Glyph Design for Temporal and Multi-Dimensional Data: Design Considerations and Evaluation

Dissertation zur Erlangung des akademischen Grades
eines
Doktors der Naturwissenschaften

vorgelegt von

Johannes Hermann Fuchs

an der

Universität Konstanz

Mathematisch-Naturwissenschaftliche Sektion
Informatik und Informationswissenschaft

Tag der mündlichen Prüfung: 12. November 2015
1. Referent: Prof. Dr. Daniel A. Keim
2. Referent: Prof. Dr. Harald Reiterer

Abstract

The goal of this thesis is to provide researchers and practitioners with guidance in designing data glyphs for temporal and multi-dimensional data. Therefore, the term ”glyph” in the context of information visualization has to be introduced and defined first, to establish a common understanding of the overall topic and motivate the need for additional support in selecting or creating data glyphs. This definition is the basis for reviewing literature about data glyph experiments, for conducting further controlled user studies, and finally for introducing new data glyph designs. In summary, the computer science contributions in the area of information visualization are threefold.

First, literature about quantitative experiments on data glyphs from the past 70 years is systematically reviewed. By sampling and tabulating the literature on data glyph studies, listing their designs, questions, data, and tasks an overview about study goals and results is provided and open research gaps are revealed. Based on this meta analysis of all results a catalog of design considerations is created, which will be further extended throughout this thesis.

Second, the previously identified research gaps are used as a motivation for conducting controlled user studies, which are introduced in this thesis. Since variations of star glyphs and radial color encodings have not received much research attention, these designs will be subject to quantitative experiments. Results indicate that, against intuition, the whisker glyph which is hardly used in practice outperforms the alternative star glyph variations. Additionally, further study outcomes suggest that radial glyph layouts making use of the visual variable orientation to separate different dimensions are the best choice for detecting specific points in time. This finding contradicts the ranking of visual variables from Cleveland and McGill where position encodings outperform orientation encodings. Based on these results the set of design considerations collected in the initial survey is extended and summarized to facilitate the guidance in creating and selecting data glyph designs.

Third, the design space of data glyphs is enriched with two new metaphoric designs tailored towards specific domains and evaluated with use cases and controlled user studies to show their applicability to real-world scenarios. The *clock glyph* representation, for example, supports the analyst in detecting specific points in time by arranging the temporal dimensions in a radial fashion. Results from quantitative experiments indicate the usefulness of this metaphoric approach outperforming well-established alternative representations like line glyphs. The *leaf glyph* technique on the other hand makes use of environmental cues to encode multi-dimensional data controlling main leaf properties like leaf morphology, leaf venation, and leaf boundary. The design is motivated by the human ability to visually discriminate natural shapes like trees in a

forest, single flowers in a flowerbed, or leaves at shrubs. Due to its aesthetically pleasing appearance, this design is suitable for being used in mass media and data journalism. A use case scenario with forest fire data reveals the strengths of this design being effectively interpretable for storytelling in environmental data analysis.

Zusammenfassung

Die vorliegende Dissertation stellt Richtlinien zur Verfügung, welche Designer und Forscher bei der Wahl anwendungsspezifischer Datenglyphen unterstützen. Zu Beginn wird der Begriff ”Glyph” im Zusammenhang mit der Datenvisualisierung näher beleuchtet, um die Thematik einzuführen und die bestehende Problematik zu motivieren. Darüber hinaus bildet diese Definition die Grundlage um Studien über Datenglyphen in der Literatur zu finden, eigene Experimente durchzuführen und neue Datenglyphen einzuführen. Zusammenfassend bereichert diese Arbeit die Wissenschaft im Gebiet Datenvisualisierung durch drei Beiträge.

Zuallererst werden sämtliche Benutzerstudien über Datenglyphen der letzten 70 Jahre untersucht. Aufgrund der systematischen Vorgehensweise können Fragen, Datentypen, Glyphendesigns und Aufgaben extrahiert und eine Übersicht erstellt werden. Dabei werden die Ziele und Ergebnisse unterschiedlicher Studien sinngemäß zusammengefasst und offene Forschungslücken aufgedeckt. Das Resultat dieser Metaanalyse fließt anschließend in einen Katalog von Designrichtlinien ein, welcher später noch weiter ausgearbeitet wird.

Als Nächstes werden die zuvor enthüllten Forschungslücken als Ausgangspunkt genommen, um weitere Studien durchzuführen. Da Variationen von Star Glyphen und Designs mit einer zirkulären Farbkodierung wenig erforscht wurden, werden diesbezüglich weitere Experimente durchgeführt. Entgegen aller Erwartungen schneiden sogenannte Whisker Glyphen besser ab als die weiter verbreiteten gewöhnlichen Star Glyphen. Darüber hinaus sind die zirkulären Farbkodierungen bestens geeignet um einzelne Dimensionen in einer Zeitserie zu erkennen. Dieses Ergebnis ist überraschend, da es dem bereits etablierten Ranking von Cleveland und McGill in Bezug auf visuelle Variablen widerspricht. Die Positionskodierung ist in diesem speziellen Fall nämlich weniger effektiv als Kodierungen mit Hilfe der Orientierung. Basierend auf diesen Ergebnissen werden weitere Richtlinien für das Design von Glyphen aufgestellt und die Liste an bereits bestehenden erweitert.

Abschließend wird der Glyphengestaltungsraum mit zwei neuen auf Metaphern basierenden Designs erweitert und mittels anwendungsspezifischer Szenarien und Benutzerstudien evaluiert. Der *clock glyph* beispielsweise unterstützt Analysten aufgrund der zirkulären Dimensionsanordnung dabei bestimmte Zeitpunkte in einer Zeitserie zu finden. Die Ergebnisse quantitativer Studien unterstreichen den Nutzen dieser Uhrmetapher, da alternative Repräsentationen wie beispielsweise kleine Liniendiagramme schlechter abschneiden. Der *leaf glyph* hingegen verwendet natürliche Formen aus der Umwelt um multidimensionale Daten darzustellen. Dies geschieht, indem die Gestalt von Blättern, deren Aderung, sowie deren Blattrand aufgrund der Daten angepasst werden. Die Fähigkeit des Menschen visuell natürliche Formen wie beispiel-

sweise Bäume in Wäldern, einzelne Blumen im Blumenbeet, oder Blätter an Büschen zu unterscheiden ist die Motivation dieses Designs. Aufgrund der künstlerischen Darstellung des Blattes ist diese Visualisierungsart auch für das breite Publikum oder in Medien geeignet. Ein Anwendungsszenario auf Basis von Waldbranddaten zeigt die Stärken des Designs durch ein erzählerisches Analysieren der Daten.

Acknowledgments

First and foremost, I want to thank Daniel Keim for giving me the opportunity to work in his group and to conduct valuable research. I would also like to thank Harald Reiterer for being my second advisor and his support at an early stage of my research. Moreover, I would like to thank Michael Grossniklaus for being part in this committee and the many discussions about various topics.

Enrico Bertini for raising the interest in quantitative evaluations. Petra Isenberg and Anastasia Bezerianos for supporting me in my research by shaping ideas, writing papers, and a lot of valuable feedback. Fabian Fischer for making the office a productive but also an enjoyable place to be. All my other colleagues at the University of Konstanz for their support, ideas, feedback, and discussions.

Last but not least, I would like to thank my family and friends who contributed to this thesis in a special way.

THANK YOU

Contents

1 Introduction **1**

2 Systematic Review of Experimental Studies on Data Glyphs **7**

2.1 Related Work . . . 9
2.2 Methodology . . . 10
2.2.1 Paper Sampling and Collection . . . 10
2.2.2 Analyzed Study Characteristics . . . 10
2.3 Results: State-of-the-Art in Glyph Evaluation . . . 12
2.3.1 Study Goals . . . 12
2.3.2 Glyph Types and Data Encoding . . . 13
2.3.3 Glyph Presentation Settings . . . 15
2.3.4 Datasets & Number of Dimensions . . . 17
2.3.5 Task Space . . . 17
2.3.6 Study Outcomes . . . 18
2.4 Discussion and Open Research Areas . . . 26
2.5 Summary . . . 29

3 Data Glyph Designs for Time-Series Data **31**

3.1 Related Work . . . 32
3.1.1 Geographic Maps . . . 32
3.1.2 Node-link Diagrams . . . 33
3.1.3 Grid Layouts, Matrix Visualizations . . . 34
3.1.4 Text Visualizations . . . 35
3.1.5 Hierarchical Visualizations . . . 35
3.1.6 Flow Visualizations . . . 35
3.1.7 Projection to 2D Space . . . 35
3.2 Clock Glyph - A Data Glyph Design to Visualize Time-Series Data . . . 36
3.2.1 Design Space for Temporal Glyphs . . . 36
3.2.2 Application-Oriented Evaluation in the Network Security Domain . . . 38
3.2.3 Conclusion . . . 46
3.3 Evaluation of Alternative Glyph Designs for Time-Series Data . . . 46
3.3.1 Experiment Design . . . 47
3.3.2 Results . . . 52

3.3.3 Discussion . . . 56
3.3.4 Design Considerations . . . 58
3.3.5 Limitations . . . 59
3.3.6 Conclusion . . . 59
3.4 Summary . . . 60

4 Data Glyph Designs for Multi-Dimensional Data 61
4.1 Related Work . . . 62
4.1.1 Many-To-One Mapping . . . 62
4.1.2 One-To-One Mapping . . . 67
4.1.3 One-To-Many Mapping . . . 68
4.2 Leaf Glyph - A Data Glyph Design to Visualize Multi-Dimensional Data with Environmental Cues . . . 69
4.2.1 Design Space for Environmental Data Glyphs . . . 71
4.2.2 Leaf Glyph Aggregation . . . 75
4.2.3 Use Case Scenarios . . . 81
4.2.4 Conclusion . . . 86
4.3 The Influence of Contour on Similarity Perception of Star Glyphs . . . 88
4.3.1 Experiment 1: Contours for Novices vs. Experts . . . 90
4.3.2 Experiment 2: Perception of Similarity . . . 98
4.3.3 Experiment 3: Improvements for Star Glyphs . . . 104
4.3.4 Design Considerations . . . 109
4.3.5 Limitations . . . 110
4.3.6 Conclusion . . . 110
4.4 Summary . . . 111

5 Design Considerations 113
5.1 Elementary Analysis Task . . . 113
5.2 Synoptic Analysis Task . . . 115
5.3 Glyph Placement Strategy . . . 116
5.4 Summary . . . 117

6 Conclusions & Future Research Directions 119

Chapter 1

Introduction

Glyph-based data visualization has a long tradition in information visualization. A lot of research has already been conducted in developing new designs, combining them with other visualization techniques, improving layout algorithms, or comparing alternative representations. However, there is only little guidance for the usage and design of data glyphs. Which glyph design is best for analyzing specific datasets? Are there changes in performance when switching to a different analysis task? Do guidelines exist for positioning data glyphs on the screen [75]? Researchers are aware of this lack in guidance. As Matthew Ward stated:

> *"Glyphs are a popular, but insufficiently studied, class of techniques for the visualization of data."*[1]

Since data glyphs consist of multiple different visual features and can be flexibly arranged on the screen the resulting design space is nearly endless. During the years many different designs have been introduced in literature and probably more are about to be developed. Without any structure and guidance this flexibility in design will be overwhelming. By providing design considerations and evaluating alternative glyph representations practitioners as well as researchers can be supported in selecting the most appropriate data glyph for specific tasks and datasets.

However, as a first step, a common understanding of the overall topic by discussing the term "glyph" in the context of information visualization needs to be established. In literature the term is used in various ways and a general definition does not exist, yet. By investigating the historic background of the term and additionally discussing definitions used in todays literature a more general definition can be contributed.

Historic background and definition: Most people may associate the term "glyph" with the Egyptian "hieroglyph", which is a sacred (*hierós*) character of the ancient Egyptian writing system (4000 BCE) engraved (*glýphõ*) in papyrus and wood. It was commonly used in a religious context to communicate god's words [5]. However, glyphs are as old as 40000 years and originate from the term "petroglyphs". Like hieroglyphs, petroglyphs were used as a means for communication. Pictograms, or logogram images were engraved in stone (*petro*) as a form of pre-writing [187]. Although, hieroglyphs and petroglyphs seem to be entirely different from nowadays data

[1] [192], page 191.

glyph visualizations they share some interesting characteristics. They are all trying to visually communicate information in a rather abstract way or by means of metaphors. Therefore, people have to first learn how to read these glyphs in order to understand their meaning. These fundamental similarities are the reason for the term glyph being eponymous for data glyph based visualizations [17].

In the area of information visualization glyph-based data visualization has a long tradition in research and application. The basic idea is to map data properties to visual properties of some appropriately designed visual structure. By the interplay of the different visual properties, each glyph then represents a data record. Many data records can be compared by appropriately laid out glyph displays. One of the first glyph designs used in information visualization was the metroglyph introduced by Anderson in 1957 [8]. His idea was to represent multi-dimensional data in a two-dimensional scatterplot using single complex representations for each individual data point. The single marks in the scatterplot, which are usually used to depict the position of the data points have, therefore, been exchanged with a composition of a circle and multiple data lines. The data lines were attached to the circle using different angles with their lengths corresponding to the value of the respective dimensions. Analysts were then able to compare different data points by looking at their overall appearance or investigating the lengths of individual data lines.

Although alternative glyph designs have been introduced in the past [7] it is interesting to note that in literature from the 60s the term "glyph" was used as a synonym for the metroglyph [52, 91]. However, with the introduction of several alternative glyph representations like star glyphs [168], or Chernoff faces [35] in the early 70s this definition changed. From this point on researchers referred to the term "glyph" in a more general way including multiple designs. As a consequence, different definitions emerged based on the current state-of-the art in glyph design and based on the subjective preference of the researchers. An objective view including the characterizing properties of a glyph visualization has not established in literature, yet. To quote from Munzner:

"*There is no strict dividing line between a region, a view, and a glyph. [...] the word glyph is used very ambiguously in the vis literature.*"[2].

To start organizing important characteristics of data glyphs, definitions and keywords were extracted from books on visualization, from papers in the literature that used the term "glyph", and from interviews with visualizations experts about their understanding of the term. The collected results can be found in Table 1.1. While Table 1.1 is certainly not exhaustive, it serves to show the wide variety of ways researchers think of, and define, data glyphs. The following discussion will contribute to a sophisticated basis for phrasing the definition, which helps to understand how the term "glyph" is used throughout this thesis.

[2] [134], page 280.

References	Data glyph characteristics			
	small, or compact	encodes different attribute dimensions	uses different visual channels	is also called: symbol, icon, or sign
[17]	X	X	X	disagree
[73]	X	X	-	X
[121]	X	X	-	-
[22]	-	X	X	-
[41]	-	X	X	sometimes
[112]	-	X	X	X
[115]	-	X	X	-
[134]	-	X	X	-
[157]	-	X	X	X
[195]	-	X	X	a symbol encoding quantities
[192]	-	X	-	-
[40]	-	-	X	X
[153]	-	-	X	-

Table 1.1: Overview of defining glyph characteristics mentioned in the literature.

Most researchers agree on the fact that data glyphs encode multiple attribute dimensions using different visual channels. Since data glyphs are basically a composition of different visual variables this definition seems to perfectly fit. Designs using e.g., a color encoding make use of a different visual channel as designs using e.g., size to represent data values.

However, there seems to be a conflict when considering simple data marks as glyphs. Ward states that "*glyphs are dictated by one or more attributes of a data record*"[3]. However, this contradicts the definition of Munzner saying that "*a glyph is made of multiple marks*"[4], thus, excluding single simple data marks. This notion of a data glyph as a multi-dimensional encoding also aligns with the majority of historical data glyph definitions. Therefore, I stick with this definition and discard the idea of single marks being equivalent to data glyphs.

Another controversial aspect is the size of a data glyph. Some researchers think that a fundamental characteristic of glyphs is their small and compact size. However, is this criteria really mandatory for defining data glyphs? In my opinion the notion of size highly depends on the context a glyph is presented in. On smaller displays like tablets, or mobile phones glyphs are perceived bigger compared to wall-sized displays or high resolution projections. Additionally, to the best of my knowledge, there exists no threshold for objects being categorized as small or big. I think more important as a certain threshold for the maximum size of a data glyph is its context information and information carrying embellishments. Data glyphs are always embedded in a context giving environment. This can either be a basic visualization like scatterplots, treemaps, node-link diagrams, geographic maps etc. or just a grid-based layout with some sort of ordering. In contrast to charts or other visualization techniques does the position of a glyph always convey a meaning. Ward refers to such kind of arrangements as data-driven or structure-driven layouts [191]. Additionally, glyph designs do not contain detailed axes or labels since they are

[3][192], page 179.
[4][134], page 280.

primarily designed to show multiple attributes in a compact way [195]. As a result less ink is necessary to plot the glyph, which will lead to smaller representations and less visual clutter or reduced overplotting. In summary, I would argue that data glyphs need not necessarily be small. However, removing labels, axes, or other detailed descriptions from the design already reduces their overall size to a minimum without the need of identifying an exact threshold for data glyphs being considered “small”.

In information visualization other terms are also used as synonyms for “glyphs”. Ropinski et al. say that symbols or iconic representations are considered glyphs [157]. However, Borgo et al. offer a detailed derivation of the individual terms showing that these expressions have a different meaning and should not be used as synonyms [17]. They clarify that icons are always metaphoric representations, symbols often take the form of characters or mathematical symbols, and signs are considered an umbrella term for all visual representations (for a more detailed explanation I refer the interested reader to Borgo’s STAR report [17]). Therefore, other terms in literature like symbols, or icons should not be mixed up with the definition of data glyphs.

As can be seen, various ideas of what a data glyph comprises exist. In this initial introduction I tried to reveal similarities and contradictions to sensitize the reader for this problem. To avoid any further confusion I, therefore, contribute my own definition of a data glyph, which is used throughout the whole thesis. As already stated, most parts of this definition were collected in interviews with visualization experts or derived from literature.

Data glyphs are *data-driven visual entities*, which make use of *different visual channels* to encode *multiple attribute dimensions*. They can be *arranged independently* on the screen and can *vary in size*. Their *position is always associated with a meaning*. Icons, symbols, and signs are *no synonyms* to data glyphs. Simple single marks such as points in a scatterplot (e.g., [127, 128]) are also *no data glyphs* because they cannot encode multiple attribute dimensions at once [134].

Design considerations and evaluations: Over the years various glyph designs were introduced. Star glyphs [168] and Chernoff faces [35] are just two examples of well-known glyph representations. Since star glyphs and metroglyphs use a similar visual mapping to represent the data values their overall appearance is similar. Chernoff faces on the other hand look entirely different because the data is mapped to various face characteristics using different visual variables like the angle of the eyebrows or the size of the ears. This simple example illustrates the nearly endless mapping possibilities of data dimensions to visual glyph encodings [134] and many more designs are certainly imaginable. This flexibility allows designers to come up with new and innovative glyph representations for specific datasets, tasks, or contexts. However, without any guidance, this freedom and large design space can become overwhelming.

Ward distinguishes between three different mapping strategies to better structure this nearly endless design space [192]. *One-to-one mappings* are designs, which visualize each dimension with a different visual variable. The aforementioned Chernoff faces fit into this category, since they encode each data value with different face characteristics. Star glyphs are an example for *many-to-one mappings* because they represent all data values with the same visual variable (i.e., length of the data line). The last category is the *one-to-many mapping*, where each data value is visualized redundantly by more than one visual variable. A possible example are colored star glyphs [103], which use the length and the color of a data line to represent the dimension value. This classification helps to better understand the visual encodings of different glyph designs, however, an indication, which category should be preferred and why is not given.

Knowledge of when and which types of designs work best or are preferred by viewers, could aid designers and practitioners in creating new designs or in selecting among existing ones. Several studies in various settings have been conducted investigating changes in performance when switching between different glyph designs. Such experiments help to identify the most suitable glyph representation for specific settings. Since many studies about glyph designs have been conducted over the years it is difficult to keep track of the results and possible implications for design choices. Till now a systematic review of this literature is missing, which would allow practitioners and researchers to find related experiments more efficiently. Additionally, such a survey would shed more light on open research gaps and reveal, which glyph designs need to be studied in more detail or which setting has not been investigated at all.

By introducing such a literature review I will provide an overview about all conducted quantitative experiments comparing different data glyphs and see how they perform according to certain tasks or datasets. Researchers interested in the performance of a specific data glyph design will be guided towards the relevant literature. Additionally, guidelines for designing data glyphs are extracted summarizing the overall outcome of all conducted experiments. The research gaps identified in this survey are the motivation for two quantitative experiments introduced later in this thesis. These user studies shed more light on the performance of different glyph designs and, therefore, contribute further design considerations.

Structure of the thesis: Based on the definition of data glyphs in the introduction, I will present a systematic review of quantitative user studies about data glyph designs in chapter 2. This survey provides the reader with an overview of study outcomes and settings and additionally identifies open research gaps. In chapter 3, I will introduce a new data glyph design for time-series data and show its applicability in the area of network security. This design will also be evaluated in a controlled experiment to compare its performance against well-known glyph alternatives and also to close some previously identified research gaps. In chapter 4 I will contribute the leaf glyph technique, which is a metaphoric glyph representation for multi-dimensional data using environmental cues. Additionally, the influence of contour lines for the well-known star glyph is evaluated to close another previously identified research gap. The design considerations found in the initial survey, as well as the guidelines retrieved from the experiments will be summarized in chapter 5. Chapter 6 concludes this work and discusses future research directions.

Chapter 2

Systematic Review of Experimental Studies on Data Glyphs

Parts of this chapter appear in the following publication:

- J. Fuchs, P. Isenberg, A. Bezerianos, and D. Keim. A systematic review of experimental studies on data glyphs. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2016[1]

As already indicated in the introduction many different data glyph designs have been introduced during the last 50 years with some of them being subjected to perceptual or comparative evaluations. Such evaluations are necessary to allow for a better understanding about changes in performance when using different glyph designs. Some researchers such as Cleveland and McGill proposed a ranking of visual variables according to their performance for different data types [43]. However, these suggestions need not necessarily be true for smaller glyph designs. Controlled user studies are, therefore, mandatory to propose design guidelines based on the results or even suggest the most suitable glyph designs for a given task or dataset. Since many different glyph designs exist it is not possible to compare all of them in a single study or paper. A survey of all studies conducted about data glyph designs helps to collect and summarize all experiments and provide researchers with an overview about the results.

In this chapter 2, I will focus on such a systematic review of the user-study literature on data glyphs focusing on quantitative controlled studies. In contrast to their qualitative counterpart, controlled experiments are more easily comparable and summarizable, as they test concrete hypotheses regarding design choices and isolate factors in the glyph designs [45]. The studies are

[1]The responsibilities for this joint publication were divided as follows: I did the literature search, categorized and characterized the papers, and spearheaded the writing. Petra Isenberg and Anastasia Bezerianos gave advice and feedback on the categorization and organization of papers, were involved in the writing and proofreading. Daniel Keim supervised the work.

categorized according to a number of criteria that are meant to help researchers and practitioners in choosing amongst the most relevant literature to read, and ultimately to make informed choices about glyph use, design, and potential future studies. These criteria include glyph types (see Figure 2.1), presentation settings, datasets, tasks, and study goals. A summary of study outcomes is extending this characterization to help practitioners select the most appropriate data glyphs according to different criteria like visual design, data density, or task. The discussion section additionally pinpoints to open research areas, some of them being tackled throughout my thesis.

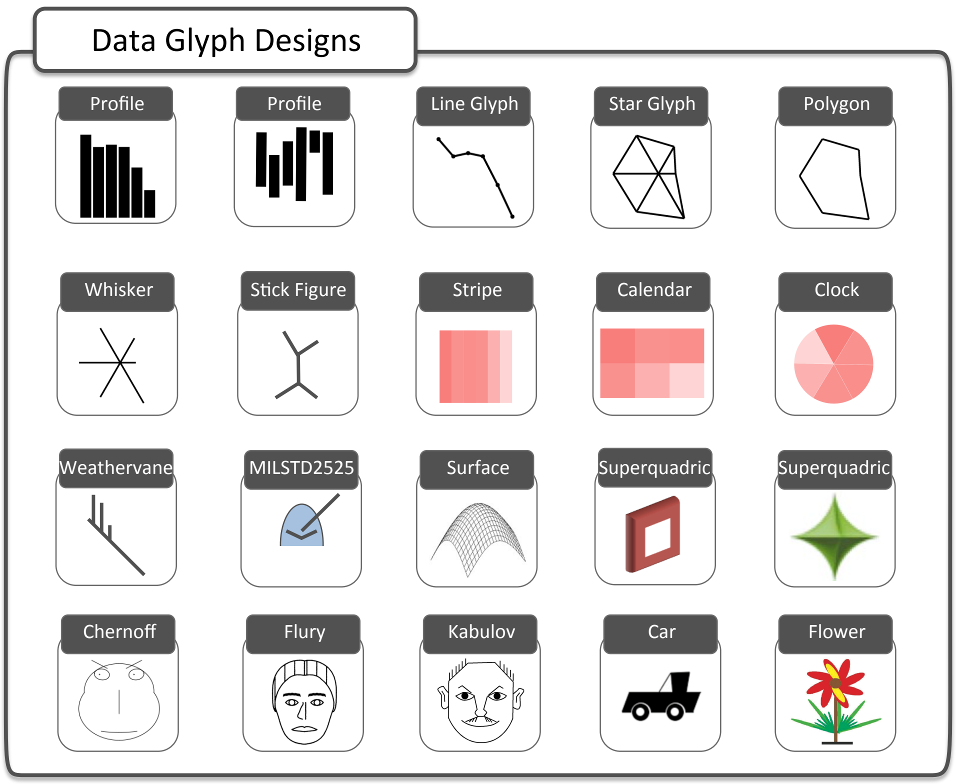


Figure 2.1: *Data glyphs:* This is a selection of the different data glyph designs used in the quantitative experiments.

The chapter is structured as follows: First, related surveys about data glyphs will be presented followed by sampling and categorization methods of the collected literature. The next section provides a summary of the collected study characteristics and outcomes and leads over to a discussion about open research gaps. The last section concludes the work and motivates the next main chapters.

2.1 Related Work

Many attempts have already been taken to structure existing data glyphs according to various criteria. Since a large number of glyph designs have been proposed in the past, a tabulation of existing individual designs is beyond the scope of this systematic review. This section focuses only on survey- or meta-papers. However, chapter 3 and chapter 4 will add additional design study papers to the related work.

Ward [191, 192] was one of the first to provide a structure of the glyph design space by classifying different layout possibilities into data-driven and structure-driven layouts. In a data-driven arrangement spatial position is determined by data: this can either be the raw data used as in a scatterplot, or a projection-based approach such as PCA. A structure-driven layout makes use of relations between the data points to calculate a layout. For example, hierarchical information can be used to lay out glyphs in a tree structure. Ward extended this work with a categorization of the visual characteristics of data glyphs. He structured data glyphs based on their mapping of data to visual attributes into three different classes: a *many-to-one mapping* where each data dimension is mapped to the same visual variable (e.g., profile glyph [52]); a *one-to-one mapping* showing each data attribute with a different visual variable (e. g., Chernoff faces [35]); and a *one-to-many mapping* representing the data dimensions redundantly with many different visual attributes (e. g., compound glyph [147]). In this survey, we[2] use this categorization to structure our own categorization of data glyph user-studies.

In contract, Chung et al. [42] proposed a categorization based on the visual channels used to represent the data and the spatial dimensionality of the glyph (2D, 2.5D, and 3D). The authors also discussed critical design aspects and guidelines for glyph visualizations, such as the normalization of data input for each dimension, the use of redundant mappings, and the visual orthogonality of different glyph components to ensure best performance. Since some of these guidelines cannot be followed for a high number of dimensions, designers have to choose between few single complex glyph designs, or many simple designs. Additionally, they suggested using halos to limit the negative effect of overplotting. In our survey we extend this list of guidelines based on our review of experimental results and provide further open research questions.

An extensive survey about data glyphs was presented by Borgo et al. [17]. The authors cover different glyph representations and propose guidelines for designing data glyphs based on a collection of design principles in the literature. While Borgo et al. also include several empirical studies in their survey, their focus is on design study papers showing the applicability of data glyphs to different data sets and tasks. In contrast to this work, we provide an overview of performance assessments from quantitative user studies.

A more data-specific survey on glyphs in the medical domain was presented by Ropinski et al. [157]. The authors classified glyph-based visualizations for medical data into two groups: pre-attentively and attentively identifiable glyph designs. Based on this grouping the authors further derive design guidelines for developing glyphs for this domain, but provide no additional empirical results from user studies.

While there is no systematic assessment of glyph user-studies that we know of, some re-

[2] In this chapter 2 the term ”we” comprises Petra Isenberg, Anastasia Bezerianos, Daniel Keim and me.

searchers have categorized subsets of the study design space. Nelson [137], for example, discusses the history of Chernoff faces [35] with its many variations such as the Flury-Rydwiel [65] or Kabulov faces [96]. She also discusses studies investigating performance changes for different data types or visual variations. We took this work as inspiration, but provide a much more comprehensive view on the study design space. Ware's [Ch. 5][195] discussion on "Glyphs and Multivariate Discrete Data" is related to our work in that he categorizes two types of user study tasks for glyphs. He focuses on tasks designed to find out which display dimensions are perceived holistically (integral) or perceived separately (separable): restricted classification tasks and speeded classification tasks. Among others, we include both types of tasks in our discussion based on slightly different terminology [10]: similarity search tasks (related to restricted classification) and lookup tasks (similar to speeded classification).

2.2 Methodology

This systematic review highlights only user studies in which participants performed controlled, quantitatively measured tasks with data glyphs. These quantitative measurements could (but did not have to) be accompanied by a subjective assessment of the tested glyphs (e.g., according to aesthetics, confidence, etc.) The categorization of the found studies is done according to the criteria discussed in the following.

2.2.1 Paper Sampling and Collection

To find relevant papers for our review we used a snowball sampling technique in which we first searched for the keyword "glyph" in the title, abstract, and keywords in the ACM digital library (leading to 80 potential results), the IEEE Xplore digital library (leading to 255 potential results), the EG digital library (leading to 66 potential results), and the DBLP computer science bibliography (leading to 134 potential results). In a next step we excluded papers that did not include at least one user study with quantitative measures or did not study glyphs that fit our definition. This filtering step removed 505 of the 535 candidate papers, leaving 30 relevant papers for our survey. From this initial set of papers we recursively scanned references for further user studies about data glyphs. Using this approach we collected 64 papers from the visualization literature as well as work from statistics and psychology.

2.2.2 Analyzed Study Characteristics

In the design of any quantitative user study several characteristics are important: the tasks to be performed, the collected measures, the presentation of the stimuli (glyphs), the size and type of data visualized, the general presentation setting, and the study goals (or main research questions) [45]. We categorized the 64 study papers using these characteristics as explained in more detail next.

Glyph Types and Data Encoding

We used Ward's data mapping taxonomy [192] to distinguish between glyphs using many-to-one and one-to-one mappings (see section 2.1). The rows and columns of Figure 2.3 give an overview of this categorization. We only found two occurrences of Ward's third group: one-to-many mappings. Thus, we do not highlight this group as a category in our result table. The two studies we found ([49] and [103]) are, however, discussed throughout this chapter.

Since the many-to-one group encodes multiple data point dimensions using the same visual variable, we further split this group into categories based on the visual variables used: position/length, color saturation, and orientation/angle (see Figure 2.3). We also distinguished whether or not a linear or circular layout was chosen to lay out the dimensions.

The category of one-to-one mapping was structured slightly differently as it includes a wide variety of design choices. As we mostly found facial glyph representations or three-dimensional designs, the result table includes these two categories: Faces and 3D Glyphs. A third category on car glyphs was added, since in one paper [170] faces were compared against unique car glyph representations. Car glyphs are abstract two dimensional representations of vehicles, which use unique characteristics (size of the trunk or hood) to encode data.

We additionally found twelve studies that tested unique glyph designs that were not compared to alternative representations: PlanningLines [3], weather vanes [123, 151], shapes [84, 206], roses [114], themes [31], arrows [200], Motifs [30, 53], flowers [33], and MILSTD2525 glyphs [172]. Rather they were either compared against textual information, tested on varying backgrounds (changes in the topological level of detail), or against different types of visualizations. Since they were not compared to other designs in the table, we positioned them slightly apart in the "One-to-One Mapping" category.

Glyph Presentation Setting

For the examined studies, we categorized how many glyphs were presented to a viewer on the screen: individual glyphs, multiple glyphs of fixed number, or multiple glyphs of varying numbers. In the category of multiple glyphs we further noted how the glyphs were arranged on the screen, as grids, scatterplots, node-link diagrams, on geographic maps, or other layouts.

Datasets

The glyphs used in the studies all encoded either multi-dimensional data of a general nature, or time-series data. Additionally, we noted how many dimensions a glyph encoded. The number of dimensions is related to the visual complexity of a glyph. Independent of data type and density we further recorded whether the data was synthetically created, or if real data was used in the study.

Tasks and Measures

Important for understanding any study results is the nature of the task participants had to perform. We group tasks in broad categories, differentiating between tasks involving the glyph as a whole (synoptic tasks [10]) and tasks where participants had to focus on single specific characteristics of a glyph (elementary tasks [10]). An elementary task is typically a lookup task during which participants focus on single dimensions of a glyph and read individual values.

We further subdivided synoptic tasks into three categories: 1) *visual search* where participants had to find a glyph differing from others, or tell whether a specific glyph is present or not; 2) *similarity search* where participants had to compare the overall structure of glyphs and group similar representations; and 3) *trend detection tasks* where participants had to keep track of the development of data values across dimensions.

Study Goals

We found three different general study goals: 1) a comparison of *various glyph designs* according to their performance and a ranking of designs based on it; 2) a comparison of different *variations* of a single glyph, to detect visual features improving a specific glyph design; and 3) a comparison of *single glyphs vs. data tables*, to motivate the use of these visual objects over textual representations.

Study Results

We summarized study outcomes on a high level, reporting findings on the impact of presentation settings, number of data points and dimensions on the tested glyphs. We further report overall ranking of different glyph types, offering explanations to seemingly contradictory results across studies. We do not enter into detail on findings regarding variations of a single glyph type. Our goal is to provide researchers and practitioners with a better grasp of the overall picture of the performance of different glyphs, and to point to individual papers for detailed study results.

2.3 Results: State-of-the-Art in Glyph Evaluation

In this section we discuss the findings from our systematic review based on the characteristics discussed in the previous section. A summary of the results is presented in individual tables and in highlighted paragraphs throughout the section. Many study descriptions did not include all information needed for our characterization and subsequently our counts do not always add up to 64–the total number of papers examined.

2.3.1 Study Goals

We found three higher-level study goals—all related to different types of comparisons: a) comparison of glyph designs, b) comparison of glyph variations, and c) comparison of glyphs with data tables or text. As can be seen in the diagonal of Figure 2.3[3], the majority of studies (39/64, 60.94%) tested case b) or c). Design variations within a glyph category were more frequently

[3]The flower [33] and theme [31] glyphs also fit into this category, however, due to their visual encoding they are not represented on the diagonal

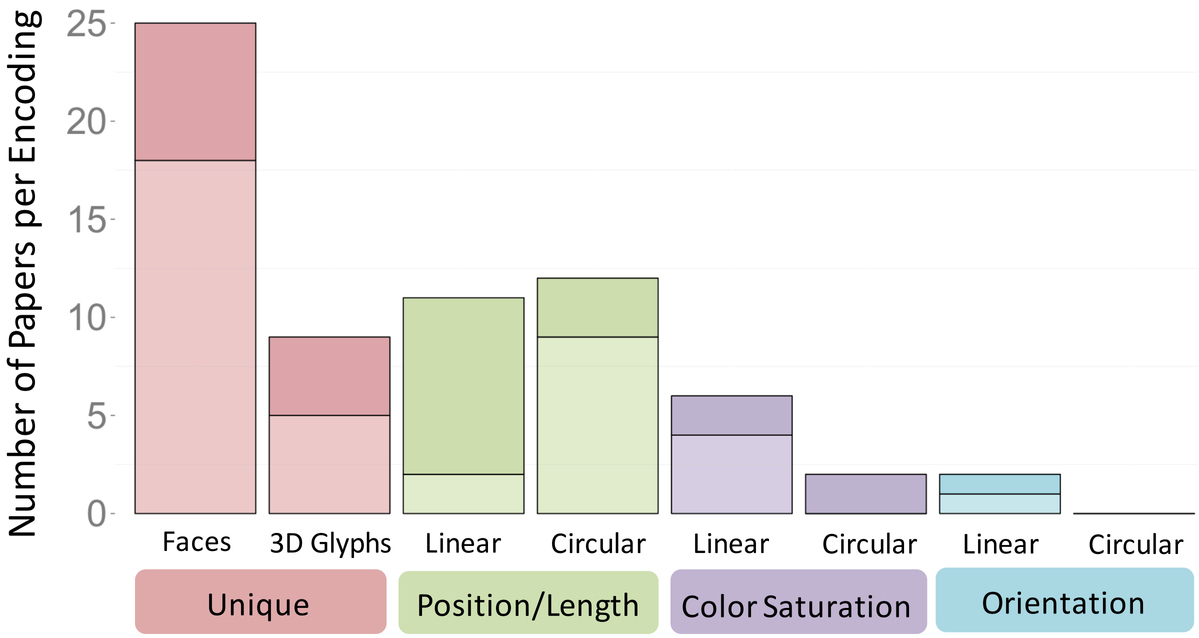


Figure 2.2: **Statistics:** Ratio of papers evaluating different visual encodings (distinguished by color). Low saturation indicates experiments evaluating design variations of this encoding, and high saturation other experiments (e.g. comparisons to other encodings).

tested against each other (32/39, 82.05%) than glyphs vs. a common data table or text description (7/39, 17.95%). The latter group was most often used to motivate the use of visuals over text descriptions or data tables [20, 91, 131, 163, 178].

To measure participant performance all studies but one [30] recorded accuracy scores, additionally 65.63% measured completion time (42/64), and 29.69% collected qualitative feedback (19/64) as well. It is interesting to note that participants' preferences did not always match with their performance [26, 61, 74, 197]. Therefore, a preferred design was not always a guarantee for a good user performance.

Summary: We found similar study goals across many experiments, yet varied were factors like number of data points and dimensions, task, or glyph design. These variations make individual study outcomes hard to compare. Thus, we will discuss the individual factors in the following sections before discussing the study outcomes in subsection 2.3.6.

2.3.2 Glyph Types and Data Encoding

Figure 2.3 summarizes evaluated glyph types and their encodings based on Ward's data mapping taxonomy outlined in section 2.2. The table is meant to be read like a matrix. The intersections of rows and columns show which glyph types and encodings a particular study compared against each other. The diagonal (top left to bottom right) of the table contains references to studies

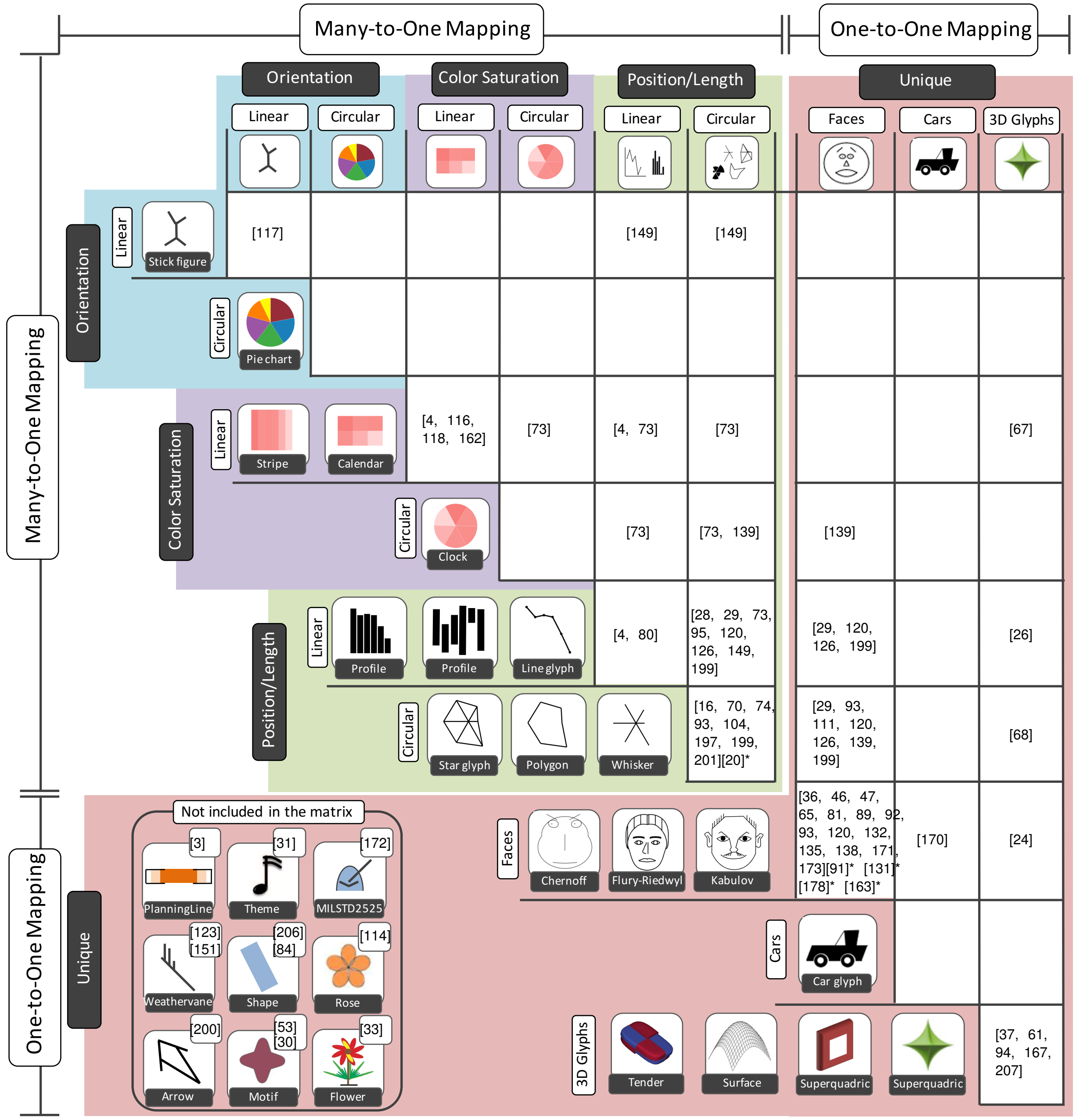


Figure 2.3: **Glyph Design Table:** Columns represent the different categories of glyph encodings, replicated in rows with glyph examples for each category. Additionally, color is used to visually separate the different categories. References refer to articles in our study bibliography that compare glyph variations from the respective row and column. Studies placed in the diagonal evaluate either variations of the same glyph type, or comparisons of the glyph with data tables (starred *). Note that papers can fall in multiple cells. Since PlanningLines [3], weather vanes [123, 151], shapes [84, 206], roses [114], themes [31], arrows [200], Motifs [30, 53], flowers [33], and MILSTD2525 glyphs [172] use a unique encoding and are not compared to other glyphs, we positioned them slightly apart in the "One-to-One Mapping" category.

that tested design variations of the same glyph category, or an evaluation of one specific design against plain text or data tables (marked with a * in the table). Empty cells indicate new research possibilities.

Figure 2.2 shows that face glyphs were evaluated most frequently (39.06%), followed by glyphs with position/length encodings (linear: 17.19%, circular: 18.75%), and 3D glyph designs (14.06%). We note that from the studies involving position/length encodings or 3D glyphs (27 in total), 8 were in fact compared to faces (Figure 2.3). Color (linear: 9.38%, circular: 3.13%) and orientation encodings (linear: 3.13%, circular: 0%) have received little research attention.

The high number–28.13%–of user studies on face variations , stands out compared to studies that only focus on other variations, e.g., circular position/length encodings (14.06%), 3D glyphs (7.81%), linear color (6.25%), or linear orientation encoding (1.56%). A possible reason for this imbalance are the many ways one can design faces and their data mappings (e.g., Chernoff faces [35], Rydwiel-Flury faces [65], Kabulov faces [96]).

We found only two studies [4, 80] that compared different linear position/length design variations. This is an interesting research gap given that profile glyphs that use this encoding are well established in practice (i. e., sparklines [185], profiles [52]).

In general, we only found three main categories of visual variables used to encode data in glyphs with many-to-one mappings (Position/Length, Color, Orientation). Almost all glyph designs in these studies mapped quantitative information to visual variables. The only exception was Lee et al.'s work [111] which compares star glyphs, faces, and 2D projected data points using bivariate data. Here bivariate information, however, was still mapped either to the length of the whiskers (star glyph) or to different face characteristics.

Summary: Faces and circular profiles have been investigated in detail, in contrast to color value and orientation encodings on glyphs that only few studies investigated. Surprisingly, we found only two studies comparing different variations of linear profiles.

2.3.3 Glyph Presentation Settings

Presentation settings can be characterized by the number of glyphs presented to viewers, as well as by how the glyphs are layed out in space. We identified three types of studies when considering the number of glyphs presented (Table 2.2): those that presented only individual glyphs to the viewers (7/64, 10.94%), those that presented a fixed number of more than one glyph at a time (46/64, 71.88%), and those in which the number of presented glyphs varied but was always higher than one (11/64, 17.19%). Seven papers did not report the exact number of glyphs represented on the screen: [3, 26, 30, 31, 33, 53, 151].

For the 46 studies that tested a fixed number of multiple glyphs at a time, we found five types of layouts. The most frequent was a common small-multiples grid (65.22%), followed by geographic maps (17.39%), scatterplots (6.52%), node-link diagrams (4.35%), and other layouts (6.52%) like different 3D environments (see Figure 2.4).

The goal of most of the studies with varying number of glyphs was to investigate changes in performance when increasing the number of visible data points in grid layouts [132, 171, 173], geographic maps [138, 200], and node-link diagrams [201]. The amount of glyphs visible to participants changed from 5–50 [132]; 5–15 [171, 173]; 6–18 [172]; 9–23 [138]; 4–300 [200];

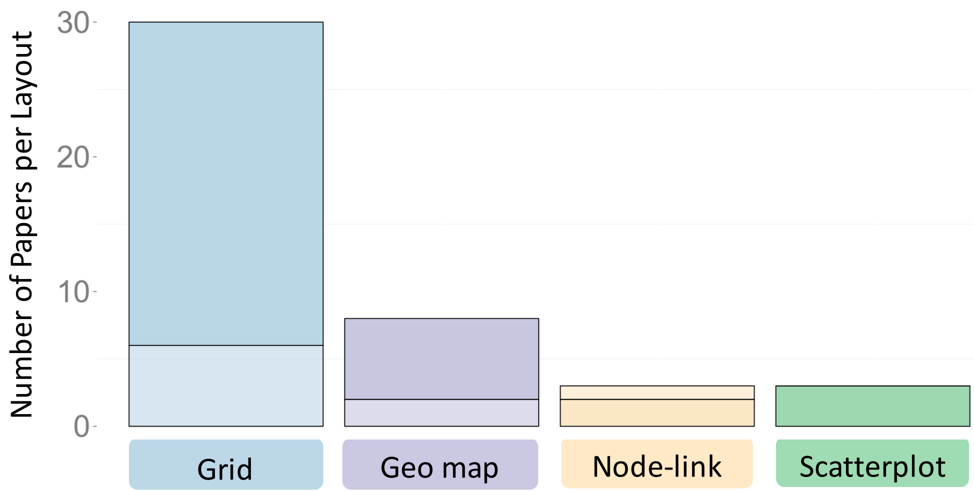


Figure 2.4: **Statistics:** Ratio of papers evaluating different glyph layouts (distinguished by color). Low saturation indicates experiments evaluating a varying number of data glyphs, and high saturation a fixed number of data glyphs.

and 30–48 [201]. In all seven studies participants were affected negatively by an increasing number of data points, as we discuss in subsection 2.3.6. In the studies conducted by Aigner et al. [3], Dunne et al. [53], Cayli et al. [30], and Zhang et al. [207] the varying number of data points was not treated as a factor in the analysis.

