paul M. Weber, Thomas P. Caudell, E. Wes Bethel, and Steve Smith. "Network Traffic Analysis With Query Driven Visualization SC 2005 HPC Analytics Results". In: *Proceedings of the 2005 ACM/IEEE SC2005 Conference on Supercomputing*. 2005. DOI: 10.1109/SC.2005.47 (cit. on p. 67).

[Sun+15] Swathy Sunil Kumar, Teenu Krishnan, Sreeja Ashok, and MV Judy. "Clutter Reduction in Parallel Coordinates using Binning Approach for Improved Visualization." In: *International Journal of Electrical & Computer Engineering (2088-8708)* 5.6 (2015). DOI: `10.11591/ijece.v5i6.pp1564-1568` (cit. on p. 68).

[SG16] Danielle Albers Szafir and Michael Gleicher. "Visualization-aware color design". In: *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Posters*. Eurographics Association. 2016, pp. 97–99 (cit. on p. 29).

[Szy34] D. Szymkiewicz. "Une contribution statistique a la géographie floristique". In: *Polskie Towarzystwo Botaniczne* (1934) (cit. on p. 83).

[Tab18] Tableau. *Tableau Software, Tableau is business intelligence software that helps people see and understand their data*. `https://www.tableau.com`, Last accessed on 2018-03-30. 2018 (cit. on pp. 110, 128, 139).

[TT78] Richard A Tapia and James R Thompson. *Nonparametric probability density estimation*. Johns Hopkins University Press, 1978 (cit. on p. 110).

[Tat+11] Andrada Tatu, Georgia Albuquerque, Martin Eisemann, Peter Bak, Holger Theisel, Marcus A. Magnor, and Daniel A. Keim. "Automated Analytical Methods to Support Visual Exploration of High-Dimensional Data". In: *IEEE Trans. on Vis. and Comp. Graph.* 17.5 (2011), pp. 584–597. DOI: `10.1109/TVCG.2010.242` (cit. on pp. 36, 57, 70, 74–76).

[Tat+09] Andrada Tatu, Georgia Albuquerque, Martin Eisemann, Jörn Schneidewind, Holger Theisel, Marcus A. Magnor, and Daniel A. Keim. "Combining automated analysis and visualization techniques for effective exploration of high-dimensional data". In: *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*. 2009, pp. 59–66. DOI: `10.1109/VAST.2009.5332628` (cit. on pp. 23, 36, 40, 41, 47, 75, 76, 140).

[Tat+10] Andrada Tatu, Peter Bak, Enrico Bertini, Daniel Keim, and Joern Schneidewind. "Visual Quality Metrics and Human Perception: An Initial Study on 2D Projections of Large Multidimensional Data". In: *Proc. Int. Conf. on Advanced Visual Interfaces*. AVI '10. ACM, 2010, pp. 49–56. DOI: `10.1145/1842993.1843002` (cit. on pp. 20, 37).

[TAS04] Christian Tominski, James Abello, and Heidrun Schumann. "Axes-based visualizations with radial layouts". In: *Proceedings of the 2004 ACM symposium on Applied computing*. ACM. 2004, pp. 1242–1247 (cit. on p. 45).

[TS20] Christian Tominski and Heidrun Schumann. *Interactive Visual Data Analysis*. AK Peters Visualization Series. CRC Press, 2020. DOI: `10.1201/9781315152707` (cit. on p. 5).

[Tor52] Warren S Torgerson. "Multidimensional scaling: I. Theory and method". In: *Psychometrika* 17.4 (1952), pp. 401–419 (cit. on pp. 134, 140).

[Tuf06] Edward R Tufte. *Beautiful evidence*. Vol. 1. Graphics Press Cheshire, CT, 2006 (cit. on pp. 49, 139).

[TG83] Edward R Tufte and PR Graves-Morris. *The visual display of quantitative information*. Vol. 2. Graphics Press, 1983 (cit. on pp. 17, 29).

[Tuf01] Edward R. Tufte. *The Visual Display of Quantitative Information*. 2nd. Cheshire, CT, USA: Graphics Press, 2001 (cit. on pp. 79, 98).

[TT85] J. W. Tukey and P. A. Tukey. "Computer Graphics and Exploratory Data Analysis: An Introduction". In: *Proc. the Sixth Annual Conference and Exposition: Computer Graphics '85, Vol. III, Technical Sessions*. Nat. Computer Graphics Association. 1985, pp. 773–785 (cit. on p. 35).

