Data Quality Visualization for Multivariate Hierarchic Data

Tatiana Tekušová*
Interactive Graphics Systems Group
TU Darmstadt, Germany

Martin Knuth†
Fraunhofer IGD
Darmstadt, Germany

Tobias Schreck‡
Interactive Graphics Systems Group
TU Darmstadt, Germany

Jörn Kohlhammer§
Fraunhofer IGD
Darmstadt, Germany

ABSTRACT

In many business applications, decision makers often have to base their decisions on large amounts of data from various sources. Often, the quality of this data varies substantially, affecting the degree of certainty the analyst can put into the analyzed data values. The data quality measures may be of qualitative or quantitative nature, and consist of one or many dimensions. In this poster paper, we first present a brief survey of currently available uncertainty visualization techniques. We then present experimental results we obtained with several techniques for visualization of multidimensional data quality information, applied on multivariate hierarchic data used in an economic data analysis scenario.

Keywords: Information Visualization, Data Quality, Uncertainty, Hierarchic Data, Multivariate Data, Economic and Financial Data.

Index Terms: I.3.8 [Computer Graphics]: Applications—; H.5.0 [Information Interfaces and Presentation]: General—;

1 INTRODUCTION

Data quality, as it may affect the degree of data (un)certainty, is a crucial aspect to consider in any data analysis. In visual data analysis applications, data certainty needs to be conveyed in an appropriate way to the users. The visualization of uncertainty information has recently been a research focus of the Information Visualization community. However, there is still a need for representing data quality and uncertainty in 3D visualizations [5], in decision support systems [3], and in visual analytics applications in general [12].

In this poster paper, we present a brief survey of currently available uncertainty visualization techniques. We concentrate on the display of multivariate data quality information for hierarchic data. In our preliminary work, we extend our 3D circular view visualization of multivariate hierarchic data [11] by applying various uncertainty visualization techniques.

2 SURVEY OF APPROACHES AND OPEN PROBLEMS

In the following, we present a brief survey of approaches to uncertainty visualization. A typology for visualizing uncertainty is presented in [13]. Methods for visualizing error and uncertainty are presented in several surveys [9, 6, 8, 4]. Available techniques include: (1) *Usage of free graphical variables*: color, size, saturation of color, position, angle, clarity, fuzziness, transparency, edge crispness; (2) *Integration of additional graphical objects*: uncertainty glyphs, labels, isosurfaces, textures; (3) *Usage of animation*: speed, duration, blinking, motion blur; (4) *Interactive representation*: clickable maps, difference images, mouse over effects, magic lenses; (5) *Addressing other human senses*: acoustic or haptic senses (e.g. sound or vibration). A user study of the methods [7] for spatial data indicates that blinking, adjacency and overlay are among the most useful techniques. At the same time, animation and saturation of color were deemed least useful.

Interestingly, none of these techniques supports a combination of qualitative and quantitative uncertainty information [3]. Although many techniques for multivariate data visualization exist, techniques for visualization of multivariate data uncertainty are still rare. One of these is the approach of Schmidt et al. [10] for Parallel Coordinate plots of environmental data. Another is the visualization technique by Davis et al. [1] for multivariate spatial uncertainty for geospatial data.

3 VISUALIZATION

We have implemented a prototype for testing various techniques to map data certainty attributes to a visualization for hierarchically structured data. We first introduce the basic technique and the used data, and then present preliminary application results.

3.1 Used Data

We consider the European consumer price data and its meta-data from the Eurostat database [2]. The categories of items defined in the so-called consumer basket form an item hierarchy. For the analysis of inflation development, the weight of items in the category hierarchy, the growth of the consumer price index and other variables are used. The data is often used as input for various macroeconomic models. With each change in the underlying dataset the models need to be updated. It is therefore desirable to display information on data quality, possible data changes, estimations, etc. to the users. As *data quality* information, Eurostat publishes additional multivariate information for each observation. 12 different binary variables (so called “flags”) are provided. Each observation can have none, one, or a combination of various flags. Therefore, there is the need to present multiple characteristics for each observation.