In seven studies (10.94%) a single data point was shown to participants at a time [4, 20, 29, 80, 94, 95, 120]. These studies tried to control all parameters and avoid possible confounding factors, so as to better reason about changes in performance when modifying specific aspects of the same design [94], or when comparing it with other representations [4, 20, 29, 95, 120], or layouts [80].

Regardless of whether a fixed or changing number of glyphs was tested, the vast majority of studies (56.25%) arranged glyphs in a grid layout, followed by geographic arrangements (15.63%), node-link diagrams (7.81%), and scatterplots (4.69%) (see Figure 2.4). The choice of a grid layout for quantitative studies is understandable. Grids can help to avoid confounding factors in visual search, comparison, or classification tasks. For example, the information provided by a background, e.g., an underlying geographic map, may influence the perception of glyphs. The background color, for example may influence the perception color hues [180], while topology (e.g., rivers, mountains, land borders) may act as grouping enclosures or as reference structures for reading data values of glyphs. We only found a single study [123] that examined the influence of reading data glyphs with different geographic backgrounds; and one [84] that studied how the reading of a glyph is affected by the presence of other glyphs around it. We discuss their results in subsection 2.3.6.

Summary: Only a small number of user studies varied the amount of data glyphs as a study factor. Most studies were conducted with a fixed number of glyphs arranged in a grid layout.

Surprisingly, only four papers investigated the influence of different background information and layout on reading data glyphs [70, 80, 84, 123].

2.3.4 Datasets & Number of Dimensions

The number of data dimensions tested can help us compare results across studies, and inform us of the imagined use-case setting for data glyphs. Only four studies (6.25%) used the number of dimensions itself as a study factor and thus varied between glyphs with different dimension counts [73, 74, 199, 201]. The remaining 60 studies tested glyphs with various fixed numbers of dimensions. Of these, 44 tested less than 10 dimensions. An overview of different dimensionality settings is provided in Table 2.3. Three papers did not report about the number of dimensions encoded by the glyph designs [53, 70, 80].

In the vast majority of studies (54/64, 84.38%) glyphs encoded general multi-dimensional data, both real and synthetic. Eight studies tested glyphs encoding time-series data (8/64, 12.5%) and in two experiments [30, 53] glyphs were used to represent network topologies.

Only a small number of studies (24/64, 37.5%) used real data to investigate the performance of different glyph designs. The respective papers and real datasets can be found in Table 2.1. For the other experiments (41/64, 64.06%) the data was created synthetically.

Summary: Overall, most studies used synthetically created multi-dimensional data (41/64, 64.06%). The majority (44/64, 68.75%) of studies used glyphs encoding less than 10 dimensions.

2.3.5 Task Space

We used the Andrienko & Andrienko task taxonomy [10] to distinguish between two higher-level tasks as discussed in section 2.2. Synoptic tasks (i.e., similarity search, visual search, trend detection) were the most common type of task used in the studies (44/64, 68.75%). This is perhaps not surprising as glyphs are often meant to provide quick overviews over a large number of varying multi-dimensional data points—and the use of synoptic tasks may reflect the authors' desire to test glyphs in a realistic use context.

As shown in Table 2.4, we found the following classes of synoptic tasks: similarity search (23/44, 52.27%), followed by visual search (14/44, 31.82%), and trend detection tasks (7/44, 15.91%). An example of a similarity search task can be found in two studies by Klippel et al.: Using a visualization tool showing 81 glyphs each representing one car, participants had to group these glyphs into different categories based on their attributes [103, 104].

In contrast to these synoptic tasks, 26/64 studies (40.63%) used elementary tasks, i.e., lookup (25/26, 96.15%) and 3D distance calculation (1/26, 3.85%). These studies focused on more perception-related questions such as the reading accuracy for visual variables used to encode a data value. In these studies, participants did not focus on reading the entire shape of the glyph, but on single glyph characteristics. For example in the user study conducted by MacGregor and Slovic [120] participants had to read the completion time of 48 marathon runners from bar chart glyphs, faces, and star glyphs. Faces performed best, followed by bar chart glyphs and star glyphs.

Dataset	Availability
Anthropometrical data about twins [65, 81]	[188]
Patients rated by psychiatrists [91]	Minnesota Multiphasic Personality Inventory
Medical data [49]	unknown
Medical images [26]	unknown
Cars dataset [170]	http://davis.wpi.edu/xmdv/datasets/cars.html
Project plans [3]	unknown
MM5 weather information [123]	unknown
Weather information [151]	NCEP forecast model
Weather information [200]	Operational Regional Atmospheric Prediction System
Modified U.S. census data [206]	unknown
Financial data [114]	Investment in education USA (2008)
Financial data [131]	Wall Street Journal Index (1974 and 1975)
Financial data [178]	Standard and Poor's firm list (1974 and 1975)
Classical music data [31]	unknown
Network data [53]	Lostpedia wiki edits
Network data [30]	10 best ranked movies (IMDb)
Google search results [33]	http://www.google.de
Marathon runners [120]	unknown
Power plant statistics [149]	unknown
Audio information [70]	One laptop per child sound library
Biological data [162]	unknown
Economic variables [89]	U.S. Department of commerce & labor
Tensor data [207]	DTI dataset

Table 2.1: **Datasets**: This table illustrates detailed information about the real datasets used in the experiments.

Summary: Most studies used a similarity search or a direct lookup task to measure the performance of glyph designs.

2.3.6 Study Outcomes

While we cannot discuss the study results individually for all 64 papers, we collected higher-level observations on study outcomes. Results on the study of factors such as number of dimensions and datapoints tested, is consistent across experiments. Nevertheless, when it comes to a general ranking, experimental results apply to a study's specific setting and should be generalized with caution. We discuss these results next.

	Layout	References
Single		[29][95][120][4][20][94]
	Text	[80]
Multiple Glyphs	Grid	[65][91][73][103][84][33][199][28][126][149] [16][74][104][93][111][68][67][117][36][46] [47][81][89][92][135][131][178][163][24][167]
	Geo map	[123][151][206][114][197][139][116][118]
	Scatterplot	[170][37][61]
	Other	[49][3][26]
	Node-link	[31][162]
Varying	Grid	[172][70][132][171][173][207]
	Node-link	[53][30][201]
	Geo map	[200][138]

Table 2.2: **Presentation Setting**: This table distinguishes between the number of data points shown to the participants during the studies and the used layout. Color is used to better distinguish between the different categories.

Influence of Background Information and Layout

Understanding the influence of layout strategies or additional context information is crucial since data glyphs can be arranged in various different ways and settings. Four studies investigated the influence of positioning or background information on the performance of data glyphs [70, 80, 84, 123].

A common setting for data glyphs was the positioning in scatterplots, or projections from a high dimensional dataset to a two dimensional space. Frisson et al. used a visual search task to examine the benefits of a two dimensional projection compared to a grid layout used in small multiple settings [70]. Performance was lower for the two dimensional projection, since after projection, some data glyphs ended up overlapping each other, which caused a loss of information making it difficult to detect the stimulus. In a follow-up study, the authors added a proximity grid [155] as an additional layout to the study setting. Results indicated that participants performed best in a visual search task when using the proximity grid.

Glyphs were also used in textual documents to communicate statistical data not only with words but visually. Sparklines are a famous example of such small visual representations [185], which are usually positioned in the reading direction next to the statistics (e. g., on the right hand side). To backup this design decision, Goffin et al. conducted a user study to compare different layout possibilities of glyphs within sentences [80]. Surprisingly, there was no significant effect on accuracy or reading performance for the different layouts. However, participants preferred the glyph being positioned above the words.

The influence of reading data glyphs with different geographic backgrounds was investigated in only one study conducted by Martin [123]. He measured the performance of participants

Number of Dimensions	References
2 & 3 Dimensions	[49][3][123][151][206][84][114][30][29] [28][197][68][67][167][94][207][54]
4 & 5 Dimensions	[91][31][200][33][172][95][120][20][139] [116][118][117][135][138][171][173][37][61]
6 & 7 Dimensions	[89][178]
8 & 9 Dimensions	[103][170][149][104][93][36][132]
10 - 15 Dimensions	[4][111][162][47][131][24]
17 - 20 Dimensions	[65][126][16][46][81][92][163]
Varying	[73][199][74][201]

Table 2.3: **Number of Dimensions**: This table illustrates the different data dimension densities used in the studies. Color is used to better distinguish between the different categories.

working with weather vane glyphs while varying the underlying geographic map. Surprisingly, his results indicated the background had no influence on the performance of reading data-glyphs. However, the glyphs in his study were arranged in a grid on top of a map, and not according to their geographic position. Using different glyph designs or an irregular layout may, nevertheless, influence their performance.

Healey and Enns conducted an experiment to compare the interaction of different visual features in the surroundings of the glyph stimulus for a visual search task [84]. Results indicated that color variations due to the presence of other glyphs in the neighborhood of the stimulus glyph, caused a significant interference effect when participants had to judge heights of glyphs or density patterns. However, different densities in the surroundings of the stimulus or heights of neighboring glyphs had no effect on the detection of colored glyphs.

Summary: The influence of background and layout on reading data glyphs has so far received little research attention. The limited evidence from this work suggests that the background and neighborhood of a glyph did not affect glyph readability. Nevertheless more work is needed to determine the perceptual difficulties of reading glyphs depending on their background and layout.

Influence of Number of Data Points

Seven studies varied the number of visible data points as a factor. The glyphs used in these experiments were either faces [132, 138, 171, 173], unique glyph designs (i.e., MILSTD2525 [172], arrow glyphs [200]), or star glyphs [201].

For the studies involving face glyphs, participants had to perform visual search tasks and find a certain stimulus in a growing set of data points. The researchers tested whether pre-attentive identification was possible, in which case search time would not have been seriously impacted by increasing the number of glyphs. Yet, in all studies the performance dropped with an increasing number of data points independent from the mapping of data to face characteristics. Based on this

Data Type	Task Description				
	Elementary Task		Synoptic Task		
	Lookup	3D Navigation (distance calculation)	Trend Detection	Similarity Search	Visual Search
Multi-dimensional	[28, 29, 33, 37, 61, 91, 94, 95, 114, 117, 120, 123, 132, 139, 149, 151, 170, 172, 197, 200, 206]	[49]	[131, 178, 206]	[16, 20, 24, 33, 36, 46, 47, 65, 74, 81, 89, 92, 93, 103, 104, 111, 126, 135, 163, 167, 199, 201, 207]	[26, 67, 68, 84, 114, 138, 139, 151, 171, 173]
Time-series data	[3, 4, 73, 80]		[73, 80, 116, 118]		[4, 70, 162]

Table 2.4: **Data and Tasks**: Most studies were conducted using a lookup or similarity search task with multi-dimensional data.

outcome Siva and co-authors concluded that participants performed a serial search and were not able to pre-attentively identify the stimulus [171, 173]. Therefore, the perception of abstract data glyph faces compared to human faces was shown to be different. This is an interesting finding, which lessens the basic motivation for using abstract faces. However, researchers could also show that a redundant visual mapping of data to face characteristics improved the performance [132].

Summary: Increasing the number of data points negatively affects search within a set of data glyphs, indicating that they— even face glyphs—cannot be read pre-attentively.

Influence of Number of Dimensions

The results of studies varying the number of dimensions as a factor showed that different designs were impacted to different extents. In a study by Fuchs et al., for example, the performance of star glyphs dropped significantly in a lookup task when increasing the number of dimensions from 24 to 96, whereas the performance of line glyphs stayed stable [73].

Wilkinson also varied the number of dimensions to investigate changes in performance for different glyph representations. His results indicated that increasing the number of dimensions had no significant effect on the *ranking* of tested glyph designs [199], although there was a drop in performance overall.

However, it is interesting to note that even slight variations of a glyph design can be affected differently by the number of dimensions. Fuchs et al. tested the effect of increasing the number of dimensions on whisker glyphs (star glyphs without a contour line), traditional star glyphs and polygon variations. Although the performance dropped for all variations, whisker glyphs were affected the least [74].

Summary: Increasing the number of dimensions negatively affects the performance of data glyphs [73, 74, 199, 201].

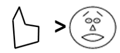
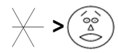
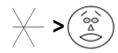
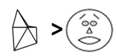
Elementary Task	Synoptic Task
 [29]	 [93, 126, 199]
 [139]	 [139]
 [120]	 [111]

Table 2.5: Studies and their result rankings: faces vs. circular profiles.

Influence of Tasks and Visual Encoding

The outcome of individual user studies often involved a ranking of data glyphs based on their performance in the study. These rankings were not always consistent for the same designs tested, and they changed, for example, based on tasks and details of the visual encoding. Table 2.5–Table 2.8 summarize the outcomes of the different experiments. The "¿" symbol indicates that the glyph on the left outperforms the design on the right (either in terms of completion time or accuracy).

Seven studies compared faces against circular position/length encodings [29, 93, 111, 120, 126, 139, 199] (Table 2.5). In four, faces performed best [93, 120, 126, 199], while circular position/length encodings performed best in the remaining three [29, 111, 139]. These seemingly contradictory results are reconciled when we consider the tasks participants had to perform and how the glyphs were designed. In five of these studies the participants performed a synoptic task [93, 111, 126, 139, 199], in the other three a lookup task [29, 120, 139]. From the five synoptic task studies, in the three where faces performed best, the circular position/length encoding was a polygon (i.e., star glyph without whiskers, but only a contour) [93, 126, 199], while in the remaining two where faces performed worst the circular encoding was a star glyph with [111] and without contour line (i.e., whisker glyph) [139]. The remaining three studies with lookup tasks also compared faces against polygons (with polygons performing best [29]), faces against star glyphs (with faces performing best [120]), and faces against whisker glyphs (with whisker glyphs performing best [139]). It seems that star glyphs compared to faces are more suitable for synoptic tasks. However, the whiskers glyph had the best performance independent from the underlying task. This finding has partially been confirmed for a similarity search [74] but not for lookup tasks.

Another example where glyph rankings change based on study characteristics can be found when comparing faces against linear profiles (Table 2.6). In three studies faces performed best [120, 126, 199], in the fourth study, profiles [29]. Again, the four studies used different tasks: lookup tasks [29, 120] and a similarity search task [126, 199]. When comparing the two lookup tasks the ranking of the two glyph designs is still different although they use a similar number of dimensions (4 [120] and 5 [29] dimensions), and just show one data point at a time. Yet, a major difference can be found when reading the task description more carefully. Although both tasks are a lookup task, participants had to either read a one-dimensional value [120] or detect when one dimension changes significantly compared to the other dimensions for a single data point [29].

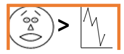
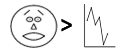
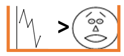
Elementary Task		Synoptic Task	
	[120]		[126, 199]
	[29]		

Table 2.6: Studies and their result rankings: faces vs. linear profiles. Conflicting results are marked with orange color.

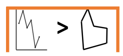
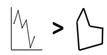
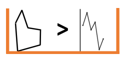
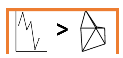
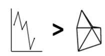
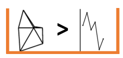
Elementary Task		Synoptic Task	
	[29]		[126]
	[28, 95]		
	[73, 120, 149]		[73]
	[73]		

Table 2.7: Studies and their result rankings: linear vs. circular profiles. Conflicting results are marked with orange color.

When comparing linear and circular position/length encodings, we found glyph ranking differences in 8 studies [28, 29, 73, 95, 120, 126, 149, 199] (Table 2.7). In four, the linear design outperformed the radial [29, 120, 126, 149], while in two, circular designs were better [28, 95], and in the last one performance varied according to the underlying task [73]. However, only 3 out of these 7 had a similar experimental setting with respect to design variations, presentation setting, number of dimensions and task [28, 29, 95]. These three all compare bar charts with polygons in a lookup task, using low dimensional data and presenting only one data point at a time. Surprisingly, the performance was still different : polygons ranked best in two of them [28, 95] and bar charts performed best in the third [29]. Again, we have to look at the studies more carefully to come to a conclusion. In the two studies where polygons performed best, the bars in the bar charts were shown without a common baseline. This was not true for the third study were bar charts outperformed the polygons. We assume that a common baseline increases the performance of the linear profiles, a finding which is proposed as a design guideline from a study by Fuchs et al. [73]. However, a user study comparing linear profiles with and without a common baseline has, to the best of our knowledge, not yet been conducted.

Additionally, it is interesting to note that there were changes in performance depending on the kind of elementary task. For reading exact data values linear profiles outperformed star glyphs, however, when reading the position of an attribute dimension (e.g., a certain point in time for time-series data) star glyphs ranked first. [73]

Data glyph designs using color saturation to encode data values have not received much attention. We only found two papers, which report on results from quantitative experiments comparing these glyphs against alternative representations [73, 139] (Table 2.8).

For overview visualizations focusing on the overall appearance of a glyph, color value en-

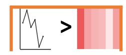
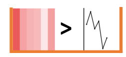
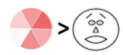
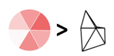
Elementary Task	Synoptic Task
 [73]	 [73]
 [4, 73]	 [4, 73]
	 [4]
 [139]	 [139]
 [139]	 [139]
 [73]	

Table 2.8: Studies and their result rankings: color saturation vs. profiles. Conflicting results are marked with orange color.

codings were not found to be effective. In three out of four user studies participants performed better using a position encoding (i. e., linear and circular profiles) in synoptic tasks. In the fourth experiment the color encodings were visually enhanced to help participants solve certain tasks and outperformed linear profiles. Only faces performed worse. However, it is more difficult to draw conclusions for elementary tasks. When pursuing a direct lookup task (e.g., reading data values) radial color value encodings have outperformed star glyphs and faces [73, 139]. Whisker glyphs on the other hand have been shown to be as accurate as color value encodings but more efficient [139]. However, linear profiles were most accurate and, therefore, the best choice for direct lookup tasks [73].

Summary: Study results differed based on individual factors like number of dimensions, task, number of data points, or slight variations to the designs. Our summary tables can be considered as a performance overview pinpointing to relevant literature.

Influence of Metaphoric Glyph Design

One goal of information visualization is to present the underlying data in a way that can be easily understood by users. Thus, researchers have tried to improve intuitive understandability of visualizations, by using metaphors when mapping data to visual representations. One such example can be found for weather forecasts. In such a scenario weather status is communicated with small icons on top of a geographic map. These icons are metaphoric representations of the real environment to facilitate their understanding. Small cloud icons represent cloudy areas, rain drops encode rainy areas, and little suns illustrate sunshine in specific regions.

While data glyphs are different from icons [17], the general concept of representing the underlying data using metaphors can also be applied here. Since the visual representation of a glyph is data driven the idea is not to use a different glyph design for each individual data point (like in the weather forecast example), but to use certain glyph characteristics to display the data while being consistent with the metaphor.

However, it is not clear whether such metaphor-based representations are better than more abstract ones. Siirtola has attempted to provide an answer to this problem by introducing metaphoric glyph designs and comparing them with more abstract ones [170]. In his experiment he visualized car related data with abstract face representations, and with metaphoric car glyphs [170]. Car glyphs were created by mapping data to parts of the glyph with related meaning. For example the attribute horsepower was mapped to the size of the engine of the car, which is metaphorically reflected in a bigger hood. In his user study participants had to answer car related questions when working with either faces or car glyphs. The metaphor helped the participants in understanding the data. As a result, they performed better when working with car glyphs compared to faces.

Li et al. [114] provided another example where metaphors were used. In their quantitative experiment they compared RoseShape glyphs against abstract polygons to visualize multi-dimensional data about the education level in the US. The glyphs were positioned on top of a geographic map and participants had to either read data values or search for certain characteristics. Results suggest that participants were more accurate and more confident of their answers when working with the metaphoric designs.

In a study conducted by Flury and Riedwyl, data collected about monozygotic and dizygotic twins, such as their height or weight was mapped to two types of face glyphs [65]. Using abstract face representations (i. e., Chernoff faces) or more realistic faces (i. e., Flury Riedwyl faces) participants had to look at a glyph for each twin and rate whether or not the two glyphs showed data about monozygotic twins. The results indicated that participants were more accurate when working with the more realistic faces.

Jacob [91] gave another example where he tested the performance of a single metaphoric glyph design. He displayed data from patients having a certain psychological condition (e.g., depression, paranoia etc.) using faces. The abstract faces were created to show facial expressions resembling those of the human faces of the patients. Participants in his study had to judge which face corresponds to which behavior without being trained or knowing the patients. The results indicated that people were able to name the correct psychological illness without knowing the mapping criteria of data to face representations.

Metaphors may help to explain the results obtained in a study conducted by Fuchs et al. [73]. The researchers ran a quantitative study using time-series data. Participants had to locate specific points in time using glyphs with either a linear dimension layout (e.g., sparklines) or a radial arrangement (e.g., star glyphs). Surprisingly, participants were more accurate when working with circular glyphs. This is interesting since the visual variable position (used in linear layouts) is considered more accurate compared to orientation (used in circular glyphs) [43]. However, participants argued that they were reminded of a clock when working with radial glyph designs, which facilitated locating certain points in time.

Summary: A small number of previous studies suggest that metaphors may help to better understand the underlying data.

Summary

While we found and reported on 64 papers, the vast design space of data glyphs and the possibility to test only a limited set of factors in a controlled user study makes it difficult to recommend a single best-of glyph design. Glyph performance depends on many different factors, such as the task used, the number of data points, or slight variations to the designs used across studies. Our analysis in subsection 2.3.6 presents a summary of rankings from the articles we analyzed, and discusses how these factors can explain seemingly contradictory results.

We were able to draw general conclusions when it comes to number of dimensions and glyphs. Some study results indicate that increasing the number of data dimensions affects the performance of glyph designs negatively [73, 74, 199, 201] with position encodings (linear and circular profiles) being more robust compared to color encodings in high-density situations [73]. As with the number of dimensions, there is evidence that performance drops with increasing the amount of visible glyphs on the screen [132, 138, 171–173, 200, 201]. This seems like a logical conclusion due to the required additional effort in visual search involving a higher number of entities. In addition, a small number of past studies indicate that metaphoric glyph designs increase performance.

Finally, it has to be noted that our analysis was made difficult by a lack of standard for reporting study details on glyphs. For example specific information (e.g., stimuli size, viewing distance, number of visible data points, etc.), that could shed light on differences across experiments, were often missing.

2.4 Discussion and Open Research Areas

In this section we identify and discuss directions for future research based on our analysis. The proposed research directions are ordered roughly according to their scope.

Types of User Studies:

Even though we focused on user studies with quantitative components for this survey, we found only a few qualitative studies that considered how glyphs are used in practice within real applications. One such exception is the experiment conducted by Sreng et al. [174] where participants used a 3D automotive assembly tool and answered questions about the perceived usefulness of the embedded glyphs. Although this study provided qualitative observations in the form of questionnaires, we can envision more elaborate field experiments and observational studies on real use of glyphs. Observers could thus gather information on how people use glyph-based visualizations in real contexts, for which tasks, and with what kind of results. Such studies could inform our understanding of how glyph-based applications are adopted and used in practice and could, thus, provide new insights on which to base design choices.

Summary: Adding qualitative evaluations observing analysts working with different glyph designs, datasets, and tasks, would help to better understand the glyphs design space. In particular, information about subjective preferences and the applicability of specific glyph designs in practice would be useful. It would be interesting to capture which design analysts choose to solve which analysis task.

Data to be Tested:

There are several pros and cons for choosing real vs. synthetic datasets for a study. On the one hand, real data has the advantage that it can demonstrate which visual representation performs best in realistic situations, providing valuable results for analysts of this data. However, real data often contains unique characteristics (e.g., size, structure, number of dimensions), that make the results noisy and hard to generalize.

On the other hand, one may argue that synthetic data does not always represent a real world scenario or problem well (ecological validity), making results again hard to generalize. However, artificial data can be easily controlled and focused on answering specific questions. Additionally, possible confounding factors due to the underlying data are excluded (e.g., visual search time according to the number of data points).

Given the above pros and cons, it seems an interesting open research question to see how glyphs behave when they undergo study using both synthetic and real data, similar to the approach taken by Caban [26].

Summary: Running quantitative experiments, using both datasets from synthetic to real world and vice versa will enhance our knowledge on the behavior of data glyphs in different situations.

Study Tasks and Measures:

In the majority of studies participants had to perform synoptic tasks (Table 2.4). This is not surprising given that glyphs are often used to provide quick overviews over a large number of multi-dimensional data points. Nevertheless, there are glyph designs (e.g., some 3D glyphs) that have not or rarely been looked at for synoptic tasks, an interesting topic for further study.

Although results from specific tasks, such as these synoptic ones, are valuable, a common visualization task is free exploration, insight generation and hypothesis forming. Inspired by recent work on insight based evaluation [161], it would be worthwhile to investigate the performance of different glyph designs in such contexts.

Summary: Adding exploration tasks or extracting insights from an unknown dataset are realistic real-world analysis tasks. They should, therefore, be added to the repertoire of user study tasks in glyph evaluation to further reason about the practical applicability of data glyphs.

Glyph Presentation Setting:

A large number of studies presented glyphs as small multiples using a grid layout. There were no studies on glyphs nested inside treemaps, or other types of representations apart from maps, scatterplots, node-link diagrams, and two 3D representations in the medical domain. This is interesting, as it is not clear that grid layouts present the most commonly assumed usage context for glyphs. For example, in the area of scientific visualization, glyphs are often used on 3D volumetric surfaces or to represent 2D flow fields in order to indicate data at specific sampling points. These glyphs are approximately uniformly spaced apart, but this relative spacing changes depending on the view's magnification factor, making them appear more or less densely packed together. There is very little to no guidance from controlled user studies on how this apparent density affects their performance.

Moreover, we know little about the influence of the background information on the performance of glyphs. Only one study investigated performance changes for glyph designs when placed on top of different geographic maps [123], and one other their performance close to neighboring glyphs [84]. Many questions remain unanswered, for example, we do not know if glyphs are perceived differently when arranged in uniform grids compared to other arrangements, such as treemaps, that vary their relative distance.

It is also unclear what effect the glyphs have on the understanding of the underlying visualization itself: for example, it would be interesting to investigate if rectangular treemaps are more effective compared to circular treemaps when adding glyph designs; or if people are distracted by the additional context information in the form of glyphs in 3D environments. There is certainly much space for further research.

Summary: Since data glyphs cannot only be positioned in small multiple grids, evaluating different arrangements of more complex layouts (i. e., treemaps, etc.), would help to better understand the influence of specific data glyph designs on the context and vice versa.

Glyph Types and Data Encodings:
Understand redundant encodings: Using Ward's glyph design categorization [192], we found only two studies that used glyphs with a one-to-many mapping (i.e., a redundant encoding). Ware [195], however, discusses interesting perceptual study approaches to learn how dimension encodings can be separable or integral. A better understanding of how redundant encodings work together, and could enforce data reading, would prove beneficial to glyph design.
Study missing mappings: In Figure 2.3 we refer to 50 of the 64 studies examined, having left out the two one-to-many mappings [49, 103] and the twelve that were not compared to other designs [3, 30, 31, 33, 53, 84, 114, 123, 151, 172, 200, 206]. Looking at the table there is still clearly an imbalance in what kind of data encodings have been comparatively tested. Many cells remain empty, and there are several sparsely populated ones. One of the first things to notice is that there is no single study on circular orientation encodings, although they are used in visualization applications: representatives of this category are the compound glyph used in network graphs [147], pie chart glyphs for analyzing multi-dimensional data (e.g., global material composition [6], or biological binding properties [145]), or as provided in visualization toolkits (e.g., JIT[4]). Perhaps this type of encoding is a-priori deemed inferior based on Cleveland and McGill's [43] work that ranks orientation low for quantitative data representation. Given past use of these encodings however, it is certainly worthwhile to confirm that Cleveland and McGill's ranking does hold for circular-orientation encodings in glyphs, in particular in the context of real multi-dimensional data.

Similarly, several other cells of Figure 2.3 are empty or populated by studies from a single paper. As discussed in subsection 2.3.6, the ranking of glyph designs or their variations often depends on tasks and encodings, and as such more studies are needed to be able to provide reliable guidance for general glyph use and design. Especially glyph designs, which have not received much research attention but are used in practice (i.e., pie chart glyphs, or variations of linear profiles) should be prioritized in future studies.
Replicate studies on face glyphs: Many studies have been conducted investigating the performance of faces. Most of these studies were conducted in the 70s, and 80s when faces were newly introduced. In recent years face glyphs have been considered inferior but there are no recent studies or replications of earlier studies to confirm this. Given that some past studies showed good performance, it may be worthwhile to try and reproduce some earlier studies to confirm that they are indeed not as good as their current reputation in the community suggests.
Test larger number of dimensions: In addition to the data encoding, the number of glyph dimen-

[4]JavaScript InfoVis Toolkit `http://philogb.github.io/jit/`

sions may highly influence performance. As we saw in Table 2.3 the vast majority of studies only examined glyphs under a fixed number of dimensions, often less than 10 data dimensions. Only four varied the number of dimensions systematically in their studies. To reliably understand how glyph performance scales, we need to further explore how glyph designs fare under different dimensions.

Summary: Quantitative user studies should be conducted to compare data glyph designs which have not yet received much research attention (i.e., pie chart glyphs). The number of dimensions should be varied during the experiment and considered as a factor for analysis, to better understand glyph scalability.

Summary:
This section motivated promising open research directions for future experiments on data glyphs. In this summary, we revisit the most important gaps we identified and most promising research directions. Firstly, we need to give priority to experiments investigating glyph designs, which have not received much research attention, yet. For example, there is only little knowledge about the performance of radial layouts, such as pie chart glyphs. Having more evaluations about data glyph designs will help to better generalize the outcomes and argue about the performance of visual variables.

Additionally, different presentation settings need to be tested in more detail, since a big advantage of data glyphs is their flexible arrangement on the screen. In most experiments the data glyphs were positioned in a regular grid layout, however, data glyphs can also be arranged in more complex layouts like treemaps. Currently, there is only little guidance whether the performance of data glyphs will change according to context information or layout.

A wider variety of experimental factors should be considered such as: multiple datasets (i. e., synthetic data and real world data), different analysis tasks (e. g., exploration or insight generation), and different study types (i. e., qualitative and quantitative) to get a deeper understanding of the utility and performance of data glyphs.

2.5 Summary

This systematic review of research papers was focusing on the evaluation of data glyphs in quantitative user studies. We organized this work using several criteria, such as glyph types, study presentation settings, datasets and tasks used. Our goal was to: first, help researchers and practitioners identify relevant previous studies that give insights into glyph design tradeoffs, and get inspired by previous study setups; second, provide a meta analysis of the study outcomes; and third, pinpoint open research directions for the study of data glyphs.

Faces and their variations were the most studied glyphs, followed by circular position encoding glyphs (e.g., star glyphs), which were often also compared to faces. Our analysis showed that at first glance performance rankings may differ across studies. Yet, we discussed how some of these seemingly contradictory results can be explained by differences in the study criteria, such as the tasks, density and variations of glyphs tested. Our categorization provides readers with references to studies with similar setups, and argues for caution when conducting a meta-analysis of past results.

Our work also aims to highlight gaps in the literature on data glyph evaluation. Few papers have evaluated variations of glyphs using color encoding, even though such glyphs are used in practice [203]. Moreover, only a few studies have compared design variations of glyphs using linear position or length encodings, that are well established in practice (i.e., sparklines [185] or profiles [52])—although some have at least been compared to faces. We were also unable to find any study on circular orientation/angle encodings as already used in applications (e.g., [147]). The visualization community cannot at this stage form general guidelines for glyphs, as existing studies do not cover the entire space—a fact that is compounded if we consider the many criteria used in our categorization, such as datasets and tasks, that can further influence the possible relationships. We see a large number of opportunities for design and evaluation and hope this work encourages researchers in pursuing them.[5]

The systematic review revealed some interesting research gaps, which are the motivation for the following chapter 3 and chapter 4. Since only a few experiments were conducted comparing linear and radial color saturation encodings I will close this gap by introducing a new glyph design called the *clock glyph* and evaluate its performance in a controlled user study described in section 3.3.

[5]Please, note that the evaluations [73, 74] introduced in chapter 3 and chapter 4 are already part of this survey.

Chapter 3

Data Glyph Designs for Time-Series Data

Parts of this chapter appear in the following publications:

- Christopher Kintzel, Johannes Fuchs, and Florian Mansmann. Monitoring Large IP Spaces with ClockView. In *Proc. of the 8th International Symposium on Visualization for Cyber Security*, VizSec '11, pages 2:1–2:10. ACM, 2011[1]
- Fabian Fischer, Johannes Fuchs, and Florian Mansmann. ClockMap: Enhancing Circular Treemaps with Temporal Glyphs for Time-Series Data. In *Proc. EuroVis Short Papers*, pages 97–101. Eurographics, 2012[2]
- Fabian Fischer, Johannes Fuchs, Pierre-Antoine Vervier, Florian Mansmann, and Olivier Thonnard. VisTracer: A Visual Analytics Tool to Investigate Routing Anomalies in Traceroutes. In *Proc. of the 9th International Symposium on Visualization for Cyber Security*, VizSec '12, pages 80–87. ACM, 2012[3]
- Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. Evaluation of Alternative Glyph Designs for Time Series Data in a Small Multiple Setting. In *Proc. CHI*, pages 3237–3246. ACM, 2013[4]

---

[1] The responsibilities for this joint publication were divided as follows: I did the writing and gave advice, Christopher Kintzel did the programming, and Florian Mansmann did some proofreading and supervised the work.

[2] The responsibilities for this joint publication were divided as follows: Fabian Fischer did the programming and the writing. Florian Mansmann and I did the proofreading and gave advice.

[3] The responsibilities for this joint publication were divided as follows: Fabian Fischer and I did the programming and the writing. Pierre-Antoine Vervier provided the data and was also involved in the writing. Florian Mansmann and Olivier Thonnard did the proofreading and gave advice.

[4] The responsibilities for this joint publication were divided as follows: Petra Isenberg and I designed the user study. Fabian Fischer and I conducted the experiment. I was also responsible for analyzing the results and writing the paper. Petra Isenberg, Florian Mansmann and Enrico Bertini gave advice and did the proofreading.

Time-series data is similar to multi-dimensional data, where each dimension corresponds to one point in time. The main difference between these two data types is the relationship between the dimensions and, therefore, the analysis task. The attributes in multi-dimensional data are most often independent from each other. Therefore, trend detection tasks across dimensions are not performed by analysts. In case of time-series data the interplay of different points in time are important and of high interest. This distinction between the two data types is mandatory since it influences the design of the visualization. An important aspect is the comparison of dimensions within one glyph design. Using a one-to-one mapping (e.g., Chernoff faces [35]) for time-series data is, therefore, not recommended because different kinds of visual variables have to be compared (e.g., angle of eyebrows, size of the nose, height of the ears, etc.).

In the following section 3.1, I will review the literature according to glyph designs for time-series data and motivate the necessity for introducing an additional glyph design namely the *clock glyph*. The development and the design choices made will be explained in section 3.2 together with use cases from the network security domain. A thorough quantitative evaluation in section 3.3 compares the *clock glyph* against well-known alternatives and proves the fact that this design is the best choice for specific analysis tasks.

3.1 Related Work

As can be seen in chapter 2 only 3 papers investigate the performance of data glyph designs for time-series data in a controlled experiment [73, 116, 118].[5] This is surprising since many different glyph designs for time-series data do exist. This related work section tries to cover application and design study papers making use of temporal glyphs. It is important to note that the focus is on data glyphs encoding temporal data with its design and not with the positioning/comparison of multiple glyphs etc. A more general time series review can be found in the survey contributed by Aigner et al. [2]. The review will be structured according to the different visualizations temporal data glyphs are combined with. It is important to note that most of the glyph designs are flexible in the way they can be arranged on the screen. Therefore, multiple layout options for data glyphs are certainly possible. The categorization is solely based on the arrangement intended by the authors of the respective research paper.

3.1.1 Geographic Maps

Whenever spatial data is included plotting data glyphs on top of geographic maps is a common technique. The "Value Flow Map" visualization [9] plots a linear profile glyph [52] on top of each country to convey changes in country characteristics over time. Since the authors did not adjust the size of the glyphs overplotting in dense areas may occur. This problem is solved in the "Icons on Maps" [71] visualization. The simple idea is to reduce the size of the glyphs in smaller country areas. Although, the problem of overplotting is solved the comparison between several glyphs is more difficult since data values with different scales have to be compared.

[5] The evaluation introduced in section 3.3 is already included in this listing.

The TimeWheel [184] is a circular glyph design for multi-variate temporal information. The time axis for each dimension is shown on the circumference of a polygon. Data values are encoded with a line connecting each point in time with a data line to the center of the polygon showing normalized data values for all attributes. This encoding is similar to a parallel coordinate plot with only 2-axis simplifying simple pattern detection like visual correlation analysis. However, based on the amount of data lines, this glyph design is not robust against occlusion.

The linear profile glyph can also be used in three dimensional spatial visualizations to convey multi-variate data, as well. Wakame [66] is a visualization, which arranges multiple linear profiles in three dimensional space by aligning their baseline at the center point of a radar chart. Each of these radar chart glyphs is then arranged on top of a geographic map to investigate multi-variate time series data for specific regions. Of course, simple navigation techniques like rotating, panning and zooming are necessary to be able to investigate all attributes over time. This is not the case for the "Data vases" visualization [182], which abstracts the linear profile glyph by using a disc metaphor for single points in time. For each timestamp a disc is drawn, which size and color encodes the underlying data value. The discs are then stacked according to their position on a geographic map. Since each disc shows only one dimension, a navigation in three dimensional space is not necessary. Profile flags [129] plot small line charts on a three dimensional banner, which can then be put on top of different basic visualizations showing spatial information. Overplotting in dense areas can be avoided by using various lengths for the flagstaff positioning the banner at different heights.

Circular profile glyphs can also be transfered to three dimensional space. Helix icons [183] for example show periodic information for spatial temporal data by plotting cylinder like glyphs on top of geographic maps. The z-axis is used to represent the time dimension and color to display the underlying data value. In order to perceive the whole display and, therefore, data space, interactively changing the perspective on the cylinder is mandatory. Same thing is true for the pencil icons [183] visualization. Their design is similar to the helix icons, however, they encode multi-variate time-series data. The z-axis illustrates the time dimension and the different planes of the pencil represent various dimensions with a color encoding for the respective data value. Perceiving all dimensions is only possible by rotating the whole view or just the three dimensional glyphs.

3.1.2 Node-link Diagrams

Visualizations like MOSAN [186] show simulation data in a node-link diagram enriched with linear profiles. The simple nodes are exchanged with more complex glyph designs to convey additional information. The graph layout displays the model structure whereas the data glyphs are used to illustrate the temporal development of an attribute over multiple runs. The linear profile glyph is also used in other network visualizations with different contexts [198, 202]. Xu et al. visualize the development of social network data using the last.fm dataset [27, 202]. Two connected glyphs indicate a friendship connection whereas the glyph itself shows the amount of interest overlap over time. Westenberg et al. introduce the expression glyph to show DNA microarrays for four points in time in a gene regularity network [198]. Since there are only four time points visible a bar chart is shown with the bars colored according to the interaction type (i.e., green $\widehat{=}$ activation, and red $\widehat{=}$ inhibition).

The cluster glyph [14] is slightly related to the sticky figure visualization [150], however, tailored towards showing changes over time. Human movement is captured and visualized using small sticky figures. The variation in motion is displayed using snapshots of the animated limbs. The opacity of each limb displays the positions, which are traversed more often. The result is a blurry picture of a sticky figure showing the degree of movement.

3.1.3 Grid Layouts, Matrix Visualizations

A straightforward arrangement of data glyphs is a common grid or matrix layout where the columns and rows refer to different attributes. An example is the "Pathline" visualization [125] showing evolutionary changes of genes in a matrix layout. Each cell encodes temporal information in a linear profile glyph with the columns representing different genes and the rows various species. Comparisons across genes and species can be easily done by scanning through the rows or columns respectively. Im et al. propose a generalized scatterplot matrix (i.e., GPLOM) [90] for continuous and categorical data by exchanging the cells with small glyph representations. For time-series data in combination with numeric data, the tool displays small bar chart glyphs using one bar for each point in time. A colored stripe glyph with a linear layout is used for temporal categorical information.

A similar color encoded glyph was used by Oelke et al. [98, 144] for displaying visual document fingerprints. Single rectangles represent different sections of the document. The fill color of each rectangle is used to encode the number of occurrences of a specific term within a section. Of course, text documents cannot be considered time-series data, however, the progress in text is somehow related to a progress in time. Borgo et al. experimented with different block sizes for the inner rectangles and how they influence the performance of the analysts [18]. Their study suggest that the size does not affect the effectiveness of the analysis significantly.

A more unique glyph visualization is the InfoBug [40]. Multiple attributes are mapped to the torso of an abstract bug representation. The wings of the bug are shaped like small line charts with the time dimension progressing from top to bottom. Therefore, symmetric wings indicate a correlation of the two attributes mapped to the individual line charts.

Gestaltlines [21, 22] are similar to sparklines [185], however, they make use of an orientation encoding to show temporal changes. For each point in time a data line is drawn. The angle of the data line corresponds to the underlying data value. The lines are stacked according to the timestamps. The temporal axis can also be arranged horizontally to deal with longer time-series making this visualization also suitable to be included in textual documents.

The timeWheel [39, 40] should not be confused with the TimeWheel [184] introduced earlier. Although, both designs encode time-series data the mapping is different. In the timeWheel visualization different temporal attribute dimensions are represented by single line charts arranged on the circumference of a circle. The single line charts are rotated according to their position with the reference point in the middle of the circle. The color and the height of the line charts encode the respective data value. This timeWheel visualization can also be extended to a three dimensional glyph design shaped as a cylinder [39]. This cylinder is, therefore, divided into slices according to the number of attribute dimensions. Each slice shows the time dimension from the center point of the cylinder to its circumference. The height of each slice encodes the corresponding data value, which may change over time resulting in a bent surface.

The response glyph [97] also uses multiple data lines to encode multi-dimensional temporal data. Each line corresponds to one attribute, which is progressing according to their data value over time. The final glyph looks like many line charts plotted on top of each other.

3.1.4 Text Visualizations

SparkCloud [110] is a text visualization combining tag clouds with linear profiles to show the development of a term over time. By adding small data glyphs the static representation is enriched with temporal information showing trends in data without using animation.

Sparklines [185] can be added to documents and add temporal information to textual contents by plotting small line charts close to the respective text section. Instead of writing single numbers for e.g., stock prices, the sparkline technique uses the space more efficiently by showing the trend of stocks in small visual representations.

3.1.5 Hierarchical Visualizations

The “SolarPlot + Aggregated TreeMap” technique [38] enriches hierarchical information with time-series data by extending a sunburst visualization with linear profiles. Each bin in the sunburst is represented by a data glyph showing temporal information. Such line chart glyphs are also embedded in treemaps to show hierarchical time-series data [164]. However, due to the varying aspect ratio of the rectangles it is difficult to compare different time-series.

In the work of Dinkla et al. [50] linear color encoded glyphs are embedded in a hierarchical tree visualization. The tree structure is used to show the semantic hierarchy of terms in a document. The color indicates the frequency of a certain term in specific text sections.

3.1.6 Flow Visualizations

Flow visualizations can also be enriched with glyphs to convey temporal information. Flow radar glyphs [86] use color and a polar coordinate system to show the development over time. The jitter of the data line along the time axis encodes the underlying data value. The technique can also be extended to three dimensional space. AmniVis [140] is a visualization bundling multiple streamlines in a widget arrow glyph to show time steps of various lengths. The glyph can be used as an overview displaying the overall trend of many streamlines in a specific region.

3.1.7 Projection to 2D Space

Steiger et al. use linear profiles to visualize power consumption for multiple sensors over time [177]. The data points are projected to two dimensional space with similar temporal patterns being combined in one data glyph prototype. Interesting events like the drop in power consumption during daytimes for certain regions can be easily spotted.

Ward and Guo project linear and circular profile glyphs to two dimensional space using a PCA [193]. According to the underlying task the user can switch between the two designs. For analyzing the shape of a time-series the authors switch to linear profiles whereas for a more compact representation circular representations are preferred.

Yang et al. arrange pixel glyphs with an MDS projection on the screen [203]. The pixels in the data glyphs correspond to single points in time with a color coding for the underlying data value. The pixel arrangement can be done in various ways. For time-series data preserving the natural temporal order in a linear layout is convenient and easy to achieve.

Summary

As can be seen in this related work section many different data glyph designs for time-series data do exist. However, it is interesting to note that there are nearly no data glyph designs making use of a clock metaphor to show temporal information. Since some quantitative evaluations [65, 91, 170] have shown that data related metaphors help to better analyze the underlying data, a new glyph design using a clock metaphor is worth pursuing. SpiraClock [51] or the spiral graph [196] are two visualization systems making use of such a metaphor to display temporal information. The idea is to convert these visualizations to small data glyph designs and investigate whether such a metaphor really works for time-series data. In the next section 3.2 I will introduce the new glyph design namely the *clock glyph* and show its applicability to the network security domain using time-series data.

3.2 Clock Glyph - A Data Glyph Design to Visualize Time-Series Data

In the following, I will introduce a new data glyph for visualizing time-series data. As explained in section 3.1 a metaphoric data glyph design is still missing. Based on the performance of other metaphoric data glyphs such an approach seems promising. After explaining the design choices made I will show the applicability of the *clock glyph* with a network security use case. A quantitative experiment conducted in section 3.3 compares the *clock glyph* against well-established data glyph designs and proves the usefulness of this metaphoric representation.

3.2.1 Design Space for Temporal Glyphs

The design space for a basic temporal glyph can be characterized by the visual variables that are used to encode two attributes of temporal data: a) the position of a timepoint on the plane and b) the data value associated with this timepoint. Different visual variables can be used to encode these two attributes. Ward [192] describes several categories of glyphs. To narrow down the design space we[6] only discuss temporal glyphs with many-to-one mappings where several or all data attributes map to a common type of graphical attribute. This is important in

[6]In this section 3.2 the term ”we” comprises Christopher Kintzel, Florian Mansmann, and me

order not to promote certain temporal dimensions and to enable easier intra-record and inter-record comparison, which is fundamental for many tasks involving time series. While many more different glyph designs exist (see section 3.1) we focus on two main types of glyphs here: profiles and stars (see [192]). Both types have the advantage that relationships between adjacent data points are easier to see than for other glyphs [192].

The focus of the design was to be able to read high data values, spot and name interesting points in time, and be able to detect positive or negative trends. To best capture the time component of the data we thought about a metaphoric design for the data glyph, thus, excluding linear profiles. Context related metaphors help to better understand the underlying data; a finding supported by a study conducted by Siirtola in 2005 [170]. Participants working with a data glyph designed to communicate the context of the underlying data performed significantly more accurate in analysis tasks compared to non related data glyphs. Therefore, we assume that arranging the time in a circular way will help analysts to better understand temporal aspects by preserving the natural order of time.

To represent the data value for each point in time we have to keep two constraints in mind. First, the clock metaphor must be preserved. The pointer for each time slot should have the same size and an equal angle to strengthen this fact. Second, dimensions must be represented with an identical visual variable. This design choice is mandatory because the analyst must be able to easily compare different dimensions against each other. As a result, we used a *many-to-one* mapping for the design [192].

Because of these constraints, there is only a limited set of visual variables which could be used to encode the respective data value. The length, the saturation/brightness, the texture, or the orientation of the pointer. Texture should not be used for numeric information because a natural order can hardly be perceived [195]. Different orientations of pointers would result in many crossing segments making it hard to read exact data values. Length, being the most accurate visual variable, seems like a reasonable choice. However, the different lengths of the pointers could harm the regular structure of a clock. Time segments with no data value would result in gaps between the pointers (see Figure 3.1). Reading the exact position of a time-slot could, therefore, be more difficult. That is why we decided to use color saturation to encode the data value (Figure 3.2).

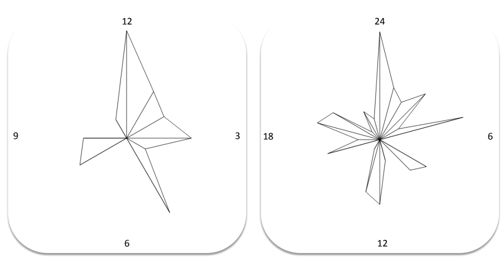


Figure 3.1: *Star glyph:* data lines are radiating from the center. The length of the lines encode the data value for different points in time. Low data values result in gaps within the design. Left: 12 points in time; right: 24 points in time.