[Tuk77] John W. Tukey. *Exploratory Data Analysis*. Pearson, 1977 (cit. on pp. 1, 108, 110, 114).

[Twe+15] James Twellmeyer, Marco Hutter, Michael Behrisch, Jörn Kohlhammer, and Tobias Schreck. "The Visual Exploration of Aggregate Similarity for Multi-dimensional Clustering". In: *In Proceedings of the 6th International Conference on Information Visualization Theory and Applications (IVAPP)*. 2015, pp. 40–50. DOI: 10.5220/0005304100400050 (cit. on p. 139).

[US09] Andrea Unger and Heidrun Schumann. "Visual Support for the Understanding of Simulation Processes". In: *Visualization Symposium, 2009. PacificVis '09. IEEE Pacific*. 2009, pp. 57–64. DOI: 10.1109/PACIFICVIS.2009.4906838 (cit. on p. 48).

[Van15] T. Van Long. "A new metric on parallel coordinates and its application for high-dimensional data visualization". In: *IEEE International Conference on Advanced Technologies for Communications*. 2015, pp. 297–301. DOI: 10.1109/ATC.2015.7388338 (cit. on pp. 74, 75).

[Var+16] Manasi Vartak, Silu Huang, Tarique Siddiqui, Samuel Madden, and Aditya G. Parameswaran. "Towards Visualization Recommendation Systems". In: *SIGMOD Rec.* 45.4 (2016), pp. 34–39 (cit. on p. 168).

[VC62] Hough Paul VC. *Method and means for recognizing complex patterns*. US Patent 3,069,654. Dec. 1962 (cit. on p. 40).

[Veg+98] Marisol Vega, Rafael Pardo, Enrique Barrado, and Luis Debán. "Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis". In: *Water Research* 32.12 (1998), pp. 3581–3592. DOI: 10.1016/S0043-1354(98)00138-9 (cit. on p. 113).

[Vil+11] Gabriele Villarini, James A. Smith, Mary Lynn Baeck, and Witold F. Krajewski. "Examining Flood Frequency Distributions in the Midwest U.S." In: *JAWRA Journal of the American Water Resources Association* 47.3 (2011), pp. 447–463. DOI: 10.1111/j.1752-1688.2011.00540.x (cit. on p. 113).

[Vil+17] Karoline Villinger, Deborah R. Wahl, Gudrun Sproesser, Harald T. Schupp, and Britta Renner. "A visual analysis of the behavioral signature of eating: The case of breakfast". In: *The European Health Psychologist* 19.Supp (2017), p. 689 (cit. on p. 135).

[Wah+20] Deborah R. Wahl, Karoline Villinger, Michael Blumenschein, Laura M. König, Katrin Ziesemer, Gudrun Sproesser, Harald T. Schupp, and Britta Renner. "Why We Eat What We Eat: Assessing Dispositional and In-the-Moment Eating Motives by Using Ecological Momentary Assessment". In: *JMIR mHealth and uHealth* 8.1 (2020), pp. e13191. DOI: 10.2196/13191 (cit. on p. 12).

[Wah+17] Deborah R. Wahl, Karoline Villinger, Gudrun Sproesser, Harald T. Schupp, and Britta Renner. "The behavioral signature of snacking : a visual analysis". In: *The European Health Psychologist* 19.5 (2017), pp. 355–357 (cit. on p. 135).

[Wal+13] Rick Walker, Philip A. Legg, Serban Pop, Zhao Geng, Robert S. Laramee, and Jonathan C. Roberts. "Force-directed parallel coordinates". In: *Information Visualisation (IV), 2013 17th International Conference*. IEEE, 2013, pp. 36–44 (cit. on p. 41).

[Wal+18] Emily Wall, Subhajit Das, Ravish Chawla, Bharath Kalidindi, Eli T Brown, and Alex Endert. "Podium: Ranking data using mixed-initiative visual analytics". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), pp. 288–297 (cit. on p. 139).

[WM18] Bing Wang and Klaus Mueller. "The Subspace Voyager: Exploring High- Dimensional Data along a Continuum of Salient 3D Subspaces". In: *IEEE Transactions on Visualization and Computer Graphics* 24.2 (2018), pp. 1204–1222. DOI: 10.1109/TVCG.2017.2672987 (cit. on pp. 6, 140).

[Wan+18a] Yunhai Wang, Kang Feng, Xiaowei Chu, Jian Zhang, Chi-Wing Fu, Michael Sedlmair, Xiaohui Yu, and Baoquan Chen. "A Perception-Driven Approach to Supervised Dimensionality Reduction for Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 24.5 (2018), pp. 1828–1840. DOI: 10.1109/TVCG.2017.2701829 (cit. on pp. 24, 33).

