Towards a Visual Data Language to Improve Insights into Complex Multidimensional Data

Jan Wojdziak[1(✉)], Bettina Kirchner[1], Dietrich Kammer[1], Martin Herrmann[1], and Rainer Groh[2]

[1] Gesellschaft für Technische Visualistik, Dresden, Germany
Jan.wojdziak@visualistik.de
[2] Technische Universität Dresden, Dresden, Germany

Abstract. Data volume is increasing steadily. Visualization helps to handle not only the volume, but the ever increasing diversity of data. Visualization gives answers faster and reveals information that would go unnoticed and therefore unused in decision making. The challenge we address in this contribution is how visualizations can be created semi-automatic without taking the individual human-centered view of the designer on an interface out of the loop. In this paper, we present a tool-supported design process to develop aesthetic and interactive data visualizations in a conceptual, guided, effective way.

Keywords: Information design · Process model · Information visualization · Aesthetics · Tool-support design

1 Introduction

Every information visualization has a very basic need – the narration of a story. Over time, new ways to visualize information were developed. Today, almost everyone is familiar with basic chart types such as line chart or pie chart. Charts are applied to present large amounts of data more understandable than spreadsheets or textual reports. Current visual interfaces that deal with big data only show different charts and graph visualizations arranged in dashboards. Hence, the user has an increased cognitive challenge to merge the visualized datasets in order to get insights into complex information. Visualizations that combine multidimensional data can emphasize answers faster than dashboards and reveal information that would go unnoticed and therefore unused in common chart visualizations. Yet, the design of suitable visual representations of complex and multidimensional datasets requires expert knowledge. The information designer has to employ the right design principles to compose a meaningful story to clarify the complexity of datasets, explaining a process, highlighting a trend, or supporting a specific argument. In consequence, a complex interactive visualization requires months of work by highly skilled professionals. On the other hand, data visualization should be a quick and easy way to convey information. But, it should be remembered that poorly designed data representations can distort the intended message, lose the user's attention, or fail to guide them toward meaningful conclusions.

In this paper, we present a novel tool-supported approach for dealing with complex datasets apart from the well-established charting perspective. The addressed challenge

S. Yamamoto (Ed.): HIMI 2016, Part I, LNCS 9734, pp. 203–213, 2016.
DOI: 10.1007/978-3-319-40349-6_20

is to reduce development time for information visualization by providing tools of the trade to realize information design in a conceptual, guided, and effective way. Common libraries, framework and tools used to create visual interfaces currently lack an overall information design process that addresses a human-centered access to multidimensional data sets. Hence, designers and developers are constrained to a manual design process.

The aim is to enhance the design process by a guided process based on visualization tools. To this end, we present the ViDaLa approach, which is able to create beautifully looking information visualizations without the expert knowledge of software engineers and information designers.

2 Related Work

Information visualization is the representation of data in a graphical form. The concept of using images to understand data has been around for centuries, from maps and graphs in the 17th century to the invention of the pie chart in the early 1800s. Today, information visualization has become a rapidly evolving blend of business, science, and art that is defined as a "visual representations of the semantics, or meaning, of information.[…] information visualization typically deals with nonnumeric, nonspatial, and high-dimensional data." [1]. Designing information visualization can be described as the practice of presenting information in a way that fosters efficient and effective understanding of the fundamental dataset.

There is a vast variety of methods and approaches in information design such as visualization patterns [2], interaction pattern [3], design guidelines in visualization [4] and in interaction design [5]. In [6] Lau and van de Moere present a model that focus on aesthetics as a conceptual influence on the technical implementation of visualizations. It reveals information aesthetics as the conceptual link between information visualization and visualization art. Lang discusses in [7] the importance of aesthetic in visualizations related to its efficiency, whereas Kosara in [8] proposes a classification of several types of information visualization based on aesthetic criteria. Based on aesthetic criteria as well as patterns and guidelines, Fry describes in [9] the process of creating a data visualization in a very accessible way and introduces a tool that simplifies the computational process for beginners. Munzner defined in [10] a layer-based workflow for the design process of visualization and validation.

Among conceptual and theoretical approaches, there are different kinds of tools and visualization libraries available today. Visualization grammars as a declarative format for creating and saving interactive visualizations like Vega [11]. Graph libraries as a set of tools to display and layout interactive graphs, for instance chartist-js or chart.js as well as business intelligence software to create interactive analysis dashboards from any data source and publish them to analyze data, high performance for large data sources, wide device support, multiple sources, and easy publishing of functions e.g. [12].