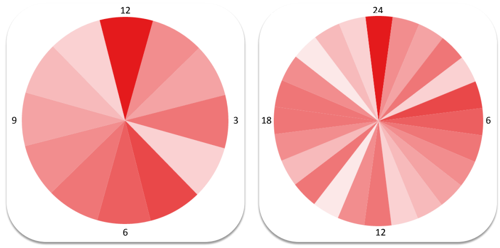


Figure 3.2: *Clock glyph:* visualizing time-series data with a clock metaphor. Slices arranged in a circular layout correspond to the time dimension. Color saturation is used to communicate the respective data value. Left: 12 points in time; right: 24 points in time.

Of course, the selection of the current encoding is only based on observations of individuals. A proper evaluation of the design choices made is still missing. In the following sections real world data from the network security domain will be used to show the applicability of the *clock glyph*. *ClockView* [100], *ClockMap* [63], and *VisTracer* [64] are three tools making use of the *clock glyph* to monitor network traffic and increase the situational awareness. For a more generalized evaluation of the *clock glyph* a controlled user study was conducted (section 3.3) to get additional information about the performance compared to other design alternatives (i.e., line glyphs, star glyphs, and stripe glyphs). Our results show that depending on tasks and data density, the chosen glyphs performed differently. Line glyphs are generally a good choice for peak and trend detection tasks but radial encodings (i.e., *clock glyph*, and star glyph) are more effective for finding specific temporal locations. The additional qualitative analysis also contributes implications for designing temporal glyphs for small multiple settings.

3.2.2 Application-Oriented Evaluation in the Network Security Domain

Detecting anomalous traffic in an entire company network is difficult because of two reasons. First, since the number of machines in a network grows at a rapid pace, many different hosts have to be monitored over time. Second, the amount of traffic leaving or entering the network grows relative to the number of new hosts. Thus, there is a need for network security tools helping the administrator to analyze the traffic. This massive amount of data cannot be effectively investigated by sequentially reading textual log files. Researchers and practitioners are aware of this fact and developed many different tools and concepts to apply filtering and visualization methods to this kind of data in the last few years. The goal is to support the administrator in dealing with this massive amount of data and in exploring anomalous traffic. Besides operationally monitoring real-time traffic to supervise a network, forensic analysis becomes an important aspect to reveal attack patterns and develop defense mechanisms against future attacks through diversifying malware aimed at circumventing traditional defense mechanisms.

To show the applicability of our data glyph design we used the *clock glyph* within a network security domain. This domain seems promising since temporal data is most often used in combination with other data types like hierarchies, or networks. Therefore, detecting anomalous traffic in an entire company network is difficult and a great research challenge. Like most data glyph designs, the introduced *clock glyph* has the advantage of being flexible in the way it can be arranged on the screen. Combinations with other visualizations like e.g., node-link diagrams to show relationships between devices, treemaps to illustrate the hierarchical order, or common matrices to understand network structures within companies, are possible. To put high demands on the design three different data sets including combinations of multiple data types are used. For each data set a different visualization system making use of *clock glyphs* is presented.

Use Case I: Monitoring NetFlows with the ClockView Application

In our *ClockView* application [100] we[7] would like to enhance the overview visualization of already existing tools to show more details about individual hosts at the same time. NVisionIP [109] for example is a software showing an entire network of hosts in a 2D matrix divided into different subnets and host IP addresses. Every host is represented as a four pixel rectangle. The color of each rectangle encodes the traffic of the host on different ports. Unfortunately, the visualization only shows one state of time at a glance. Anomalous behavior over time cannot be discovered on one sight. A more detailed perspective of the network is provided by the *Small Multiple View*, which uses two bar charts to visualize further information about the hosts. However, the overview is lost because only a limited amount of hosts can be displayed on the screen in this detailed way. To obtain more information, the analyst can dig deeper and investigate a single host by looking at its raw traffic data in the *Machine View*.

NVisionIP was inspiring because of the way the network is monitored in a matrix visualization using small representations for every single host with the possibility to get details on demand. However, the way in which the hosts were displayed was not satisfying. With the aforementioned representation it was only possible to code a single parameter within each host (e.g. number of ports used). Therefore, the *clock glyph* would be a better way to display single machines in the network to have the possibility to code more parameters without loosing the overview. As a consequence, we embed our *clock glyph* in a matrix visualization to show the network hierarchy as well as the amount of traffic for each individual device over time. Analysts are able to monitor thousands of hosts with our *ClockView* prototype.

To get a global picture of the servers and workstations used in the network, it is useful to visually encode each host individually in a network overview. The hosts are represented in a way the user can easily notice, if a specific machine's behavior matches more a server with 24 hours of traffic or a client with traffic only on the working hours. Therefore, we want to show all internal hosts with their traffic at a granularity of one hour for a timespan of one day. For this purpose we need to display up to 65536 (256*256 possible IP addresses for a /16 network) time series, each with 24 (one per hour) data values. This leads to a maximum of 1572864 data points.

Each host is represented as a *clock glyph*, which is subdivided into 24 segments, each of them showing the traffic of one hour of the day encoded with color saturation. 0:00 o'clock is at the top, 6:00 o'clock at the right side, 12:00 o'clock at the bottom and 18:00 o'clock at the left side (Figure 3.2 (right)). As a clock metaphor is used here, this segmentation is more intuitive as the segmentation into rectangles, even if the clock is transformed from 12 to 24 hours. Also the

[7] In this subsection 3.2.2 the term "we" comprises Christopher Kintzel, Florian Mansmann, and me

natural order of time is better preserved, since there are no line breaks between the data points. The time representing segments are not only at the same position for every host, but also have the same orientation. Corresponding hours of different hosts are displayed in parallel and thus at a glance can be recognized as group. Since the separation between the glyphs is already achieved due to the circular shape, no additional spacing has to be added. Because of this, the glyph is more space-efficient on smaller screen resolutions.

The amount of traffic is represented by a fixed diverging color scale from blue (negative, only used for comparison showing a decrease in traffic) over white (0) to red (positive). Due to the fixed color scale hosts remain comparable on different days. Otherwise a host with the same amount of traffic on different days could be perceived entirely different.

Figure 3.3 shows the *ClockView* application run on a Powerwall display with a resolution of 5224 x 2160 pixels.[8] Without changing the setting (i.e., level of detail), analysts can get an overview picture of many time series when looking at the visualization as a whole to spot interesting patterns. However, when moving closer to the display individual glyphs can be investigated in more detail to really compare the amount of traffic for specific points in time.

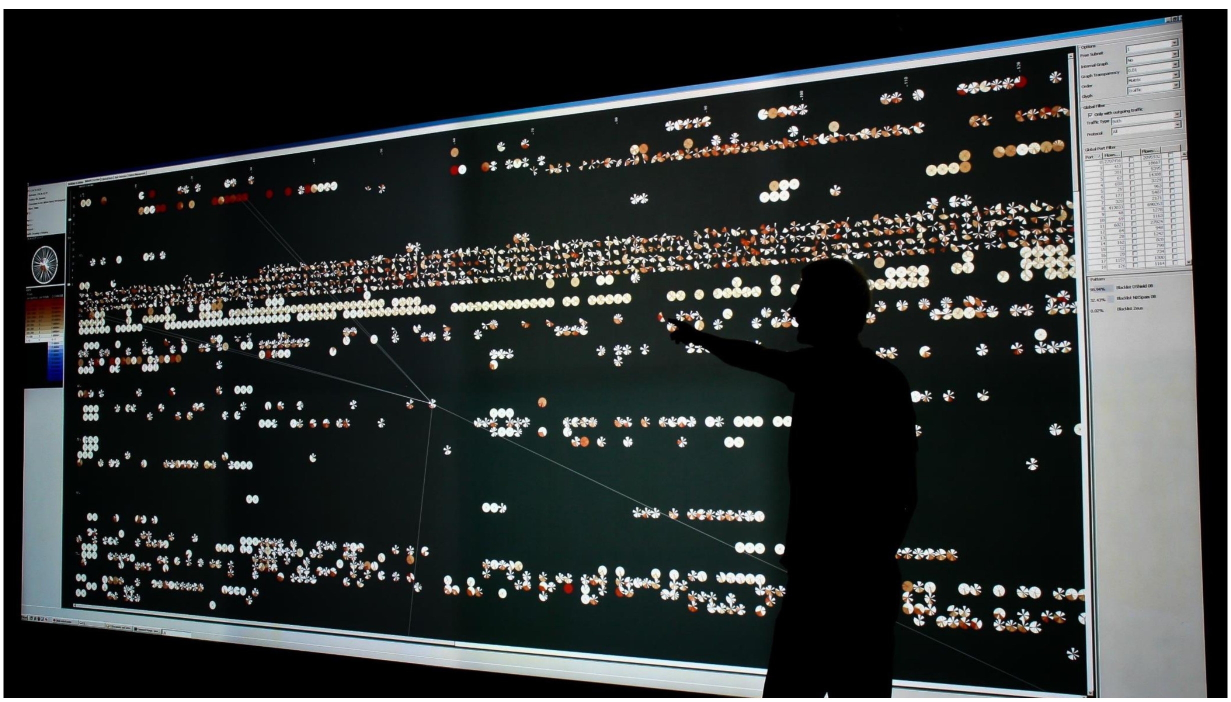


Figure 3.3: *ClockView:* The application is used on a Powerwall display with 8 HD projectors. *Clock glyphs* are arranged in a matrix layout showing all network devices of one big company. The position of each glyph is based on its IP address in the network.

To evaluate the *ClockView* visualization and show its operational usage we apply it to our university's network. To spot suspicious behavior the traffic of a whole day for all network

[8]http://www.vis.uni-konstanz.de/en/powerwall/, retrieved 02.02.2015.

devices is monitored and displayed using *clock glyphs*. To better identify abnormal behavior we color the glyphs according to the change in traffic from the current day compared to the previous five days.

As expected, most of the glyphs are colored white thus signalizing nearly no change, except for a partial red pattern in one single subnet (Figure 3.4). To take a closer look at the single glyphs we enlarge the visual representations by zooming in this exact area. With the additional space for each circle the traffic distribution over time is getting more obvious. Basically on the second half of the day the amount of traffic rises. It seems that some new machines have been added to the network causing extra traffic. This is suspicious because the monitored dataset was a Sunday where there is no regular daily work in the university. After investigation, we discovered that the corresponding subnet of the university is assigned to the vpn connections. A computer connecting to the university from an external network gets an IP address in this specific subnet. With this additional information the suspicious pattern can be explained as a common occurrence.

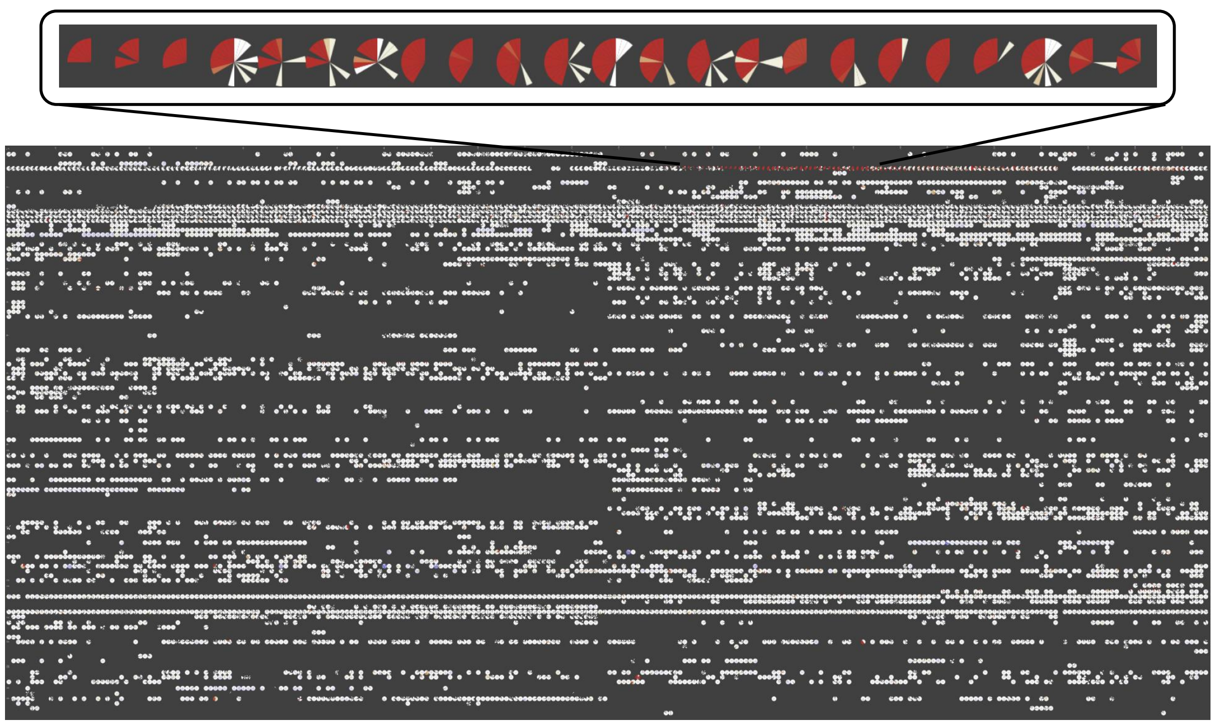


Figure 3.4: *Suspicious behavior:* The amount of traffic for all network devices is compared over multiple days. The change of traffic for each single time slot is mapped to color (i.e., white $\widehat{=}$ low change, red $\widehat{=}$ high change). Most devices have a similar behavior, however, several hosts in one particular subnet show a quite diverse behavior compared to the previous days (i.e., artificially highlighted).

Use Case II: Exploring Network Traffic with the ClockMap Application

The general idea of the *ClockMap* application [63] is to show hierarchical time-series data with nested circles and glyphs. A circular treemap is used to convey the hierarchical structure of the network, whereas, each time-series is represented with *clock glyphs* (see Figure 3.5).

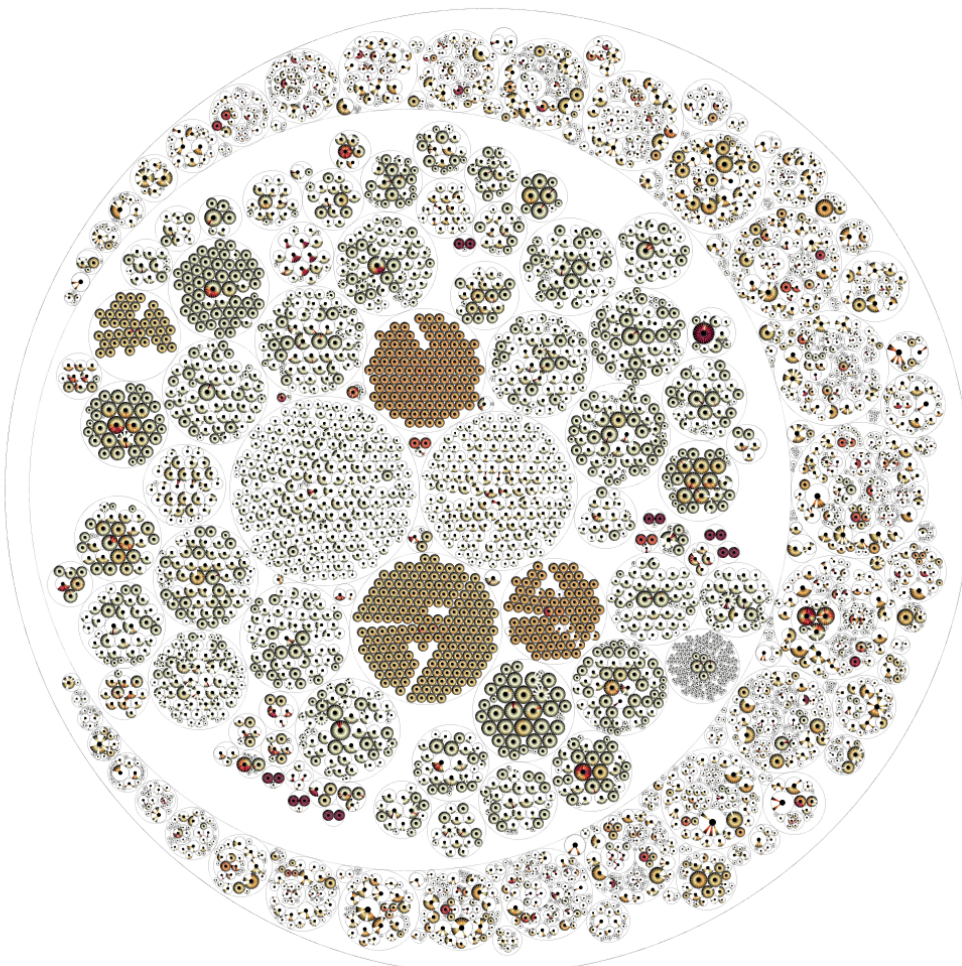


Figure 3.5: *ClockMap without aggregation:* The data glyphs are embedded in a circular treemap. Each device in the network corresponds to one time-series and is, therefore, represented by one *clock glyph*. The nested circles convey the hierarchy information of the underlying network structure.

Of course, a rectangular treemap would be more space efficient, however, the radial layout of the data glyphs perfectly fits into the circles of the circular treemap. This is true for each level in the hierarchy since the radial layout of the glyphs scales with the circular design of the treemap. This flexibility enables a highly interactive exploration process with panning and zooming possibilities. In the *ClockMap* application each hierarchy level is represented by one *clock glyph* showing the aggregated network traffic of all children in a specific branch. On the highest level of the hierarchy *ClockMap* displays only a single *clock glyph* (i.e., root node), which visualizes the aggregated time-series of all underlying nodes. However, analysts can switch between different levels of hierarchies by zooming into this information space. After passing a certain threshold the root node is split into multiple *clock glyphs* each representing one branch of the current hierarchy level. Again, each *clock glyph* shows the aggregated time-series of all its children. The size of each glyph encodes the number of children for this specific branch. This recursive behavior can be repeated until the analyst reaches the lowest level of the hierarchy where each leaf is represented by one *clock glyph*. Figure 3.6 illustrates the aggregation of the underlying hierarchies in a static screenshot. The dashed rectangle is artificially included showing the next hierarchy level in the circular treemap after the threshold is passed.

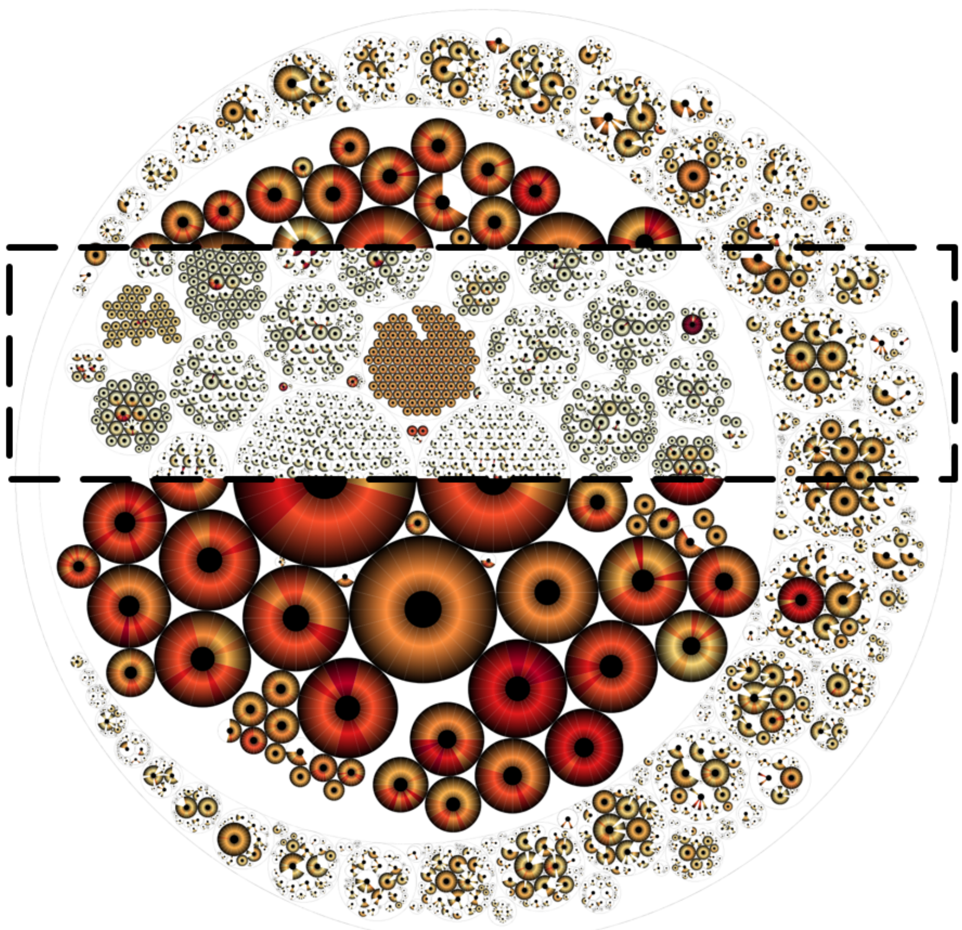


Figure 3.6: *ClockMap with aggregation:* The time-series of different hierarchy levels are aggregated in one *clock glyph* per branch. After zooming into certain areas the visualization switches and shows all underlying children again represented with data glyphs.

People may argue that rectangular treemaps could be enriched with linear profiles, too, to visualize the same information. This may be true, however, I would like to stress that the clock metaphor of the glyph design would then be violated.

As already shown in the *ClockView* application network data contains an inherent hierarchical structure with temporal information. Since *ClockView's* matrix overview can only show the relation between two different levels of hierarchies (i.e., x-axis, and y-axis) the *ClockMap* application is able to reveal network traffic patterns for multiple hierarchy levels.

Like in the previous use case we[9] consider NetFlow data of 24 hours. The data set contains 6048 hosts belonging to the same /16 IPv4 address block. On the first hierarchy level several different *clock glyphs* are visible each encoding a single branch, which corresponds to the second block of the IP address space (Figure 3.7). Besides the big clock representation one smaller glyph caught our attention because it is entirely colored in red, thus, signalizing a high amount of traffic. To get additional information about this subnet we zoom into this region to trigger the semantic zoom. Consequently, the *clock glyphs* are replaced by multiple smaller clock representations, which belong to the respective subnet. Interestingly, the branch with the small red clock glyph contains a further entirely red *clock glyph*. This means that only this address space is responsible for the high amount of traffic. Another zoom into this region reveals three network devices, which have a high amount of traffic especially in the night hours. Since their behavior seems to be entirely different compared to the other devices in this address space these three hosts should be investigated in more detail.

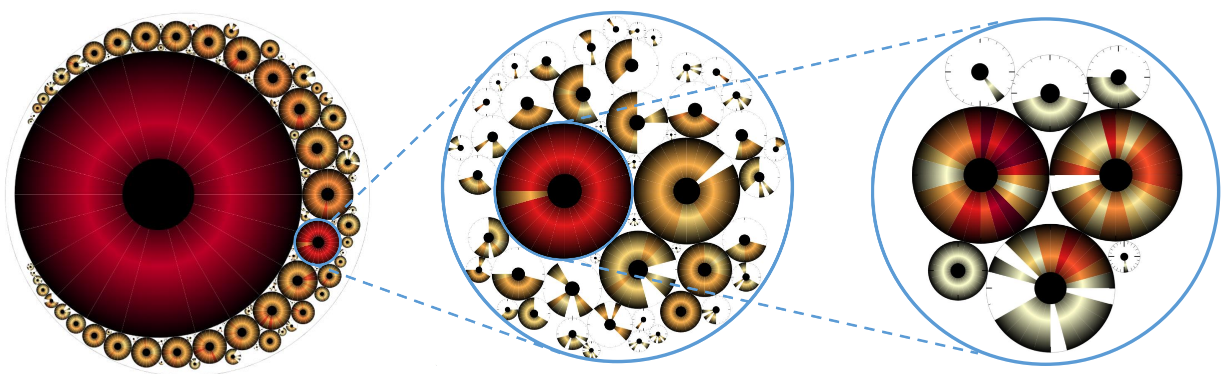


Figure 3.7: *Suspicious hosts:* A semantic zoom is used to investigate a smaller subnet with high traffic (i.e., entirely red) in more detail (left). Only one branch of this subnet seems responsible for this high amount of traffic (middle). Another semantic zoom reveals three hosts having high traffic especially in the night times (right).

Use Case III: Analyzing Temporal Network Changes with the VisTracer Application

The use cases I and II focus on the temporal and hierarchical aspect of the data not including the communication between different devices. In the third use case the *VisTracer* application

[9] In this subsection the term "we" comprises Fabian Fischer, Florian Mansmann, and me

[64] considers network information like routing changes over time using traceroutes. The tool, therefore, combines node-link diagrams with *clock glyphs* to detect anomalies during the communication.

The graph layout of the node-link diagram reads as follows. The communication starts on the left side of the screen and progresses to the right until it reaches the final destination. Whenever a different route is taken a new layer is added on top of the actual route showing the new direction. The nodes in the graph represent the different hops while the edges show the connections with each other. The width of an edge depends on the amount of traces using this exact connection. The nodes are exchanged with *clock glyphs* with equally sized slices and small flags reflecting the country of the hop as can be seen in Figure 3.8.

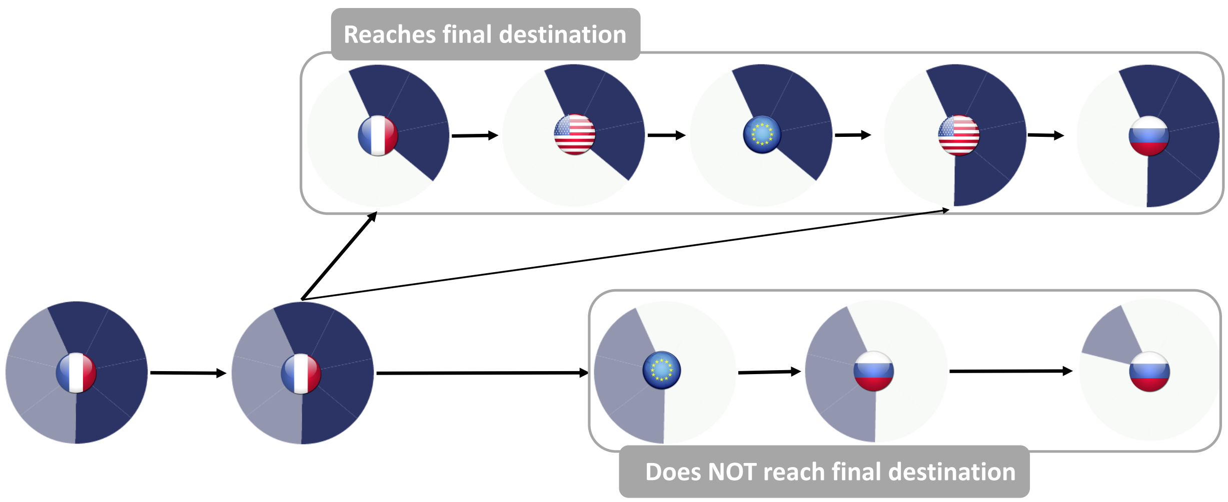


Figure 3.8: *Routing anomaly*: Two different paths are taken to reach the final destination in Russia. After 4 traceroutes a different route is taken. It is interesting to note that only the route displayed at the top reaches the destination.

Because of the aspect ratio, the circular glyphs can be directly integrated into the graph nodes without wasting additional space for this temporal information or requiring animation. The number of slices depends on the amount of traceroutes shown in the graph. The clockwise arranged slices represent the different traceroutes for one day. When a hop was used in a traceroute the respective slice is colored in dark blue or gray, otherwise it is filled with white color. The color (i.e., dark blue or gray) depends on whether the traceroute reaches its destination or not. This encoding supports the analyst in detecting the main route (i.e., based on the path's width), the usage of hops (i.e., the proportion of colored slices), the reachability of the destination (i.e., the hue of the colored slices) and the temporal development of the route (i.e., the partition of the slices). Additionally, the geographic location of the corresponding country can be taken into account by looking at the flag in the center of the *clock glyphs*. This additional information may highlight possible route flappings between different countries along the route in the graph.

Figure 3.8 shows an interesting use case scenario. Seven traceroutes have been initiated from a computer located in France (node at the left) with the destination being somewhere in Russia (node at the right). The single traceroutes are visualized with different slices in the *clock glyph* and color is used to show whether the traceroutes successfully traverse the path to the destination or not. It is interesting to note that the first three traceroutes are looping between hosts in the US and Europe before reaching the final destination. This is an interesting finding since the traceroutes do not to take the shortest path according to the geographic location. Additionally, after the forth initiated traceroute command the path is changing entirely not traversing through the US anymore but taking a more direct route through Russia. However, the gray slices in the *clock glyphs* illustrate that this route is not reaching its destination. This may be an indication for an attack rerouting network packages through hops in Russia never reaching the final destination. Combining *clock glyphs* with node-link diagrams helps to better understand routing issues in networks over time. Temporal changes can be investigated in a static view without the need of additional timeline navigations or animations.

3.2.3 Conclusion

As shown in section 3.1 there exists a great variety of glyph designs for time-series data. Since metaphor based designs for temporal data were missing the *clock glyph* was introduced as a suitable alternative. To show its applicability to real world scenarios the data glyph was embedded in three network security tools (i.e.,*ClockView*, *ClockMap*, and *VisTracer*) showing network data over time. The familiar clock design helped analysts to identify temporal patterns and spot suspicious behavior as shown in the use cases.

To investigate the performance of this glyph in more detail we[10] further conducted a controlled user study with well-known design alternatives. Based on quantitative (e.g., the effectiveness and efficiency) and qualitative (e.g., confidence) results we were able to rank all design alternatives for different tasks and data densities. In the following section (section 3.3)I will describe this experiment in more detail and finally come up with design considerations for glyphs encoding time-series data.

3.3 Evaluation of Alternative Glyph Designs for Time-Series Data in a Small Multiple Setting

The following section contributes to two research areas. First, as mentioned in chapter 2 there are open research gaps when investigating the performance of different glyph designs (see Figure 2.3). For some alternative representations no quantitative experiments have been conducted. Thus, practitioners as well as researchers cannot rank these designs or refer to design considerations. Second, the *clock glyph*, which was introduced in section 3.2 was not evaluated in a controlled user study. Measuring its performance and comparing it to alternative representations

[10]In section 3.3 the term ”we” comprises Petra Isenberg, Fabian Fischer, Florian Mansmann, Enrico Bertini, and me

helps to better understand the influence of different visual variables in glyph designs, to propose design guidelines for glyph representations showing time-series data, and to confirm that a clock metaphor works for time-series data.

As shown in subsection 3.2.2 analyzing many time series at once is a common yet difficult task. This is not only true in the network security domain. Time-series data is the basis for decision making in many different application domains, as well–such as finance, or traffic management. Detecting trends, spotting peaks, or investigating single points in time from a visual representation are daily analysis tasks of vital importance [2, 100, 108, 147].

Since different visual variables such as length, color, or position can be used to encode two aspects of temporal data in one glyph: a) the location of a data point in time, and b) the quantitative data value, a multitude of designs have been proposed (see section 3.1). When confronted with the task of choosing an appropriate glyph design, a visualization designer or practitioner currently has little guidance on which encodings would be most appropriate for which tasks and on which visual features and factors influence people's perception of data encoded in glyphs. While one could follow Cleveland and McGill's ranking of elementary perceptual tasks [43] and try to predict the performance of glyphs based on these results, it is not clear whether their results will hold. Temporal glyphs include dual encodings, are used in specific temporal analysis tasks, and come in many different sizes and densities.

In order to address this lack of guidance on the use of temporal glyphs, we ran a controlled experiment to compare four carefully selected glyphs using two different data densities.

As a starting point we took the *clock glyph* [100] and the *sparklines* [185] technique as famous representatives for time-series data. Additionally, we extended the design space to a total of four glyph designs including a stripe glyph [108] and the well-known star glyph [168]. These alternatives were chosen for their use of different combinations of visual variables to encode temporal position and quantitative value of a data point. We evaluated all glyph designs in a small multiple setting as small multiple is the most common usage scenario for temporal glyphs and the regular layout reduces confounding factors due to the positioning.

3.3.1 Experiment Design

The purpose of our experiment was to compare the performance of different, potentially powerful, temporal glyphs in a small multiple setting. Our three tasks are inspired from our work with network analysts but generalize to other domains in which temporal data has to be compared and analyzed.

Experiment Factors

Our experimental factors were *glyph*, *task*, and *data density*.

Glyphs: Since we wanted to compare our *clock glyph* against the *sparklines* technique we thought about additional alternative representations to bridge the gap between the two designs. Our *clock glyph* (CLO) uses a radial layout and a color saturation encoding to visualize time-series data. The *sparklines* technique (LIN) has a linear layout for the time dimension and a position/length encoding for the data value. To be able to better reason about changes in performance we included a stripe glyph (STR) with linear layout and a color saturation encoding, and a star glyph (STA) with a circular layout and a position/length encoding. We chose to test STA for its similar value encoding to LIN and STR for its similar value encoding to CLO. When comparing glyphs visually, the distance between the representations matters. We chose to keep the distance for the different designs identical and, therefore, to have the same uniform small multiple layout. As a consequence it was important to set a fixed aspect ratio for each glyph. To maximize display space for circular glyphs for a fairer comparison we chose a square aspect ratio for each glyph.

For the color encoded glyphs (CLO and STR) we chose a heatmap colorscale, which was motivated by the yellow to red colorscale from ColorBrewer [23]. This scale takes advantage of the fact that the human visual system has maximum sensitivity to luminance changes for the orange-yellow hue [113] and it is also suitable for color blind people.

For each trial, the same type of glyph—but showing different data—was drawn on the screen in a small multiple layout of $8 \times 6 = 48$ glyphs in total (Figure 3.9). Each glyph was drawn at a resolution of 96 $\times$ 96 pixel.

Tasks: Many different tasks exist that can be performed on time-oriented data [2, 10, 119]. We chose our tasks taking two criteria into account: (1) their ecological validity, i. e. how commonly they are performed in environments where the quick comparison of multiple time series is needed. (2) their heterogeneity in terms of the elementary perceptual tasks, i. e. we picked tasks that involve the comparison of visual variables for encoding data values, investigating different layouts for time and the combination of the two. In terms of ecological validity our tasks were inspired by our work with network security analysts from a large university computer center who had to monitor large amounts of network devices. The analysts had to be able to efficiently detect anomalous traffic patterns (e.g., peak values in none working hours) to be able to quickly react on the possible threat. Our three tasks were:

Task 1—Peak Detection: Amongst all small multiple glyphs, participants had to select the glyph that contained the highest data value (Figure 3.9). This task, thus, involved scanning all glyphs for its highest value and comparing across glyphs using length (LIN, STA) or saturation (STR, CLO) judgements.

Task 2—Temporal Location: Among all small multiples, participants were asked to select the glyph with the highest value at a predefined time-point. This time-point was textually shown to the participant in advance (e.g. “3am”). This task, thus, involved first identifying the location of a time-point by making positional (LIN, STR) or angular judgements (STA, CLO) and then comparing the peaks as in Task 1.

Task 3—Trend Detection: Among all small multiples, participants had to select the glyph with the highest value decrease over the whole displayed time period (Figure 3.10). This task, thus, involved first detecting all decreasing trends and then comparing the first and the last value.

Data Density: In order to test the scalability of each glyph in terms of the number of datapoints it can encode, we tested two data densities. The smaller density consisted of 24 data values (1

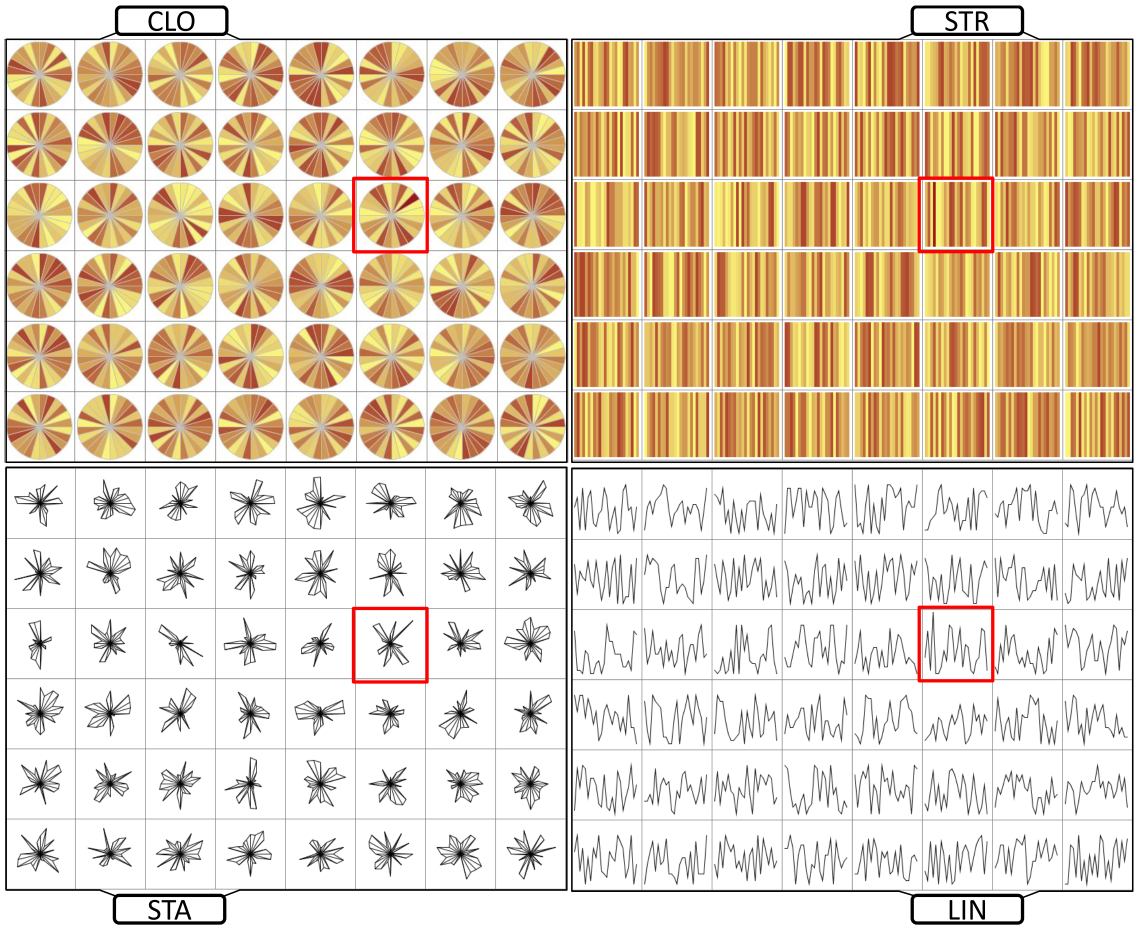


Figure 3.9: *Peak detection:* Illustration of the different glyphs with one high data value at a random point in time. For a better understanding the correct glyph is artificially highlighted.

for each hour), and the larger of 96 data values (1 for each 15 minutes). The rendered size of the glyphs holding these data points was not varied between each density (Figure 3.11).

Hypotheses

We previously conducted two exploratory pilot studies with similar glyphs and tasks. From these and the related literature [43, 192] we derive the following hypotheses:

H1: *For tasks involving primarily a value judgement **LIN** & **STA** (position/length encodings) are more accurate and efficient than **CLO** & **STR** (color encodings). This effect is strongest for **LIN**.* This hypothesis is based on Cleveland and McGill's experiments [43] on the perception of position, length, and color. We expect the results to hold for both data densities.

H2: *For tasks involving primarily a value judgement, **CLO** & **STR** (color enc.) are more impacted by higher data density than **LIN** & **STA** (position/length enc.).* Color perception

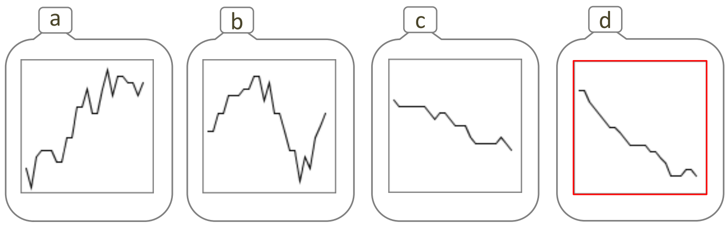


Figure 3.10: *Trend detection:* The four glyphs demonstrate different kinds of trends. From left to right: (a) visualizes a positive trend; (b) contains a positive and negative value development but for the whole displayed time interval there is no clear trend visible; (c and d) picture a negative trend over the whole displayed time period with (d) having the higher decrease. The glyph with the highest decrease over the whole displayed time period is artificially highlighted.

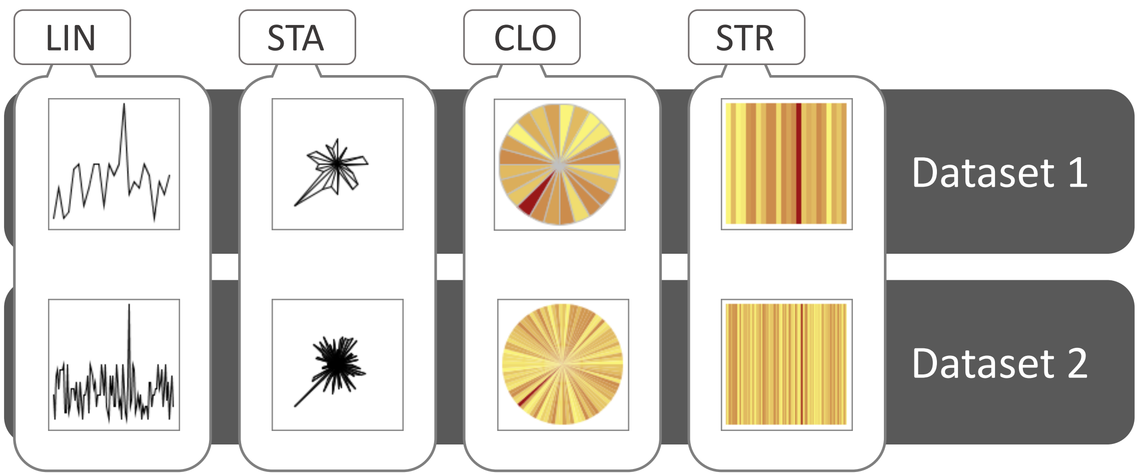


Figure 3.11: *Data density:* Differences between the two datasets for each glyph design.

may change drastically with varying context colors and size of the object being viewed [179, 195]. We expect color perception to be more impacted than visual acuity on dense line and position encodings.

H3: *When detecting temporal positions, **STA** & **CLO** (angular enc.) outperform **LIN** & **STR** (position enc.).* Using the familiar clock metaphor, we expect that circular glyphs allow the perception of specific points in time to be more accurate. This effect is stronger for **CLO** than **STA** as the clock shape is more clearly retained.

H4: *When detecting temporal positions, increasing data density will negatively impact performance with each glyph..* This is because color judgements are impacted by the size of the object being viewed [179] and angular as well as positional judgements by visual acuity. We expect **CLO** & **STA** to perform best as they spread out values towards the circumference of the circle giving additional space for perceiving color and position.

H5: *For trend detection, **LIN** & **STA** (position and length enc.) are most effective.* In trend detection, two mental sub-tasks have to be integrated by the participant: a) analysis of data development over time (characterized by the slope) and, b) comparison of the first and last data value (trend steepness). We expect the first sub-task to be performed equally well with all glyphs but expect that the comparison of distances between two data values is more difficult with color compared to position/length.

H6: *For trend detection tasks, the participants' performance for each design is not influenced by data density*. For detecting a trend comparing the overall shape rather than single data values is necessary. We expect that increasing the data density will not influence the trend shape and, thus, has no effect on task performance.

Experiment Design

We used a mixed repeated-measures design with the between-subjects variable *task* and the within-subjects independent variables *glyph* and *data density*. The dependent variables were *error*, *time* and *confidence*. Each participant conducted one task with all four glyphs, two densities, and four trial repetitions.

Data: To control the data values and their resulting visual representations, we created synthetic data for the experiment. In total, we created 48 data instances (glyphs) for each repetition, task, and data density. The data was created such that just one glyph represented the correct answer. The glyphs with smaller density held 24, the ones with large density 96 data values. In previous pilot experiments these two values were established as being sufficiently different from one another. Data for each task was created as follows:

Task 1: Each glyph was filled with random noise to a threshold of 80% of its value range according to our experience from pilot studies. For the target glyph a peak value at 100% of the value range was added to the dataset at a random point in time.

Task 2: Each glyph was filled with random noise as in Task 1. A peak value at 100% of the value range was added to the target glyph at a predefined point in time. For the distractor glyphs, peak values of the same value were integrated but at wrong temporal positions.

Task 3: We designed different decreasing trends by varying the values of the first (0–25% of value range) and last data point (75–100% of value range). The target trend decreased 75% of the value range from first to last data value while the distractor glyphs included a decrease of 55%. Along the trend line each data point was varied by zero, one, or two values using a probabilistic function.

Participants: We recruited 24 participants (12 male, 12 female) mainly from the local student population. All participants had normal or corrected-to-normal vision and did not report color blindness. Their age ranged from 19–56 years (median age 24). Each participant had at least finished high school, eight held a Bachelor's, two a Master's degree, and one a Ph. D. The academic background of the participants was quite diverse with no one having a computer science background. 34% of the participants reported to use the computer for more than 30 hours per week and 50% less than 20 hours.

Procedure: The experiment took place in a quiet closed room at our university. In addition to the study participant, the experimenter was the only person present. The participant sat in front of a table at a distance of approx. 50cm from a 24in screen set to a resolution of 1920×1200. Participants interacted with the study software using only a mouse.

The experimenter began by explaining the data, the single task, and the design of the different glyphs. The data was presented as financial stock data to provide context.

Only when the participant was familiar with the current glyph design and task, he/she was allowed to proceed. For each glyph and density tested, the participant stepped through four practice trials followed by the four actual study trials. After each trial, the participant entered a confidence score for their answer on a 5-step Likert scale.

The task question was visible on the screen at all times. The presentation order of each glyph was randomized in a Latin square fashion between participants. The glyphs were presented in a 6×8 matrix layout (Figure 3.9). Each participant saw the same glyphs per trial in different random configurations.

3.3.2 Results

We report on significant results ($p < .05$) from our quantitative analysis in this section and refer to the qualitative feedback in the discussion section afterwards.

Data Analysis

Task completion time, error rate, and confidence score were recorded for the analysis. We used a repeated-measures ANOVA for the analysis of completion time. Time in our experiment was log-transformed where it did not follow a normal distribution. For the error rate as well as for the confidence score, a non-parametric Friedman's test was used.

Except for the second task we did not observe a strong learning effect between trials. Therefore, we analyzed all four trials for the first and third task, glyph and dataset for each participant. For the second task we analyzed the results of the last three trials. In addition, single answers were marked as outliers when each metric (time, error) was beyond two standard deviations from the mean for a given task and glyph per participant. Outliers were replaced with the closest value two standard deviations from the mean for each participant according to standard procedure. The tasks used in the study differed in their characteristics, so we analyzed the results of each task and dataset independently. Finally, we analyzed the feedback and subjective preference from the post-session interview for a qualitative analysis.

Task 1: Peak Detection

Task 1 consisted of four training repetitions and 2 densities $\times$ 4 repetitions with an increasing difficulty for each repetition block. This setting was used for each glyph design. For the analysis we only considered the more difficult repetition block since the results reveal more interesting insights (see Figure 3.12).