[Wan+18b] Yunhai Wang, Fubo Han, Lifeng Zhu, Oliver Deussen, and Baoquan Chen. "Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series". In: *IEEE Transactions on Visualization and Computer Graphics* 24.2 (2018), pp. 1141–1154. DOI: 10.1109/TVCG.2017.2653106 (cit. on p. 24).

[WG11] Matthew O Ward and Zhenyu Guo. "Visual Exploration of Time-Series Data with Shape Space Projections". In: *Computer Graphics Forum*. Vol. 30 - 3. Wiley Online Library. 2011, pp. 701–710 (cit. on p. 48).

[War08] Matthew O. Ward. "Multivariate Data Glyphs: Principles and Practice". In: *Handbook of Data Visualization*. Springer, 2008, pp. 179–198. DOI: 10.1007/978-3-540-33037-0_8 (cit. on pp. 56–58).

[War63] Joe H Ward Jr. "Hierarchical grouping to optimize an objective function". In: *Journal of the American statistical association* 58.301 (1963), pp. 236–244 (cit. on p. 81).

[War20] Colin Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2020. DOI: 10.1016/B978-155860819-1/50001-7 (cit. on pp. 5, 27, 58, 90, 98).

[War+02] Colin Ware, Helen Purchase, Linda Colpoys, and Matthew McGill. "Cognitive measurements of graph aesthetics". In: *Information visualization* 1.2 (2002), pp. 103–110 (cit. on p. 29).

[Wat05] Martin Wattenberg. "A note on space-filling visualizations and space-filling curves". In: *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE. 2005, pp. 181–186 (cit. on p. 42).

[Weg90] Edward J. Wegman. "Hyperdimensional Data Analysis Using Parallel Coordinates". In: *Journal of the American Statistical Association* 85.411 (1990), pp. 664–675. DOI: 10.1080/01621459.1990.10474926 (cit. on p. 73).

[Wer23] Max Wertheimer. "Untersuchungen zur Lehre von der Gestalt. II". In: *Psychologische Forschung* 4.1 (1923), pp. 301–350. DOI: 10.1007/BF00410640 (cit. on pp. 5, 90).

[Wil05] Leland Wilkinson. *The Grammar of Graphics (Statistics and Computing)*. Berlin, Heidelberg: Springer-Verlag, 2005. DOI: 10.1007/0-387-28695-0 (cit. on p. 115).

[WAG05] Leland Wilkinson, Anushka Anand, and Robert L. Grossman. "Graph-Theoretic Scagnostics". In: *IEEE Symposium on Information Visualization*. 2005, pp. 157–164. DOI: 10.1109/INFVIS.2005.1532142 (cit. on pp. 20, 26, 35, 36, 40, 78).

[WW08] Leland Wilkinson and Graham Wills. "Scagnostics Distributions". In: *Journal of Computational & Graphical Statistics* 17.2 (2008), pp. 473–491. DOI: 10.1198/106186008X320465. eprint: https://doi.org/10.1198/106186008X320465 (cit. on pp. 35, 36).

[WEG87] Svante Wold, Kim Esbensen, and Paul Geladi. "Principal component analysis". In: *Chemometrics and Intelligent Laboratory Systems* 2.1 (1987), pp. 37–52. DOI: 10.1016/0169-7439(87)80084-9 (cit. on pp. 33, 35).

[Won+16] Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand, Jock D. Mackinlay, Bill Howe, and Jeffrey Heer. "Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations". In: *IEEE Trans. Vis. Comput. Graph.* 22.1 (2016), pp. 649–658. DOI: 10.1109/TVCG.2015.2467191 (cit. on pp. 20, 110, 167).

[Won+17] Kanit Wongsuphasawat, Zening Qu, Dominik Moritz, Riley Chang, Felix Ouk, Anushka Anand, Jock D. Mackinlay, Bill Howe, and Jeffrey Heer. "Voyager 2: Augmenting Visual Analysis with Partial View Specifications". In: *CHI*. ACM, 2017, pp. 2648–2659 (cit. on pp. 20, 110, 168).

[Wri+12] Jim A. Wright, Hong Yang, Ulrike Rivett, and Stephen W. Gundry. "Public perception of drinking water safety in South Africa 2002–2009: a repeated cross-sectional study". In: *BMC Public Health* 12.1 (2012), p. 556. DOI: 10.1186/1471-2458-12-556 (cit. on p. 113).