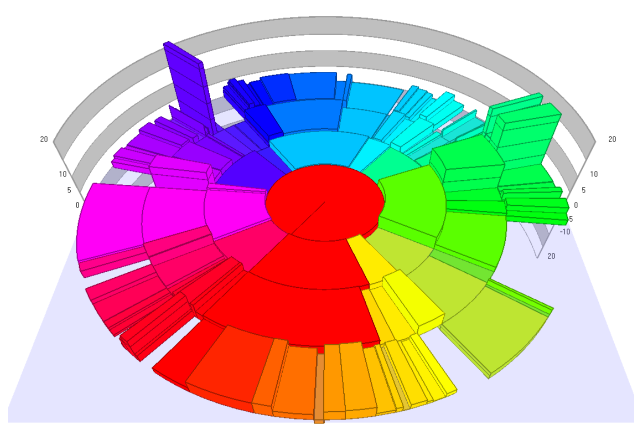


Figure 1: Visualization of multidimensional hierarchic data [11].

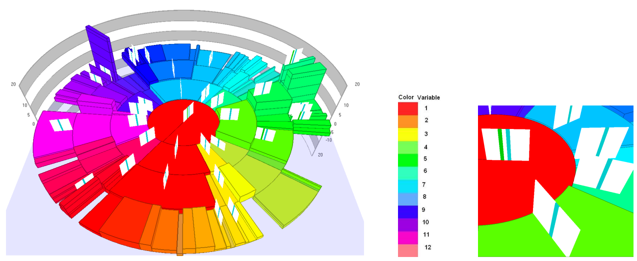
*E-Mail: tatiana.tekusova@gris.informatik.tu-darmstadt.de
†E-Mail: martin.knuth@igd.fraunhofer.de
‡E-Mail: tobias.schreck@gris.informatik.tu-darmstadt.de
§E-Mail: joern.kohlhammer@igd.fraunhofer.de

3.2 The Basic Visualization Technique

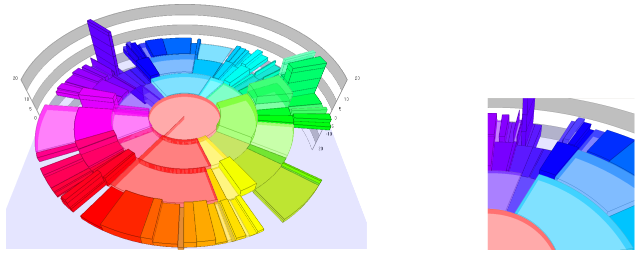
We base our data quality visualization on the technique for presentation of multivariate hierarchic data [11]. In the circle view (see Figure 1), the center of the plot represents the root of the hierarchy. Each ring of the circle represents one layer of the item hierarchy. Each segment represents a subset of items, or ultimately, a single item from the item hierarchy. For each node, the children are displayed as a segment in the next layer, adjacent to the parent segment. The width of the child segment is calculated as the child's weight divided by the parent's weight multiplied by the width of the parent. The height and the color of the segments can be mapped to additional dimensions of the data.

3.3 Data Quality Visualization

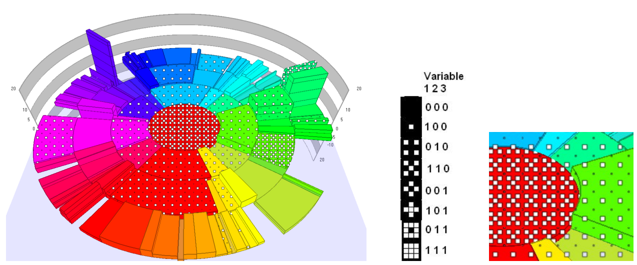
We extended the presented data visualization with multidimensional data quality information in several ways. We apply (1) *multivariate glyphs*, (2) *transparency*, and (3) *texture* overlays of different textures for showing data quality in a single view.



(a) Multi-dimensional glyphs. Glyph legend. Detailed view.



(b) Transparency. Detailed view.



(c) Texture overlay. Texture legend. Detailed view.