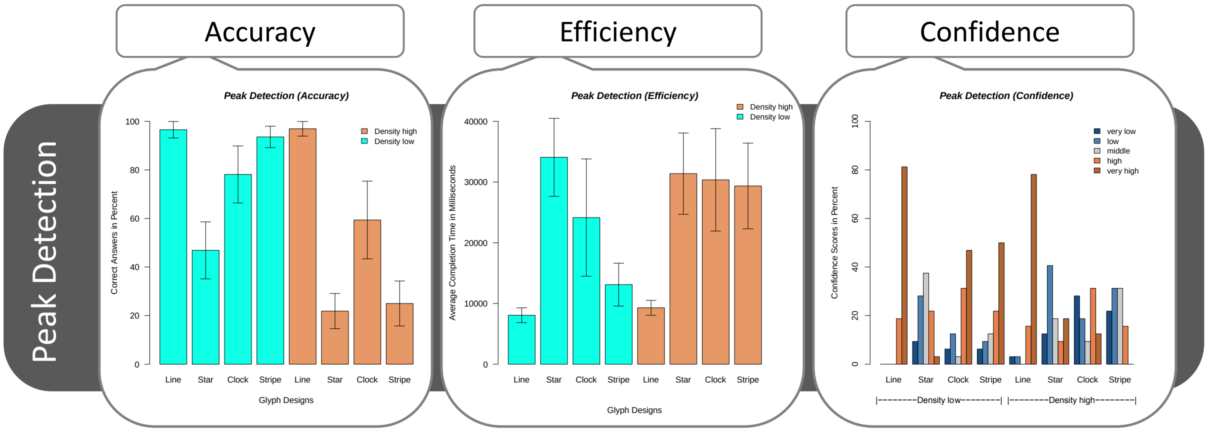


Figure 3.12: Bar charts with mean and standard deviation showing the results for the peak detection task and factor. The x-axis represents the different dependent variables.

Accuracy: There was a significant effect of *glyph* on *error* for both the low density ($\chi^2(3, N = 32) = 11.62, p < .01$) and the high density condition ($\chi^2(3, N = 32) = 17.59, p < .001$). In the low density condition pair-wise comparisons showed that errors in judgement were significantly worse for STA (46.9%) than all other designs ($p < .05$). LIN (96.5%) and STR (93.6%) both showed high accuracy with LIN nearly at 100% accuracy. In the high density condition LIN (96.9%) significantly outperformed the other designs by staying at nearly 100% accuracy (all $p < .05$). In addition, CLO (59.4%) performed significantly better than STR (25%) and STA (21.9%) with $p < .01$ in each case. With an increasing data density, STR (from 93.6% to 25%) and STA (from 46.9% to 21.9%) significantly lost accuracy (all $p < .05$).

Efficiency: There was an overall effect of *glyph* on *time* in the low density $(F_{3,21} = 12.1, p < .0001)$ and the high density $(F_{3,21} = 11.5, p < .001)$ condition. Post-hoc comparisons showed that completion time was significantly higher for STA (34.1 sec.) compared to STR (13.1 sec) and LIN (8 sec.) for the low densities (all $p < .01$). For the higher densities LIN had the fastest completion time (9.3 sec.) compared to the other designs (nearly 30s per repetition on average) ($p < .05$). There was also a significant effect of *glyph* across densities $(F_{3,21} = 4.7, p < .05)$. From low to high densities STR (from 13.1 sec. to 29.4 sec.) and CLO (from 24.1 sec. to 30.4 sec.) worsened ($p < .05$), whereas the mean for LIN stayed relatively stable (from 8 sec. to 9.3 sec.).

Confidence: There was an overall effect of *glyph* on *confidence* for both the low density ($\chi^2(3, N = 32) = 15.47, p < .01$) and the high density ($\chi^2(3, N = 32) = 16.28, p < .001$) condition. In the low density condition participants using STA (56.3%) reported a significantly lower confidence score with their answers than for all other designs (all $p < .01$). LIN (96.3%) received the highest confidence with significantly better ratings compared to CLO (80%, $p < .05$)

and STA (56%, $p < .001$). In the high density condition LIN (92.5%) is significantly better than the other designs ($p < .001$) and STA (56.3%) better than STR (48.1%) ($p < .05$). From low to high densities STR (from 80% to 48.1%, $p < .05$) and CLO (from 80% to 56.3%, $p < .001$) worsened.

Task 2: Temporal Location

Task 2 consisted of four training repetitions and four real trials for both densities. After the initial training trials we asked participants to detect a different temporal location for the peak value. Therefore, the first real trial was discarded due to the mental recalibration necessary by the participants (see Figure 3.13).

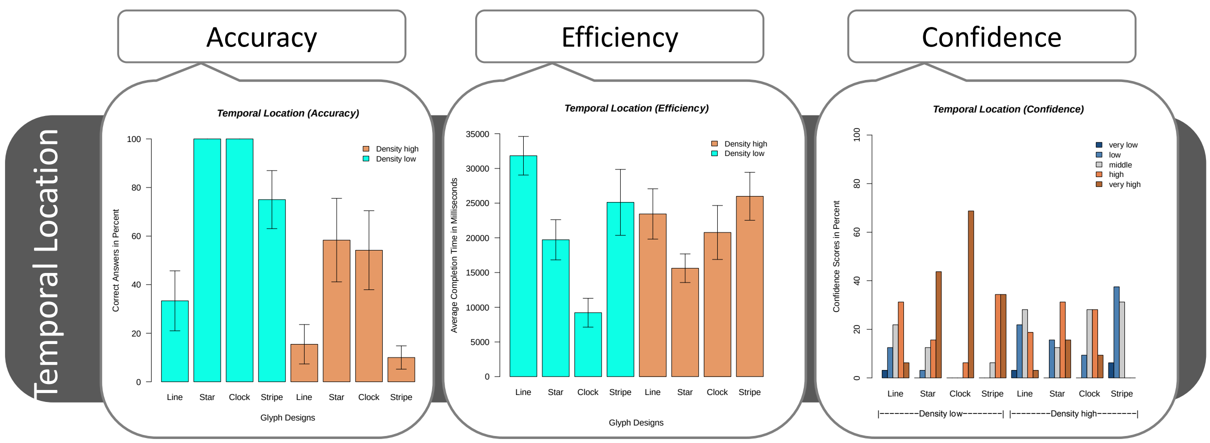


Figure 3.13: Bar charts with mean and standard deviation showing the results for the temporal location task and factor. The x-axis represents the different dependent variables.

Accuracy: There was a significant effect of *glyph* on *error* for both the low density ($\chi^2(3, N = 32) = 17, p < .001$) and the high density condition ($\chi^2(3, N = 32) = 7.81, p = .05$). In the low density condition pair-wise comparisons showed that errors in judgement were significantly worse for LIN (33.3%) compared to CLO (100%) and STA (100%) (both $p < 0.01$) and STR (75%) compared to CLO (100%) and STA (100%) (both $p < 0.001$). In the high density condition STA (58.3%) significantly outperformed LIN (15.5%) and STR (10%) (both $p < 0.05$). With an increasing data density, STA (from 100% to 58.3%), CLO (from 100% to 54.2%) and STR (from 75% to 10%) significantly lost accuracy with $p < .05$ in each case.

Efficiency: For the completion time there was only an overall effect of *glyph* on *time* in the low density ($F_{3,21} = 9.1, p < .001$) condition. Post-hoc comparisons showed that CLO (9.2 sec.) significantly outperformed LIN (31.8 sec.) ($p < .01$). There was another significant effect of *glyph* across densities ($F_{3,21} = 5.45, p < .01$). From low to high densities CLO (from 9.2 sec. to 20.8 sec.) deteriorated significantly ($p < .05$).

Confidence: There was an overall effect of *glyph* on *confidence* for both the low density ($\chi^2(3, N = 32) = 13.78, p < .01$) and the high density ($\chi^2(3, N = 32) = 12.12, p < .01$) condition. For the low density condition the results showed a clear picture for the confidence of the participants. The users were significantly more confident when using CLO (73.8%, $p < .05$), and had least confidence with LIN (50%, $p < .05$). For the high density condition the subjects were nearly equally confident using CLO (52.5%) or STA (54.4%), whereas LIN (44.4%, $p < 0.05$) and STR (35%, $p < 0.001$) are ranked worst. From low to high densities STA (from 65.6% to 54.4%, $p < .05$), CLO (from 73.8% to 52.5%, $p < .001$) and STR (from 65.6% to 35%, $p < .001$) worsened.

Task 3: Trend Detection

Task 3 consisted of four training repetitions and four real trials for both densities. For the analysis we discarded the training repetitions and focus only on the real trials (see Figure 3.14).

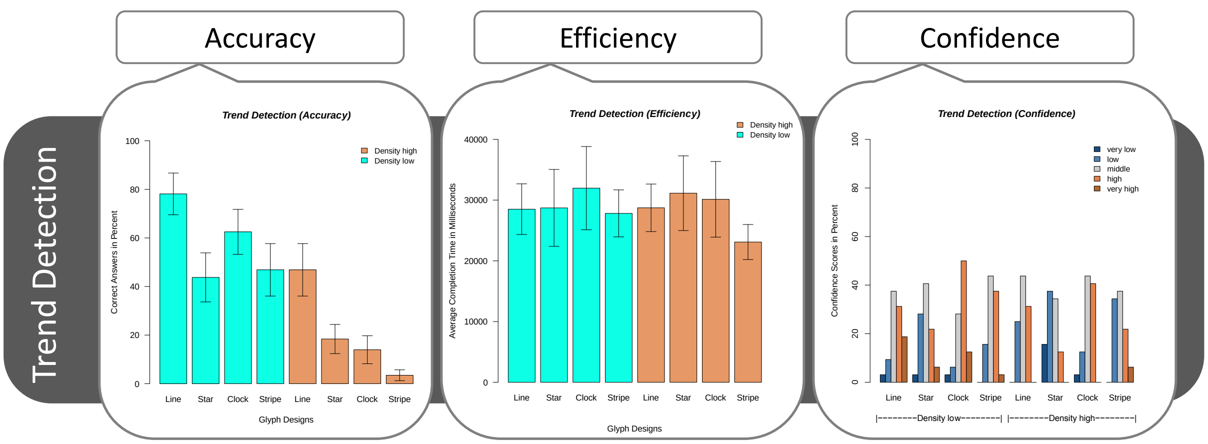


Figure 3.14: Bar charts with mean and standard deviation showing the results for the trend detection task and factor. The x-axis represents the different dependent variables.

Accuracy: There was a significant effect of *glyph* on *error* for both the low density ($\chi^2(3, N = 32) = 7.43, p = .05$) and the high density condition ($\chi^2(3, N = 32) = 8.9, p < .05$). In the low density condition pair-wise comparisons showed that errors in judgement were significantly better for LIN (78.1%) compared to STA (43.8%) and STR (46.9%) ($p < .05$). In the high density condition LIN (46.9%) significantly outperformed CLO (14%, $p < .05$) and STR (3.5%, $p < .01$). With an increasing data density, LIN (from 78.1% to 46.9%, $p < .05$), CLO (from 62.5% to 14%, $p < .01$) and STR (from 46.9% to 3.5%, $p < .05$) significantly lost accuracy (all $p < .05$).

Efficiency: For both densities no significant differences can be shown. The participants needed around 30 seconds on average. This was expected to be the maximal amount of time per repetition.

Confidence: There was an overall effect of *glyph* on *confidence* for both the low density ($\chi^2(3, N = 32) = 8.06, p < .05$) and the high density ($\chi^2(3, N = 32) = 7.6, p = .05$) condition. For the low density condition STA (60%) had lower ratings compared to CLO (72.5%, $p < 0.01$) and LIN (70.6%, $p < 0.05$). Same is true for the high density as well with STA (48.8%) being worse compared to CLO (64.4%, $p < 0.01$) and LIN (61.3%, $p < 0.05$). With an increased data density STA (from 60% to 48.8%, $p < 0.01$) and CLO (from 72.5% to 64.4%, $p < 0.01$) lost significantly confidence.

3.3.3 Discussion

In this section we combine both quantitative and qualitative data collected in our study to explain the varying performance of the different glyph designs according to our hypotheses. An overview of the quantitative results for each task is given in Table 3.1 where values highlighted in orange signify the best result compared to the other designs.

Task	Measure	**LIN**	**STA**	**CLO**	**STR**
Peak Detection	accuracy	96%	34%	69%	60%
(value comparison)	efficiency	8s	28.2s	18.6s	16.9s
Temp. Location	accuracy	24%	79%	77%	43%
(time comparison)	efficiency	27.6s	17.7s	15s	25.5s
Trend Detection	accuracy	63%	31%	39%	25%
	efficiency	26.2s	25.5s	27.1s	23.7s

Table 3.1: *Glyph performance for different tasks:* This table illustrates the percentage of correct answers (accuracy) and the average time needed (efficiency) for each of the tasks for both densities combined. The orange background signifies the best result compared to the other designs.

Peak Detection

In H1 we conjectured that LIN & STA would outperform CLO & STR due to their position and length encodings for value. The analysis of *error*, however, revealed that nearly no mistakes were made with LIN and only few with STR and that STA had the lowest accuracy followed by CLO. Apparently, the participants had more problems reading value with the circular layouts. This becomes obvious by comparing the most with the least accurate glyph design (i. e., LIN with STA). Both use the same value encoding but differ in the layout of the time dimension. This effect did not change across the two density conditions. STA and STR had a similarly high error rate across densities, CLO deteriorated only slightly, whereas LIN still performed best.

We can, thus, only partially confirm H1. We conclude that polar coordinates must have an effect on *error* for value judgements when the value is encoded with length. The same effect seems not to take place when the value is encoded with color. This can perhaps be explained by the different baselines of the designs. Comparing position/length in a radial design perhaps involves mental rotation to transfer the overall design to a comparable linear layout. This is not true for color encodings, since color does not need an identical baseline.

Another notable effect is the one between CLO and STR: while accuracy was not significantly different for low data density, CLO outperformed STR with high data density. This suggests that CLO is more resilient with respect to data density than STR. We believe this to be due to the fact that the slices in the circular design get more space near the circumference, wheras the slices in the stripe get too small, making the comparison more difficult. This only partially confirms H2: while STR is strongly affected by data density, LIN and CLO are either not affected by data density or affected to a smaller extent (decrease CLO: 18.8%; decrease STR: 68.7%).

The confidence score of the participants for this task was unambiguous with LIN having the highest ratings. In the final interview the participants had to rank the different glyph designs according to their subjective preference. LIN was the most preferred glyph type which matches the performance results of the quantitative analysis.

In the post-session interview, some participants argued that color was better than position/length for data value comparison especially when the distance between the values was very large. Of course, this depends on the color scale used, but seems plausible when the color value is entirely different, which may lead to a preattentive recognition effect. With smaller distances most of the participants commented that they would prefer the position/length encoding. When explaining their performance with STA (i. e. angle/length encoding), participants argued that they had problems comparing lengths with different orientation which further supports our hypothesis that mental rotations may be necessary for comparison and make values harder to compare in these glyphs. Especially in a small multiple setting this is an interesting finding and has to be further tested and considered when arranging glyphs.

Temporal Location

Our results partially support H3. In terms of accuracy both polar designs (CLO and STR) outperformed the linear designs when data density was low. To find an explanation for this result, we looked at the selections made by our participants and discovered an interesting side effect. The data sets corresponding to these wrongly answered questions were enriched with distractors very similar to the correct data instances by showing the same high value but at a different point in time. Participants seemed less likely to select such distractors when using the circular layouts for the time dimension. Participants were significantly more confident and made significantly less mistakes with the polar designs. The participants also reported to like the clock metaphor. Some suggested, however, to visualize only 12 hours at a time for a more intuitive encoding.

When data density was high we observed the same trend, even though only STA showed significant differences with respect to STR and LIN. The good performance of STA can be explained with the combination of the encodings. The length encoding for the data values makes it possible to easily spot the highest value even with lots of datapoints. With the color encodings, participants had problems spotting the peak value. The circular layout performed better than the linear one and worked for estimating the correct point in time.

We saw almost no significant differences between the designs for efficiency (only CLO was better than LIN with low data density and STA better than STR with high data density). Nonetheless, we observed that the overall trend for efficiency did not contradict the trend we found in terms of accuracy.

A significant decrease in performance between the two data densities can only be seen for accuracy. All designs had an increased error rate except for LIN. However, LIN's accuracy had been very low for the low density, thus, a significant decrease was nearly not possible. In terms of efficiency only CLO has a higher completion time, whereas, the other designs remained stable. These investigations partially support our hypothesis H4 where we had conjectured that the performance for detecting temporal positions would drop for an increased data density.

Trend Detection

In H5 we had conjectured that LIN & STA would be most effective for this task with the required value judgement as the bottleneck of the two required subtasks. As we expected, in terms of accuracy, the participants performed best using LIN independent from the data density. There was no significant difference between STA, CLO and STR on *error* and no significant results for *time* and, thus, H5 can only be partially confirmed. Independent from the designs, the participants needed around 30 seconds to complete the task.

With an increased data density the accuracy of LIN, CLO and STR dropped significantly. The completion time remained stable with no changes between the two density conditions. Our hypothesis H6 stating that the performance will not change by increasing the data density can, therefore, not be confirmed. Interestingly, participants commented that subjectively the task difficulty was not impacted by higher data density. The qualitative feedback almost matched the quantitative results. Nearly all participants reported to prefer LIN (i.e., position/length encoding) for solving the task.

3.3.4 Design Considerations

With the results gained from the analysis and discussions we derive the following design considerations.

- **To improve value comparison, use a linear layout or switch to color encoding for value:**
  As can be seen in the results for the first and third task, LIN and STA's performance are quite diverse although the value encoding is similar. The polar design has a strong effect on the perception of the position/length encoding.

- **For value encoding, position/length encodings should be preferred to a color encoding:**
  As can be seen in the results gained from Task 1 and 3 where a value comparison was necessary, LIN performs best. Even with an increased data density values could still be compared.

- **Triangular shapes rather may be better than rectangular shapes for color encoding:** The slices used in CLO for encoding single data values form a triangular shape because of the circular layout. As can be seen in the results for CLO compared to STR, having more space near the circumference increased participants' performance. Designers could experiment with adding triangular shapes in a linear encoding.

- **Color encodings for higher data densities should be used with caution:** The results from task 1 and 3 illustrate, that the performance of the color encoded designs (CLO and STR) depends on the data density. Having a higher data density leads to a decreased performance.

- **Circular layouts rather than linear ones should be preferred for detecting temporal locations:** Polar designs are better for detecting specific points in time. This guideline results from the analysis of the second task. Participants performed significantly better using CLO and STA compared to LIN and STR. The clock metaphor increases users' chronological orientation.

- **For time-dependent tasks, sufficient space should be assigned to the designs:** Whereas, for solely value comparison tasks the performance of the best design (LIN) is not affected, the accuracy for tasks including temporal information decreases. This is independent from the combination of visual variables used as can be seen for task 2 (STA and CLO) and 3 (LIN). The designs performing best for these tasks are encoded differently but still show the same behavior.

3.3.5 Limitations

As stated at the beginning, we were inspired by time series data for a daily monitoring task. Especially CLO and STA with their 24 hour clock metaphor profit from this data arrangement. The performance may change with different lengths of time series.

The same is true for the aspect ratio and the size of the single glyphs. The aspect ratio was chosen in order not to greatly disadvantage the circular designs in terms of display space used. However, especially STR would profit from an aspect ratio with more horizontal space. With varying sizes of glyphs, the performance of the designs could change. In our setting we used the minimal space possible to be able to assign one pixel to one data value for the higher data density.

3.3.6 Conclusion

The goal of this experiment was to compare the performance of the *clock glyph* against well-established alternative data glyph designs. Therefore, we quantitatively measured accuracy and efficiency, and qualitatively surveyed user confidence and preferences for four glyph types based on three tasks important to our domain experts: peak detection, peak detection at a certain point in time, and trend detection. The results show that depending on tasks and data density, the chosen glyphs performed differently. We show that the line glyph is generally a good choice for peak and trend detection tasks but that radial encodings of time (star glyph and *clock glyph*) were more effective when one had to find a particular temporal location. Participants' subjective preferences support these findings and underline the fact that the clock metaphor helped in detecting specific temporal locations. Thus, our study shows that both accuracy and efficiency of tasks such as ours can be boosted when carefully choosing the most appropriate design.

3.4 Summary

In this chapter 3, the literature about different data glyph designs has been carefully reviewed for time-series data in various settings. Structuring the glyphs according to the basic visualization techniques they were combined with illustrated the great flexibility in positioning them on the screen.

Based on this related work I motivated the necessity for a metaphoric *clock glyph* design to convey temporal information in an easy to understand way. This design was implemented and applied to real world data in the network security domain. Three different prototypes were introduced (i. e., *ClockView*, *ClockMap*, and *Vistracer*) using *clock glyphs* to visualize complex temporal data structures. Each prototype was evaluated with a use case scenario to show the applicability of the design especially in combination with different visualization techniques.

To generalize the findings and get more concrete information about the performance of the *clock glyph* a controlled user study was conducted. Since there was a lack of quantitative experiments for data glyph designs using color saturation to encode data values this evaluation additionally closed some previously identified research gaps, which were revealed in chapter 2. Based on the results of this experiment I further confirmed the usefulness of metaphoric designs for information visualization.

Chapter 4

Data Glyph Designs for Multi-Dimensional Data

Parts of this chapter appear in the following publications:

- Johannes Fuchs, Roman Rädle, Dominik Sacha, Fabian Fischer, and Andreas Stoffel. Collaborative Data Analysis with Smart Tangible Devices. In *IS&T/SPIE Electronic Imaging*, pages 90170C–90170C. International Society for Optics and Photonics, 2013[1]

- Johannes Fuchs, Petra Isenberg, Anastasia Bezerianos, Fabian Fischer, and Enrico Bertini. The Influence of Contour on Similarity Perception of Star Glyphs. *IEEE TVCG*, 20(12):2251–2260, Dec 2014[2]

- Johannes Fuchs, Dominik Jäckle, Niklas Weiler, and Tobias Schreck. Leaf Glyph - Visualizing Multi-dimensional Data with Environmental Cues. In *Digital Library Scitepress*, pages 195–206, March 2015[3]

- Johannes Fuchs, Dominik Jäckle, Niklas Weiler, and Tobias Schreck. *Leaf Glyphs: Story Telling and Data Analysis Using Environmental Data Glyph Metaphors*, pages 123–143. Springer International Publishing, Cham, 2016[4]

[1]The responsibilities for this joint publication were divided as follows: I spearheaded the writing of the paper, Roman Rädle, Fabian Fischer and I did the programming and conducted the user study, Dominik Sacha and Andreas Stoffel formalized the problem and did the proofreading.

[2]The responsibilities for this joint publication were divided as follows: Petra Isenberg and I designed the user study. Fabian Fischer and I conducted the experiment. I was responsible for analyzing the results and writing the paper. Petra Isenberg, Anastasia Bezerianos and Enrico Bertini gave advice and did the proofreading.

[3]The responsibilities for this joint publication were divided as follows: I spearheaded the writing of the paper, Niklas Weiler did the programming, Dominik Jäckle was involved in the writing, Tobias Schreck and I did the proofreading and gave advice.

[4]The responsibilities for this joint publication were divided as follows: I spearheaded the writing of the paper, Niklas Weiler and Dominik Jäckle did the programming, Tobias Schreck and I did the proofreading and gave advice.

Multi-dimensional data can be considered as an n x m matrix with n being the different data points and m the corresponding attribute dimensions. In contrast to time-series data the relationship between the attributes may be dependent or independent from each other. This is reflected in the analysis tasks. Since attributes need not necessarily be related trend detection tasks across dimensions like in an intra-record comparison are not this likely. As can be seen in the systematic review of user studies in chapter 2 participants performing a trend detection task with multi-dimensional data compared single dimensions across different entities (inter-record comparison) and not within a single design (intra-record comparison) like with time-series data.

The design of the data glyph is, therefore, more flexible. Restrictions like the comparability of the different attribute dimensions like in time-series data are not necessarily given. The possibility for mapping data values to visual variables is, therefore, much more flexible. Based on Ward's categorization [192] data glyph designs may comprise many-to-one, one-to-one, and one-to-many mappings, which results in a much bigger design space compared to temporal data. Of course, the data glyph designs can also be combined with different visualizations (e.g., geographic maps, node-link diagrams etc.), as well.

In the following section 4.1 I will review the literature about data glyphs for multi-dimensional data. Motivated from the related work section I will introduce a new data glyph design namely the *leaf glyph* making use of environmental cues to visualize multi-dimensional data. This design will be evaluated in a use case analyzing the forest fire data set form the UCI machine learning repository [44]. The dataset was carefully chosen to show the usefulness of the context related design and, therefore, the benefit of a metaphoric representation. After introducing this new data glyph I will focus on the well-known star glyph design with all its variations used in literature. Since only little guidance exists, which star glyph variation works best for similarity search tasks this research gap will be closed by conducting a controlled user study and additionally trying to further improve on the design.

4.1 Related Work

Since data glyph designs for multi-dimensional data can be created in an entire flexible way, I will structure this section according to Ward's classification of data glyphs [192]. In his categorization he distinguishes between three different ways a data value can be mapped to a glyph representation.

4.1.1 Many-To-One Mapping

All data dimensions and their respective values are mapped to a common visual variable. Therefore, these designs can be systematically created by choosing the most effective visual variable for a certain task. Additional guidance is given by Cleveland et al. with a ranking of visual variables [43].

Position/Size Encoding

A well-known example are linear profiles like small bar charts [52]. Each data point is represented by one data glyph and the different bars correspond to one attribute dimension. The height of the bars reflect the respective data value. A similar encoding is used in dot plot glyphs[205] exchanging the bars with simple dots. The vertical position of each dot communicates the underlying data value. Depending on the number of dimensions these designs profit from a more rectangular like aspect ratio to have enough space visualizing the single dimensions. These linear profiles are also used in combination with different visualization techniques.

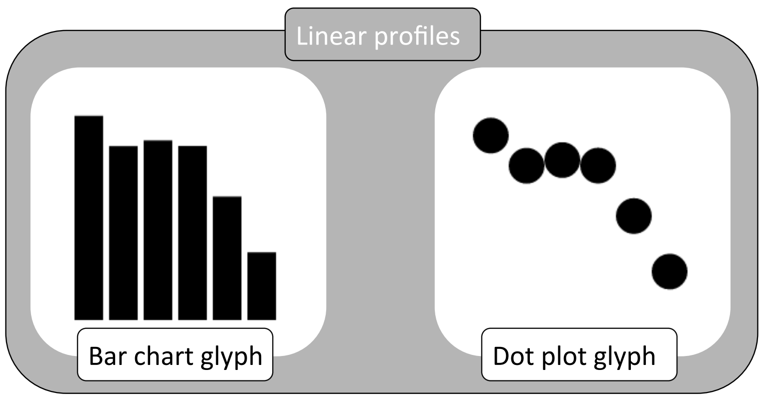


Figure 4.1: *Linear profiles:* Whereas bar chart glyphs use a length encoding to represent the underlying data, dot plot glyphs use the visual variable position. Both designs make use of a similar baseline to facilitate data comparison.

Bernard et al. enriches node-link diagrams with bar chart glyphs to visualize cancer data [15]. The patients are clustered according to their attribute values. Similar data points are connected with an edge and positioned close to each other. Additionally color is used to make the cluster membership more obvious.

Similar bar chart glyphs are also used in scatterplots to represent the cars dataset from UCI machine learning repository [124]. Each data glyph represents one car. Two attributes of the data are mapped to the x-and y-axis of the scatterplot to position the data points. The remaining dimensions are represented in the data glyph as a colored bar chart. Because of many similar car characteristics a lot of overplotting may occur in this visualization due to the data-driven layout.

For displaying a file system with additional file attributes McDonnel and Elmqvist embed bar chart glyphs in treemaps [124]. Additional shaders help to better perceive the single hierarchy levels. Color saturation is used to distinguish the different attribute dimensions like file size, or last time modified etc.

Ward and Lipchak focused on radial layouts (e.g., spirals) to position linear and circular profile glyphs [194]. These layouts are especially useful for communicating temporal periodic information. A possible example is the comparison of different stock prices during the last years and whether there are periodic specifics. Therefore, the glyphs are arranged in a spiral layout. Each cycle in this spiral corresponds to one year. The year is further divided into months using

12 equal distant anchor points for each cycle. On each anchor point a glyph is drawn showing different stock prices for this specific point in time. Temporal patterns can be easily perceived by either scanning along the spiral, or along the anchor points using the same angle.

Instead of comparing different lengths alternative glyph designs use the visual variable area to encode the data value. Due to Cleveland and McGill's ranking of visual variables, glyph designs using length are more accurate compared to area [43]. However, area communicates smaller changes between data values more effectively [195]. Fischer et al. [64] make use of such an area encoding to show anomalous behaviour of BGP routes over time. Each rectangle correspondence to one timestamp and incorporates four additional rectangles (one for each anomaly). The size of each inner rectangle encodes the proportion of anomalous behavior. Using the visual variable area helps in this case to perceive even slight differences between the single anomaly groups. Volume encodings on the other hand are considered poor choices for communicating data values [57, 195]. However, some glyph designs make use of them anyway [83].

Circular profiles also use a position/length encoding for visualizing data values and are, therefore, quite similar to linear profiles. These designs usually have a quadratic aspect ratio to make sure to introduce no bias towards certain dimensions. A well-known representative is the star glyph [168] with its variations whisker and fan plots [150, 195], and sensitivity star glyphs [32]. Star glyphs use data lines radiating from the center to display the different dimensions. The length of the data line corresponds to the underlying data value. Finally, the end points of the lines are connected to create a "star-like" shape. Whisker and fan plots use the very same encoding, however, the endpoints of the data lines are not connected. The sensitivity star glyph only shows the contour line of the star glyph without the data lines. As part of my thesis, I will evaluate these variations in a controlled user study for similarity search tasks.

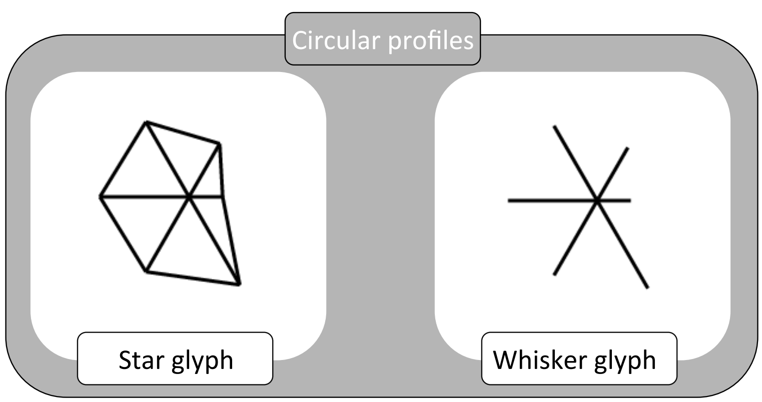


Figure 4.2: *Circular profiles:* Star glyphs and whisker glyphs use a similar encoding to represent multi-dimensional data. The only difference is the surrounding contour line.

Star glyphs and all its variations are used in various applications. Elmqvist et al., for example, connect different star glyphs with directed edges to visualize data flows [60]. Dynamic queries can be incrementally refined by adding more filters to the node-link diagram. The star glyphs are then adjusted according to the filters set.

Friendly used colored star glyphs to represent different characteristics for regions in France [69]. Each region was represented by one data glyph, which were positioned on top of geographic map. Areas with e.g., high crime rate can be easily detected by searching for star glyphs with a peak value for this attribute. The color helped to distinguish between different regions within France.

Since clutter is a major drawback of data glyphs Yang et al. and Peng et al. introduced automatic algorithms to reduce these effects [148, 204]. Whereas Peng et al. focuses on dimension reordering techniques with a grid based layout [148], Yang et al. used additional dimension spacing and filtering techniques to position star glyphs in scatterplot matrices [204]. These filtering steps help to reduce clutter for large numbers of dimensions. Additionally, the user can also be involved in this optimization by steering certain parameter.

Besides these applications making use of star glyphs and its variations there are other radial designs making use of a position/length encoding, as well. Metroglyphs [8] for example look similar to whisker and fan plots and are used in scatterplots. Data lines are connected to a circle with different angles and lengths. Again, the length of the data lines is used to encode the respective data value. The different orientations help to better distinguish the single dimensions.

Clustnails [181] use a similar visual encoding as whisker glyphs. Each cluster is represented by one data glyph. The data lines represent different sub-clusters. The length of the data lines encodes the importance of each sub-cluster for the whole cluster according to a certain measure. The different clustnails are arranged in a grid to facilitate the visual comparison between several clusters.

Color Saturation Encoding

In comparison to the visual variables position and size, color saturation is considered less accurate [43]. Of course, color cannot convey the data as accurate as a position/length encoding [73], however, for certain tasks like spotting outliers the color saturation encoding is a reasonable choice. Therefore, several glyph designs making use of color saturation to represent the data value do exist.

The *clock glyph* introduced in section 3.2 and evaluated in section 3.3 can also be used with multi-dimensional data like in the experiment conducted by Nelson and Gilmartin [139]. Instead of representing time dimensions the single slices are used to communicate different attribute dimensions. Still, color saturation is used to communicate the underlying data value. A slight variation of this radial design are color icons [112] making use of a square instead of a circle to arrange the slices. The dimensions are, therefore, represented as triangles positioned in a circular fashion.

Linear designs, on the other hand, make use of small rectangles or stripes with a fill color to represent data values. Several different variations and applications have been proposed in literature. One example are calendar view like glyph designs. Each data point is represented by one square or rectangle, which contains several smaller squares. The inner squares correspond to the attribute dimensions of the respective data point. Their arrangement is done linearly and they are colored according to the underlying data value. Beddow used such glyph designs to communicate changes across thirteen parameters of magnetosphere and solar wind data over time [13]. Each glyph represents one point in time and the inner rectangles refer to the different parameters. Single glyphs can be easily compared to get an overall idea of the temporal changes. Additionally, the analyst can focus on only single inner rectangles to compare the development of certain parameters over time.

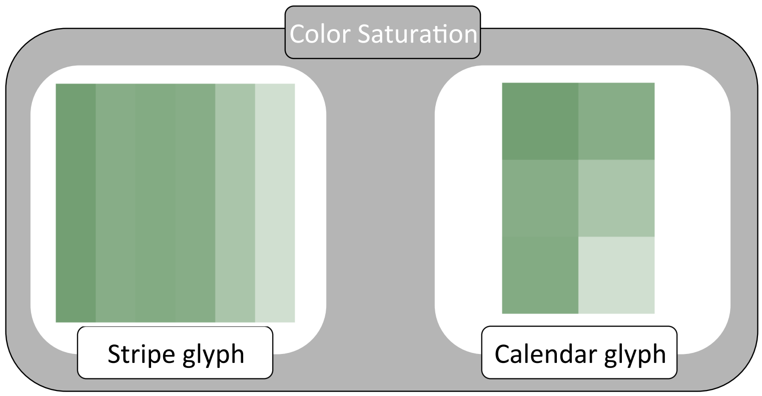


Figure 4.3: *Color saturation:* The stripe glyph and the calendar glyph both use a linear layout in combination with a color saturation encoding for the data value.

A similar glyph design is used in the work of Abdul-Rahman et al. to better understand and explore the tongue position in the mouth while speaking [1]. The authors divided the month in nine regions using a 3 x 3 grid. The single grid cells correspond to a certain location in the month. Colors are used to highlight regions within this grid where the tongue was located during the articulation. Such a glyph is created for each vowel in a text. A comparison of different texts based on the tongue position is, therefore, easily possible by scanning along the different colored glyphs.

Orientation Encoding

Sticky figures [150] are prominent representatives of multi-dimensional data glyphs, where different data lines are representing the attribute dimensions. These data lines are then connected to a common stem. To communicate the data value for each dimension the visual variable orientation is used, which is considered not as accurate as position or size in communicating exact data values [43]. However, in overview visualizations the single designs are perceived as a whole approximating the underlying data points. The analyst does not necessarily need to check single data lines but compare entire shapes.

Gestaltlines [22] consist of single data lines, each encoding one attribute dimension. The slope of the lines conveys the underlying data value. Due to their linear layout, gestaltlines are close related to sparklines [185] and can also be embedded in text sections but need not necessarily communicate time-series data.

Pie chart glyphs are famous examples of radial orientation encodings [6, 194]. Pearlman and Rheingans use a slight variation of these pie chart glyphs to visualize network traffic. Therefore, they introduced the compound glyph, which is embedded in a node-link diagram [147]. The

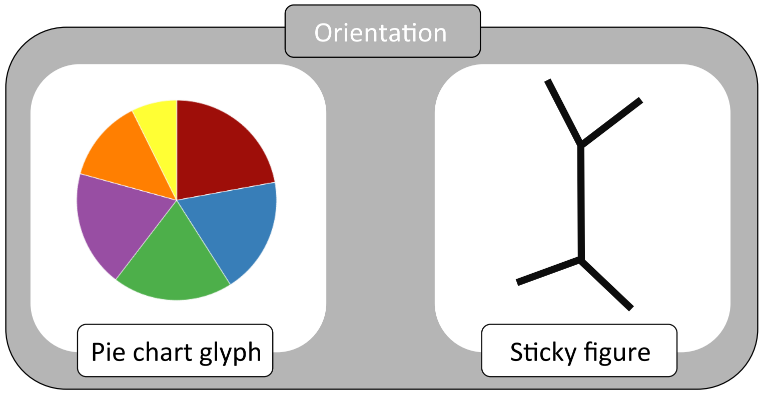


Figure 4.4: *Orientation:* Pie chart glyphs and sticky figures are two well-known representatives using an orientation encoding to visualize data values.

glyph uses the angle of different slices to encode the amount of traffic for the respective services. Additionally, multiple compound glyphs can be stacked to visualize different points in time. The graph layout helps to not only communicate the traffic of the different services but also to show, which network devices are communicating with each other.

4.1.2 One-To-One Mapping

The one-to-one mapping category is the most flexible way on how to map attribute dimensions to visual variables. There are basically no restrictions on how to do the mapping except that redundancy is not allowed. This means that each dimension can be represented by different visual variables creating a design space, which is nearly endless. Although, the mapping can be done arbitrarily some data glyph designs have received more research attention than others.

Probably, the most well-known representation in this category are Chernoff faces [35]. As already shown in chapter 2 this glyph design has been extensively studied in literature. The whole design space is restricted to an abstract face representation. Each data value is mapped to a specific face characteristic, like the height of the nose, the size of the ears, or the angle of the eyebrows etc. Over the years different variations have been introduced. The most famous examples are Flury-Riedwyl faces [65], which look more realistic and Kabulov faces [96], which group the data into three or four different classes to improve the visual mapping. The face symbol has been used in many different application areas like cartography [130, 190], multi-dimensional comparison [25], or multi-variate trend development [25].

Another well-researched design space are data glyph designs for flow visualizations. Weather vanes [208] are famous examples to visualize the different characteristics of wind data for geographic regions. Wind direction for example is mapped to the orientation of the data glyph, whereas the wind magnitude is represented by the amount of small whiskers attached to the vane. This glyph design has been established in literature and was used several times in the "*Monthly Weather Review*" in 2008 [165]. However, Ware argues that the small whiskers interfere with the perception of wind direction making it difficult to read the wind speed [195].

Other data glyph designs have been proposed to visualize multi-dimensional data in tensor fields. Most of them use orientation, size and color to map multiple attribute dimensions to a specific region [34, 160]. There are several three dimensional designs, which further enhance the spatial perception by adding a third dimension [55, 56, 87, 99, 106, 136, 166, 167, 169, 209].

Other more exotic designs like bugs [40] (changing the shape, length or color of wings, tails and spikes), or hedgehogs [101] (manipulating the spikes by changing the orientation, thickness and taper) generally encode multi-dimensional data and can be used in various applications and context. A metaphoric glyph design was developed to understand the influence of climate change on the cultivation of maize. Nocke et al. designed a mosaic data glyph shaped like a corncob and positioned the design on top of a geographic map [142]. By adjusting the visual appearance of the glyph analysts can easily distinguish drought and fertile regions. A use case scenario has shown that analyst were able to easily understand the underlying data because of the metaphoric design.

However, there are also quite specific glyph designs, which are likely to be used in certain domains. Maguire et al. introduced a taxonomy for data glyph designs to communicate workflows [121, 122]. This taxonomy is based on the connection of the hierarchy of concept categorization and the ranking of the visual channels with additional domain specific metaphors. In the area of health monitoring the VIE-VISU glyph [88] helps to represent 15 health-related patient parameters by changing the width, height, or color of different connected shapes. By plotting multiple glyphs next to each other a temporal comparison of several parameters can be achieved. In sport analytics MatchPad glyphs [41] can be used to visualize football events and additional meta information by adding circles and rectangles to a square, which contains an icon for each related football event. As a result the analyst need not necessarily watch the entire game to extract useful information about the match but can use the MatchPad software to gain insights. The respective data glyphs are, therefore, positioned along a timeline to see how the match evolves over time.

The major drawback of these kinds of glyph representations is that they are often sensitive to the order by which the data dimensions are mapped to visual variables. Changing the order could significantly influence the final glyph representation and its visual perception by users since some visual variables are more dominant than others. Additionally, measuring differences between single dimension values within a data point is typically a difficult task, as the analyst has to compare different kinds of visual variables with each other (e.g., compare length with color saturation or orientation, etc.)

4.1.3 One-To-Many Mapping

Unlike the other two categories, the data attributes are mapped redundantly. Each dimension is represented by a combination of at least two visual variables. This redundant mapping can be useful to strengthen the perception of individual dimensions. Theoretically each data glyph design can be changed to represent a one-to-many mapping.

Data glyph designs making use of a position or size encoding can be modified using color saturation, as well, like the color encoded star glyph [103]. Of course, the other way round is also possible. The *clock glyph* can make use of an additional length encoding for the single colored slices to encode the underlying data values more accurately. Such a design would be highly related to the Nightingale chart [141]. The Nightingale glyph is, for example, used in hierarchical node-link diagrams to convey the structure of file systems [124]. The position of the glyphs is used to understand the relationships between the different files. Additionally, the data glyph conveys information about file characteristics like time since created etc. Another application area is the comparison of different countries and their characteristics. The OECD Better Life index [143] shows flower like glyphs to visualize different countries. The petals are used to represent the attribute dimensions making use of a size and color saturation encoding. Big colorful flowers illustrate countries having high values throughout all dimensions.

Summary

In this related work section I have only focused on data glyph designs for multi-dimensional data and grouped the glyphs according to Ward's visual mapping categorization [192]. It is interesting to note that data glyph designs for one-to-many mappings have only received little research attention. This finding is also supported in chapter 2 based on the small number of quantitative experiments conducted for one-to-many mappings.

In comparison to glyph designs for time-series data metaphors were used more often to communicate the underlying data [35, 142, 170, 208]. As already indicated in chapter 2 studies have shown that glyph designs based on metaphors help to better understand the underlying data [65, 73, 91, 170]. Since there is a lack of metaphoric glyph designs to communicate environmental data I will introduce the *leaf glyph* in the following section 4.2 as a new data glyph design and show its applicability in a use case based on the forest fire data set from the UCI machine learning repository [44].

Another research gap will be tackled in section 4.3. As the systematic review of user studies on data glyph designs has revealed (chapter 2) some more research on the performance of different star glyph variations needs to be conducted. Till now no guidance exists whether the surrounding contour line of star glyphs is really beneficial or not. Therefore, I will introduce three quantitative user studies investigating changes in performance for similarity search tasks when removing or adding the contour line.

4.2 Leaf Glyph - A Data Glyph Design to Visualize Multi-Dimensional Data with Environmental Cues

The *leaf glyph* is a novel data glyph design for visualizing multi-dimensional data based on an environmental metaphor. The rationale of introducing such a new design is threefold. First, as can be seen in chapter 2 studies suggest that metaphors help to better understand the represented data. Especially for multi-dimensional data many different metaphoric designs have been suggested (see section 4.1 for more information) and evaluated in different use case scenarios. Such designs seem to be highly suitable in communicating context specific information.

Second, the design space is large, giving ample opportunities for the visualization expert to map data variables to visual variables. As will be discussed, the variable space amounts to more than 20 different visual variables that can be controlled. While we[5] have not formally evaluated the effectiveness of these variables or their combinations, we presume this is a large design space from which appropriate effective selections can be found.

And third, there is reason to believe that the human visual sense, due to long evolutionary processes, is highly trained in recognizing, distinguishing and comparing natural forms. Shapes like trees in a forest, single flowers in a flower-bed, or leaves at shrubs can be easily discriminated by humans. These visual recognition processes typically work well even in low illumination conditions, or in presence of partial occlusion of natural objects. By background knowledge and experience, humans are able to efficiently recognize natural shapes, also often in cases where only parts of the shape or their boundary are visible.

A subset of the designs studied in information visualization to date make use of the aforementioned benefits and are inspired by nature. For example, tree structures have inspired hierarchical node-link diagrams. Stefaner, for example, uses an abstract tree layout to show the editing history of Wikipedia entries represented as single branches [176]. The branches grow to the right whenever people decided to delete an article or to the left in the other case. The resulting tree nicely summarizes 100 articles with the longest discussion whether to keep them or not. A 3D application is the botanical tree [102], which uses a 3D tree layout to represent hierarchical information. The single nodes are represented as fruits. The authors argue that people can more easily identify single nodes in this visualization compared to a more abstract representation because they are used to detect fruits or leaves on shrubs or trees. A 2D visualization using a botanical tree metaphor are so-called ContactTrees [159], which show relationships in data, e.g., contacts between persons. The branches consist of single lines representing an attribute in the data, e.g., a longer line refers to an older tie between people. Finally, fruits or leaves are added to the tree according to some data property, e.g., the kind of relation between people (friends, coworkers etc.). However, the fruits and leaves are highly abstract representations (mainly colored dots) and their shape does not change according to some data characteristics.

Also some environmental glyph designs have been introduced in literature. The OECD's Better Life Index visualization [143], for example, systematically changes the appearance of flower glyphs to represent data. Stefaner uses these environmental cues to visualize multi-dimensional data about country characteristics. Each country is represented by one flower. The petals encode the different economic branches with varying sizes and lengths for the corresponding values. The flowers are arranged according to their weighted rank across all dimensions. People can change the layout by changing the weights of the dimensions or simply focusing on just one dimension. Müller invented a leaf glyph to visualize poems in a more artistic way [133]. The branches of the tree are invisible just dealing as an anchor point to arrange the glyphs. Each word in the poem is represented with a leaf glyph and attached along the tree structure. The work is not eligible of representing the text data accurately but tries to illustrate a creative unique picture or fingerprint of the underlying poem.

As can be seen, literature also suggest that environmental designs work well in representing data. However, we did not address research in the area of computer graphics, since this work

[5] In section 4.2 and all corresponding subsections the term "we" comprises Niklas Weiler, Dominik Jäckle, Tobias Schreck and me

mainly focuses on photo-realistic representation of the environment. We refer the interested reader to a summary work about this topic by Deussen and Lintermann [48].

In the following subsection 4.2.1, we investigate the design space for leaf shapes as natural metaphors for data glyphs encoding multi-dimensional data. From observing leaves in nature, it is clear that there is a large variability in the different types and forms of leaves that exist. Overall leaf shape, shape boundary, and shape interior all comprise several visual parameters that can in principle, be used to map data to generate glyphs.

4.2.1 Design Space for Environmental Data Glyphs

According to Biological literature, leaves may be categorized by their function or usage in the environment [12]. For our purposes, we divide leaves according to their shape (or morphology). The overall appearance of a leaf consists of the combination of (1) the overall shape type, (2) the boundary details, and (3) the leaf venation. We consider these three aspects as the main dimensions for controlling the leaf glyph by mapping data. As a result we come up with a design space structured along the overall leaf shape.

Leaf Shape Design Space

Following Palmer who pointed out: "Shape allows a perceiver to predict more facts about an object than any other property" [146], this visual variable should be used for the most important data dimension. In the environment, there exists a nearly endless amount of different leaf shapes since each leaf is unique. However, it is possible to distinguish leaves according to their overall shape [48]. A first categorization can be done between conifer and deciduous leaves.

Conifer leaves can be found for example at fir or pine trees and have a thin long needle-like shape. Therefore, they do not offer much space for a venation pattern, which we want to use later for mapping additional attributes (e.g., Acicular leaves). Since the differences in shape are quite small for the different kinds of this group and the provided area is limited due to the distorted aspect ratio, we do not consider them in our design space.

Deciduous leaves cover a large group of different shapes and can again be further divided into four sub-categories [48].

Pinnate and *palmate* compound leaves are shapes, which consist of several smaller leaflets attached to a shared branch (e.g., Alternate, or Odd and Even Pinnate leaves etc.). In order to avoid any misinterpretation between single leaflets at a branch and individual leaves, we discard this group from our final design space. However, these kinds of leaves seem an appropriate representation to visually summarize multiple data points where one leaflet corresponds to a single leaf.