[Xia+18] Jiazhi Xia, Fenjin Ye, Wei Chen, Yusi Wang, Weifeng Chen, Yuxin Ma, and Anthony K. H. Tung. "LDSScanner: Exploratory Analysis of Low-Dimensional Structures in High-Dimensional Datasets". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), pp. 236–245. DOI: 10.1109/TVCG.2017.2744098 (cit. on pp. 6, 140).

[Xia+12] Yang Xiang, David Fuhry, Ruoming Jin, Ye Zhao, and Kun Huang. "Visualizing Clusters in Parallel Coordinates for Visual Knowledge Discovery". In: *Advances in Knowledge Discovery and Data Mining*. Springer, 2012, pp. 505–516. DOI: 10.1007/978-3-642-30217-6_42 (cit. on pp. 74, 75).

[Xu+20] Kai Xu, Alvitta Ottley, Conny Walchshofer, Marc Streit, Remco Chang, and John Wenskovitch. "Survey on the Analysis of User Interactions and Visualization Provenance". In: *Computer Graphics Forum* (2020). DOI: 10.1111/cgf.14035 (cit. on p. 167).

[Yan+03] Jing Yang, Wei Peng, Matthew O. Ward, and Elke A. Rundensteiner. "Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration of High Dimensional Datasets". In: *IEEE Symposium on Information Visualization*. 2003, pp. 105–112. DOI: 10.1109/INFVIS.2003.1249015 (cit. on pp. 39–41, 50, 56, 57, 68, 73–76, 78).

[YF00] Julie Yang-Peléz and Woodie C Flowers. "Information content measures of visual displays". In: *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*. IEEE. 2000, pp. 99–103 (cit. on p. 29).

[Yua+13] Xiaoru Yuan, Donghao Ren, Zuchao Wang, and Cong Guo. "Dimension Projection Matrix/Tree: Interactive Subspace Visual Exploration and Analysis of High Dimensional Data". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2625–2633. DOI: 10.1109/TVCG.2013.150 (cit. on pp. 6, 35, 140).

[Zha+16] Leishi Zhang, Chris Rooney, Lev Nachmanson, William Wong, Bum Chul Kwon, Florian Stoffel, Michael Hund, Nadeem Qazi, Uchit Singh, and Daniel A. Keim. "Spherical Similarity Explorer for Comparative Case Analysis". In: *IS&T Electronic Imaging Conference on Visualization and Data Analysis*. 2016, pp. 1–10. DOI: 10.2352/ISSN.2470-1173.2016.1.VDA-496 (cit. on p. 13).

[ZMM12] Zhiyuan Zhang, Kevin T. McDonnell, and Klaus Mueller. "A Network-Based Interface for the Exploration of High-Dimensional Data Spaces". In: *IEEE Pacific Visualization Symp.* IEEE Computer Society, 2012, pp. 17–24. DOI: 10.1109/PacificVis.2012.6183569 (cit. on p. 57).

[Zho+09] Hong Zhou, Weiwei Cui, Huamin Qu, Yingcai Wu, Xiaoru Yuan, and Wei Zhuo. "Splatting the Lines in Parallel Coordinates". In: *Computer Graphics Forum* 28.3 (2009), pp. 759–766. DOI: 10.1111/j.1467-8659.2009.01476.x (cit. on pp. 94, 96).

[Zho+08] Hong Zhou, Xiaoru Yuan, Huamin Qu, Weiwei Cui, and Baoquan Chen. "Visual Clustering in Parallel Coordinates". In: *Comput. Graph. Forum* 27.3 (2008), pp. 1047–1054. DOI: 10.1111/j.1467-8659.2008.01241.x (cit. on pp. 73, 140).

[ZH16] Liang Zhou and Charles D Hansen. "A survey of colormaps in visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 22.8 (2016), pp. 2051–2069 (cit. on p. 143).

[Zho+18] Zhiguang Zhou, Zhifei Ye, Jiajun Yu, and Weifeng Chen. "Cluster-aware arrangement of the parallel coordinate plots". In: *J. Vis. Lang. Comput.* 46 (2018), pp. 43–52. DOI: 10.1016/j.jvlc.2017.10.003 (cit. on pp. 74–76).

[ZSK12] Arthur Zimek, Erich Schubert, and Hans-Peter Kriegel. "A survey on unsupervised outlier detection in high-dimensional numerical data". In: *Statistical Analysis and Data Mining* 5.5 (2012), pp. 363–387 (cit. on p. 166).