In conclusion, there are many expert tools and libraries as well as specific methods available that are used to train professionals in the field of information design. Without such training and tool knowledge, it is almost impossible to design and create sophisticated information visualizations of multidimensional data.

3 Approach

This section outlines the concepts of our Visual Data Language (ViDaLa). ViDaLa is the process model to consolidate data into one collective, illustrative, and interactive visualization. Our conceptual model consists of six parts that are built on one another. The ViDaLa design process that is shown in Fig. 1 is based on the steps found in visualization design processes (cp. [9, 10, 13]).

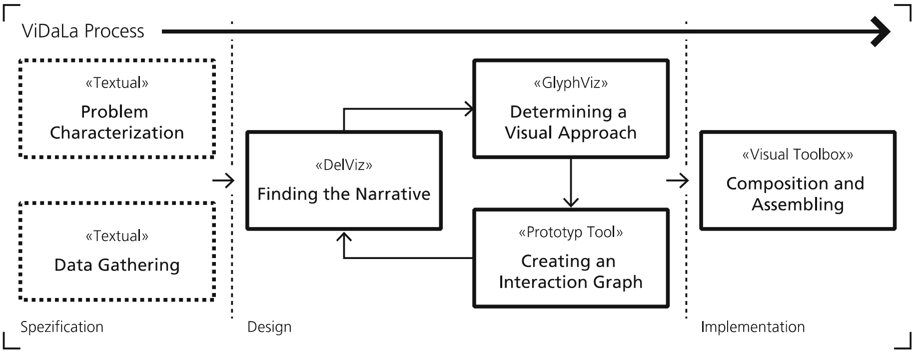

Fig. 1. ViDaLa design process starts with a specification stage, continues with an iterative design procedure and finishes with an implementation of an interactive information visualization.

3.1 Preliminary Considerations

As shown in the related work section, several models of the visualization process have outlined that there is no one way to create visualizations. There are many different paths through the process and most of them can produce useful results. The most remarkable characteristic of the process is that it is always explorative and iterative. The ViDaLa Process Model is intended to serve as a process template based on the conceptual layers of Munzners workflow for creating visualizations. The concept consists of four nested layers: characterize the task and data in the vocabulary of the problem domain, abstract into operations and data types, design visual encoding and interaction techniques, and create algorithms to execute techniques efficiently [10]. Our Process Model enriches the model, excepting the fourth layer by concrete methods and tools to help users to select appropriate design steps based on the intended usage. The following subsection describe these steps of the ViDaLa approach in detail.

3.2 Problem Characterization and Data Gathering

The first step of the ViDaLa process starts with a period of discovery. Preliminary questions should be asked to adequately develop a strategy for the information visualization in order to create a successful design that suits the needs. Only when domain challenges, user needs, and stakeholder requirements are analyzed during the early stage of the information design process, the whole development can be completed successfully and therefore create information designs that fulfil complex interweaved requirements.

In some situations, the story is well-established, but the required data are raw or messy. Similarly, the gathering of information requires additional research of randomly distributed data sources. User-centered methods for getting data and information can take different forms: surveying of users, obtaining data from existing APIs, conducting in-person interviews, or digging on network drives. This can easily be the most time-consuming part of the process. Yet, the full picture of the story is almost always assembled by multiple sources, not isolated in one dataset alone. The objective in this context is to retrieve data and information that answers the following aspects:

- Identify the quantity of data and datasets
- Find out the existing file formats and data sources and whether data need to be scraped
- Discover the complexity of the dataset
- Analyze the data set regarding to numerical range and dimensionality
- Find out if the structure of the dataset is directly usable or if it needs some restructuring
- Identify the quality of the data

In other cases, a visualization projects starts just with a dataset and the initial question is: "What is the best strategy to create a visual representation of the given dataset?" In that case, it is necessary to analyze the data in search of interesting stories and insights. Analyzing data may also require some data mining or statistics in order to come up with interesting insights. Developing a visualization strategy starts with a unique intent and an essential purpose. It might be clarifying a complex set of data, explaining a process, highlighting a trend, or supporting some kind of argument. Finding the narrative of the information visualization results from this challenge. With an existing and familiar dataset, the following basic conditions should be identified to develop a strategy:

- Identify the target audience
- Find out the type of content
- Discover time and place for the information visualization
- Analyze reasons for information need
- Identify the manner of use of the information

To discover necessary requirements and answering initial questions, different kinds of mind mapping tools as well as text editor tools can be used to quickly add records and notes to any part of the problem characterization as well as data gathering step and to reorganize aspects on the fly.