Lance-like leaves have a parallel venation and are thin and long, similar to conifer leaves. Therefore, it is difficult to distinguish different kinds of these leaves since the differences in the overall shape are limited. Like the conifer leaves, we do not keep them in our design space because of the limited area to map a venation pattern, and because of possible confusion of different lance-like shapes.

Leaves with *net veins* or *reticulate* venation patterns encompass the largest group of deciduous leaves with a big diversity in shape. We restrict ourselves to the most common leaf shapes for this category to avoid misinterpretation of intermediate structures, which could not clearly be distinguished. Additionally, we focus on leaves with a big surface to show venation patterns and small stems to save space. Leaves similar to Flabellate, Unifoliate, etc. will, therefore, not be considered.

The most important requirement for shapes in visualizations is that they should be easily distinguishable. Therefore, our final design space covers wave-like (e.g., Pinnatisect), circular (e.g., Orbicular), triangular (e.g., Deltoid), heart-like (e.g., Cordate, Deltoid etc.), arrow-like (e.g., Hastate, Spear-shaped etc.), two variations of tear-drop like (e.g., Acuminate, Cuneate etc.), elliptic (e.g., Ovate, Obtuse, Obcurdate etc.), and star-like (e.g., Pedate, etc.) shapes. Figure 4.5 illustrates the nine different leaf shape categories covered by our design space. In subsection 4.2.3 we will introduce a heuristic to map data points to leaf shapes, based on the idea of representing outlying points by the more jagged leaf shapes; conversely, non-outlying points will be represented by the more regular or smooth leaf shapes.

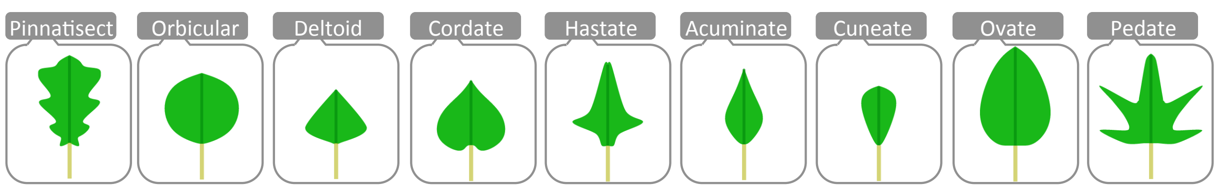


Figure 4.5: *Leaf shapes*: Selected from our overall design space, these are the shapes used in our final glyph design. From left to right: Wave-like shape, circular, triangular, heart-like, arrow-like, tear drop up, tear drop down, elliptic, and star-like shape.

We take these categories as a starting point and further extend them by mapping additional attribute dimensions to the width and the height of the glyph, scaling the overall shape. Therefore, similar shapes according to a certain data characteristic can look different because of the varying aspect ratio. However, the individual shape categories can still be distinguished (Figure 4.6). Because of this decision, we will deviate from the precise environmental reference, where leaves typically show a homogeneous aspect ratio. However, we thereby are able to encode additional data dimensions. Note that we do not want to represent leaves as accurate as possible (or even photo realistic), but use their expressiveness to visualize data.

Leaf Boundary Design Space

Basically, the boundary (or margin) of a leaf can be described as either serrated or unserrated. *Unserrated* boundaries have a smooth contour adapting to the overall leaf shape. *Serrated* boundaries are toothed with slight variations depending on the size of teeth, their arrangement along the boundary, and their frequency. Of course, there are more detailed differences and variations in nature. However, especially in overview visualizations (the major domain of data glyphs), distinguishing between small variations of the contour line of a leaf shape is nearly impossible.

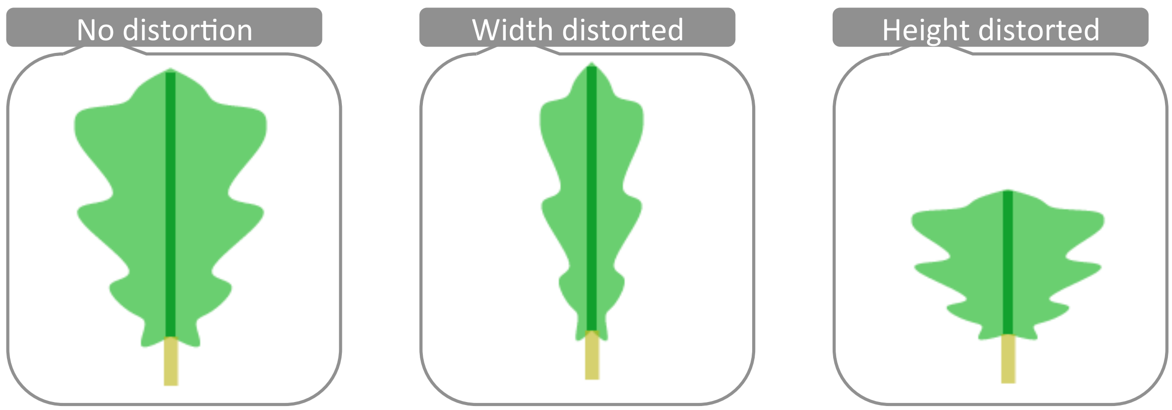


Figure 4.6: *Leaf scaling*: The Lobate leaf shape is scaled using either the width (middle), or the height (right) of the glyph. Even after scaling, the glyph can still be recognized as a wave-like leaf, although the precise environmental reference to the Lobate leaf is reduced.

We therefore focus on just the two main boundary categories of teethed or smooth (serrated or unserrated). For mapping data values to the leaf boundary, we distinguish between a smooth and a toothed contour line and vary the width, height, and frequency of the teeth according to the underlying data value (Figure 4.7).

Leaf Venation Design Space

We also control the leaf venation pattern as to map additional data variables to the glyph. Several main leaf venation patterns exist, which differ in their overall structure within the leaf. A rough distinction can be made between single, not intersecting (e.g., Parallel), paired (e.g., Pinnate), or net-like (e.g., Reticulate) veins. The venation is perceived as an additional texture for the glyph and further increases the glyph expressiveness. Since it is hard to find a natural order within this texture, we propose to use the venation type for visualizing qualitative (or categorical) data, similar than the overal leaf shapes discussed in Section 4.2.1. Within a given venation type, we may also encode numeric data. This works as follows. Generally, the leaf is split in the middle by a main vein, with small veins growing from there in a given direction (angle). For mapping numerical data, we may either control this *angle of the veins* branching out from the main vein. An alternative is to control the *number of veins* shown on the surface Figure 4.8. As a result, we come up with a venation texture able of encoding categorical and numerical data.

Summary

Besides modifying the leaf shape given by morphology, boundary and venation, further dimensions can be assigned to the color hue or saturation of the glyph. Of course, the designer has to

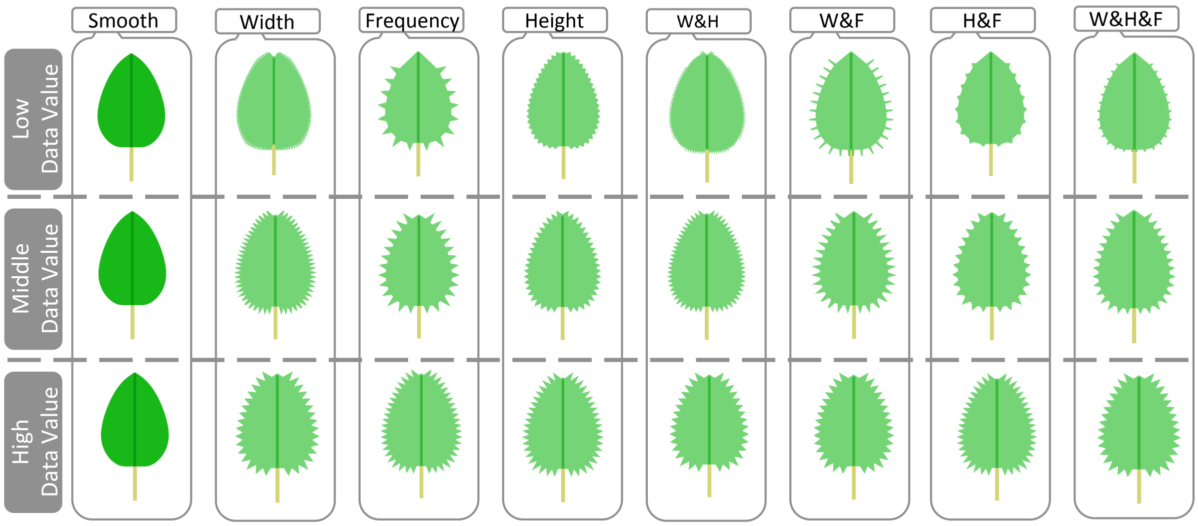


Figure 4.7: *Leaf boundary*: Modifying the boundary in our design is realized by changing the frequency, the height, or the width of the boundary serration (teeths). Combinations of these three variables are possible and increase the expressiveness of the glyph. The figure illustrates all possible combinations for low, middle, and high data values for an elliptically shaped leaf glyph.

pay attention to the contrast between the venation texture and the background color. Additionally, orientation of the glyph in the display can be used to encode further numeric information. We draw a short stem to each leaf shape, showing its orientation. Finally, it is also possible to modify the stem's width or height as well.

This represents a comprehensive design space for mapping data to leaf glyphs, controlled by 12 categorical and 14 numeric parameters, summing up to 26 variables altogether (see Table 4.1 for an overview of all variables.) We propose this design space as a toolbox from which the designer may select visual variables as appropriate. The number of 26 parameters is considered more a theoretical upper limit of data variables that we can show. We expect not all visual parameters in this design space to be of the same expressiveness; but some variables may be more effective than others, and may not all be orthogonal to each other. Careful choice should be done in selected and prioritizing the variables. An option is of course always, to redundantly code data variables to different glyph variables, to emphasize perception of important data variables. In subsection 4.2.3, we will illustrate by practical examples, how glyph variables can be combined to form data displays.

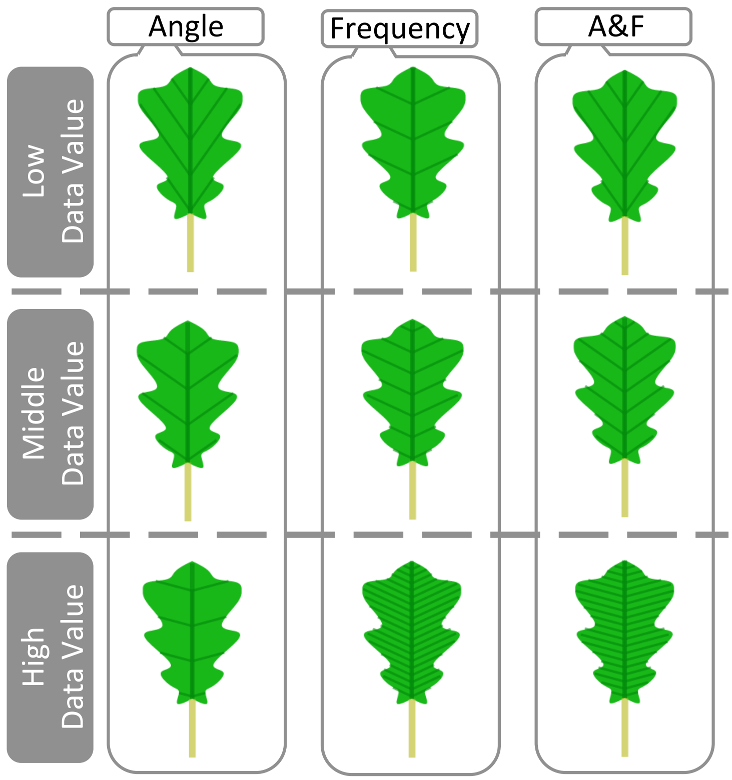


Figure 4.8: *Leaf venation*: The texture for the venation system can either be created by mapping data values to the angle or frequency of the veins separately, or by combining the two. The figure illustrates all possible combinations for low, middle, and high data values for a wave-like leaf shape.

Leaf Design	Numeric Variables	Categorical Variables
Shape	2 (x/y scale)	9 (selected morphologies)
Boundary	3 (frequency, width, height of teeth)	–
Venation	2 (number, angle of child veins)	3 (parallel, paired, net)
Other	7 (hue, saturation, orientation, x/y position, stem width/height)	–
Sum	14	12

Table 4.1: Summary of the parameters of our glyph design. It comprises 14 numeric and 12 categorical variables, which form the theoretic upper limit for the expressiveness of our glyph. Note that in practice, these variables are expected to not all be orthogonal, and comprise different perceptional performance, depending also on the data.

4.2.2 Leaf Glyph Aggregation

Leaf glyphs are prone to overlap since our design intention is to use big shapes for adding e.g., venation patterns. As a result analysts may not able to identify single leaf glyphs in dense areas. Figure 4.9 exemplifies such a scenario.

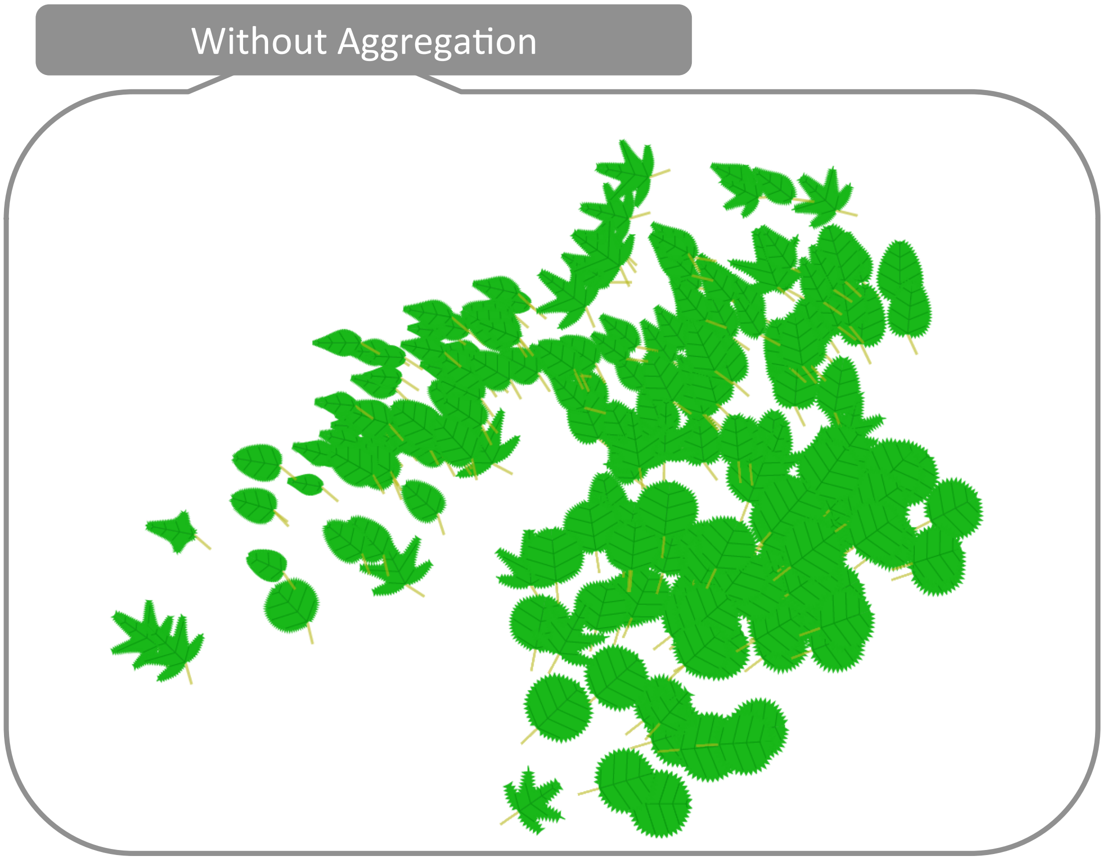


Figure 4.9: *Overplotting:* When multiple leaves overlap or coincide, we are not able to distinguish properly between their shapes and related characteristics.

The easiest solution for this problem is to apply transparency to the surface of the data glyphs. This Alpha Compositing technique [152] can also been seen in nature. When looking at shrubs or trees the single leafs or leaflets are not entirely opaque. At least the structure of some covered leaf shapes can be perceived (see Figure 4.10). By applying this technique to our leaf glyph design, we get a more realistic representation, which might help the analyst in pursuing his tasks.

However, this approach has its limits. In highly dense areas the opacity of all data glyphs will sum up until the individual leaf shapes cannot be discerned anymore. Therefore, we use another approach inspired by the environment. In our data glyph design space we discarded *palmate and pinnate* compound leaves, since their overall structure of adding multiple leaflets to a single branch might be confused with several individual leaves. This characteristic is useful for a design specific visual aggregation method. The idea is to combine individual leaf glyphs in highly dense areas to a combined compound leaf by reducing their size and attaching them to a common branch. As a first step, a grid is applied to the information space to detect dense

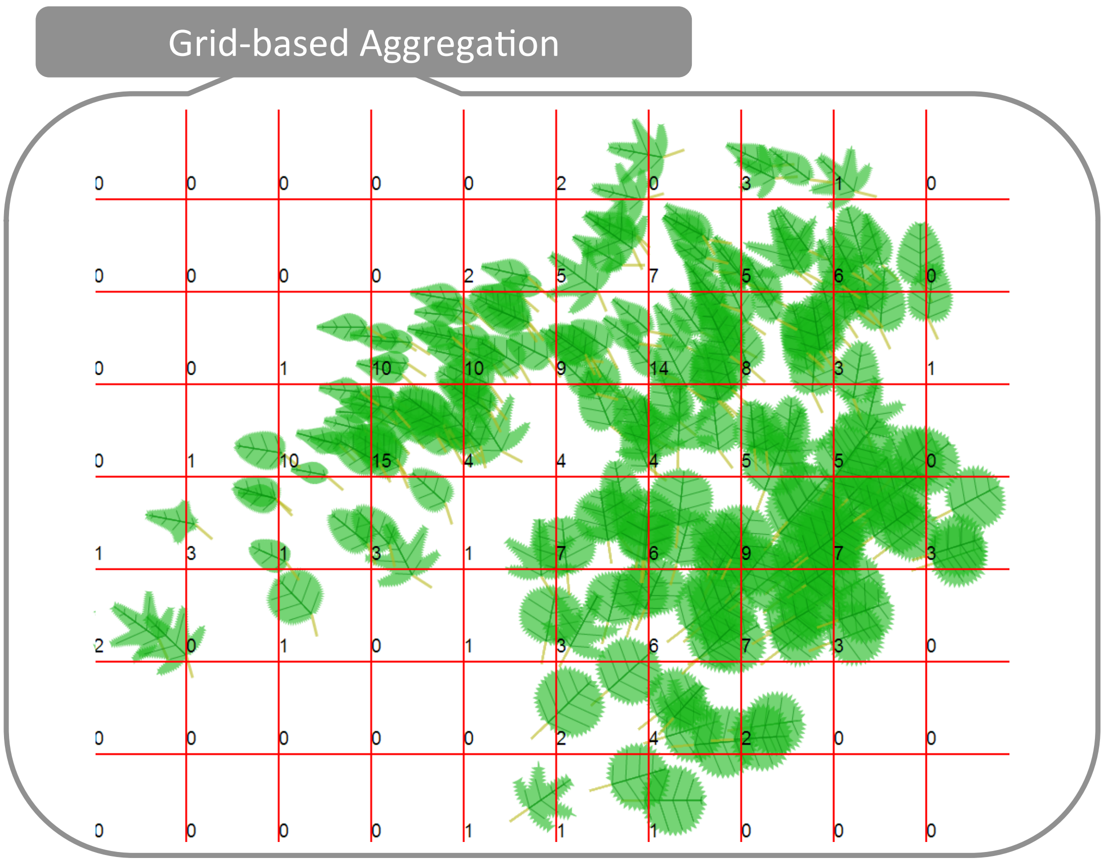


Figure 4.10: *Alpha Compositing:* Transparency is used to make covered leaf shapes distinguishable. This is especially true for dense areas where many data points are located.

areas. When too many data glyphs are located in the same grid cell a single compound leaf is created with its leaflets reflecting the aggregated data glyphs (see Figure 4.11). Of course, this technique has its limitations. Since the prototype has just one single branch, it is only possible to orient it towards one direction loosing the orientation of all the other leaves. Additionally, the size of the data glyphs has to be reduced when combining all of them in a single prototype. These compromises lead to a loss of information.

To avoid changing the size of the leaves we invented another grid based prototype. Again inspired by the environment we propose a stacked bouquet for representing dense areas. The method stacks all similar leaf shapes on top of each other and applies transparency to all data points. Different leaf shapes are repositioned and arranged in a certain angle to make them distinguishable from the others. Again, this visual aggregation technique makes use of the orientation of the data glyphs, however, the size of the leaves does not change (see Figure 4.12).

The previously introduced aggregation techniques are not only suitable to visualize dense

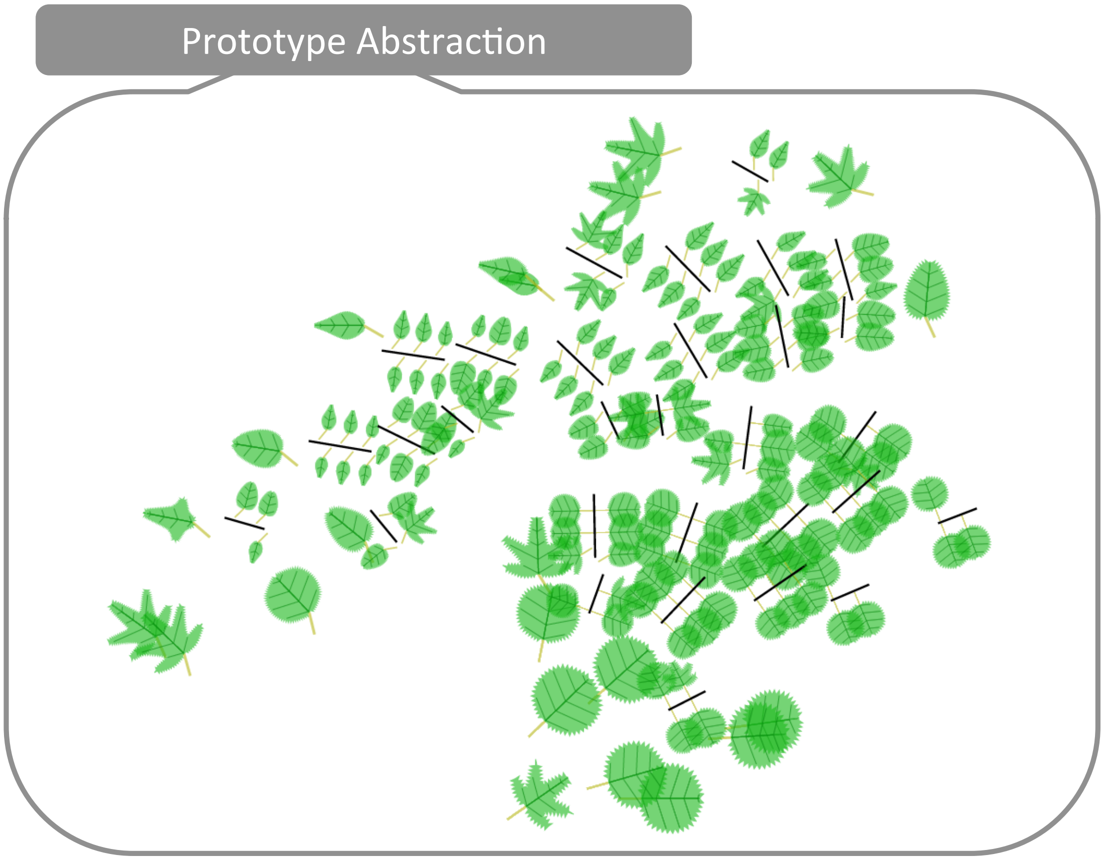


Figure 4.11: *Compound leaf:* Data glyphs are abstracted by combining single shapes in one compound leaf. This technique is especially useful in highly cluttered regions when transparency cannot be used anymore.

areas in 2D projections. Another design alternative is to use hierarchical arrangements, which can convey aggregate information and therefore, help with scalability. The relevant concept is that of a dendrogram (see Figure 4.13). Each parent node in a dendrogram may be represented by an aggregate prototype showing properties of the represented data partitions. Basic hierarchical visualizations can, therefore, be enriched with additional information like the composition of data points for individual clusters.

In Figure 4.13 we clustered the Iris dataset from the UCI Machine learning repository and represented the hierarchical structure in a radial dendrogram. The class attribute is used to assign different leaf shapes to the data. Other visual features like color, venation, and margin represent different attribute dimensions of the dataset. In each level, the nodes have been replaced with aggregated leaf glyphs using alpha composition together with a position bundling. The leaf glyph positioned in the middle of the visualization (*#1*) aggregates the dimension values of all

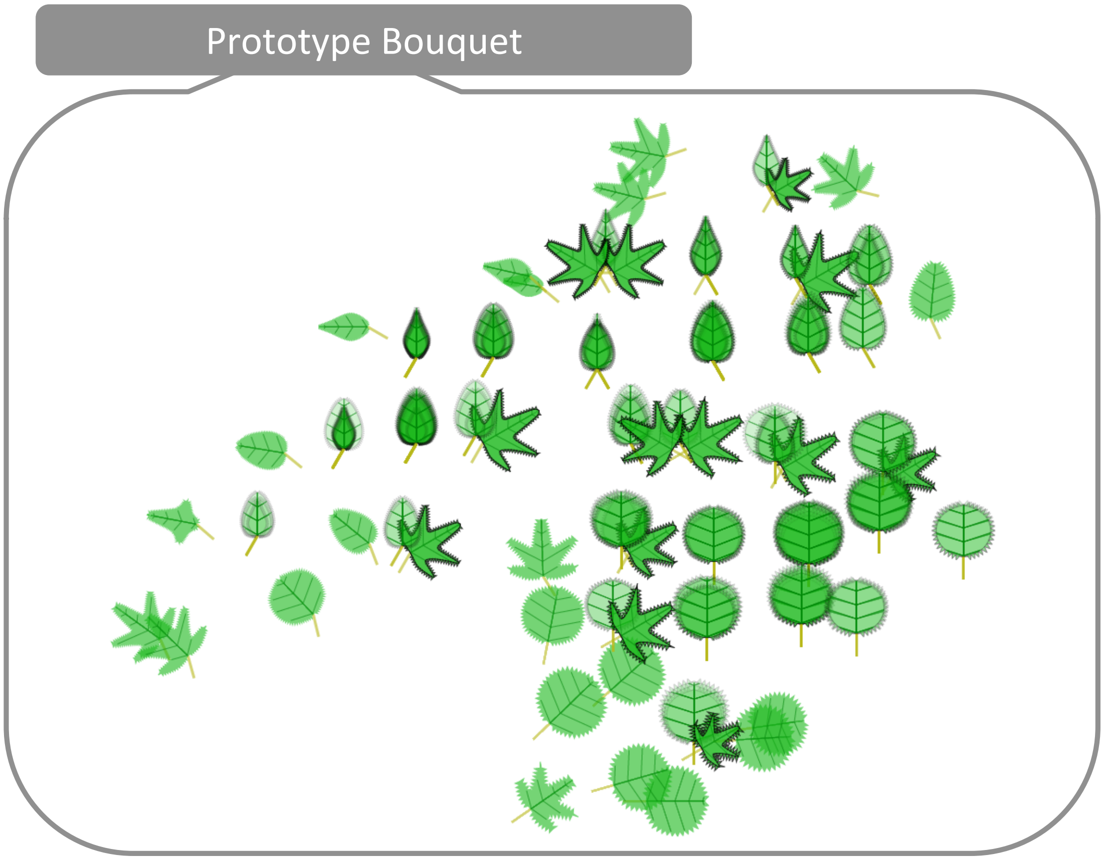


Figure 4.12: *Bouquet:* Leaf glyphs with an identical shape are stacked when they are located in the same grid cell. Different leaf shapes are oriented towards another direction to make them distinguishable from the other data points.

nodes in the diagram. It, therefore, contains many different sub-clusters as can be seen in Figure 4.13. When traversing the single branches to the lower levels (from inside out) the prototype representations of lower aggregate levels are getting more homogeneous. For example, after the first hierarchical split two main clusters are separated (*a* and *b*). The node labeled with *b* shows only green ovate leaves thus representing a homogeneous group of data points. The other aggregated prototype labeled with *a* seems to be more heterogeneous showing two different kinds of leaf shapes (hastate leaves and maple leaves). However, after descending to the next hierarchy level these two sub-clusters are separated. The inner node labeled with *#2* represents only maple leaves, whereas the other node labeled with *#3* contains hastate leaves. By traversing along the different branches the inner node is getting more and more homogeneous (e.g., similar colored leaves). Step by step different sub-clusters are divided till the lowest level of the hierarchy is reached.

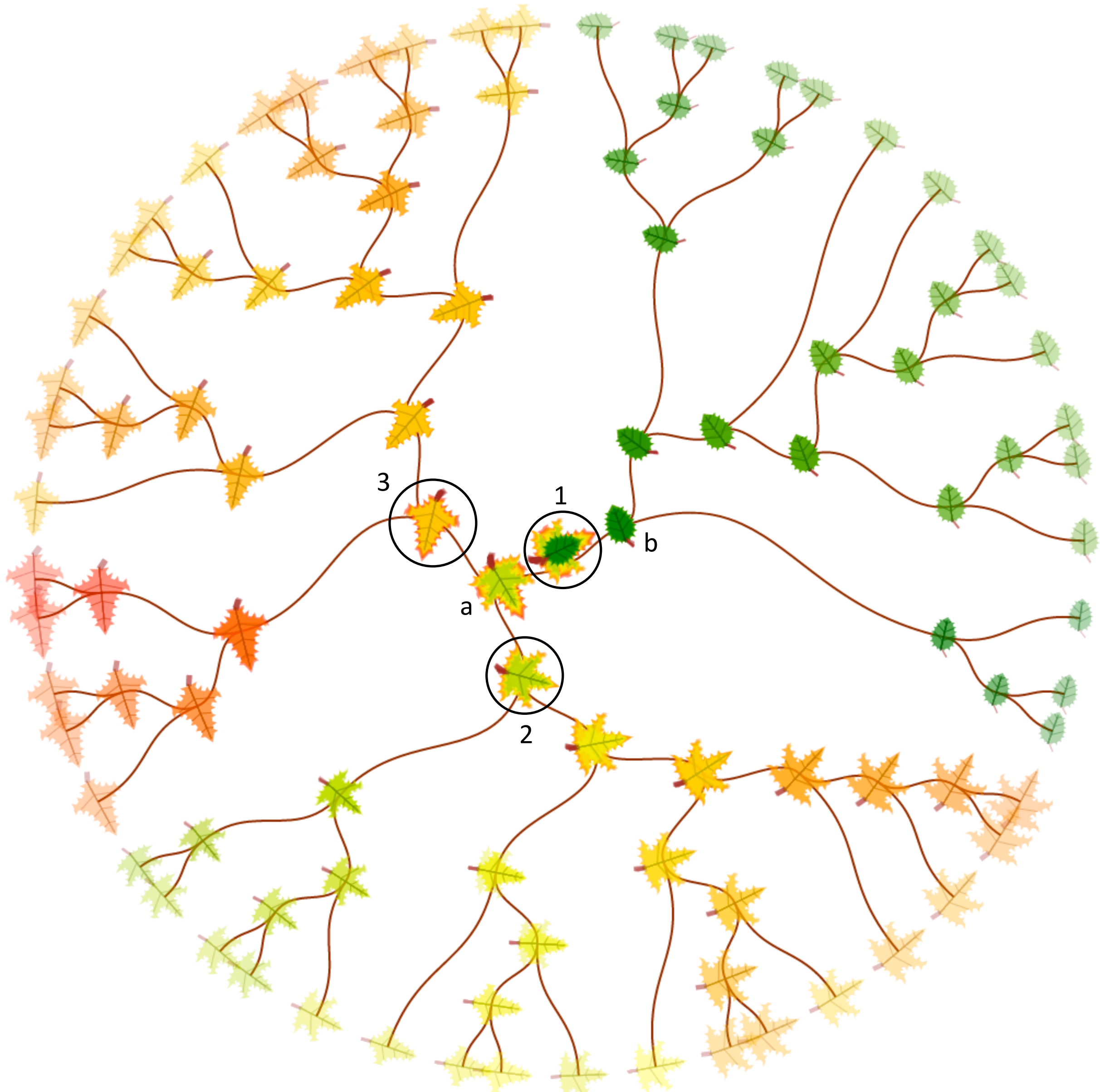


Figure 4.13: *Enhanced dendrogram:* A selection of data points from the iris dataset has been hierarchically clustered and the structure represented in a radial dendrogram. Leaf glyphs are used to visualize the groups and individual data points along the hierarchy. As can be seen, the visual structure of the leaf glyph is getting more and more precise when approaching the leaf nodes illustrating the homogeneity of the lower levels in the dendrogram.

For our visual aggregation we tried to keep a metaphorical approach in mind to help analysts understanding the visual representation without any training or explanation. In the following subsection 4.2.3 we evaluate our design space and the introduced aggregation techniques by applying the leaf glyph to a realistic data set from the environmental domain.

4.2.3 Use Case Scenarios

We defined an encompassing scheme to generate leaf glyph-based data visualizations for large data sets. We implemented the above described designs in an interactive system. We here exemplify results we obtained with three data sets. These results aim to show the principle applicability. Note that a thorough comparison against alternative glyph designs and user testing remain to be done in future work.

Forest Fire

The forest fire data set is available in the UCI machine learning repository [44] and called *forest fire*. It contains data about burned areas of forests in Portugal on a daily basis for one year. Additionally, weather information is included, e.g., temperature, humidity, rain and wind conditions at respective points in time. This data set does not contain any categorical data which could be mapped to the leaf shape. Therefore, we initially clustered the data points with the DBSCAN algorithm [82] and assign local or global outliers to different glyph shapes (Figure 4.14). Our idea is to map outliers to the more jagged leaf shapes, while non-outlier points get mapped to more regular or smooth shapes, thereby providing a first visual assessment of the degree of outlyingness for the data. Our analysis task is to find similarities between burned areas to be able to predict fires due to certain weather conditions.

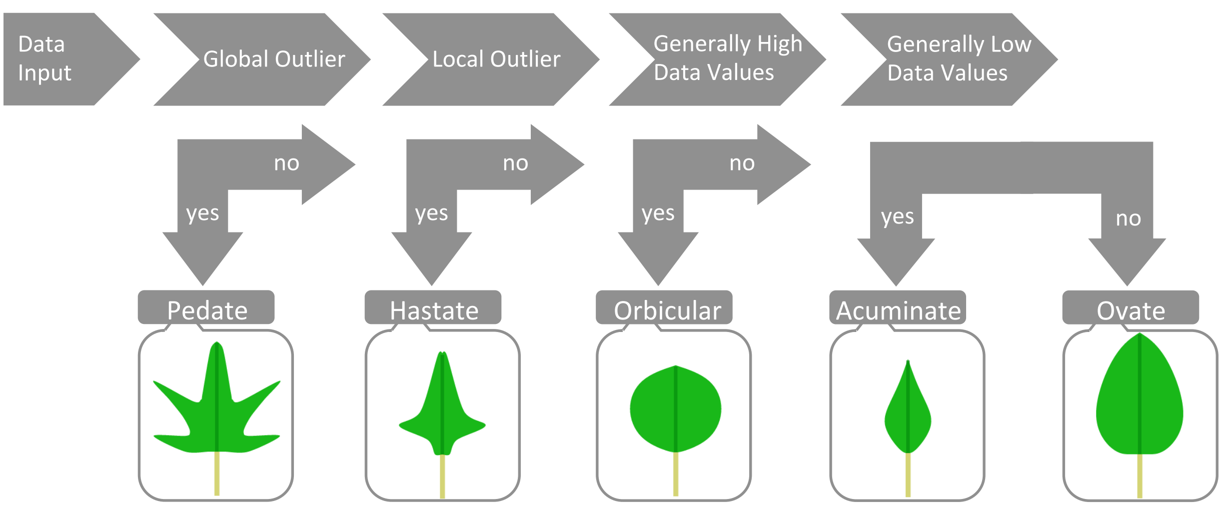


Figure 4.14: *Shape categories*: Based on the results of the clustering we assign different leaf shape templates according to the data characteristics.

First, we wanted to get an idea about the data distribution. We used one data glyph for each data point and positioned the leaf glyphs in a common scatterplot layout. The x-axis is reflecting the temperature and the y-axis the humidity. By intention we swapped the y-axis showing low data values at the top and high data values at the bottom. This reflects our background knowledge that possible indicators for forest fires are a high temperature and a low humidity. Potentially vulnerable areas are, therefore, positioned at the top right corner of the scatterplot. Figure 4.15 allows a first look on the data. At a first glance there seems to be a positive correlation between temperature and humidity. However, because of the high number of data points a lot information gets lost due to overplotting.

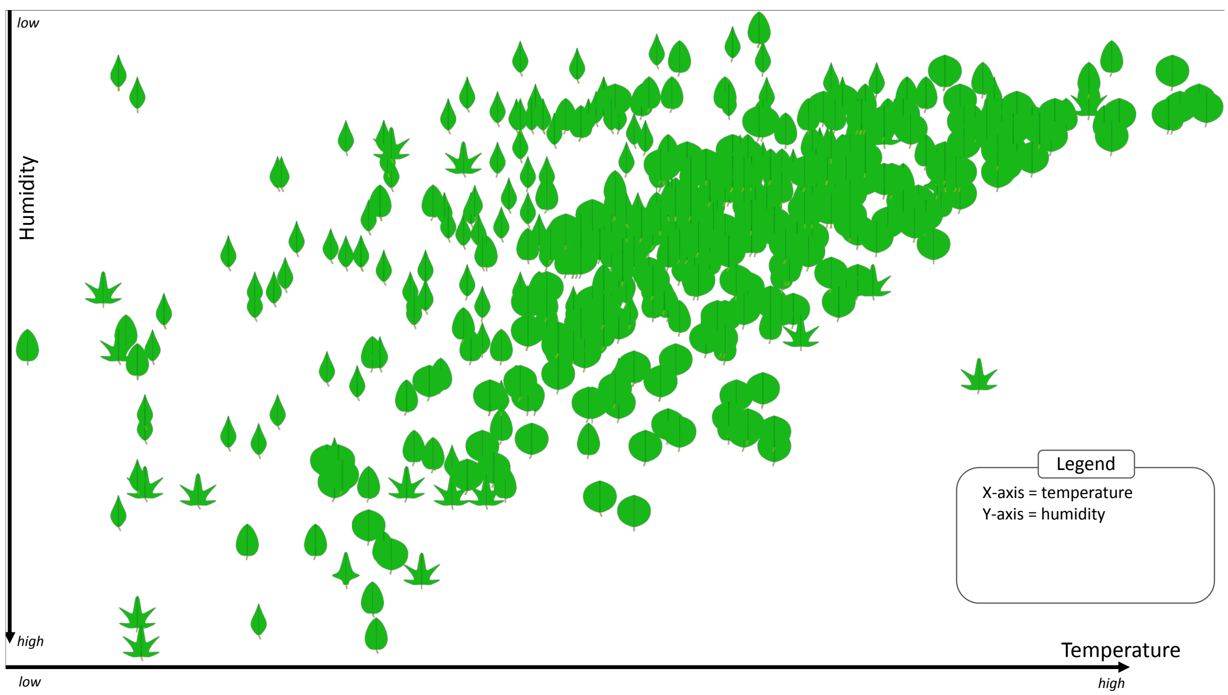


Figure 4.15: *Scatterplot layout*: Leaf glyphs are positioned in a scatterplot according to their temperature and humidity. Since no aggregation technique is applied on the data a lot of overplotting occurs.

As a next step, we applied transparency to the data points and also use color to show temporal information and orientation to encode the wind speed. The alpha compositing technique helps to detect some more leaf shapes, however, especially in the dense area on the diagonal still a lot of overplotting exists. For the color encoding we decided to use a metaphoric approach to help analyst understand the encoding without a color legend. We try to associate the seasons (i.e., winter, spring, summer, autumn) with the leaves. During winter and autumn the leaves in nature have a brownish or reddish color, whereas the color hue changes during spring and summer getting more green. Therefore, we colored our leaf glyphs accordingly. As can be seen in Figure 4.16 the data points are divided into 2 main clusters. Brown and red leaf glyphs are located above the diagonal and the more greener leaves are positioned on the diagonal. It seems as if humidity and temperature are both lower during autumn and winter times compared to spring or summer.

Another metaphoric approach was used to represent the magnitude of wind. The orientation of the leaf glyphs is changing according to the wind speed. Data points with low speed are oriented to the left. With an increasing wind speed the angle changes pointing right. The idea was to simulate a blast blowing from left to right catching all leaves and changing their direction accordingly. However, no additional visual pattern can be perceived. The leaf glyphs are pointing to various directions showing no correlation between wind magnitude and temperature, humidity, or time.

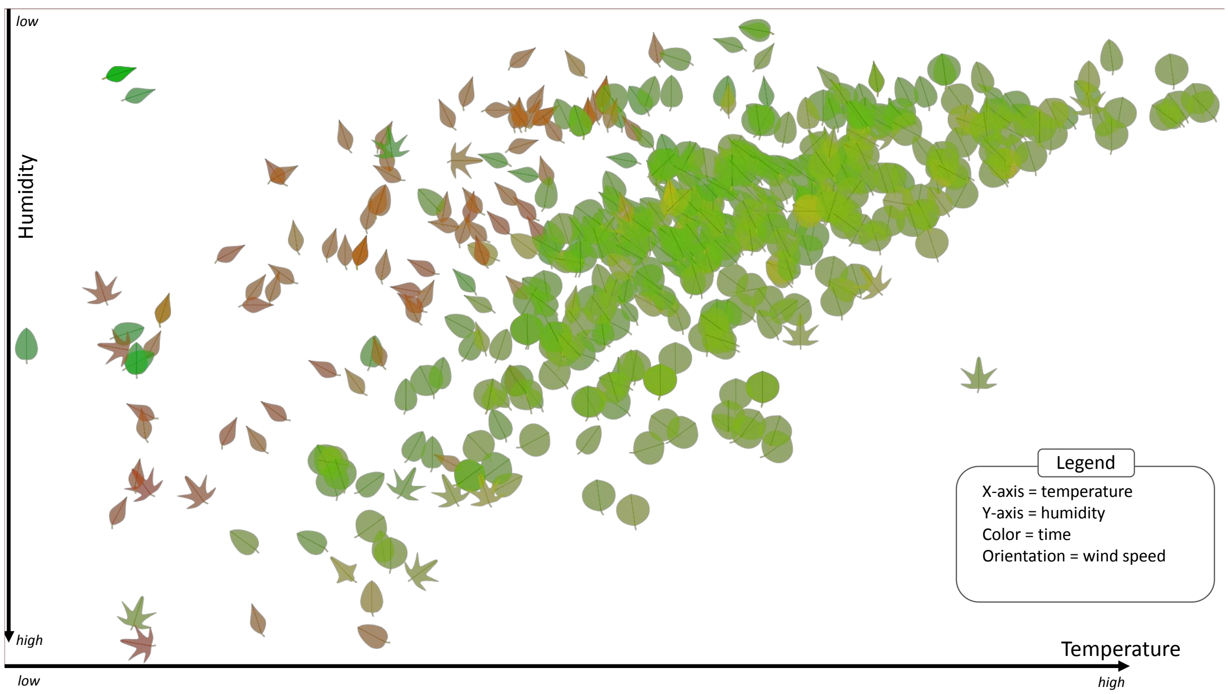


Figure 4.16: *Alpha Compositing*: Transparency is used to better perceive the data in cluttered areas. Since too many data points are located in the dense regions this aggregation technique does not provide the best view on the data.

To find similarities between burned forest areas we map the size of the burned regions to the size of the glyphs. Of course, this encoding is not a metaphoric representation, however, it helps to associate the information with the respective visual dimension. When inspecting Figure 4.17 it seems as if all leaf glyphs were reduced in size, and differences according to size cannot be perceived. This is surprising, since we would expect the size of burned forest areas to be different. One possible explanation is that some data points with different size are located in the cluttered area on the diagonal.

To get a different perspective on the data and to further reduce the overplotting, we switch to an alternative aggregation technique to better understand the highly cluttered area (Figure 4.18). Due to the design of the bouquet prototype generation, the visual attribute of orientation is lost,

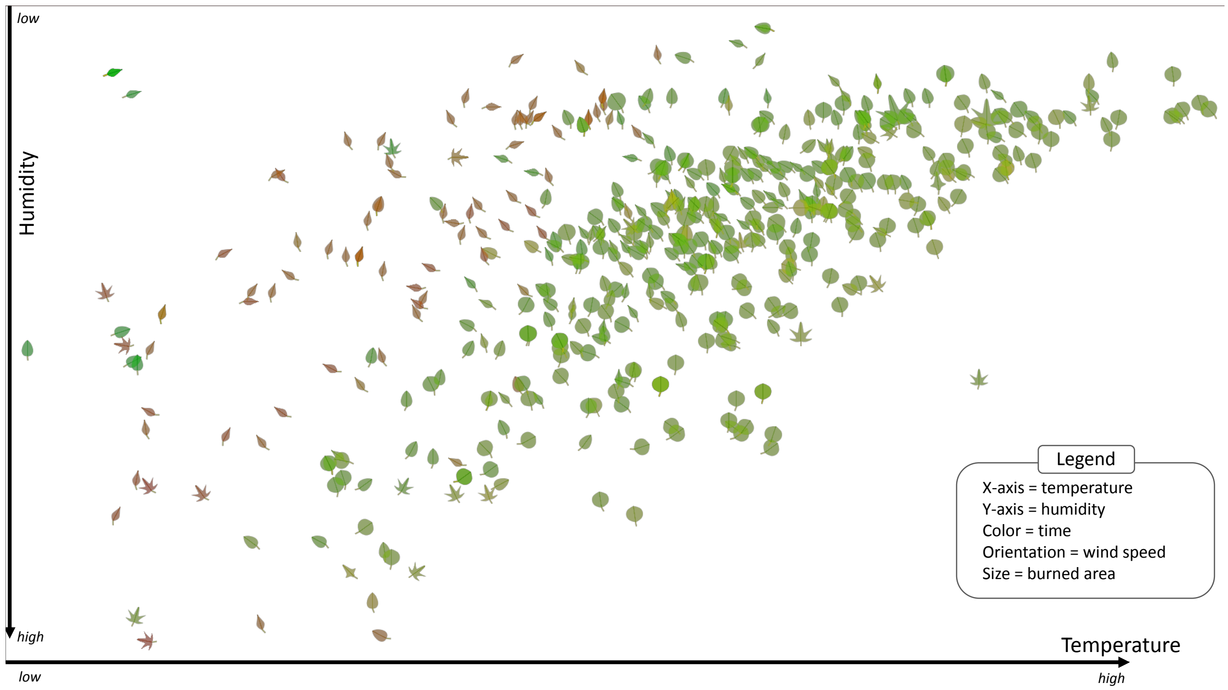


Figure 4.17: *Forest fire data set*: We applied alpha compositing as an aggregation technique to get a first overview of the data set. We used the following mapping to represent the multi-dimensional data: Shape ≙ local/global outlier, x-position ≙ temperature, and y-position ≙ humidity, color hue/saturation ≙ time (i.e., month), size ≙ area of burned forests, orientation ≙ magnitude of wind.

and therefore, we cannot map the wind magnitude to this variable anymore. In the highly cluttered area in the middle of the plot, several different maple leaf shapes become apparent. These refer to outliers detected by our previous clustering algorithm. However, more interesting are the two big maple leaf shapes located at the top right corner. They represent huge areas of burned forests during the summer time with high temperature and low humidity. When switching to Figure 4.17 and keeping in mind the concrete location of these data points, we can further extract the wind magnitude, which seems to be medium. With this understanding of the data, it is plausible why the burned forest areas are large. High temperature, medium winds, and low humidity all support the spread of a forest fires. However, since there are more smaller data points with similar data characteristics these features are not necessarily an indication for large forest fires. Perhaps, other factors, e.g., the area or the coverage of fire stations, which is not covered in the data might be an additional factor.