Figure 2: Options for visualization of data quality in a hierarchic chart.

Multi-dimensional glyphs (see Figure 2a) offer the possibility to show all data quality information for each data point. The used glyphs are rectangles divided into colored stripes. This shape allows to accurately locate the glyph also on very narrow segments. We also experimented with circular glyphs. However, glyphs close to each other cause occlusion with the typical visibility problems. The occlusion of glyphs, in general, can be overcome by user interaction (rotation, zooming, details on demand). For qualitative data quality information, the stripes can only have two states: colored for uncertain state, or white otherwise. In the quantitative case, the data value is mapped to the color saturation.

Transparency (see Figure 2b) is a natural way of conveying uncertainty. For qualitative data, the level of uncertainty can be mapped to the transparency level. In case of quantitative data, the transparency shows whether or not the data has special data uncertainty features. In case of multi-dimensional data uncertainty, only one value can be mapped to the transparency. For qualitative data, the number of dimensions with uncertain information can be used for determining the transparency level. For quantitative data, a suitable data aggregation can be used. In this case, further details on demand can be used to reveal additional information.

Depending on the definition of the texture pattern, *texture overlays* (see Figure 2c) can be used for qualitative or quantitative information. In the first case, predefined textures can be used, in the latter the texture pattern (color, frequency, direction) can be mapped to the data. For multi-dimensional data, we use texture overlay with several textures. The texture design maps each uncertainty dimension to a separate texture. The texture overlay creates the final pattern encoding the data quality information. The texture approach is not very suitable for narrow segments, where only a close-up zoom level can reveal the actual texture pattern.

4 Conclusions and Future Work

We reviewed various approaches to uncertainty visualization, and discussed possible applications of uncertainty visualization for hierarchically structured multidimensional data, relying on circular 3D plots. In the future, we like to extend the system with additional techniques, considering also more abstract visual metaphors such as fog, color weaving or particle systems. Conducting a user study comparing the effectiveness of the techniques would be interesting.

References

[1] T. J. Davis and C. P. Keller. Modelling and visualizing multiple spatial uncertainties. *Computers & Geosciences*, 23(4):397 – 408, May 1997.
[2] Eurostat. *http://epp.eurostat.ec.europa.eu*. Eurostat, 2007.
[3] H. Griethe and H. Schumann. Visualizing uncertainty for improved decision making. In *Proc. of the 4th International Conference on Perspectives in Business Informatics Research (BIR 2005)*, 2005.
[4] H. Griethe and H. Schumann. The visualization of uncertain data: Methods and problems. In *Proc. SimVis 2006*, pages 143–156, 2006.
[5] C. Johnson. Top scientific visualization research problems. *IEEE Computer Graphics and Applications*, 24(4):13–17, July 2004.
[6] C. R. Johnson and A. R. Sanderson. A next step: Visualizing errors and uncertainty. *IEEE CG&A*, 23(5):6–10, Sept. 2003.
[7] J. Kardos, A. Moore, and G. Benwell. The visualisation of uncertainty in spatially-referenced attribute data using trustworthy data structures. In *Proc. 15th Annual Colloquium of the Spatial Information Research Centre (SIRC 2003: Land, Place and Space)*, pages 11–26, 2003.
[8] A. M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3):139–160, July 2005.
[9] A. Pang, C. Wittenbrink, and S. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, Nov. 1997.
[10] G. S. Schmidt, S.-L. Chen, A. N. Bryden, M. A. Livingston, L. J. Rosenblum, and B. R. Osborn. Multidimensional visual representations for underwater environmental uncertainty. *IEEE Computer Graphics and Applications*, 24(5):56– 65, Sept. 2004.
[11] T. Tekušová and T. Schreck. Visualizing time-dependent data in multivariate hierarchic plots - design and evaluation of an economic application. In *12th Int. Conf. on Information Visualisation (IV08)*, 2008.
[12] J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, 2005.
[13] J. Thomson, B. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel. A typology for visualizing uncertainty. In *Proc. SPIE conf. on Visualization and Data Analysis*, volume 5669, pages 146–157, 2005.