Of course, these findings would need to be substantiated by additional data considerations. Further information, e.g., the amount of firemen fighting the fire, the exact kind and amount of trees, or the time until the fire was recognized are important side factors not covered within the data. However, with our new glyph approach we were able to easily identify timely patterns,

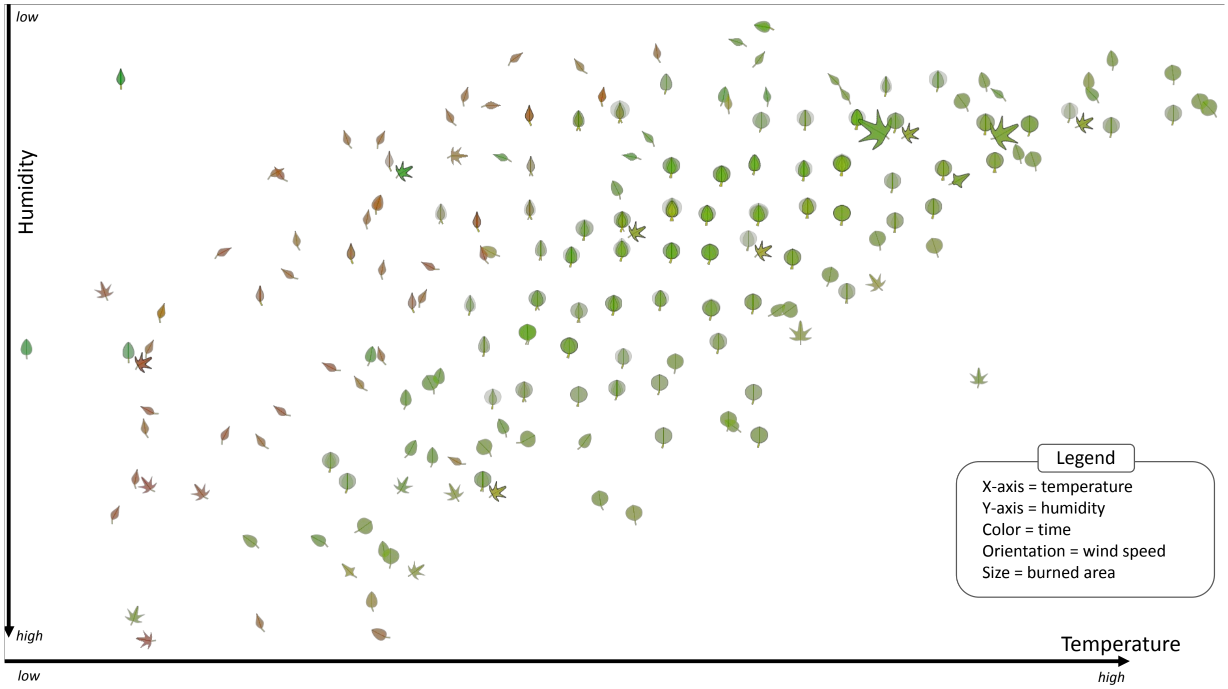


Figure 4.18: *Forest fire data set*: We applied a prototype aggregation technique to reveal insights to the highly cluttered areas in the plot. Interesting to note are the relatively big outlier leaf shapes, which were not visible beforehand.

outliers, and similar behavior of data points. Of course, other glyph designs (i.e., star glyphs etc.) might also be suitable to represent the data, however, our leaf glyph technique helps to easily associate the appearance of the data point with its attribute dimensions.

Iris and Seeds

Figure 4.19 and Figure 4.20 illustrate two well-known data sets (i.e., *iris* and *seeds*) from the UCI machine learning repository as an infographic representation. For both data sets, an initial k-means clustering is performed based on the number of classes within the data set. The clusters are then mapped to unique leaf shapes and projected to 2D space by Principal Component Analysis (PCA). As a last step the data dimensions are mapped to leaf glyph properties providing insights of the data. Due to the projection, some classes can already be distinguished. However, additionally assigning the clusters to different shapes helps to characterize the data more easily.

By mapping all data dimensions to glyph features, it is possible to extract more detailed information. In the seeds data set, there is a visual correlation between orientation (length of the grain) and venation frequency (width of the grain). The same thing is true for the color hue (asymmetry coefficient) and the y-position (1st principal component). The size (compactness) seems to slightly reflect the x-position (2nd principal component) (see Figure 4.19)

The iris data set is clearly divided into two different clusters by performing a PCA projection.

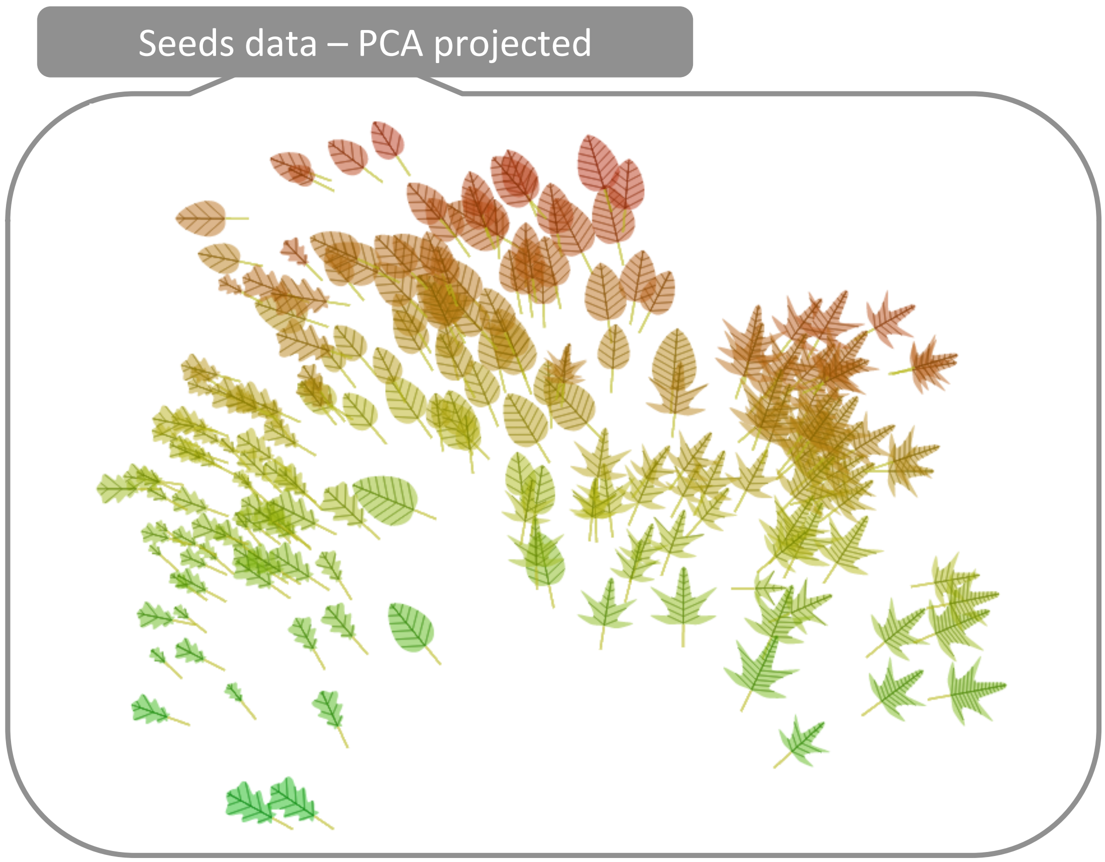


Figure 4.19: *Infographic representation - seeds data*: The well-known seeds dataset from the UCI machine learning repository is visualized using a 2D projection, and an appropriate mapping of data dimensions to leaf shape characteristics.

However, the data contains three classes, which are mapped to the shape by performing a k-means clustering. The visualization clearly shows two classes within the single cluster on the left. There seems to be a high correlation between the sepal height and length, which are mapped to the height and length of the glyph respectively. Since no leaf shape gets rescaled, the ratio between the two is read similar. Within the three classes, there is an almost equal distribution of the petal length mapped to the color hue. Finally, the orientation represents the petal width, which highly correlates to the x-position (2nd principal component) (see Figure 4.20).

4.2.4 Conclusion

Reviewing the literature has shown that metaphor based data glyph designs are suitable for conveying multi-dimensional data of a certain domain. Since no metaphor based glyph design for environmental data existed we introduced the *leaf glyph*. The design is based on a naturally

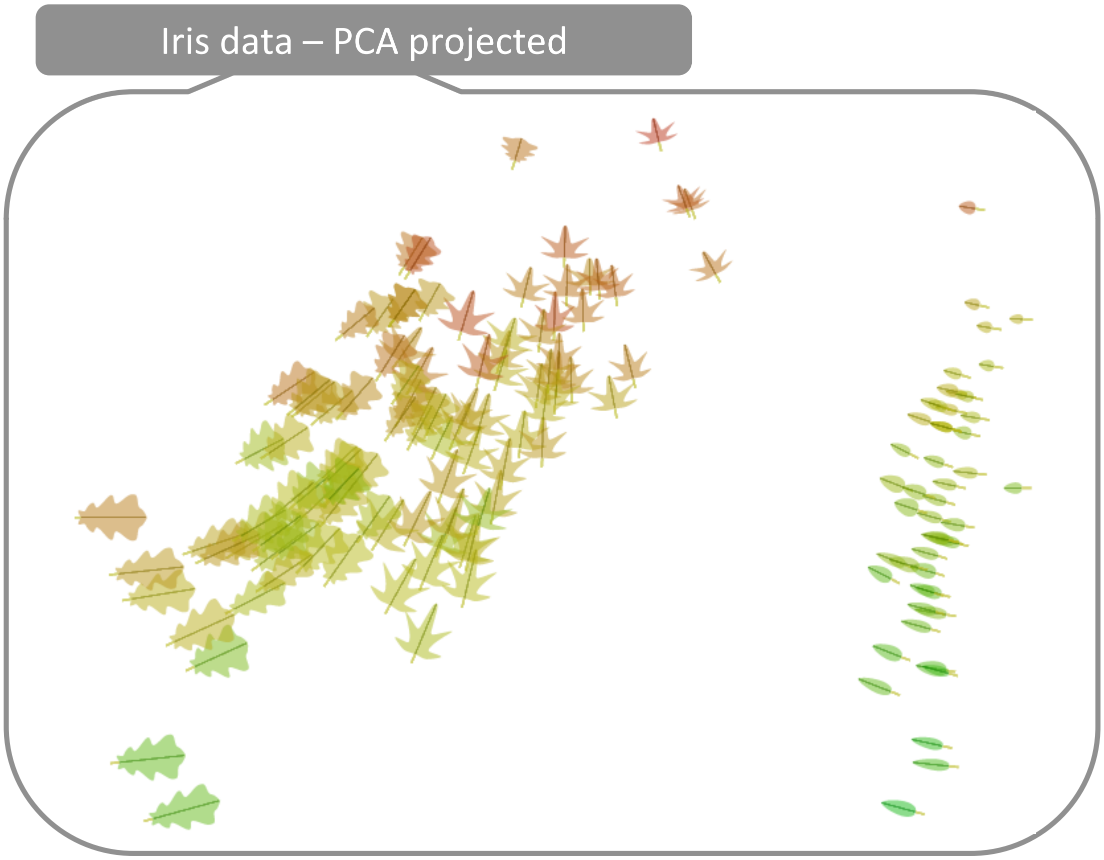


Figure 4.20: *Infographic representation - iris data*: The well-known iris dataset from the UCI machine learning repository is visualized using a 2D projection, and an appropriate mapping of data dimensions to leaf shape characteristics.

prominent shape, which should connect well to human perception, supposedly also under conditions of partial overlap. To come up with a well defined visual mapping we systematically structured the leaf glyph design space. Specifically, we mapped data to the main properties of the leaf glyph: leaf morphology, leaf venation, and leaf boundary. Furthermore, we defined a custom visual aggregation to scale the glyph for large numbers of data records with respect to its counterpart in nature. The applicability and effectiveness of our approach is evaluated by exploring three different multivariate datasets by expert users. A quantitative evaluation comparing the design against well-established alternatives is missing, however, we believe that our approach is aesthetically pleasing and may spark interest by a wider audience, for use, e.g., in mass media communication. Possible application areas are infographics in newspapers or publicly available websites communicating environmental information.

However, for lookup tasks including intra record comparisons the *leaf glyph* is certainly not the optimal choice. Same thing is true for similarity search tasks, where data similarity is most important. For such a scenario many-to-one mappings like the well-known star glyph should be preferred. Interestingly different variations of the star glyph are used quite often in literature and only little advice exists, which one performs best. To shed more light on this topic I will introduce in the following section 4.3 a controlled user study testing the star glyph and the influence of its contour line on similarity perception.

4.3 The Influence of Contour on Similarity Perception of Star Glyphs

In the related work section 4.1 several different data glyph designs have been introduced. One of those techniques is the well-established star glyph [168], which is used in various application domains: maps showing the geographical distribution of multi-dimensional objects (e.g., comparison of indicators such as crime rate or suicides for different regions of France [69]), multi-dimensional scaling visualizations exposing relationships between scaling algorithms and data distributions (e.g., election patterns to show political party proportions by region [175]), data objects organized in a grid layout to show how multi-dimensional objects distribute across sets of predefined categories (e.g., food nutrients in different food categories), or others [32, 60, 103, 104, 143, 168, 194, 195]

As can be seen in these examples, there exists a great variety of alternative designs for star glyphs that differ in the amount of reference structures used, the use of additional visual variables on the "rays," or whether or not the individual rays are connected to form a contour for the glyph [191]. The version of the star glyph with unconnected rays is also sometimes called *whisker or fan plot*, while the connected version also carries the name *star plot* [195]. Star glyphs are frequently used but very little advice exists on how to choose between different star glyph encodings. The question arises to what degree changes in the design of a star glyph influence its perception and, thus, the effectiveness of the glyph in certain tasks.

Some researchers have studied the perception of glyphs in the context of similarity tasks—yet with a variety of methodological approaches. For a more detailed overview I refer to chapter 2, which reviews the literature in more detail. In the following only the most relevant work is hightlighted.

Wilkinson [199] conducted a user study comparing star glyphs, castles, Chernoff faces and blobs. Participants had to sort 8 glyphs of each type—varied by a variety of factors—according to increasing dissimilarity. Their findings indicate that judgments on Chernoff faces were closer to the actual factor distances, followed by star glyphs, castles and blobs.

A similar sorting-based task was used by Borg and Staufenbiel [16] in their comparison of snowflakes (similar to star glyphs), suns, and factorial suns. Participants had to sort $3 * 44$ shuffled cards showing data points of one type of glyph into four categories according to their similarity. Factorial suns—that make use of some preprocessing of the data—were most easily discriminated and star glyph performed the worst in this respect.

Lee et al. [111] showed participants several datasets represented by one of: small-multiples

Chernoff faces, star glyphs, and two plots produced with multi-dimensional scaling. For each dataset participants were given eight questions to answer, some of which included similarity judgments based on pairwise comparisons. The authors did not perform an analysis on the basis of individual similarity questions. Instead, they found that participants performed best and were most confident with one of the 2D spatial plots, in particular on global questions where the whole set of data points has to be considered.

Klippel et al.'s study [104] is probably the most related to our evaluation as it also studied the influence of shape on glyph perception based on similarity judgments. Yet, instead of the influence of contour, as in our case, they varied shape by reordering the dimensions in a star glyph with contour. The authors studied how shape changes influenced the interpretation of data points in a similarity-based grouping task. They found that differences in shape influenced cognitive processing of the data and that perceptually salient features (such as spikes) strongly influenced how people thought about a data point.

One important task for glyphs in small-multiple settings is the comparison of the encoded data points to one-another. Such a comparison task may be conducted to find data points that are very close over all dimensions, very different, or similar in just a subset of dimensions. We focus on the first task: finding data points encoded as star glyphs that are very similar to a target glyph. We are interested in this task because if it is well supported, it should improve people's ability to perform the other two types of comparison tasks, too. We hypothesized that the ability to perceive a star glyph as a coherent and closed shape would strongly influence the correctness of data similarity detection tasks—as it would potentially be easier to compare a single shape than having to compare individual rays. This hypothesis was motivated by prior research showing that a closed contour has an influence on the perception of a coherent shape [58]. As Palmer noted: *"Shape allows a perceiver to predict more facts about an object than any other property"* [146].

Contour closure is a phenomenon related to the Gestalt principle of closure [105] according to which we[6] perceive visual objects as grouped together if they seem to complete a visual entity. Visual objects can, thus, have an "open contour" of unconnected marks (dots, lines, etc.) or a completely "closed contour" that forms one continuous boundary. Several researchers have tested open vs. closed contours and found that closed contours were perceptually superior to open contours in a variety of different tasks:

Elder and Zucker [58] showed that the efficiency of visual search was better for shapes with a closed contour. Similar to our studies, participants were shown a stimulus object, which had to be found amongst a larger set of distracters. In later work [59], the authors found supporting evidence that geometric region boundaries, such as contours, are processed much earlier by our perceptual system than other surface features, such as a region's texture.

Garrigan [79] investigated the recognition accuracy of stimuli with a closed contour compared to those with an open contour. Participants had to learn a set of stimuli and later recall whether a newly presented stimulus had been in the previously learned set of objects or not. The experiment showed that the closed contour shapes were recognized more easily. The authors argue that this is due to a better encoding of the stimuli in the human visual system. Finally,

[6] In this section 4.3 and all corresponding subsections the term "we" comprises Petra Isenberg, Anastasia Bezerianos, Fabian Fischer, Enrico Bertini, and me

Saarinen et al. [158] measured the precision of shape perception. Participants had to judge whether the aspect ratio of the stimulus (i.e., a rectangle) was tall or wide. Their results showed that for discrimination tasks, closed contour shapes were more effective than non-closed shapes.

In summary, a large body of literature exists on the effect of contours on shape perception. We highlighted a few that have particularly inspired our hypotheses that the presence of contours may be of particular importance to the effective use of glyphs in visual data analysis tasks. We contribute a further study focused on a concrete application in visual data analysis, with glyphs showing multi-dimensional data points. We examine in particular whether glyph shape, emphasized by a closed contour, is an important predictor for the effectiveness of glyph designs in a similarity detection task.

We are not aware of any previous studies on the importance of glyph contour on tasks with multi-dimensional data. We contribute three studies, showing the differences in performance when adding a contour structure to well-known glyph designs. In the first we investigate the effect of contours on the perception of data similarity with data visualization experts and novices on star glyphs. Our results indicate that contours influenced expert the most, whereas novices had poor performance across the board. Nevertheless, we observed that some types of shape-related similarities overpowered the participants' ability to read data. Based on this result we conducted a second study on the nature of similarity perception for a larger set of glyphs with and without contours. We found that, even without any participant training, glyphs without contours led to similarity judgments that are close to data reading, thus making them better candidates for visualization. In our third and final study, we added simple reference structures—tick marks and grids—to the designs, and examine how they affect similarity judgments. Our results show that reference structures aid data similarity comparisons in star glyphs with contours, but have little effect on ones without. Based on these results we present considerations for the design of more effective glyphs for similarity detection tasks.

4.3.1 Experiment 1: Contours for Novices vs. Experts

In our first study we were interested in the fundamental question: does contour affect people's perception of data similarity with star glyphs? Data similarity judgments are cognitive tasks, where the viewer has to judge the absolute difference in all dimension data values between two data points. This differs from other types of similarity judgments, such as detecting shape similarity e.g., under rotation or scale.

Detection of data similarity is a synoptic task according to the Andrienko & Andrienko [10] task taxonomy. Synoptic tasks are very common and important for glyphs in small-multiple settings. Analysts have to visually compare data points to detect outliers or to identify similar groups of data points, by referring to the whole data set or a subset of the data (e.g., finding countries with similar characteristics).

We were interested in the effect of contour, as we hypothesized—based on previous perception studies [58]—that a contour would impact the rapid perception of shapes and, thus, aid in tasks that require the data point to be perceived in its entirety. Finally, we hypothesized that there would be a difference between experts' and novices' ability to make accurate data similarity judgments, and thus chose to conduct a between-subjects experiment with these two groups of participants.

Design and Procedure

Glyphs: We used three different variations of the star glyph (Figure 4.21). The first, also called whiskers or fan plot [150, 195] uses "rays" to encode quantitative values for each dimension through the length of each ray. We refer to this variation as "Data lines only (D)". The second variation, "Data lines + Contour ($D + C$)", connects the end of each ray with a line to add a closed contour [168]. In the third variation the radial rays are removed, and only the contour line is presented [32]. We use the term "Contour only (C)" for this design variant. All three star glyph contour variations have been used in real-world context and in the scientific literature, thus adding external validity of our glyph choice.

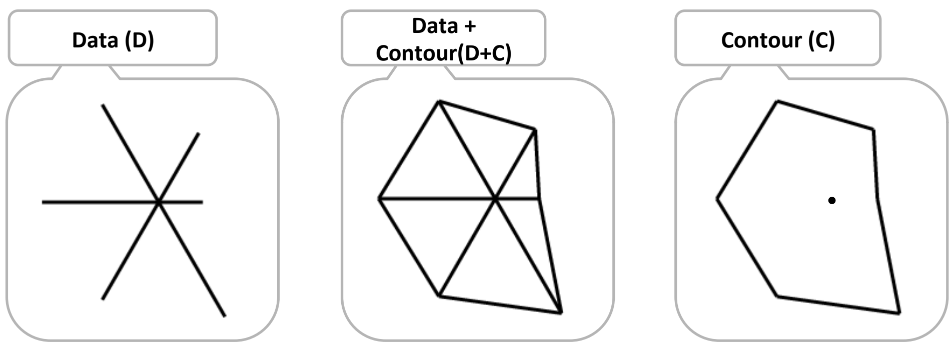


Figure 4.21: *Experiment 1 contour variations*: (from left to right) star glyph with rays and no contour (D); common star glyph ($D + C$); only the contour line of the star glyph (C).

Dimensionality: To investigate the effect of contours on different data densities we varied the number of dimensions shown in the glyphs. The *low* dimension density consisted of three data dimensions with corresponding data values, while the *high* density consisted of ten data dimensions. We considered ten dimensions to be *high*, as glyphs used in the literature rarely visualize more than ten dimensions; also to our knowledge there is no study investigating the maximum number of perceivable dimensions in a single star glyph to use as a basis.

Task, Procedure and Apparatus: Participants were shown a highlighted stimulus glyph surrounded by 8 more glyphs in a $3 * 3$ matrix configuration (Figure 4.22). One of these glyphs was closest in data space (lowest absolute data distance) while the rest were distracters further away in data space. The participant had to select the glyph closest to the stimulus in terms of data value. For each contour variation, participants were given training explaining how data was encoded and the notion of similarity in data space. They were then given four practice trials where the correct answer was revealed to help learning. During the actual experiment the correct answer was no longer provided.

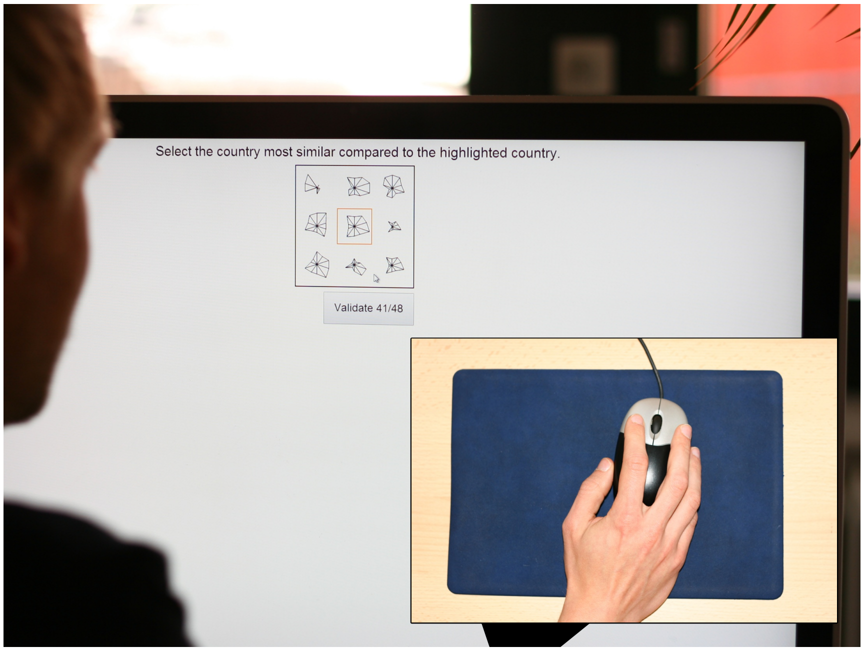


Figure 4.22: *Experiment setting*: The participant was seated in front of a 24” Screen with a resolution of 1920x1200. The only input device was a computer mouse.

The three glyph variations were presented in an order randomized using a latin square. The position of the correct answer as well as the different distracters was also randomized. Similarly, the exact glyph values were randomized. Each participant repeated 4 training and 4 real trials for each contour variation.

The study took place in a lab setting in the presence of an experimenter. The experiment was conducted on a 24 inch screen with a resolution of $1920 * 1200$ and took around 25 minutes. The only input device was a common computer mouse to make selections.

Participants: Twelve novices (7 female) and twelve experts (2 female) participated in our study. The age of novice participants ranged from 18–23 years (mean & median age 20), and from 26–38 years (mean 30.3 and median 29) for experts. All participants reported normal or corrected-to-normal vision. All novice participants reported no experience in reading glyphs, but were familiar with common chart visualizations seen in print (e.g., bar and pie charts). All 12 experts were visualization researchers and students who reported a strong background in data visualization with at least basic knowledge of reading glyphs (1 Bachelor; 8 Master; 3 PhD).

Overall our experiment consisted of:

3	contour variations (*D,D+C,C*)	$*$
2	dimensionalities (*high*, *low*)	$*$
4	repetitions	=
24	trials per participant	$*$
24	participants (12 per expertise)	=
576	**trials in total**	

Expertise was the between-subjects factor.

Hypotheses

H1: Novices are less accurate in judging data similarity than experts.
H2: Both experts and novices make more accurate judgments in the low dimensional than the high dimensional condition.
H3: For both experts and novices, contour variations ($D + C$,C) improve the accuracy of data similarity judgments.
H4: This effect will be stronger in novices who have no prior glyph reading experience.
H5: Contour variations ($D + C$,C) lead to more accurate judgments mostly in the high dimensional condition, while the low dimensional condition is be less affected overall.

Data Generation and Distracters

Our data was synthetically created: 3 dimension values for the low and 10 for the high dimensional case. For each dimension we consider data values ranging from 0 to 5, partitioned in three value categories: low [0,1], middle [2, 3], high [4, 5]. We avoided larger value ranges as we were not interested in studying visual acuity.

The stimulus (i.e., central highlighted glyph) was created randomly by assigning either a middle or a high data value to the different dimensions with an equal chance of 25% (i.e., 50% for each value categories and 50% for the final data value). This was done once for all repetitions. To avoid learning effects, the stimulus was rotated between repetitions, keeping the values and the neighboring dimensions identical.

Each trial also contained a *target* glyph—the correct answer, thus the most similar to the stimulus in terms of data closeness (minimum data value distance). To generate it, we changed the data values of the stimulus randomly up to a maximum of 7 changes in data distance for the high dimensional condition, and 1 for the low. This was done by sequentially scanning the dimensions with a probabilistic function, which first decided to change the dimension or not (50%), second to increase or decrease the corresponding data value (50%) and third by how much (i.e., 1 or 2)(50%). At the end we ensured that the resulting data values fit into one of the three categories (i.e. low, middle, and high) and that the sum of all changes meet the predefined criteria.

Besides the stimulus and target glyph, we created 3 types of *distracters*. First, a *rotated* version of the stimulus, keeping the data values identical, but switching the dimensions one step either to the left or to the right. Second, a *scaled* version of the stimulus where we reduced the data values of each dimension by 1. Since the data values of the stimulus reach from 2 to 5 it is not possible to end up with negative values. Third, a close *alternative* of the target glyph. This alternative takes the data values from the stimulus and changes the values randomly up to a maximum of 8 changes in data distance for the high dimensional case, or 3 for the low. Values were chosen to ensure that the alternative glyph is not too different from the stimulus, while the target glyph continues to be the most similar in data distance. The remaining distracters were created randomly by assigning a data value to each dimension with an equal chance (see

Figure 4.23). For each trial we ensured that the sum of all differences between stimulus and all distracters was higher to that between stimulus and target glyph.

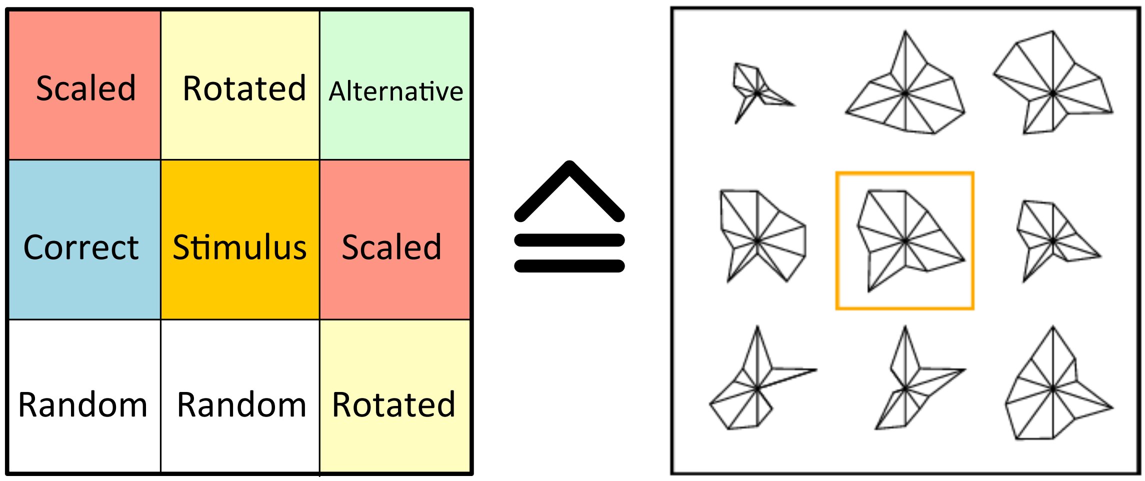


Figure 4.23: *Experiment setting*: For each trial glyphs were arranged in a $3 * 3$ matrix. The stimulus is highlighted and positioned in the middle to assure an equal distance to the other glyphs. This setting is used in all three experiments.

Results

We report only statistically significant results ($p < .05$) for accuracy. Given the non-normal nature of our data we used a non-parametric Friedman's test for the analysis of correct answers between glyph variations and a Kruskal-Wallis test for comparisons between expertise (between group factor). Figure 4.24 shows overall correct answers, and Figure 4.25 (low dimensional) and Figure 4.26 (high dimensional) which type of distracters participants chose under the different experimental conditions. Although completion time was logged, we found no differences across variations and user groups, with low dimension trials taking on average $11sec$ ($D = 12.7sec$, $D + C = 11.3sec$, $C = 9,7sec$) and high ones $18sec$ ($D = 19.7sec$, $D + C = 16.9sec$, $C = 16.7sec$).

Overall accuracy for experts across variations was 79.1% for the low and 44.4% for the high dimensional glyphs, and for novices 74.3% and 36.8% respectively. However, there was no significant effect of *expertise* on *accuracy*. Figure 4.24 illustrates more high level results.

Dimensionality: There was a significant effect of *dimensionality* on *accuracy* ($\chi^2(1, N = 288) = 23, p < .001$).

Post-hoc tests revealed that participants were more accurate in the low dimensional condition (76.7%) compared to the high dimensional condition (40.6%, $p < .001$).

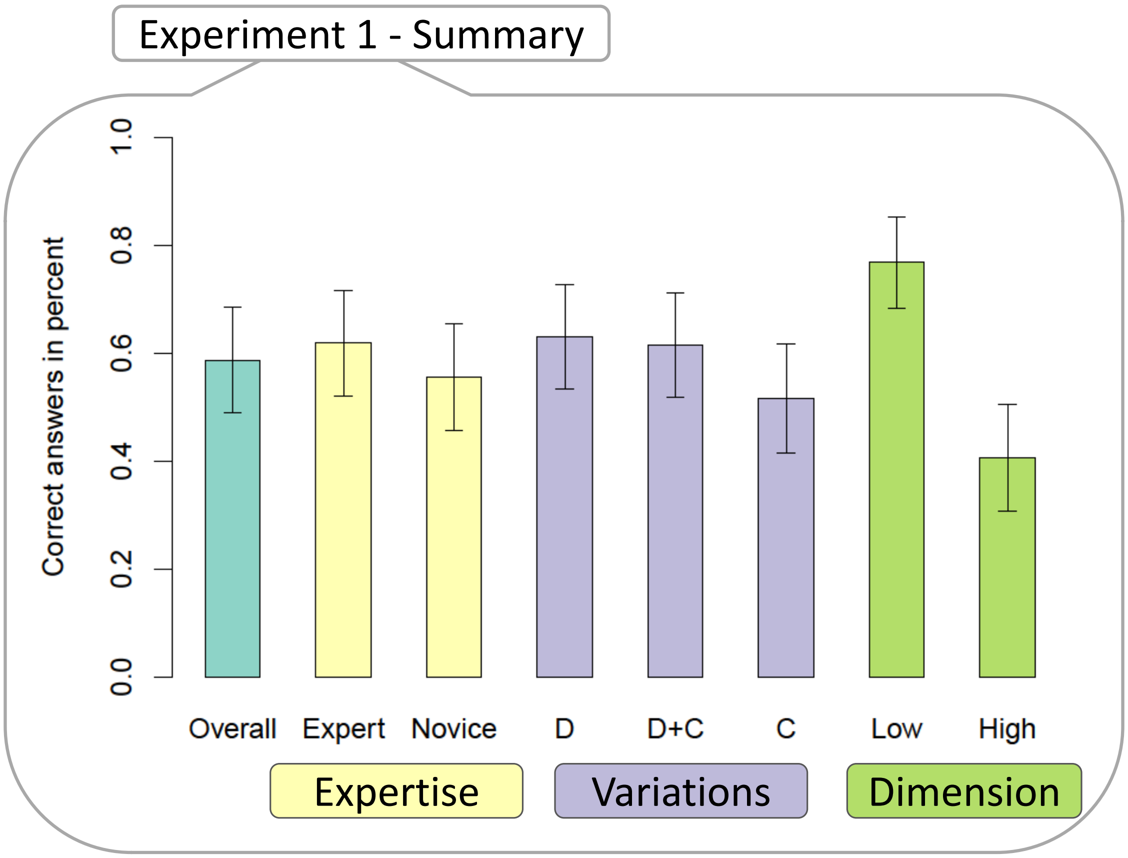


Figure 4.24: *Experiment 1 summary*: The bar charts illustrate the percentage of correct answers and the standard deviation.

Contour variation: There was a significant effect of *contour variation* on *accuracy* ($\chi^2(2, N = 192) = 7.9, p < .05$). Participants using variation C performed significantly worse (51.6%) compared to D (63%, $p < .05$) and $D + C$ (61.5%, $p < .05$). For experts, there was a significant effect of *contour variation* on *accuracy* in the high dimensional condition ($\chi^2(2, N = 48) = 12, p < .001$). A pairwise comparison revealed a significant higher accuracy with the D variation (66.7%) compared to both $D+C$ (41.7%, $p < .05$) and C (25%, $p < .001$). No significant results were found for novice participants.

When comparing the accuracy of the two participant groups, we found that for the variation D, there was a significant effect of *expertise* on *accuracy* in the high dimensional condition ($\chi^2(1, N = 96) = 5.85, p < .05$). Experts performed significantly better (66.7%) using the D variation compared to novice participants (39.6%, $p < .05$).

When selecting a wrong answer, both experts and novices most frequently selected the second closest data point to the stimulus (17.7%, 20.5% respectively), followed by a scaled version of the stimulus (16%, 16.3%) and to a lesser extent rotated versions (2.4%, 4.1%), mostly in the high dimension case of the contour variations ($C + D, C$).

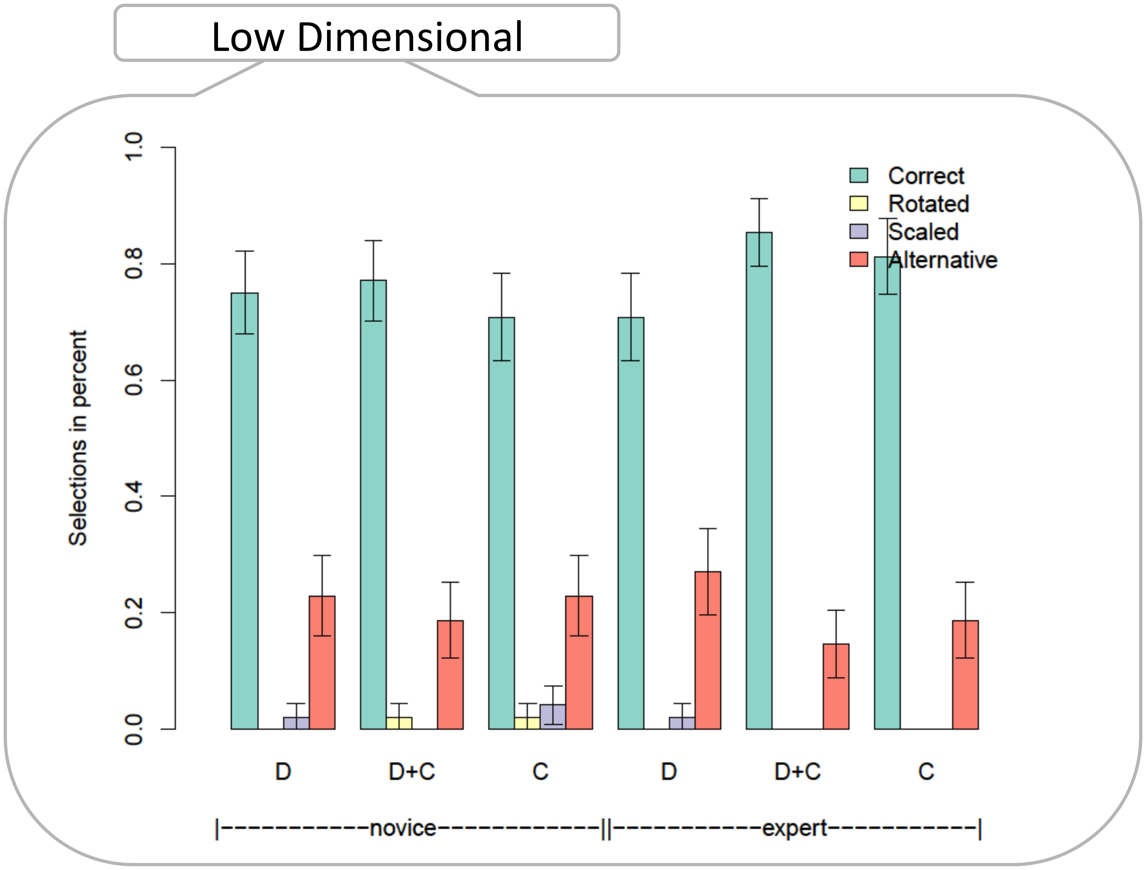


Figure 4.25: *Experiment 1 results low dimensionality*: The bar charts illustrate the percentage of selections and the standard deviation for each factor. In the low dimensional condition there were no significant differences across expertise and design variation.

Summary and Discussion

Overall we cannot confirm H1, our experts were not significantly more correct than novices on average. This is especially true for the low dimensional condition where both user groups had a good performance ($\approx 80\%$ correct). However, for higher dimensionalities experts using variation D were significantly more accurate compared to novices (partially confirming H1).

When comparing the two dimensionalities, similarity judgments were significantly more accurate for both user groups in the low dimensional condition compared to higher dimensionalities, confirming H2. With an increasing number of dimensions more data values have to be visually compared, leading to more complex mental calculations resulting in a higher error rate.

Contrary to intuition from previous work that contour can improve similarity judgments [58, 79], we found that contour affected the accuracy of judgments negatively for experts. Thus we cannot confirm H3. As no significant effects were found for novice participants, we could also not confirm H4, however, mean accuracy for C (50%) was lower compared to $D + C$ (59.4%) and D (57.3%).

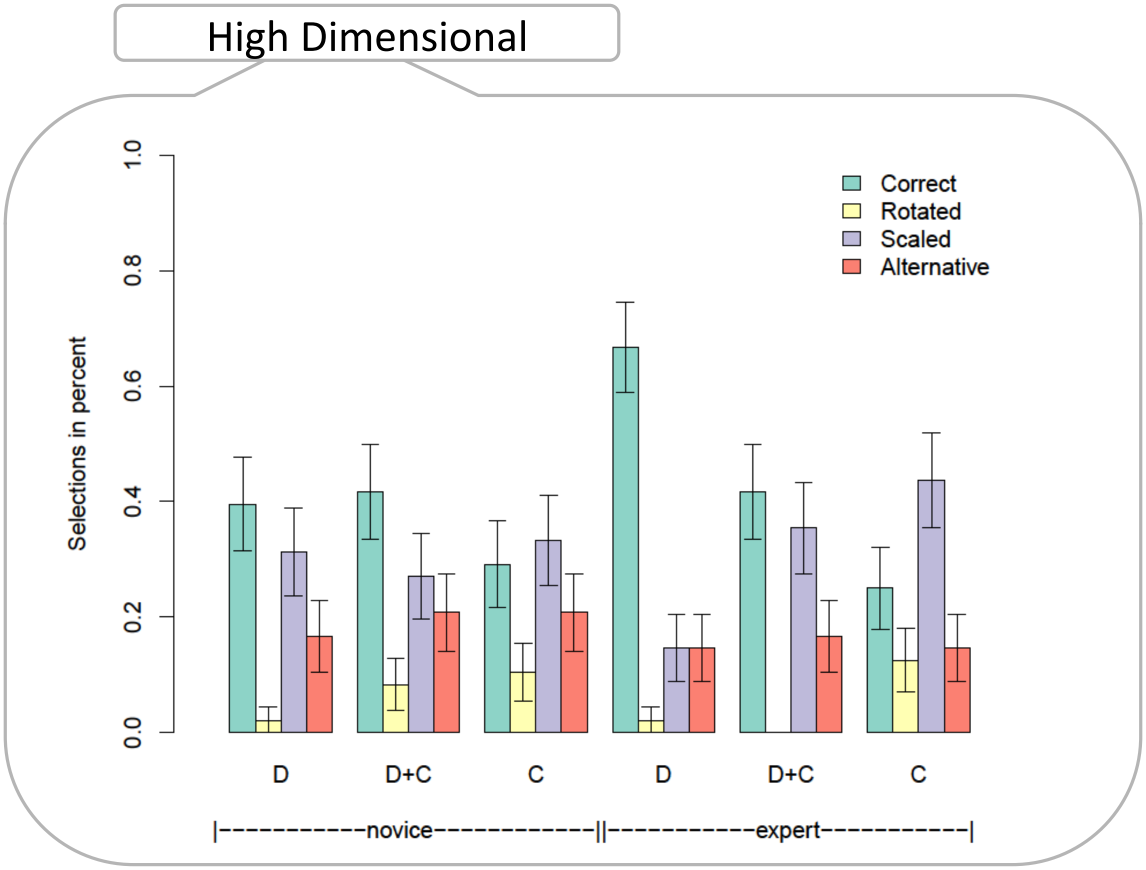


Figure 4.26: *Experiment 1 results high dimensionality*: The bar charts illustrate the percentage of selections and the standard deviation for each factor. In the high dimensional condition experts using variations $D+C, C$ were lead to judge shape similarity rather than data similarity whereas the accuracy of novices was low for all three variations.

We also could not confirm H5. Contrary to expectations, the variation without a contour (D) led to significantly more correct answers for high-dimensional glyphs. The effect was not visible in the low dimensionality case where all participants were overall approx. 80% accurate with all variations.

Trying to explain the unexpected negative effect of contour on experts, especially in high dimensional cases, we noted that at least half of the erroneous answers in the contour variations ($C+D, C$) were in the form of scaled versions of the stimulus glyph, and to a lesser extent rotated versions, i.e., glyphs that have a geometric form similar to the stimulus glyph. In retrospect, this negative effect of contour could be explained by the fact that contour, and closure in general, is one of the factors promoting the notion of unity according to Gestalt psychology [105]. In our case contours led our experts to erroneously consider glyphs as coherent shapes when judging similarity, rather than data points. This resulted in judgments and comparison of geometrical shapes rather than data, with experts being led to consider as more similar data points that were either scaled or rotated versions of the stimulus, rather than the one closest in data space.

Given the overall poor performance of novices in the high dimensional case we conjecture that due to their lack of familiarity and experience they tended to fall back to judging shape rather than data similarity for all star glyph variations. This is evidenced by the fact that at least half of their errors were a combination of scaled and rotated versions of the stimulus glyph.

4.3.2 Experiment 2: Perception of Similarity

Results from Experiment 1 indicated that in high dimensional cases contours mislead even experts to perceive rotated or scaled versions of the stimulus as more similar, rather than the one closest in data space. Based on this finding, we conducted a second experiment to better understand what type of similarity different kinds of data glyphs naturally support. To this end, participants were not given any training or explanation of what similarity means, and we did not inform them that the glyphs encoded multi-dimensional data. Their only instruction was to select the most similar glyph. Our goal in this experiment was to examine what viewers naturally perceive as similar in different data glyph variations, without being instructed on how to judge similarity. Based on our results we hoped to identify the data glyph variations, if any, that naturally promote data similarity rather than shape similarity and, therefore, are more suitable for data visualization.

Design and Procedure

Glyphs. We included another two common profile glyph designs in the study to extend the applicability of our results [52]. They are a small bar chart (we refer to this representation as Bar glyph) and a linear version of the star glyph, which we call MirrorBar glyph. By adding these designs we can investigate possible differences in similarity perception between directional (i.e., Star glyph) and positional laid dimensions (i.e., Bar and MirrorBar glyph). We additionally include a filled version of each design to examine whether variations of glyphs that are filled, reinforce the notion of a closed shape due to foreground/background contrast [105]. We conjectured that fill color may lead to shape rather than data similarity choices. In total we had the 6 glyph designs illustrated in Figure 4.27.

Task. We again used a synoptic task, where participants selected the most similar glyph compared to a stimulus glyph. Participants were shown a highlighted stimulus surrounded by another 8 glyphs in a $3 * 3$ matrix configuration. The positions of the surrounding glyphs were randomized around the stimulus. Again, we wanted to explore the notion of similarity and examine if some glyphs are naturally judged in a manner that approaches data rather than shape comparison. We thus gave no explanation as to what the glyphs represented and provided our participants with no training. Participants were free to interpret the word “similar” as they saw fit.

Data, Target Types and Dimensionality. Our data was generated as in Experiment 1, and again we tested *low* and *high* dimensionality. However, we included slightly different glyph choices to our participants, that we call ”Target Types” (they are no longer distracters, as there is no correct answer). To balance the selection likelihood between each target type, we included two of each shape similarity and two glyphs that were closest to the stimulus in data space (we refer to this kind of target as “data”). As a result we had 2 data, 2 rotated and 2 scaled versions of the stimulus, and 2 randomly generated targets.

Participants and Procedure. Our study was conducted on Amazon Mechanical Turk (AMT), inspired by previous graphical perception experiments [19, 85]. We accepted 185 participants in total, and subjects were paid $0.50$$ per Human Intelligence Task (HIT). Given the simple nature of our perceptual study, no qualification tests were required to complete our HITs. In accordance with AMT guidelines, however, only workers with 95% or more HIT approval rate were allowed to participate. Furthermore, we added control questions (3 in total) throughout the study, where one of the targets was identical to the stimulus and the answer was, therefore, obvious. We dismissed workers who did not get all the control questions correctly and their data was not included in the analysis. As a result we ended up with 108 participants (18 per fill type). Each participant worked on 4 trials for each variation and dimensionality, and viewed either the fill or the no-fill types. The order of presenting the glyph variations was randomized.

Overall our experiment consisted of

3	glyphs (*Bar, MirrorBar, Star*)	$*$
2	filling types (*Fill, No-Fill*)	$*$
3	contour variations (*D, D+C, C*)	$*$
2	dimensionalities (*high, low*)	$*$
4	repetitions	=
144	trials per participant	$*$
18	participants per glyph and fill type	=
2592	**trials in total**	

for a between subjects design (glyph and fill type being the between subjects factor).

Hypothesis

Given the results from Experiment 1, and our conjecture on filling, we formulated the following hypothesis.

H1: Given Exp 1, glyphs without contours and filling will promote data similarity comparison rather than shape, even when participants are unaware they are viewing data.

H2: Familiar glyphs such as chart visualizations (e.g. Bar glyph) will promote data similarity more than unfamiliar ones (e.g. MirrorBar or Star glyph).

H3: Low dimensionality will lead more often to data similarity judgements in cases where shape is not enforced (no contour, non-filled), even with no instruction about what the glyphs represent and what is similarity.

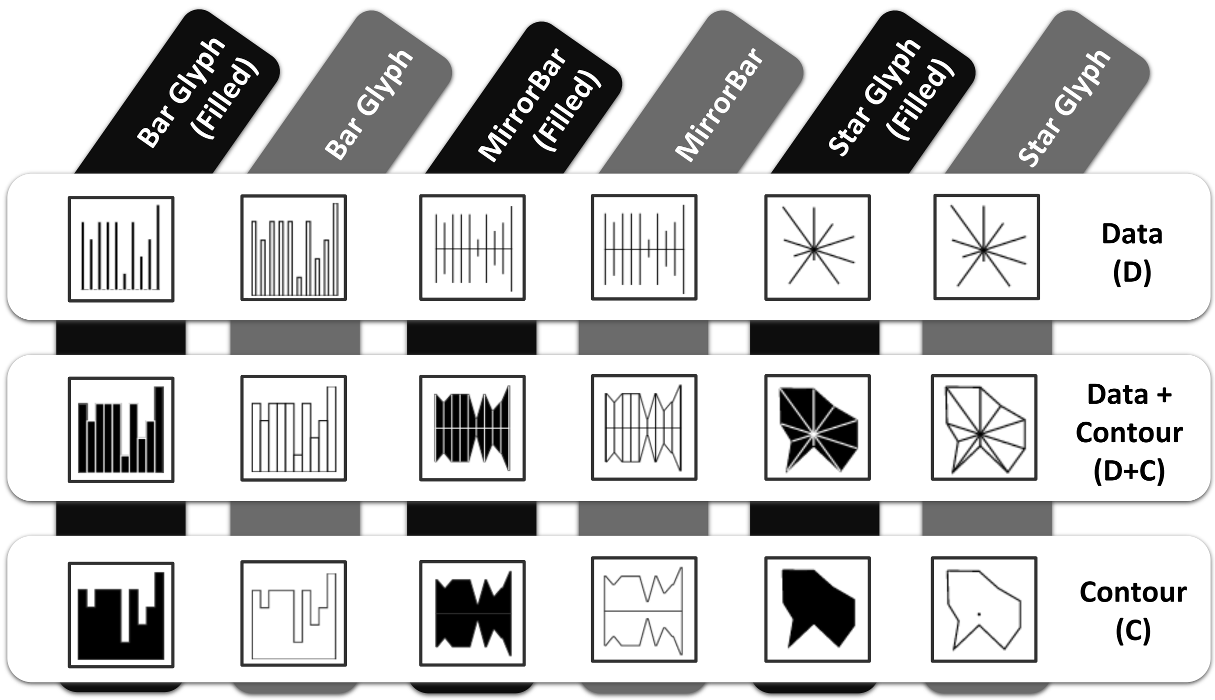


Figure 4.27: *Experiment 2 design space*: We enriched the design space from our previous study by adding bar chart like glyphs (i.e., Bar glyph) and a linear variant of the star glyph (i.e., MirrorBar glyph). All designs also appear in a filled version. The different design variations of the first study (i.e., D, $D + C$, C) are applied to all designs.

Results

We only report statistically significant results ($p < .05$) for the quantitative data. We used a non-parametric Friedman's test for the analysis of the selections between the glyph variations (within-subjects) and a Kruskal-Wallis test for comparisons between glyph designs (between group factor).

We first report overall averages for all selections made. For **Star** glyphs the most common selection was the data distracter (44.6%). The second was the rotated distracter (37.3%), particularly in higher dimensions and $C, D + C$ contour variations. The scaled distracter came third (17.8%), mostly in the $C, D + C$ filled versions of the Star glyph.
For the **Bar** the most common selection was the scaled distracter (69.5%), particularly in the high dimensional cases and filled versions of this glyph. Nevertheless, the data distracter was the most often selected one in the D variation.
For the **MirrorBar** the most common selection was the scaled (44.1%) then the data distracter (36.4%). Again, the data distracter was more common in low dimensions for D.

We next treat each distracter as a separate dependent variable but focus on results for *data* distracter selections, to help us compare glyph variations for data visualization. We first examine each glyph design and data dimensionality independently for the different variations (D, $D + C$,

C). We then compare glyph designs between them.

Bar glyph: For the Bar glyph, there was a significant effect of *variation* on *distracter* in the low ($\chi^2(2, N = 72) = 8.64, p < .05$) and high dimensional case ($\chi^2(2, N = 72) = 7.53, p < .05$) (see Figure 4.28). Pairwise comparisons showed that participants considered significantly more often the data distracter as similar for variation D in the low and high dimensional case (52.8%; 9.7% respectively), compared to both $D + C$ (34.7%; 1.4%, all $p < .05$) and C (31.9%; 0%, all $p < .01$).

The filled Bar glyph only showed a significant effect of *variation* on *distracter* for the low dimensionality ($\chi^2(2, N = 72) = 14.62, p < .001$). Again, the data distracter was selected significantly more often with variation D (50%) compared to $D + C$ (31.9%, $p < .01$) and C (26.4%, $p < .001$).

There was no significant effect of *filling types* for the Bar glyph design.

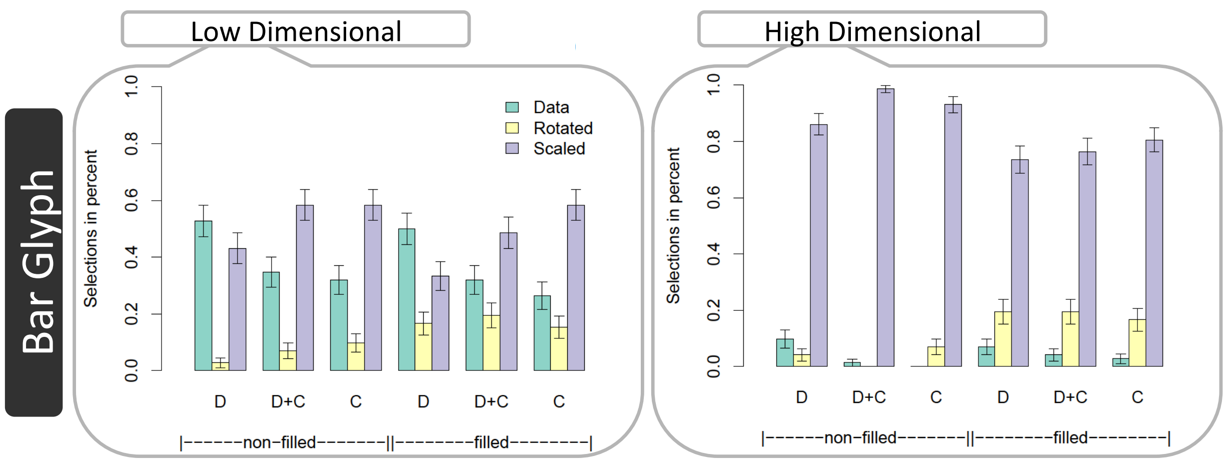


Figure 4.28: *Experiment 2 results bar glyph*: The bar charts illustrate the percentage of selections and the standard deviation for each factor.

Star glyph: The Star glyph showed a significant effect of *variation* on *distracter* for both low ($\chi^2(2, N = 72) = 8.21, p < .05$) and high dimensional cases ($\chi^2(2, N = 72) = 28.25, p < .001$) (see Figure 4.29). Post-hoc tests revealed a significantly higher selection rate for data distracters in variation D for the low and high dimensional case (75%; 62.5%) compared to $D + C$ (61.1%; 15.3%, all $p < .05$) and C (59.7%; 9.7%, all $p < .01$).

The filled Star glyph only had a significant effect of *variation* on *distracter* in the high dimensional case ($\chi^2(2, N = 72) = 17.33, p < .001$). Participants working with variation D selected significantly more often the data distracter (41.7%) compared to $D + C$ (11.1%, $p < .001$) and C (9.7%, $p < .001$).

When comparing the filled Star glyph with the non-filled version there is a significant effect on *filling types* with variation D in the high dimensional case ($\chi^2(1, N = 144) = 6.22, p < .05$). Participants working with the non-filled Star glyph selected the data distracter significantly more often (62.5%) compared to the filled design (41.7%, $p < .05$).

MirrorBar glyph: We found no significant results for the MirrorBar glyph, but the filled MirrorBar glyph showed a significant effect of *variation* on *distracter* in the low density case

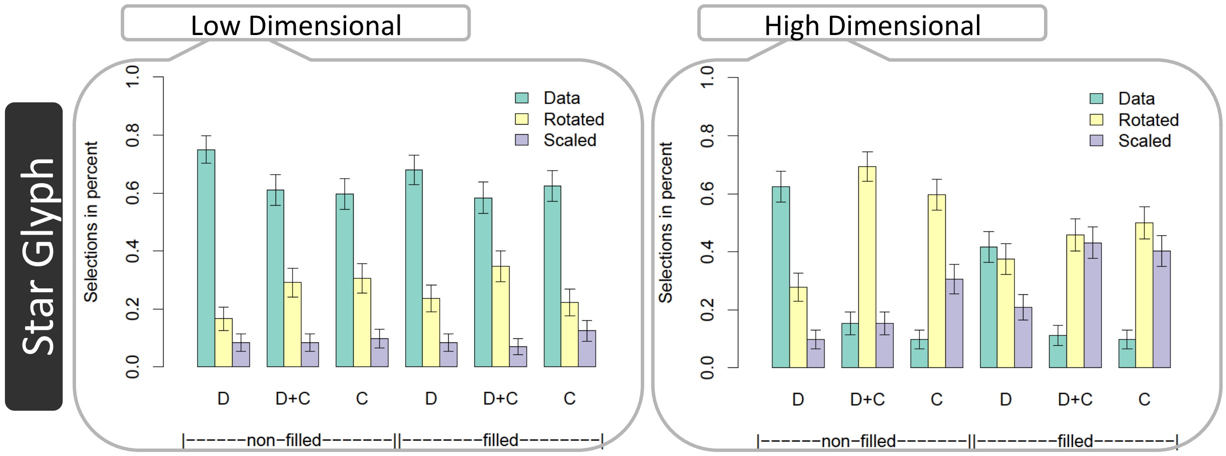


Figure 4.29: *Experiment 2 results star glyph*: The bar charts illustrate the percentage of selections and the standard deviation for each factor.

$(\chi^2(2, N = 72) = 6.77, p < .05)$ (see Figure 4.30). Pairwise comparisons revealed that the data distracter was selected significantly more often with variation D (52.8%) than with C (33.3%, $p < .01$).

A significant effect on *filling types* can be seen for the MirrorBar with variation C in the low dimensional condition $(\chi^2(1, N = 144) = 8.05, p < .01)$. With the non-filled version participants tend to select the data distracter more often (56.9%) compared to the filled design (33.3%, $p < .01$).

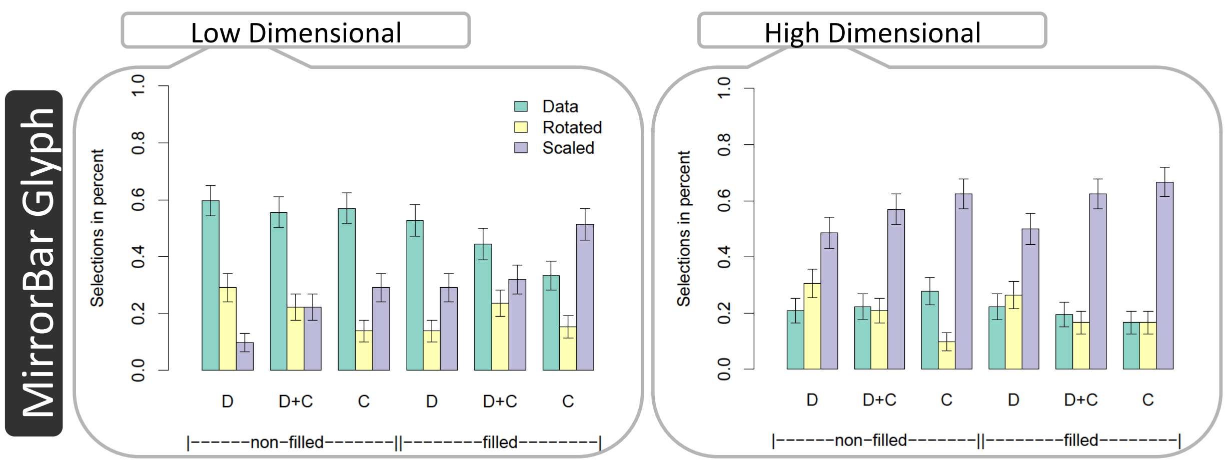


Figure 4.30: *Experiment 2 results mirrorBar glyph*: The bar charts illustrate the percentage of selections and the standard deviation for each factor.

All glyphs: When comparing the different glyphs (i.e., Bar glyph, Star glyph, MirrorBar glyph) with each other we only consider the version of each design (i.e., filled or non-filled), which performed better. We have thus chosen the common non-filled Star glyph, the non-filled MirrorBar and the filled Bar glyph. Although there was no significant difference between the filled and non-filled versions of the Bar glyph, we consider the filled version as it is most often seen in print in with fill colors. We also focus on the non-contour variation D that gave the best results.

Variation D: There was an overall effect of *glyphs* on *distracter* in both the low ($\chi^2(2, N = 216) = 9.62, p < .01$) and high dimensional case ($\chi^2(2, N = 216) = 56.95, p < .001$). In the low dimensional case, pairwise comparison showed a better performance for the Star glyph (75%) compared to the Bar glyph (50%, $p < .01$). When switching to the high dimensional condition the Star glyph still performed best (62.5%) compared to the MirrorBar (20.8%, $p < .001$) and the Bar glyph (6.9%, $p < .001$), with the MirrorBar having a higher performance compared to the Bar glyph ($p < .05$).

Summary and Discussion

We found strong evidence that glyphs without contours promote data similarity comparison rather than shape (H1). This was particularly the case for Star glyphs across dimensionalities, and less for Bar and MirrorBar in low dimensionalities. Data similarity judgments were also more common in the non-filled variations of Star and MirrorBar (H1). Thus overall factors enforcing perceptual unity of shape[105], such as contour containment and emphasis between glyph and background (figure and ground), lead viewers to naturally make shape judgements of similarity rather than data, while open variations of the glyphs lead to similarity choices closer to data comparisons, even without any training.

Even though the lack of contours increases data judgements, the majority of selections for the Bar and MirrorBar glyph, especially in high dimensionality, were based on shape similarity, mostly scaled. This is contrary to our expectation that glyphs resembling familiar visualizations will lead to data similarity (H2). We feel this finding merits further study. One explanation for this effect, is that we are accustomed to reading such charts to see trends (e.g. increasing and decreasing tendencies) and thus scaled versions look similar as they preserve trends. On the other hand, for Star glyphs the second most popular selection in high dimensions (after the data similarity), was the rotated version of the stimulus. This is not unexpected, as the Star glyph layout is directional and thus small rotation shifts look indeed very similar.

For all glyphs in low dimensionality, when no contours and no filling were present, data similarity was more common than shape, but in higher dimensions shape similarity was more common for Bar and MirrorBar (H3). We note again that in this study participants were never told they were viewing data visualizations, they were just asked to find the most similar shapes without further instructions. Thus, our results indicate the natural tendency of people to judge glyphs instinctively in a more “data-centric” manner in low dimensionalities, and in high ones when factors that enforce coherent shapes are absent. It is clear that with training we can further enforce data similarity judgments–but given that some glyphs and glyph variations seem to be naturally well suited for data judgments, we focus on those star glyph designs and try to further improve their performance with small design variations.

4.3.3 Experiment 3: Improvements for Star Glyphs

The first experiment showed that people judge data similarity with non-contour designs more accurately while the second experiment showed that non-contour designs also lead to data similarity judgments to be made more naturally. Yet, accuracy in the high-dimensional case was quite low for all main design variations we tested previously. In this last experiment, we thus explore whether we can improve the accuracy of data similarity judgments by adding simple reference structures—tickmarks and grids—to the designs. We focused on static reference structures to learn how much these general approaches would aid data comparison before considering the design of interactive aids.

Star Glyph Reference Structures

Reference structures such as grids and tickmarks are frequently recommended for data charts to aid in relating content to axes [107]. We, thus, hypothesized that they could provide similar reading aids for star glyphs despite their smaller footprint. Tickmarks and grids use two different types of reference mechanisms. While tickmarks add information to each individual data line only, grids connect the overall glyph design. While there are many different ways to draw grids and tickmarks we settled on the following designs:

Tickmarks T: Whenever a data line exceeds a certain threshold we draw a short orthogonally oriented tickmark on the data lines using the same stroke color. Tickmarks are spaced to be 17 pixels apart. The resulting $D + T$ glyph (see Figure 4.31) resembles the snowflake glyph previously mentioned in literature [16] and is also close to how tickmarks are used on axes in many data charts.

Grid G: We draw three circles in the background of the glyph using a gray value of #ccc in RGB color space chosen according to design considerations by Bartram et al. [11]. The circles are spaced 16.6 pixels apart. The resulting design resembles radar graphs or spider plots [189]. As an alternative we considered drawing a gridline at the end of each data line. Doing so would create an underlying texture that could help to identify the overall data distribution across all dimensions. Yet, we chose not to use this design as this texture can be misleading since rotated star glyphs with similar data values would have the same texture, although they have entirely different data values.

Of course, the readability of glyphs could further be improved by adding double encodings (e.g., additionally using color to distinguish dimensions or data values), dimension ordering [148], or sorting the glyphs on the display. Yet, all of these encodings have limitations: use of color is limited to glyphs with a small number of dimensions, dimension ordering may not improve legibility for a large number of variable glyphs in a small-multiple setting, and sorting glyphs may disrupt a pre-defined layout based on other meta-data such as time. We, thus, did not consider these encodings for the study.

Design and Procedure

Glyphs: We tested the two star glyph variations that performed best in the first experiments: the data-only glyph (D) and the star glyph with data lines and a contour line ($D + C$). The

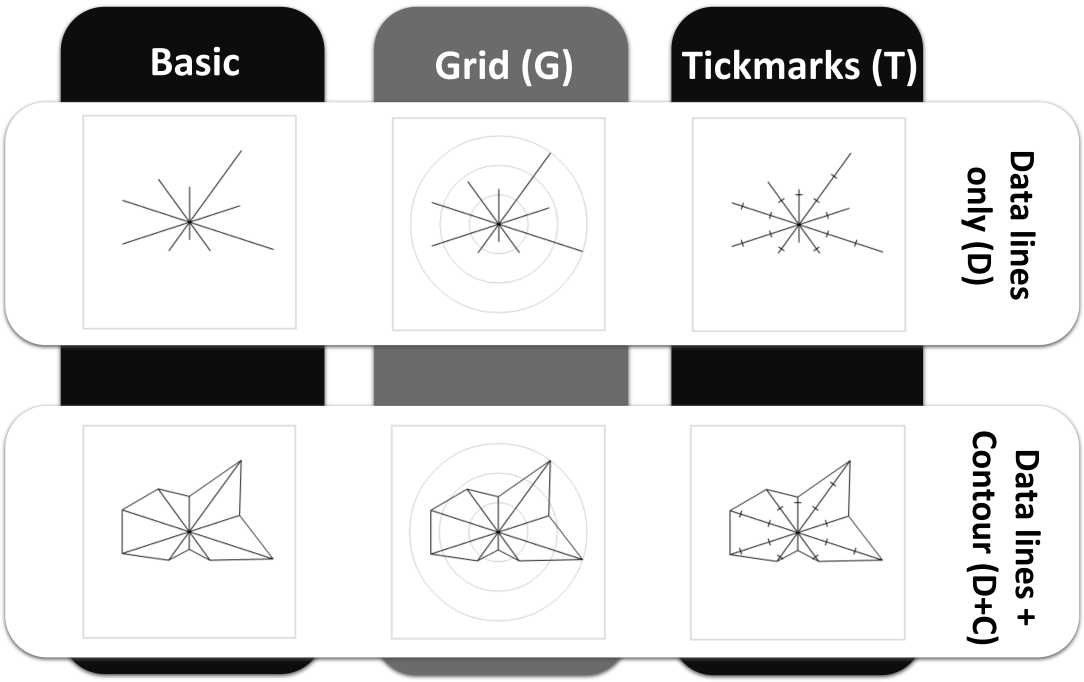


Figure 4.31: *Experiment 3 design space*: We have chosen the star glyph only with data whiskers (D) and with an additional contour line ($D + C$) and applied tickmarks (T) and gridlines (G) to these designs.

reason for discarding the contour only design (C) is the bad performance for previous similarity judgments, the lack of ability to place tickmarks, and the minimal number of real-world examples of this glyph type in use.

For baseline comparisons we kept the originally tested versions of the star glyph (D, $D + C$) and added two types of reference structures (T, G). The experiment, thus, compared the six different designs (D, $D + T$, $D + G$, $D + C$, $D + C + T$, $D + C + G$) in Figure 4.31.

Participants: We recruited 12 data visualization experts (3 female). The age ranged from 23–40 years in age (mean (29.75) & median age (30)). All participants reported normal or corrected-to-normal vision. All experts focused during their studies on data visualization (4 Bachelor; 5 Master; 3 PhD) or a related topic and were familiar with reading data glyphs. They had not participated in the first study.

Task and Procedure: Participants completed data similarity search trials with all 6 designs. The order of the designs was randomized using a latin square. For each design there was a short introduction of the visual encoding and the similarity search task with 5 test questions.

The participants had to complete those simple test trials with 80% accuracy in order to continue the experiment. The purpose of the test was to first check the participants' ability to read the visual encoding of the glyph and second to test their data similarity judgments. All participants passed the test section. The introduction was followed by 4 training trials to help the participants develop a strategy for solving the task. For training trials, the correct answer was shown to participants after they had made a choice. Finally the four study trials were shown without any visual feedback of the correct answer.

The experiment took place in a lab setting using a 24" screen with a resolution of $1920 * 1200$ pixels. The experimenter was present during the study. After the study, 11 of the 12 participants filled out a questionnaire for subjective feedback on aesthetics of the designs and strategies used to answer the questions.

Data, Distracters and Dimensionality: Since participants were already $\approx 80\%$ correct in the low dimensional condition in Experiment 1, we only used high-dimensional glyphs in Experiment 3. We generated the data the same way as in Experiment 2 and balanced selection likelihood between distracters. To reduce the chance of a successful random guess we generated only one data point closest in data space (target) and another one second closest in data space (alternative) as in Experiment 1. The experiment included 2 rotated, 2 scaled, 2 random, 1 alternative and 1 target glyph. The stimulus was highlighted and positioned in the middle of the $3 * 3$ matrix as in the two previous experiments. The distracters were randomly arranged around the stimulus.

Overall our experiment was a within-subjects design with the following factors, participants, and trials:

1	glyph (*Star*)	$*$
2	contour variations (*D, D+C*)	$*$
3	improvements (*Basic, T, G*)	$*$
4	repetitions	=
24	trials per participant	$*$
12	participants per glyph	=
288	**trials in total**	

Hypotheses

Based on our previous experiments and the frequent use of reference structures to aid chart reading, we tested the following hypotheses:

H1: Tickmarks (T) in star glyphs improve the accuracy of data similarity judgments for both (D) and ($D + C$) variations compared to the variations without the tickmarks. The additional anchor points help to better read and compare line distances.

H2: An underlying grid (G) in the background of the star glyph provides additional orientation and facilitates more accurate comparison of data values for both (D) and ($D+C$) variations than the variations without the grid.

H3: The contour variation $D + C$ benefits more from the additional reference structures than the D variation since contour has previously shown to lead to shape comparison rather than data similarity comparisons.

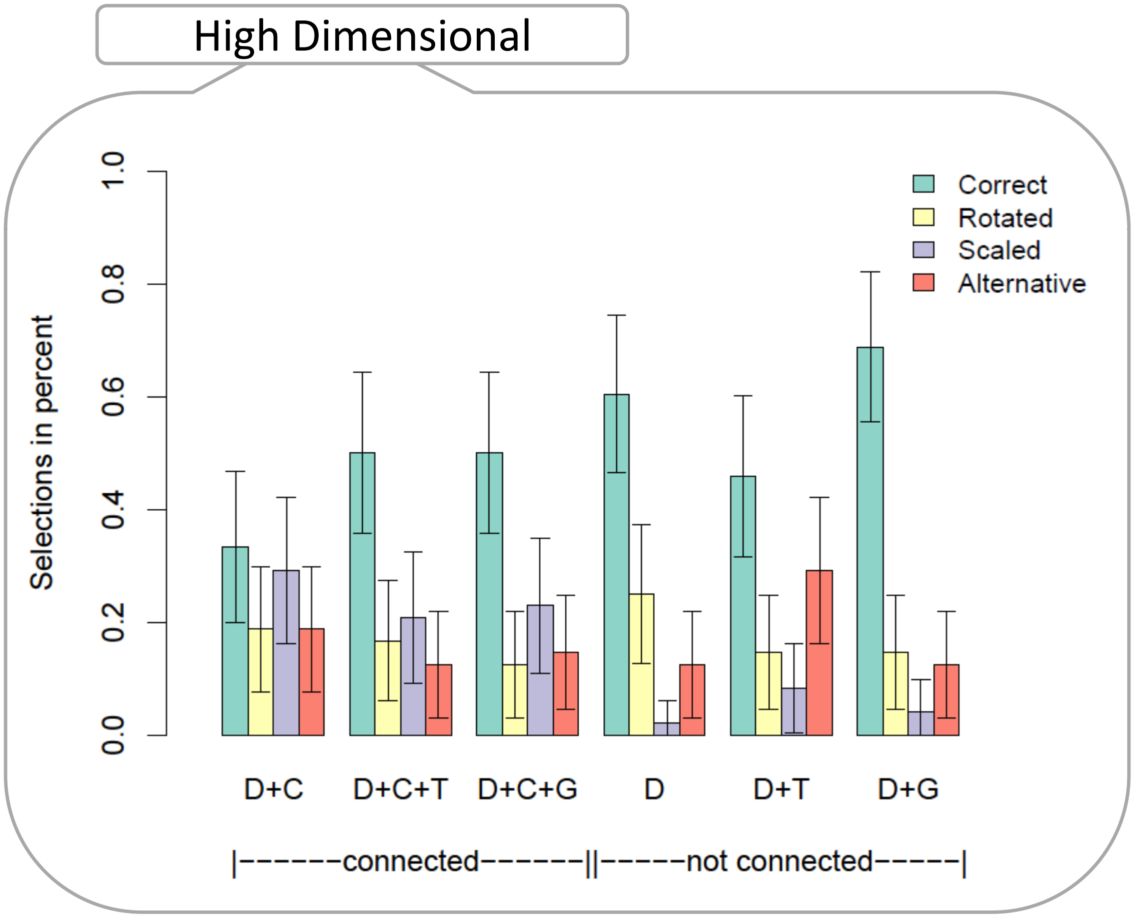


Figure 4.32: *Experiment 3 results* of the percentage of selections and the standard deviation for each factor. Design improvements (T, G) do not significantly increase the accuracy of the two star glyph variations $(D + C, C)$.

H4: Completion time is higher for designs enriched with reading marks (T or G). The viewer has to invest more mental effort to process the additional visual information.

Results

Similarly to Experiment 1 we used a non-parametric Friedman's Test on the data to analyze accuracy, and a one-way ANOVA for the completion time. We only report statistically significant results ($p < .05$).

The overall accuracy was 51.4%, with designs with grids (G) being more accurate (59.4%), followed by the tickmark designs (T) (47.9%) and then designs without additional marks (46.9%). There was a statistical trend for different types of reference structures on accuracy ($p < .1$), with glyphs with grids being more accurate than with tickmarks. There was no difference between designs with reference structures and the baseline design.

Next, we compared the different glyph variations without contour (D) and with contour ($D+$

C). As in Experiment 1, participants were significantly more accurate with variation D (60.4%) than when the contour was present $D + C$ ($33.3\%, p < .01$).

Reference structures on glyphs without contours (the D glyphs) did not significantly improve accuracy over the glyph without the reference structure. Participants were 60.4% accurate with D, 68.8% accurate with ($D + G$), and 45.8% accurate with ($D + T$). Nevertheless, we note that the mean accuracy of the ($D + G$) variation is indeed higher than for D only. We also found that for the two variations using reference structures, grids ($D + G$) were significantly more accurate than tickmarks ($D + T$) ($45.8\%, p < .05$).

For the contour variations, we have a statistical trend ($p < .1$) indicating that the accuracy of both the contour variation with a grid ($D + C + G$) and the one with tickmarks ($D + C + T$) tend to be more accurate (both 50%) than that of simple glyph with contour ($D + C$) with accuracy 33.3% ($p = .06$ and $p = .08$ respectively).

Looking at differences across variations, we also found that $D + G$ (68.8%), which had the highest overall mean accuracy, performed significantly better than $D + C$ ($33.3\%, p < .001$) and had a statistical trend to perform better than $D + C + G$ ($p = .1$) and $D + C + T$ ($p = .8$).

The mean number of selections per distracter type are shown in Figure 4.32. We found a significant effect of *variation* on *distracter* ($\chi^2(5, N = 48) = 12.68, p < .05$). Participants using variations with contour lines most often selected the scaled distracter (24%) followed by the rotated (16%) and the alternative (15%) distracter. For the non-contour variations participants chose the alternative and the rotated distracter equally often (18%) followed by the scaled distracter (5%).

No significant results can be reported for the completion time, thus we cannot confirm that additional marks influenced comparison times. However, participants needed approx. $2sec$ longer when working with designs using additional marks. Average completion time was $22sec$ per trial ($D = 21.7sec$, $D + G = 24.8sec$, $D + T = 26.1sec$, $D + C = 17.9sec$, $D + C + G = 21.5sec$, $D + C + T = 22sec$).

The questionnaire showed that the glyph variations with contours ranked highly amongst participants' aesthetic preferences. The mostly strongly preferred glyph variation was $D+C+G$ (5/11 participants), followed by $D+C$ (3/11 participants). Interestingly, no participants preferred the D variation even though its mean accuracy (60.4%) was higher than $D + C + G$ (50%). Participants also ranked the D variation as hard to use (median=6 on a 7-point Likert scale) with all other designs ranking at least between median 4–2. The $D+C+T$ and $D+C+G$ variations were both found easy to use (median=2). We report on the results of the questions regarding strategy in our discussion section.

Discussion

Adding reference structures to the star glyph did not have the effect on accuracy we were expecting for our data similarity search task. Additional anchor points on the data line (i.e., tickmarks) did not significantly improve the comparison of data points. Therefore, we cannot accept H1. Nevertheless, there was a statistical trend indicating that an overall reference in the background (i.e., gridlines) may increase accuracy, especially in the case of contour star glyphs, providing some evidence for H2.

This lack of strong significant effects is surprising, especially given that most participants mentioned in the questionnaire that for the simple star glyph D, gridlines (81%), and to a lesser extent tickmarks (72%), helped them find the most similar data point. Although the mean accuracy for the $D + G$ variation was indeed higher, the effect was not significant, perhaps due to the already very good performance of the D variation. The value of gridlines and tickmarks in general may warrant further research. As Few notes [62], gridlines may be useful only in specific cases, e.g., when small differences have to be compared. Therefore, it is possible that for other tasks, such as direct lookup, these additional reference marks could help more strongly.

For the star glyph with contour ($D + C$), only 54% of our participants reported using tickmarks and 36% gridlines to complete the task. From their reports they felt (erroneously) that glyphs with contours are easier to compare and, thus, did not make conscious use of the additional improvements. Thus, in the contour case, participants were not only more error prone, but also misled to feel confident in their choices, ignoring the marks that could help them improve their performance. Nevertheless, it is highly likely that the addition of reading marks was taken into account, even if unintentionally, explaining the trend we see for both the tickmark and grid variation to be more accurate than simple contour glyphs (H3).

Finally, we could not confirm H4 due to a lack of significant results when comparing task performance time.

Even though participants using variation (D) performed very well, it is interesting that they did not like this design variation. On a 7-step Likert scale 63% of the participants rated the design with either 6 (difficult to use) or 7 (very difficult to use). Most participants (46%) preferred the star glyph with contour and gridlines, with only 1 participant rating it with a 5 (slightly difficult to use) and the others with 3 or better.

Given the results of this experiment the benefit of using reference structures for star glyphs is limited. Especially since in real world scenarios when multi-dimensional glyphs are projected to two dimensional surfaces, there is the possibility of over-plotting, and adding marks or gridlines could worsen this effect due to the additional ink introduced.

4.3.4 Design Considerations

With the results gained from the analysis and discussions we derive the following design considerations.

When judging data similarity avoid contours in glyph designs. Viewers have a natural tendency to judge data similarity in star glyphs without contours. In all our experiments viewers were tricked into making shape-based, rather than data-based judgments when using contours. This is especially true if glyphs in the visualization are scaled or rotated versions of each other.

For low number of dimensions (around 4) any glyph variation can safely be used for data similarity judgments. In the first and second experiment viewers naturally leaned towards data similarity for each glyph variation in low dimensions, even without training.

When there is a need for contours, add data lines to the design to strengthen data similarity judgments. Participants independent of glyph design (fill or no-fill) judged data similarity better using the $D + C$ variation compared to C in the first two experiments. Although, there was no statistical significance, mean data comparisons for contour + data variations were always higher than contour only.

When there is a need for contours, the designer can decide whether or not to use fill color. Our Experiment 2 gave no indication that fill color degrades the performance of glyphs with contour.

When clutter is an issue avoid reference structures in non-contour star glyphs for similarity search tasks. Results of Experiment 3 illustrate that even though participants preferred using tickmarks or grids they did not perform significantly better with them, especially for glyphs without contours. Nevertheless, there is a statistical trend that shows that tickmarks and grids improve glyphs with contours.

If references are required use grids rather than tickmarks. Independent from the design (i.e., with or without contour) gridlines always increased mean accuracy, which is not true for tickmarks.

4.3.5 Limitations

As stated at the beginning, we focused in our study on the similarity perception of multi-dimensional data points and, therefore, the comparison over all displayed dimensions. Pre-studies suggest that different tasks like reading exact data values of individual dimensions may yield different results (e.g., additional orientation like grids or tickmarks should improve performance) [154]. However, since glyphs are meant to be compact representations of data points their main advantage is to give an overview of the data rather than a highly accurate representation of single data values. This also influenced our decision to only investigate a value range of 6 different data values. Increasing this number would result in too small visual differences, which due to the compact glyph representation could be hardly perceived. In order to conduct a controlled lab experiment we restricted ourselves to variations of star glyphs. Generalizing the results for designs not tested in our studies must be done with caution.

4.3.6 Conclusion

We investigated the effect of contours on the perception of similarity for star glyphs. In a first controlled experiment with 24 participants, we examined the influence of contours for novice and expert users. We found that experts can be tricked into making similarity judgments based on shape, rather than data closeness, when viewing glyphs with contours. For novices the effect was less pronounced. To better understand how people naturally judge similarity, we conducted a second online study with 108 participants and asked about intuitive notions of similarity. We found that removing contours and fillings from star glyphs, naturally increased the perception of data similarity (rather than shape) even when viewers were not trained or aware they are viewing data.

As a next step we tried to improve the accuracy in judging data similarity. We added two types of reference structures to the star glyph, gridlines and tickmarks, and tested these alternatives in a third experiment. Surprisingly, the star glyph without contour line and reference structure still performs best for similarity search tasks. Based on our findings we provide a set of glyph design considerations, the most important being that visualization designers should avoid contours when representing similar data points to analysts.

In summary, our work has provided insights as to the effect of contours on similarity perception for star glyphs. Similarity perception is an important task especially since glyphs are mostly used for quick overviews, and to detect trends and similarities [195], rather than to provide highly accurate value representations [73]. Other tasks performed on glyphs, however, such as exact data value reading, may yield different results from ours, e.g., adding grids or tickmarks could improve performance [154].

Given our experimental results and our provided guidelines, we would like to focus on two future research directions. First, we would like to examine whether our findings can be applied to different glyph designs (e.g., profile glyphs [52]), as it is unclear if contours promote shape similarity rather than data in glyphs that already resemble familiar data charts. Thus we can derive a more generalized set of design considerations. Second, based on our results on possible pitfalls in data similarity judgments, we plan to introduce a more task specific training, focusing on rotated and scaled distracters that seem to mislead viewers the most. Given our results, both novices and experts would profit from such specific training, especially when using glyphs with contours.

4.4 Summary

This chapter 4 introduces a review on multi-dimensional data glyph designs used in practice. The papers are organized according to Ward's categorization simplifying the search for individual designs. In contrast to data glyphs for time-series data more uniquely shaped designs do exist. This is due to the different nature of the underlying data and, therefore, the different analysis tasks.

Unlike designs for time-series data metaphoric designs are used more often [35, 142, 170, 208]. This is also due to the possibility of having more flexible designs. The survey has shown that no metaphoric data glyphs for environmental data exists. This research gap is closed by introducing the *leaf glyph* as a suitable representation. Its usefulness is shown in an analysis of the forest fire dataset from the UCI machine learning repository. Since the *leaf glyph* design is a one-to-one mapping according to Ward's classification a lot of possibilities for a visual mapping do exist. Further studies need to be conducted to come up with the most effective mapping strategy, which is not a trivial task as can be seen in the many studies conducted about variations of Chernoff faces.

To shed some more light on the performance of the well-established star glyph a controlled user study was conducted to reason about changes in performance when switching between designs. Results indicate not to use the surrounding contour line for similarity search tasks to increase performance. The suggested visual improvements of the star glyph did not help to further decrease the error rate. Against intuition the simple star glyph without a contour line is best for performing a similarity search task.

Chapter 5

Design Considerations

Throughout my dissertation I have conducted several controlled user studies to shed more light on changes in performance when switching between different data glyph designs. As a result, I have suggested various design considerations to help practitioners and researchers in choosing the most appropriate data glyph alternative.

In this chapter 5, I will summarize the design considerations proposed in literature and provide the interested reader with an overview on how to design effective data glyphs. Additionally, I will indicate whether a guideline is the result of an experiment or a subjective opinion from researchers. This may help to better assess the quality of each guideline. The design considerations are categorized according to different analysis tasks.

5.1 Elementary Analysis Task

Simple lookup tasks are typical examples of elementary analysis. The analyst is focused on individual parts of the data glyph and tries to read data values as accurately as possible. Therefore, the analyst needs to first identify the respective dimension and in a next step extract the necessary information.

Lookup Data Values: An important aspect of reading data values is the support of the quantitative analysis in the attentive phase [157]. Therefore, elementary tasks not focusing on the pre-attentive phase should be visually supported with a glyph legend showing the different scales or mapping criteria. For example, having a color legend is crucial when encoding data values with color to incorporate the range of values. This general design consideration proposed by Ropinski et al. is only based on observations and wasn't confirmed in a user study. For reading single data values Ware proposes to "*Ideally, use glyph length or height, or vertical position, to represent quantity. If the range of values is large, consider using glyph area as an alternative. Never use the volume of a three-dimensional glyph to represent quantity.*"[1]. For this design consideration Ware refers to studies conducted by Ekman and Junge [57] and Cleveland and McGill [43]. In

[1] [195], page 169

general representing quantity can be effectively done by mapping the data also to: "*size, lightness (on a dark background), darkness (on a light background), vividness (higher saturation) of color, or vertical position in the display*"[2]. However, as the studies conducted in section 3.3 have shown the ranking of Cleveland and McGill's visual variables seems to hold for smaller glyph representations [43]. At least this is true for linear data glyph designs where position/length encodings should be preferred to a color encoding. Additionally, Ropinski et al. suggest to avoid perspective projections when using the glyph size to convey data values [157].

Interestingly, color encodings seem to be a reasonable choice for circular designs because with position/length encodings the mental rotation required for comparisons affects participants more strongly than color comparisons that can be conducted without a common axis. For linear designs it is important to arrange the different dimensions along a common baseline. Three studies indicate that linear designs are most effective when the dimensions are aligned along a common baseline [28, 29, 95].

Having an intuitive mapping based on semantics helps to read data values [17]. For example, encoding low temperature values to a blueish color and high values to reddish colors facilitates the information extraction. Important variables should be encoded redundantly to reduce the risk of a possible information loss [17]. This design consideration is based on thoughts and ideas discussed by Lie et al. and Ropinski et al [115, 156].

Detect Dimensions: For detecting individual dimensions independently within a data glyph design the orthogonality and the separability of the different glyph components must be pursued [17, 195]. This design consideration is based on thoughts and ideas discussed by Lie et al. [115] and studies conducted by Garner [78] and can be applied to data glyph designs for temporal as well as multi-dimensional data. For time-series data circular layouts seem to be more appropriate compared to linear ones. As the experiment conducted in section 3.3 has shown participants were able to read the correct point in time more accurate with circular glyph designs. This may be true because participants tend to read circular designs like clocks. However, I assume that this clock metaphor only works for a certain number of dimensions, which can be mapped to the layout of common clocks. Additionally, it is important to use sufficient space for communicating the different dimensions since there is a significant drop in performance when increasing the number of dimensions keeping the size of the glyph stable.

Conclusion: Based on the aforementioned guidelines two different fundamental data glyph designs should be pursued. For reading data values a linear position/length encoding meets the predefined criteria Figure 5.1. When detecting certain dimensions in time-series a color encoded circular design is best.

[2] [195], page 168

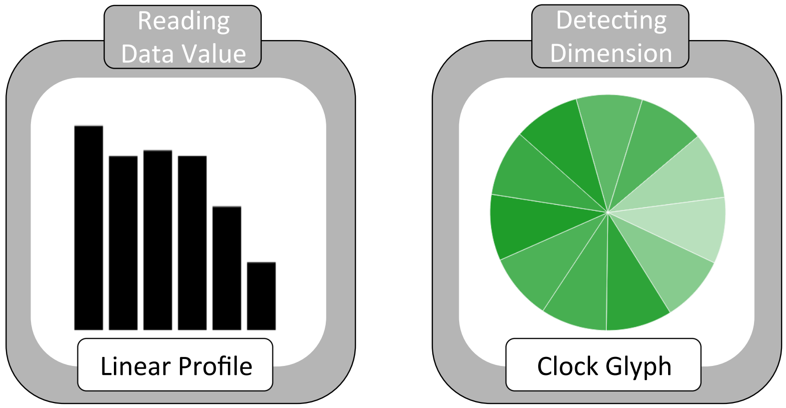


Figure 5.1: *Designs for elementary tasks*: When reading data values a linear position/length encoding should be preferred. For detecting specific dimensions radial designs with a color encoding are more appropriate.

5.2 Synoptic Analysis Task

Analysts performing synoptic tasks focus on grouping or comparing data points according to their similarity, trying to detect trends, or searching for elements. Therefore, it is important to consider the entire data glyph and not just single dimensions. For all tasks, these "*glyph shapes should be unambiguously perceivable independent of the viewing direction.*"[3]. This design consideration is based on observations and needs to be further evaluated. However, it seems reasonable especially in three dimensional visualizations that the viewing angle may change. In such a case the glyph needs to be readjusted to match the new position/angle of the analyst's perspective. Otherwise the perception of data will be distorted. An additional halo surrounding data glyphs in three dimensional space facilitates depth perception and helps to distinguish between individual glyphs [17, 115]. In order to detect even minor shape changes, which is beneficial for all tasks, Borgo et al. suggest to use simple symmetric shapes like the wings of InfoBugs [17, 40].

Similarity Search: For similarity and grouping tasks, a surrounding contour line has a big influence on the perception of data glyphs. The experiments conducted in section 4.3 investigate the influence of a contour line on star glyphs. When data points should be compared according to their shape similarity adding a contour line to data glyphs is beneficial. However, for data similarity judgments the contour line should be removed. This design consideration is especially important for higher number of dimensions. Till now it is not possible to tell whether a linear or a circular layout is more appropriate. Only one study recommends that linear layouts should be preferred [126].

A suitable alternative representation are faces. Especially for similarity judgments faces performed well [93, 126, 199]. The more detailed or realistic the faces look like, the better [65].

[3] [157], page 399

Although they are currently not used in practice designers should rethink using faces more often.

Trend Detection: For trend detection tasks multiple data variables should be encoded with the same visual variable to improve the comparison of different data variables for a single data element [17]. Therefore, one-to-one mapping strategies should be avoided as discussed by Ward [192]. Based on the outcome of the systematic review in chapter 2 linear designs outperform circular arrangements [73, 126]. Therefore, simple line chart glyphs are an appropriate choice for trend detection tasks. Additionally, results from the experiment introduced in section 3.3 indicate that for time-dependent tasks sufficient space should be assigned to the designs. As a consequence, data glyphs must grow in size with an increasing number of dimensions. Especially, for color encodings sufficient space is mandatory to avoid blurring effects and be able to extract information. Circular designs profit from the additional space near the circumference. As a result the single slices depicting the different dimensions can be read more easily, which has partially been proven in section 3.3.

Visual Search: Only little advice exists, which glyph design is best for visual search tasks. One quantitative experiment concludes to use star glyphs without a contour line compared to faces or circular color encodings. However, if three dimensional visualizations are possible Forsell et al. recommend to use surface glyphs, which outperformed star glyphs without a contour line and also linear color encodings [67, 68].

It is interesting to note that faces were often used in visual search tasks, however, not compared to alternative designs. Therefore, it is not possible to draw general conclusions. An interesting fact is that faces cannot be pre-attentively identified [171, 173].

Metaphors: Whenever possible a metaphoric design should be preferred. Compared to abstract data glyph designs, metaphors performed always better [65, 73, 91, 170]. This is true for similarity search and lookup tasks. Of course, additional studies need to be conducted to further generalize these results.

Conclusion: The performance of data glyphs changes according to the underlying task. Therefore, different glyph designs need to be considered (see Figure 5.2). For similarity search or grouping tasks star glyphs without a contour line are a suitable choice. Detecting trends is easiest with simple line charts and three dimensional surface glyphs help to perform a visual search.

5.3 Glyph Placement Strategy

According to Ward there are basically two different glyph placement strategies [191]. Data-driven layouts and structure-based arrangements. Data-driven layouts make use of the data values to position data glyphs. This can either be a direct usage e.g., in a scatterplot, or a computation based on these values like a two-dimensional projection. Structure-based arrangements make use of inherent data characteristics like hierarchies, temporal sequence, geographic location, or

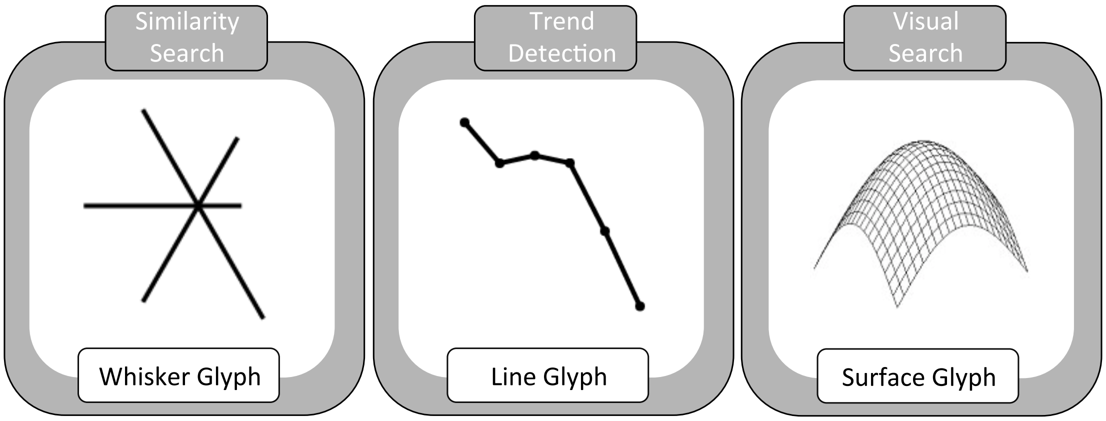


Figure 5.2: *Designs for synoptic tasks*: Different designs should be used for different tasks.

any kind of ordering. Independent from these two placement strategies Ropinski et al. further suggests to avoid unwanted glyph aggregations in image space [157]. Due to the layout and the distribution of the data several data glyphs might be positioned in a small area. Therefore, overplotting may occur in these dense regions. In such a case different relaxation procedures or jitter techniques should be applied to avoid these effects. This can be done in a static approach by slightly repositioning the data glyphs or interactively with different kinds of lens based techniques.

Since glyph designs can be flexibly arranged on the screen and used in various contexts understanding the different interconnections is crucial. Results from a study conducted by Martin et al. indicate that the reading capabilities of a data glyph do not change according to the background information [123]. Therefore, designers can think of showing more detailed context information without influencing the performance of glyphs. Of course, this suggestion is only based on one experiment and has to be considered with caution.

5.4 Summary

This chapter was meant as a summary of design considerations to guide practitioners as well as researchers to a data glyph design most suitable for a certain analysis task and dataset. The design considerations are based on the results of quantitative experiments but also on thoughts and ideas from researchers. Since there is still a lack of evaluations to be conducted for certain designs and presentation settings this guidance in designing data glyphs is not exhaustive. Most of the design considerations mentioned are based on results from quantitative experiments conducted under strictly controlled conditions making it difficult to generalize their outcomes.

Chapter 6

Conclusions & Future Research Directions

This thesis has shed more light on the performance and usage of data glyphs in information visualization for different analysis tasks and datasets. Therefore, several contributions in the area of information visualization and visual analytics were provided. At the beginning, a common definition of the term ”data glyph” was introduced to establish a common understanding of the overall topic. Clarifying the usage of the term was necessary to avoid contradictions or misinterpretations of the whole topic. Based on this definition the literature was systematically reviewed to summarize existing research and extract the insights of quantitative experiments to formulate design considerations (chapter 2). Moreover, open research gaps were identified within this survey, which were partially closed by conducting several controlled user studies (section 3.3 & section 4.3). The insights gained from these experiments contributed to a catalog of design considerations comprising former conclusions and new findings (chapter 5). Since results from user studies suggest that metaphoric designs perform well, two new data glyph designs were introduced, which make use of metaphors to encode the underlying data (section 3.2 & section 4.2). In the future, more evaluations need to be conducted based on the research gaps revealed in the initial survey. Additionally, some contributions introduced throughout this thesis would profit from a more in-depth analysis. In the following paragraphs, I will discuss possible research directions for future work structured along the thesis outline.

Definition of data glyphs: In the introduction of this thesis I raised awareness that different definitions of the term ”glyph” in the context of information visualization do exist. By contributing a more general definition with a summary of ideas and concepts from various sources I wanted to establish a common understanding of the term.

However, this definition is partially build on subjective opinions about certain characteristics of data glyphs. There is potential to further analyze the different properties of data glyphs in a more structured way. In the future, a qualitative user study with information visualization experts should be conducted comparing common charts with abstract data glyphs. As a factor the level of detail for certain visual characteristics is varied and participants have to tell whether they call the representation a data glyph or a chart. The threshold when the experts change their opinion would be an indicator for the respective visual feature being an important part of the definition of a data glyph.

Systematic literature review: Literature about quantitative evaluations and the practical usage of data glyphs has been systematically reviewed to provide practitioners and researchers with guidance on how to create or choose an appropriate data glyph. Based on this survey design considerations were extracted and several open research gaps were identified, which provide space for further controlled experiments. Besides these revealed gaps other research directions are also worth pursuing.

To keep the study outcomes comparable and to better structure the study design space only quantitative experiments were reviewed in this thesis. However, results from qualitative user studies could also contribute to a more complete catalog of design considerations. Subjective preferences based on the aesthetics, the ease of use, or the learnability of different data glyphs are also important indicators, which have not been investigated in detail, yet. Another research direction could be the analysis of data glyphs used in practice. Which designs have been chosen to accomplish certain use cases and why? Is there a reason why some designs are used more often than others? By analyzing the practical usage of data glyphs additional insights can be gained, which might be interesting for analysts, since these application oriented examples better reflect real world scenarios.

Data glyphs for temporal data: A new data glyph representation (i.e., *clock glyph*) was introduced to fill a research gap visualizing time-series data with metaphors. The new design looked like a common clock to help analysts in identifying single points in time intuitively. Additionally, this *clock glyph* was integrated in three different visualization tools using several layout techniques to communicate varying context information. A quantitative user study was conducted to compare the *clock glyph* against well-established alternative representations like line glyphs or star glyphs in a small multiple setting. Results indicate that the *clock glyph* facilitates the detection of certain temporal dimensions and for this specific task outperforms the alternative representations.

It is important to note that these results are restricted to a specific context. The data glyphs were arranged in a grid layout to avoid confounding factors due to the positioning. However, a major advantage of data glyphs is the flexibility in arranging them on the screen and in combination with some context information. It would be interesting to see, whether the layout or the additional context information is influencing the reading performance of *clock glyphs*. Keeping the design identical and just switching between different layouts opens new space for further experiments. Currently, only one quantitative user study has been conducted investigating a similar topic with weather vanes and varying context information [123]. Results suggest that the context information is not influencing the reading performance of data glyphs.

Data glyphs for multi-dimensional data: To visualize multi-dimensional data from the environmental domain more intuitively a new metaphoric data glyph design was introduced, namely the *leaf glyph*. In combination with domain specific aggregation techniques this data glyph is applicable to larger datasets, as well. A use case scenario focusing on the exploration of the well-known forest fire dataset from the UCI machine learning repository [44] showcased the usefulness of the metaphor.

However, this evaluation does not help to prove the fact, whether metaphors are really beneficial for displaying domain specific data. As future research, I would like to see more studies about metaphoric glyph designs alternating between domain specific datasets and the respective glyph representations.

Additionally, a quantitative experiment was conducted to investigate the influence of a contour line on the well-known star glyph. This study focuses solely on similarity search tasks using varying number of dimensions. Of course, it would be interesting to know whether the findings can be generalized to other tasks, as well. Are the designs also influenced by the surrounding contour line when performing a visual search, or reading exact data values? Furthermore, the effect of a contour line could also be studied on different glyph designs like linear profile glyphs, or size encoded pixel glyphs. Getting more information about various settings would facilitate the generalization of findings concerning the influence of a contour line.

Summary: Literature has been systematically reviewed, quantitative experiments have been conducted, and new data glyph designs were introduced within this thesis. All these research directions contributed to the catalog of design considerations proposed for different tasks and datasets in chapter 5. Practitioners and researchers can now easily follow the suggested design considerations to systematically create the most appropriate glyph design for their analysis task and dataset.

However, as previously discussed, there are still many open research gaps. Because of the huge design space of data glyphs and the various analysis tasks and different datasets only a few of them could be tackled within this thesis. A list of promising research topics in the area of quantitative experiments on data glyph designs was also identified in section 2.4. Researchers can use this list as a starting point for future research or refer to the discussion in this chapter 6.

Bibliography

[1] Alfie Abdul-Rahman, Julie Lein, Katherine Coles, Eamonn Maguire, Miriah Meyer, Martin Wynne, Chris R Johnson, A Trefethen, and Min Chen. Rule-based Visual Mappings–with a Case Study on Poetry Visualization. In *Computer Graphics Forum*, volume 32, pages 381–390. Wiley Online Library, 2013.

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer-Verlag, 2011.

[3] W. Aigner, S. Miksch, B. Thurnher, and S. Biffl. Planninglines: novel glyphs for representing temporal uncertainties and their evaluation. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pages 457–463, July 2005.

[4] Danielle Albers, Michael Correll, and Michael Gleicher. Task-driven evaluation of aggregation in time series visualization. In *Proc. CHI*, CHI '14, pages 551–560, New York, NY, USA, 2014. ACM.

[5] James P Allen. *Middle Egyptian: an Introduction to the Language and Culture of Hieroglyphs*. Cambridge University Press, 2000.

[6] Artem Amirkhanov, Bernhard Frhler, Johann Kastner, Eduard Grller, and Christoph Heinzl. InSpectr: Multi-Modal Exploration, Visualization, and Analysis of Spectral Data. *Computer Graphics Forum*, 33(3):91–100, 2014.

[7] Edgar Anderson. Maize in Mexico a Preliminary Survey. *Annals of the Missouri Botanical Garden*, pages 147–247, 1946.

[8] Edgar Anderson. A Semigraphical Method for the Analysis of Complex Problems. *Proceedings of the National Academy of Sciences of the United States of America*, 43(10):923–927, 1957.

[9] 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, pages 417–420, New York, NY, USA, 2004. ACM.

[10] Natalia Andrienko and Gennady Andrienko. *Exploratory Analysis of Spatial and Temporal data*. Springer, 2006.

[11] Lyn Bartram, Billy Cheung, and Maureen C Stone. The Effect of Colour and Transparency on the Perception of Overlaid Grids. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1942–1948, 2011.

[12] Charles B Beck. *An Introduction to Plant Structure and Development: Plant Anatomy for the Twenty-First Century*. Cambridge University Press, 2010.

[13] Jeff Beddow. Shape Coding of Multidimensional Data on a Microcomputer Display. In *Proceedings of the 1st conference on Visualization'90*, pages 238–246. IEEE Computer Society Press, 1990.

[14] J. Bernard, N. Wilhelm, B. Kruger, T. May, T. Schreck, and J. Kohlhammer. MotionExplorer: Exploratory Search in Human Motion Capture Data Based on Hierarchical Aggregation. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2257–2266, Dec 2013.

[15] Jrgen Bernard, Martin Steiger, Sven Widmer, Hendrik Lcke-Tieke, Thorsten May, and Jrn Kohlhammer. Visual-interactive Exploration of Interesting Multivariate Relations in Mixed Research Data Sets. *Computer Graphics Forum*, 33(3):291–300, 2014.

[16] Ingwer Borg and Thomas Staufenbiel. Performance of Snow Flakes, Suns, and Factorial Suns in the Graphical Representation of Multivariate Data. *Multivariate Behavioral Research*, 27(1):43–55, 1992.

[17] 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 *EG STAR*, pages 39–63. Eurographics, 2013.

[18] Rita Borgo, Karl Proctor, Min Chen, Heike Janicke, Tavi Murray, and Ian M Thornton. Evaluating the Impact of Task Demands and Block Resolution on the Effectiveness of Pixel-based Visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):963–972, 2010.

[19] Nadia Boukhelifa, Anastasia Bezerianos, Tobias Isenberg, and Jean-Daniel Fekete. Evaluating Sketchiness as a Visual Variable for the Depiction of Qualitative Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2769–2778, December 2012.

[20] Margery D. Boulette, Bruce G. Coury, and Nadeem A. Bezar. Classification of Multidimensional Data under Time Constraints: Evaluating Digital and Configural Display Representations. *Proc. Human Factors and Ergonomics Society Annual Meeting*, 31(1):116–120, 1987.

[21] U. Brandes and B. Nick. Asymmetric Relations in Longitudinal Social Networks. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2283–2290, Dec 2011.

[22] Ulrik Brandes, Bobo Nick, Brigitte Rockstroh, and Astrid Steffen. Gestaltlines. In *Proc. EuroVis*, pages 171–180. Eurographics, 2013.

[23] Cynthia A. Brewer. Colorbrewer—Color Advice for Maps. Accessed online September, 2012, `http://www.colorbrewer.org/`.

[24] R. L. Brown. Methods for the Graphic Representation of Systems Simulated Data. *Ergonomics*, 28(10):1439–1454, 1985.

[25] Lawrence A Bruckner. On Chernoff Faces. *Graphical Representation of Multivariate Data*, pages 93–121, 1978.

[26] J.J. Caban, P. Rheingans, and T. Yoo. An Evaluation of Visualization Techniques to Illustrate Statistical Deformation Models. *Computer Graphics Forum*, 30(3):821–830, 2011.

[27] Ivan Cantador, Peter Brusilovsky, and Tsvi Kuflik. Second Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec2011). In *RecSys*, pages 387–388, 2011.

[28] Melody C. Carswell and Christopher D. Wickens. Information Integration and the Object Display: An Interaction of Task Demands and Display Superiority. *Ergonomics*, 30(3):511–527, 1987.

[29] Elizabeth J. Casey. Visual Display Representation of Multidimensional Systems: The Effect of System Structure: And Display Integrality. *Proc. Human Factors and Ergonomics Society Annual Meeting*, 31(1):112–115, 1987.

[30] Merve Cayli, Murat Can Cobanoglu, and Selim Balcisoy. Glyphlink: An interactive visualization approach for semantic graphs. *Journal of Visual Languages & Computing*, 24(6):463 – 471, 2013.

[31] Wing-Yi Chan, Huamin Qu, and Wai-Ho Mak. Visualizing the semantic structure in classical music works. *Visualization and Computer Graphics, IEEE Transactions on*, 16(1):161–173, 2010.

[32] Yu-Hsuan Chan, C.D. Correa, and Kwan-Liu Ma. The Generalized Sensitivity Scatterplot. *IEEE TVCG*, 19(10):1768–1781, October 2013.

[33] Michael Chau. Visualizing Web Search Results Using Glyphs: Design and Evaluation of a Flower Metaphor. *ACM TMIS*, 2(1):1–27, March 2011.

[34] Guoning Chen, Darrel Palke, Zhongzang Lin, Harry Yeh, Paul Vincent, Robert S Laramee, and Eugene Zhang. Asymmetric Tensor Field Visualization for Surfaces. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):1979–1988, 2011.

[35] Herman Chernoff. The Use of Faces to Represent Points in K-Dimensional Space Graphically. *Journal of the American Statistical Association*, pages 361–368, 1973.

[36] Herman Chernoff and M. Haseeb Rizvi. Effect on Classification Error of Random Permutations of Features in Representing Multivariate Data by Faces. *Journal of the American Statistical Association*, 70(351):548–554, 1975.

[37] E.B. Chlan and P. Rheingans. Multivariate Glyphs for Multi-object Clusters. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 141–148, Oct 2005.

[38] M.C. Chuah. Dynamic Aggregation with Circular Visual Designs. In *Information Visualization, 1998. Proceedings. IEEE Symposium on*, pages 35–43, 151, Oct 1998.

[39] M.C. Chuah and S.G. Eick. Managing Software with new Visual Representations. In *Information Visualization, 1997. Proceedings., IEEE Symposium on*, pages 30–37, Oct 1997.

[40] Mei C Chuah and Stephen G Eick. Information Rich Glyphs for Software Management Data. *Computer Graphics and Applications, IEEE*, 18(4):24–29, 1998.

[41] David H. S. Chung, Philip A. Legg, Matthew L. Parry, Rhodri Bown, Iwan W. Griffiths, Robert S. Laramee, and Min Chen. Glyph Sorting: Interactive Visualization for Multi-dimensional Data. *Information Visualization*, Online Nov. 2013.

[42] David HS Chung, Robert S Laramee, Johannes Kehrer, and Helwig Hauser. Glyph-based multi-field visualization. In *Scientific Visualization*, pages 129–137. Springer, 2014.

[43] W.S. Cleveland and R. McGill. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, pages 531–554, 1984.

[44] Paulo Cortez and Aníbal de Jesus Raimundo Morais. A Data Mining Approach to Predict Forest Fires Using Meteorological Data. 2007.

[45] Douglas W. Cunningham and Christian Wallraven. *Experimental Design: From User Studies to Psychophysics*. CRC Press, 2012.

[46] Geen De Soete and Wilfried Do Corte. On the Perceptual Salience of Features of Chernoff Faces for Representing Multivariate Data. *Applied Psychological Measurement*, 9(3):275–280, 1985.

[47] Geert De Soete. A Perceptual Study of the Flury-Riedwyl Faces for Graphically Displaying Multivariate Data. *International Journal of Man-Machine Studies*, 25(5):549–555, 1986.

[48] Oliver Deussen and Bernd Lintermann. *Digital Design of Nature*. Springer, 2005.

[49] C. Dick, R. Burgkart, and R. Westermann. Distance Visualization for Interactive 3D Implant Planning. *IEEE TVCG*, 17(12):2173–2182, December 2011.

[50] K. Dinkla, M. A. Westenberg, H. M. Timmerman, S. A. F. T. van Hijum, and J. J. van Wijk. Comparison of Multiple Weighted Hierarchies: Visual Analytics for Microbe Community Profiling. In *Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'11, pages 1141–1150, Aire-la-Ville, Switzerland, Switzerland, 2011. Eurographics Association.

[51] Pierre Dragicevic and Stéphane Huot. SpiraClock: a Continuous and Non-Intrusive Display for Upcoming Events. In *CHI'02 extended abstracts on Human factors in computing systems*, pages 604–605. ACM, 2002.

[52] Stephen HC Du Toit, A Gert W Steyn, and Rolf H Stumpf. *Graphical Exploratory Data Analysis*. Springer, 1986.

[53] Cody Dunne and Ben Shneiderman. Motif simplification: Improving network visualization readability with fan, connector, and clique glyphs. In *Proc. CHI*, CHI '13, pages 3247–3256, New York, NY, USA, 2013. ACM.

[54] David S Ebert, Randall M Rohrer, Christopher D Shaw, Pradyut Panda, James M Kukla, and D Aaron Roberts. Procedural Shape Generation for Multi-Dimensional Data Visualization. *Computers & Graphics*, 24(3):375–384, 2000.

[55] David S Ebert, Christopher D Shaw, Amen Zwa, and Cindy Starr. Two-handed Interactive Stereoscopic Visualization. In *Proceedings of the 7th Conference on Visualization'96*, pages 205–ff. IEEE Computer Society, 1996.

[56] Matt Edmunds, Robert S Laramee, Rami Malki, Ian Masters, TN Croft, Guoning Chen, and Eugene Zhang. Automatic Stream Surface Seeding: A Feature Centered Approach. In *Computer Graphics Forum*, volume 31, pages 1095–1104. Wiley Online Library, 2012.

[57] Gösta Ekman and Kenneth Junge. Psychophysical Relations in Visual Perception of Length, Area and Volume. *Scandinavian Journal of Psychology*, 2(1):1–10, 1961.

[58] James Elder and Steven Zucker. The Effect of Contour Closure on the Rapid Discrimination of Two-Dimensional Shapes. *Vision Research*, 33(7):981–991, 1993.

[59] James H Elder and Steven W Zucker. Evidence for Boundary-Specific Grouping. *Vision Research*, 38(1):143–152, 1998.

[60] Niklas Elmqvist, John Stasko, and Philippas Tsigas. DataMeadow: a Visual Canvas for Analysis of Large-Scale Multivariate Data. *Information visualization*, 7(1):18–33, 2008.

[61] David Feng, Yueh Lee, Lester Kwock, and Russell M. Taylor, II. Evaluation of Glyph-based Multivariate Scalar Volume Visualization Techniques. In *Proc. APGV*, pages 61–68. ACM, 2009.

[62] Stephen Few. Grid Lines in Graphs are Rarely Useful. In *DM Review*. Perceptual Edge, February 2005. `http://www.perceptualedge.com/articles/dmreview/grid_lines.pdf`.

[63] Fabian Fischer, Johannes Fuchs, and Florian Mansmann. ClockMap: Enhancing Circular Treemaps with Temporal Glyphs for Time-Series Data. In *Proc. EuroVis Short Papers*, pages 97–101. Eurographics, 2012.

[64] Fabian Fischer, Johannes Fuchs, Pierre-Antoine Vervier, Florian Mansmann, and Olivier Thonnard. VisTracer: A Visual Analytics Tool to Investigate Routing Anomalies in Traceroutes. In *Proc. of the 9th International Symposium on Visualization for Cyber Security*, VizSec '12, pages 80–87. ACM, 2012.

[65] Bernhard Flury and Hans Riedwyl. Graphical Representation of Multivariate Data by Means of Asymmetrical Faces. *Journal of the American Statistical Association*, 76(376):757–765, 1981.

[66] Clifton Forlines and Kent Wittenburg. Wakame: Sense Making of Multi-dimensional Spatial-temporal Data. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI '10, pages 33–40, New York, NY, USA, 2010. ACM.

[67] C. Forsell, S. Seipel, and M. Lind. Simple 3D Glyphs for Spatial Multivariate Data. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 119–124, Oct 2005.

[68] Camilla Forsell, Stefan Seipel, and Mats Lind. Surface Glyphs for Efficient Visualization of Spatial Multivariate Data. *Information Visualization*, 5(2):112–124, 2006.

[69] Michael Friendly. A.-M. Guerry's "Moral Statistics of France": Challenges for Multivariable Spatial Analysis. *Statistical Science*, pages 368–399, 2007.

[70] Christian Frisson, Stéphane Dupont, Willy Yvart, Nicolas Riche, Xavier Siebert, and Thierry Dutoit. Audiometro: Directing search for sound designers through content-based cues. In *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound*, AM '14, pages 1:1–1:8, New York, NY, USA, 2014. ACM.

[71] G. Fuchs and H. Schumann. Visualizing Abstract Data on Maps. In *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, pages 139–144, July 2004.

[72] J. Fuchs, P. Isenberg, A. Bezerianos, and D. Keim. A systematic review of experimental studies on data glyphs. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2016.

[73] Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. Evaluation of Alternative Glyph Designs for Time Series Data in a Small Multiple Setting. In *Proc. CHI*, pages 3237–3246. ACM, 2013.

[74] Johannes Fuchs, Petra Isenberg, Anastasia Bezerianos, Fabian Fischer, and Enrico Bertini. The Influence of Contour on Similarity Perception of Star Glyphs. *IEEE TVCG*, 20(12):2251–2260, Dec 2014.

[75] Johannes Fuchs, Dominik Jäckle, Niklas Weiler, and Tobias Schreck. Leaf Glyph - Visualizing Multi-dimensional Data with Environmental Cues. In *Digital Library Scitepress*, pages 195–206, March 2015.

[76] Johannes Fuchs, Dominik Jäckle, Niklas Weiler, and Tobias Schreck. *Leaf Glyphs: Story Telling and Data Analysis Using Environmental Data Glyph Metaphors*, pages 123–143. Springer International Publishing, Cham, 2016.

[77] Johannes Fuchs, Roman Rädle, Dominik Sacha, Fabian Fischer, and Andreas Stoffel. Collaborative Data Analysis with Smart Tangible Devices. In *IS&T/SPIE Electronic Imaging*, pages 90170C–90170C. International Society for Optics and Photonics, 2013.

[78] Wendell R Garner. *The Processing of Information and Structure*. Lawrence Erlbaum Associates, 1974.

[79] Patrick Garrigan. The Effect of Contour Closure on Shape Recognition. *Perception*, 41(2):221–235, 2012.

[80] Pascal Goffin, Wesley Willett, Anastasia Bezerianos, and Petra Isenberg. Exploring the effect of word-scale visualizations on reading behavior. In *Proc. CHI Work-In-Progress*, pages 1827–1832. ACM, 2015.

[81] C. G. Hamner, D. W. Turner, and D. M. Young. Comparisons of Several Graphical Methods for Representing Multivariate Data. *Computers & Mathematics with Applications*, 13(7):647–655, 1987.

[82] J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. Elsevier Ltd, Oxford, 3rd edition, 2011.

[83] John A Hartigan. Printer Graphics for Clustering. *Journal of Statistical Computation and Simulation*, 4(3):187–213, 1975.

[84] Christopher G Healey and James T Enns. Large datasets at a glance: Combining textures and colors in scientific visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 5(2):145–167, 1999.

[85] Jeffrey Heer and Michael Bostock. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings Human Factors in Computing Systems (CHI)*, pages 203–212. ACM, 2010.

[86] M. Hlawatsch, P. Leube, W. Nowak, and D. Weiskopf. Flow Radar Glyphs - Static Visualization of Unsteady Flow with Uncertainty. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):1949–1958, Dec 2011.

[87] Mario Hlawitschka and Gerik Scheuermann. HOT-lines: Tracking Lines in Higher Order Tensor Fields. In *Visualization, 2005. VIS 05. IEEE*, pages 27–34. IEEE, 2005.

[88] Werner Horn, Christian Popow, Lukas Unterasinger, et al. Support for Fast Comprehension of ICU Data: Visualization Using Metaphor Graphics. *Methods Inf Med*, 40(5):421–424, 2001.

[89] David L Huff and William Black. A Multivariate Graphic Display for Regional Analysis. In *Graphical Representation of Multivariate Data*, pages 199–218. Academic Press New York, 1978.

[90] J.-F. Im, M.J. McGuffin, and R. Leung. GPLOM: The Generalized Plot Matrix for Visualizing Multidimensional Multivariate Data. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2606–2614, Dec 2013.

[91] Robert Jacob. Facial Representation of Multivariate Data. In *Graphical Representation of Multivariate Data*, pages 143–168. Academic Press, 1978.

[92] Robert Jacob. Investigating the Space of Chernoff faces. *Recent Advances in Statistics*, pages 449–468, 1983.

[93] Robert Jacob and Egeth Howard. The Face as a Data Display. *Human Factors*, 18(2):189–200, 1976.

[94] T.J. Jankun-Kelly, Y.S. Lanka, and J.E. Swan. An Evaluation of Glyph Perception for Real Symmetric Traceless Tensor Properties. *Computer Graphics Forum*, 29(3):1133–1142, 2010.

[95] Patricia M. Jones, Christopher D. Wickens, and Stuart J. Deutsch. The Display of Multivariate Information: An Experimental Study of an Information Integration Task. *Human Performance*, 3(1):1–17, 1990.

[96] BT Kabulov. A Method for Constructing Chernoff Faces Oriented Toward Interval Estimates of the Parameters. *Soviet Journal of Computer and Systems Sciences*, 30(3):94–97, 1992.

[97] A Kachkaev, J Wood, and J Dykes. Glyphs for Exploring Crowd-sourced Subjective Survey Classification. In *Computer Graphics Forum*, volume 33, pages 311–320. Wiley Online Library, 2014.

[98] Daniel A Keim and Daniela Oelke. Literature Fingerprinting: A New Method for Visual Literary Analysis. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*, pages 115–122. IEEE, 2007.

[99] Gordon Kindlmann and C-F Westin. Diffusion Tensor Visualization with Glyph Packing. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):1329–1336, 2006.

[100] Christopher Kintzel, Johannes Fuchs, and Florian Mansmann. Monitoring Large IP Spaces with ClockView. In *Proc. of the 8th International Symposium on Visualization for Cyber Security*, VizSec '11, pages 2:1–2:10. ACM, 2011.

[101] R Victor Klassen and Steven J Harrington. Shadowed Hedgehogs: A Technique for Visualizing 2D Slices of 3D Vector Fields. In *Proceedings of the 2nd conference on Visualization'91*, pages 148–153. IEEE Computer Society Press, 1991.

[102] E Kleiberg, H van de Wetering, and JJ van Wijk. Botanical Visualization of Huge Hierarchies. In *Information Visualization, 2001. INFOVIS 2001. IEEE Symposium on*, pages 87–94. IEEE, 2001.

[103] Alexander Klippel, Frank Hardisty, Rui Li, and Chris Weaver. Colour-Enhanced Star Plot Glyphs: Can Salient Shape Characteristics Be Overcome? *Cartographica*, 44(3):217–231, 2009.

[104] Alexander Klippel, Frank Hardisty, and Chris Weaver. Star Plots: How Shape Characteristics Influence Classification Tasks. *Cartography and Geographic Information Science*, 36(2):149–163, 2009.

[105] K. Koffka. *Principles of Gestalt Pschology*. Lund Humphries, London, 1935.

[106] Zoltán Konyha, Kresimir Matkovic, and Helwig Hauser. *Interactive 3D Visualization of Rigid Body Systems*. IEEE, 2003.

[107] S.M. Kosslyn. *Graph Design for the Eye and Mind*. Oxford University Press, New York, 2007.

[108] Sven Krasser, G. Conti, J. Grizzard, J. Gribschaw, and H. Owen. Real-time and Forensic Network Data Analysis Using Animated and Coordinated Visualization. In *Information Assurance Workshop, 2005. IAW '05. Proceedings from the Sixth Annual IEEE SMC*, pages 42–49, June 2005.

[109] K. Lakkaraju, W. Yurcik, R. Bearavolu, and A.J. Lee. NVisionIP: An Interactive Network Flow Visualization Tool for Security. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 3, pages 2675–2680. IEEE, 2005.

[110] Bongshin Lee, Nathalie Henry Riche, Amy K. Karlson, and Sheelash Carpendale. SparkClouds: Visualizing Trends in Tag Clouds. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1182–1189, November 2010.

[111] Michael D. Lee, Rachel E. Reilly, and Marcus E. Butavicius. An Empirical Evaluation of Chernoff Faces, Star Glyphs, and Spatial Visualizations for Binary Data. In *Proc. APVis*, pages 1–10. Australian Computer Society, Inc., 2003.

[112] H. Levkowitz. Color Icons—Merging Color and Texture Perception for Integrated Visualization of Multiple Parameters. In *IEEE VIS*, pages 164–170, Oct 1991.

[113] H. Levkowitz and G.T. Herman. Color Scales for Image Data. *Computer Graphics and Applications, IEEE*, 12(1):72–80, 1992.

[114] Yi-Na Li, Dong-Jin Li, and Kang Zhang. Metaphoric transfer effect in information visualization using glyphs. In *Proceedings of the 8th International Symposium on Visual Information Communication and Interaction*, VINCI '15, pages 121–130, New York, NY, USA, 2015. ACM.

[115] Andreas E. Lie, Johannes Kehrer, and Helwig Hauser. Critical Design and Realization Aspects of Glyph-based 3D Data Visualization. In *Proc. SCCG*, pages 19–26. ACM, 2009.

[116] M.A. Livingston and J.W. Decker. Evaluation of Trend Localization with Multi-Variate Visualizations. *IEEE TVCG*, 17(12):2053–2062, Dec 2011.

[117] Mark A. Livingston, Jonathan Decker, and Zhuming Ai. An Evaluation of Methods for Encoding Multiple 2D Spatial Data. *Proc. SPIE*, 7868:78680C–78680C–12, 2011.

[118] Mark A. Livingston and Jonathan W. Decker. Evaluation of Multivariate Visualizations: A Case Study of Refinements and User Experience. *Proc. SPIE*, 8294:82940G–82940G–12, 2012.

[119] A.M. MacEachren. *How Maps Work*. Guilford Press, 1995.

[120] Donald MacGregor and Paul Slovic. Graphic Representation of Judgmental Information. *Human Computer Interaction*, 2(3):179–200, 1986.

[121] E. Maguire, P. Rocca-Serra, S.-A. Sansone, J. Davies, and Min Chen. Taxonomy-Based Glyph Design with a Case Study on Visualizing Workflows of Biological Experiments. *IEEE TVCG*, 18(12):2603–2612, December 2012.

[122] Eamonn Maguire, Philippe Rocca-Serra, S-A Sansone, Jim Davies, and Min Chen. Visual Compression of Workflow Visualizations with Automated Detection of Macro Motifs. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2576–2585, 2013.

[123] Joel P. Martin, J. Edward Swan, Robert J. Moorhead, Zhanping Liu, and Shangshu Cai. Results of a User Study on 2D Hurricane Visualization. *Computer Graphics Forum*, 27(3):991–998, 2008.

[124] B. McDonnel and N. Elmqvist. Towards Utilizing GPUs in Information Visualization: A Model and Implementation of Image-Space Operations. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1105–1112, Nov 2009.

[125] M. Meyer, B. Wong, M. Styczynski, T. Munzner, and H. Pfister. Pathline: A Tool for Comparative Functional Genomics. In *Proceedings of the 12th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'10, pages 1043–1052, Aire-la-Ville, Switzerland, Switzerland, 2010. Eurographics Association.

[126] Juan E Mezzich and David R L Worthington. A Comparison of Graphical Representations of Multidimensional Psychiatric Diagnostic Data. *Graphical Representation of Multivariate Data*, pages 123–141, 1978.

[127] L. Micallef, P. Dragicevic, and J. Fekete. Assessing the Effect of Visualizations on Bayesian Reasoning through Crowdsourcing. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2536–2545, Dec 2012.

[128] Jameson Miller, Cory Quammen, and Matthew Fleenor. Interactive Visualization of Intercluster Galaxy Structures in the Horologium-Reticulum Supercluster. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1149–1156, September 2006.

[129] M. Mlejnek, P. Ermest, A. Vilanova, R. van der Rijt, H. van den Bosch, F. Gerritsen, and M.E. Groller. Profile Flags: A Novel Metaphor for Probing of T2 Maps. In *Visualization, 2005. VIS 05. IEEE*, pages 599–606, Oct 2005.

[130] Daniel R Montello and M Gray. Miscommunicating With Isoline Preference Maps: Design Principles for Thematic Maps. *Cartographic Perspectives*, (50):24–33, 2005.

[131] Shane Moriarity. Communicating Financial Information Through Multidimensional Graphics. *Journal of Accounting Research*, 17(1):205–224, 1979.

[132] Christopher J. Morris, David S. Ebert, and Penny L. Rheingans. Experimental Analysis of the Effectiveness of Features in Chernoff Faces. *Proc. SPIE*, 3905:12–17, 2000.

[133] Boris Müller. Poetry on the Road, 2014. http://www.esono.com/boris/projects/poetry05/. Retrieved July 2014.

[134] Tamara Munzner. *Visualization Analysis & Design*. CRC Press, 2015.

[135] Moshe Naveh-Benjamin and Robert G. Pachella. The Effect of Complexity on Interpreting ”Chernoff” Faces. *Human Factors*, 24(1):11–18, 1982.

[136] Alisa Neeman, Boris Jeremic, and Alex Pang. Visualizing Tensor Fields in Geomechanics. In *Visualization, 2005. VIS 05. IEEE*, pages 35–42. IEEE, 2005.

[137] Elisabeth S Nelson. The Face Symbol: Research Issues and Cartographic Potential. *Cartographica*, 42(1):53–64, 2007.

[138] Elisabeth S Nelson, David Dow, Christopher Lukinbeal, and Ray Farley. Visual Search Processes and the Multivariate Point Symbol. *Cartographica*, 34(4):19–33, 1997.

[139] ES Nelson and P Gilmartin. An Evaluation of Multivariate, Quantitative Point Symbols for Maps. *Cartographic design: Theoretical and practical perspectives*, pages 191–203, 1996.

[140] M. Neugebauer, K. Lawonn, O. Beuing, P. Berg, G. Janiga, and B. Preim. AmniVis - A System for Qualitative Exploration of Near-Wall Hemodynamics in Cerebral Aneurysms. *Computer Graphics Forum*, 32(3pt3):251–260, 2013.

[141] Florence Nightingale. *Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army: Founded Chiefly on the Experience of the Late War. By Florence Nightingale. Presented by Request to the Secretary of State for War*. Harrison and Sons, St. Martin's Lane, WC, 1858.

[142] Thomas Nocke, Stefan Schlechtweg, and Heidrun Schumann. Icon-Based Visualization Using Mosaic Metaphors. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pages 103–109. IEEE, 2005.

[143] OECD Better Life Index. Your Better Life Index, 2011. `http://www.oecdbetterlifeindex.org/`. Retrieved July 2014.

[144] D. Oelke, D. Kokkinakis, and D. A. Keim. Fingerprint Matrices: Uncovering the Dynamics of Social Networks in Prose Literature. In *Proceedings of the 15th Eurographics Conference on Visualization*, EuroVis '13, pages 371–380, Aire-la-Ville, Switzerland, Switzerland, 2013. Eurographics Association.

[145] Steffen Oeltze, Wolfgang Freiler, Reyk Hillert, Helmut Doleisch, Bernhard Preim, and Walter Schubert. Interactive, graph-based visual analysis of high-dimensional, multi-parameter fluorescence microscopy data in toponomics. *IEEE TVCG*, 17(12):1882–1891, 2011.

[146] Stephen E Palmer. *Vision Science: Photons to Phenomenology*, volume 1. MIT press Cambridge, MA, 1999.

[147] J. Pearlman and P. Rheingans. Visualizing Network Security Events using Compound Glyphs from a Service-oriented Perspective. In *Proc. VIZSEC*, pages 131–146. Springer, 2007.

[148] W. Peng, M.O. Ward, and E.A. Rundensteiner. Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering. In *IEEE InfoVis*, pages 89–96, 2004.

[149] Rohn J. Petersen, William W. Banks, and David I. Gertman. Performance-based Evaluation of Graphic Displays for Nuclear Power Plant Control Rooms. In *Proc. CHI*, pages 182–189. ACM, 1982.

[150] Ronald M Pickett and Georges G Grinstein. Iconographic Displays for Visualizing Multidimensional Data. In *Proceedings of the 1988 IEEE Conference on Systems, Man, and Cybernetics*, volume 514, page 519, 1988.

[151] David HF Pilar and Colin Ware. Representing flow patterns by using streamlines with glyphs. *Visualization and Computer Graphics, IEEE Transactions on*, 19(8):1331–1341, 2013.

[152] Thomas Porter and Tom Duff. Compositing Digital Images. In *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '84, pages 253–259, New York, NY, USA, 1984. ACM.

[153] W. Ribarsky, E. Ayers, J. Eble, and S. Mukherjea. Glyphmaker: Creating Customized Visualizations of Complex Data. *Computer*, 27(7):57–64, July 1994.

[154] Naomi B Robbins. *Creating More Effective Graphs*. Wiley-Interscience, Hoboken, NJ, USA, 2005.

[155] Kerry Rodden, Wojciech Basalaj, David Sinclair, and Kenneth Wood. Does organisation by similarity assist image browsing? In *Proc. CHI*, CHI '01, pages 190–197, New York, NY, USA, 2001. ACM.

[156] T. Ropinski and B. Preim. Taxonomy and Usage Guidelines for Glyph-Based Medical Visualization. In *Proceedings of the 19th Conference on Simulation and Visualization (SimVis08)*, pages 121–138, 2008.

[157] Timo Ropinski, Steffen Oeltze, and Bernhard Preim. Survey of Glyph-based Visualization Techniques for Spatial Multivariate Medical Data. *Computers & Graphics*, 35(2):392–401, 2011.

[158] Jukka Saarinen and Dennis M Levi. The Effect of Contour Closure on Shape Perception. *Spatial Vision*, 12(2):227–238, 1999.

[159] Arnaud Sallaberry, Yang-Chih Fu, Hwai-Chung Ho, and Kwan-Liu Ma. ContactTrees: Ego-Centered Visualization of Social Relations. Technical report, 2012.

[160] Allen R Sanderson, Chris R Johnson, and Robert M Kirby. Display of Vector Fields Using a Reaction-Diffusion Model. In *Proceedings of the conference on Visualization'04*, pages 115–122. IEEE Computer Society, 2004.

[161] Purvi Saraiya, Chris North, and Karen Duca. An Insight-Based Methodology for Evaluating Bioinformatics Visualizations. *IEEE TVCG*, 11(4):443–456, jul 2005.

[162] Purvi Saraiya, Chris North, and Karen Duca. Comparing benchmark task and insight evaluation methods on timeseries graph visualizations. In *Proc. BELIV*, BELIV '10, pages 55–62, New York, NY, USA, 2010. ACM.

[163] P C Saxena and K Navaneetham. Comparison of Chernoff-type Face and Non-Graphical Methods for Clustering Multivariate Observations. *Computational Statistics & Data Analysis*, 15(1):63–79, 1993.

[164] Tobias Schreck, Daniel Keim, and Florian Mansmann. Regular Treemap Layouts for Visual Analysis of Hierarchical Data. In *Proceedings of the 22nd Spring Conference on Computer Graphics*, pages 183–190. ACM, 2006.

[165] D M Schultz. *Monthly Weather Review*, volume 136. American Meteorological Society, 2008.

[166] Thomas Schultz and Gordon L Kindlmann. Superquadric Glyphs for Symmetric Second-Order Tensors. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1595–1604, 2010.

[167] Christopher D Shaw, James A Hall, David S Ebert, and D Aaron Roberts. Interactive Lens Visualization Techniques. In *Proceedings of the conference on Visualization'99: celebrating ten years*, pages 155–160. IEEE Computer Society Press, 1999.

[168] J.H. Siegel, E.J. Farrell, R.M. Goldwyn, and H.P. Friedman. The Surgical Implications of Physiologic Patterns in Myocardial Infarction Shock. *Surgery*, 72(1):126–141, 1972.

[169] Andreas Sigfridsson, Tino Ebbers, Einar Heiberg, and Lars Wigström. Tensor Field Visualisation Using Adaptive Filtering of Noise Fields Combined with Glyph Rendering. In *Proceedings of the conference on Visualization'02*, pages 371–378. IEEE Computer Society, 2002.

[170] H. Siirtola. The Effect of Data-Relatedness in Interactive Glyphs. In *Proc. IV*, pages 869–876, July 2005.

[171] Navaneethan Siva, Alex Chaparro, Duy Nguyen, and Evan Palmer. Visual Search Efficiency for Features in Chernoff Faces. *Journal of Vision*, 13(9):689–, 2013.

[172] Navaneethan Siva, Alex Chaparro, and Evan Palmer. Evaluation of MILSTD 2525 Glyph Features in a Visual Search Paradigm. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1):1189–1193, 2014.

[173] Navaneethan Sivagnanasundaram, Alex Chaparro, and Evan Palmer. Evaluation of the Presence of a Face Search Advantage in Chernoff Faces. *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, 57(1):1610–1614, 2013.

[174] J. Sreng, A. Lecuyer, C. Megard, and C. Andriot. Using Visual Cues of Contact to Improve Interactive Manipulation of Virtual Objects in Industrial Assembly/Maintenance Simulations. *IEEE TVCG*, 12(5):1013–1020, September 2006.

[175] Moritz Stefaner. Wahlland, 2013. `http://moritz.stefaner.eu/projects/wahlland/`. Retrieved March 2014.

[176] Moritz Stefaner. The Deleted, 2014. http://notabilia.net/. Retrieved July 2014.

[177] Martin Steiger, Jürgen Bernard, Sebastian Mittelstädt, Hendrik Lücke-Tieke, Daniel Keim, Thorsten May, and Jörn Kohlhammer. Visual Analysis of Time-Series Similarities for Anomaly Detection in Sensor Networks. In *Computer Graphics Forum*, volume 33, pages 401–410. Wiley Online Library, 2014.

[178] Duane Stock and Collin J. Watson. Human Judgment Accuracy, Multidimensional Graphics, and Humans Versus Models. *Journal of Accounting Research*, 22(1):192–206, 1984.

[179] Maureen Stone. In Color Perception, Size Matters. *IEEE Computer Graphics and Applications*, 32(2):8–13, March/April 2012.

[180] Maureen Stone. *A Field Guide to Digital Color*. CRC Press, 2013.

[181] Andrada Tatu, Leishi Zhang, Enrico Bertini, Tobias Schreck, Daniel Keim, Sebastian Bremm, and Tatiana von Landesberger. ClustNails: Visual Analysis of Subspace Clusters. *Tsinghua Science and Technology*, 17(4):419–428, 2012.

[182] S. Thakur and A.J. Hanson. A 3D Visualization of Multiple Time Series on Maps. In *Information Visualisation (IV), 2010 14th International Conference*, pages 336–343, July 2010.

[183] C. Tominski, P. Schulze-Wollgast, and H. Schumann. 3D Information Visualization for Time Dependent Data on Maps. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pages 175–181, July 2005.

[184] Christian Tominski, James Abello, and Heidrun Schumann. Axes-based Visualizations with Radial Layouts. In *Proceedings of the 2004 ACM Symposium on Applied Computing*, SAC '04, pages 1242–1247, New York, NY, USA, 2004. ACM.

[185] E.R. Tufte. *Beautiful Evidence*. Graphics Press, 2006.

[186] A. Unger and H. Schumann. Visual Support for the Understanding of Simulation Processes. In *Visualization Symposium, 2009. PacificVis '09. IEEE Pacific*, pages 57–64, April 2009.

[187] unknown. Petroglyph. Accessed online March, 2015, `https://en.wikipedia.org/wiki/Petroglyph`.

[188] Fr Vogel and GG Wendt. Zwillingsuntersuchung über die erblichkeit einiger anthropologischer maße und konstitutionsindices. *Z. Menschl. Vererb. Konstitutionslehre*, 33:425–446, 1956.

[189] Georg Von Mayr. *Die Gesetzmässigkeit im Gesellschaftsleben*, volume 23. R. Oldenbourg, Munich, 1877.

[190] H Wainer. Graphic Experiment in Display of Nine Variables Uses Faces to Show Multiple Properties of States. *Newsletter of the Bureau of Social Sciences Research*, 13:2–3, 1979.

[191] Matthew O. Ward. A Taxonomy of Glyph Placement Strategies for Multidimensional Data Visualization. *Information Visualization*, 1(3/4):194–210, December 2002.

[192] Matthew O. Ward. Multivariate Data Glyphs: Principles and Practice. In *Handbook of Data Visualization*, pages 179–198. Springer, 2008.

[193] Matthew O. Ward and Zhenyu Guo. Visual Exploration of Time-series Data with Shape Space Projections. In *Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'11, pages 701–710, Aire-la-Ville, Switzerland, Switzerland, 2011. Eurographics Association.

[194] Matthew O. Ward and Benjamin N. Lipchak. A Visualization Tool for Exploratory Analysis of Cyclic Multivariate Data. *Metrika*, 51(1):27–37, 2000.

[195] Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2012.

[196] M. Weber, M. Alexa, and W. Muller. Visualizing Time-Series on Spirals. In *Information Visualization, 2001. INFOVIS 2001. IEEE Symposium on*, pages 7–13, 2001.

[197] C. Weigle and R.M. Taylor. Visualizing Intersecting Surfaces with Nested-surface Techniques. In *IEEE VIS*, pages 503–510, Oct 2005.

[198] M. A. Westenberg, S. A. F. T. van Hijum, O. P. Kuipers, and J. B. T. M. Roerdink. Visualizing Genome Expression and Regulatory Network Dynamics in Genomic and Metabolic Context. In *Proceedings of the 10th Joint Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'08, pages 887–894, Aire-la-Ville, Switzerland, Switzerland, 2008. Eurographics Association.

[199] Leland Wilkinson. An Experimental Evaluation of Multivariate Graphical Point Representations. In *Proc. CHI*, pages 202–209. ACM, 1982.

[200] Craig M Wittenbrink, Alex T Pang, and Suresh K Lodha. Glyphs for visualizing uncertainty in vector fields. *Visualization and Computer Graphics, IEEE Transactions on*, 2(3):266–279, 1996.

[201] Yingxin Wu and M. Takatsuka. Visualizing multivariate networks: A hybrid approach. In *Visualization Symposium, 2008. PacificVIS '08. IEEE Pacific*, pages 223–230, March 2008.

[202] Panpan Xu, Fan Du, Nan Cao, Conglei Shi, Hong Zhou, and Huamin Qu. Visual Analysis of Set Relations in a Graph. In *Proceedings of the 15th Eurographics Conference on Visualization*, EuroVis '13, pages 61–70, Aire-la-Ville, Switzerland, Switzerland, 2013. Eurographics Association.

[203] Jing Yang, Anilkumar Patro, Huang Shiping, Nishant Mehta, M.O. Ward, and E.A. Rundensteiner. Value and Relation Display for Interactive Exploration of High Dimensional Datasets. In *IEEE InfoVis*, pages 73–80, 2004.

[204] 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 InfoVis*, pages 105–112. IEEE, 2003.

[205] Nathan Yau. *Data Points: Visualization That Means Something*. John Wiley & Sons, 2013.

[206] Beth Yost and Chris North. Single complex glyphs versus multiple simple glyphs. In *Proc. CHI*, CHI EA '05, pages 1889–1892, New York, NY, USA, 2005. ACM.

[207] Changgong Zhang, Thomas Schultz, Kai Lawonn, Elmar Eisemann, and Anna Vilanova. Glyph-based comparative visualization for diffusion tensor fields. *Visualization and Computer Graphics, IEEE Transactions on*, 22(1):797–806, 2016.

[208] Xiaoyan Zhang, Qingnong Xiao, and Patrick J Fitzpatrick. The Impact of Multisatellite Data on the Initialization and Simulation of Hurricane Lili's (2002) Rapid Weakening Phase. *Monthly Weather Review*, 135(2):526–548, 2007.

[209] Leonid Zhukov and Alan H Barr. Heart-muscle Fiber Reconstruction from Diffusion Tensor MRI. In *Proceedings of the 14th IEEE Visualization 2003 (VIS'03)*, page 79. IEEE Computer Society, 2003.