3.3 Finding the Narrative

After describing domain problems and data characteristics, additional design strategies will be applied in the process model to solve the identified problem. In almost any information design approach, there is a central aspect or fact that leads the story.

This piece of data or information will be the key element. If the element is identified, the objective is to organize data structure and story line as well as solidify the hierarchical structure of the information design. Storyboards focus on organizing, structuring, and identifying content in an effective and viable way. The determination of the story can be supported by creating personas and examining scenarios. The goal is to find user journeys to develop an information architecture. The user journey can be developed by merging the captured story pieces to create the larger picture, of how items relate to each other within the given context. The following questions will help to find the narrative:

- How does the designer categorize and structure information?
- How does the designer represent information?
- How does the user interact with information?
- How does the user search information and on which level of detail?

In order to create the story of the visualization, it is necessary to understand the correlations between users, content, and context. To address this problem, an exploration tool for existing information design was developed that illustrates characteristics of visualizations as well as the usage context. The software tool DelViz (Deep Exploration and Lookup of Visualizations) [14] supports searching for information visualizations from various points of view The application allows search and analysis in a collection of information visualizations. The data set used by DelViz currently contains 700 visualization projects which are characterized and stored with title, description, preview picture, and link to a demo of the information visualization or to a related website. DelViz is intended to support different search tasks: finding suitable information visualizations for a given context, and analysis of the underlying structured data set to discover relationships between the visualizations. Due to different search strategies, suitable visualizations and forms of depiction for the story can explored (Fig. 2).

At the end of this step, supporting elements, the level of visual detail to determine degree of interaction, the number of views, and type of views are identified and characterized to tell the story. This becomes a kind of mood board of story points and the picture of the final information design will begin to appear.

3.4 Determining a Visual Approach

After creating the storyboard with the aid of DelViz, the concrete data has to be turned into visuals. Creating a complex information design means evolving a visualization. For this purpose, visual representations for the dataset have to be discovered and determined to tell the given story. In this step of the ViDaLa process, it is an almost 1-to-1 relationship between data and visual representation.

Ware terms the basic building blocks of the visualization as “Preattentive Attributes” [15]. These attributes come into play when we determine the visual representation of data. For instance, position, and length can used to perceive quantitative data with precision. Other attributes are useful for perceiving other types of data such as categorical, or relational data. The choice of representation should address the simplest possible form that conveys the most relevant aspects of the data set. Whatever the case,

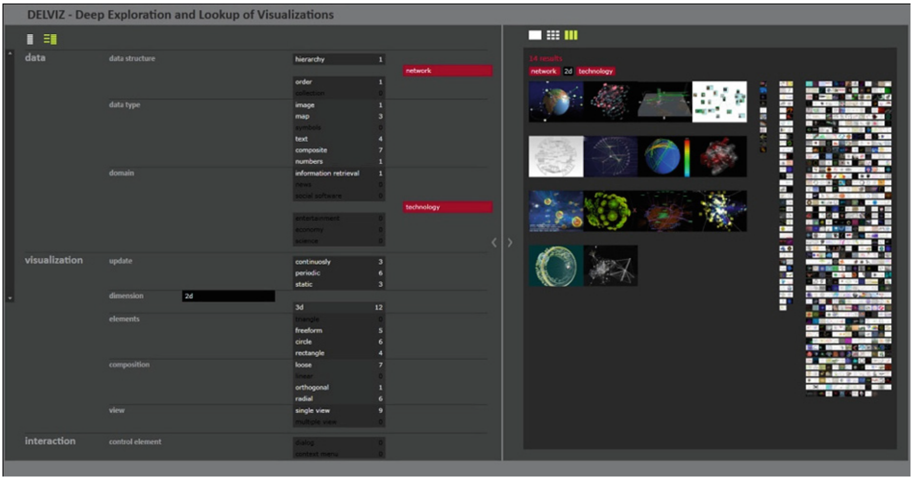

Fig. 2. DelViz interface for exploratory search. On the left-hand side the selected characteristics of information design are displayed, the right-hand side shows the resulting subset of visualizations.

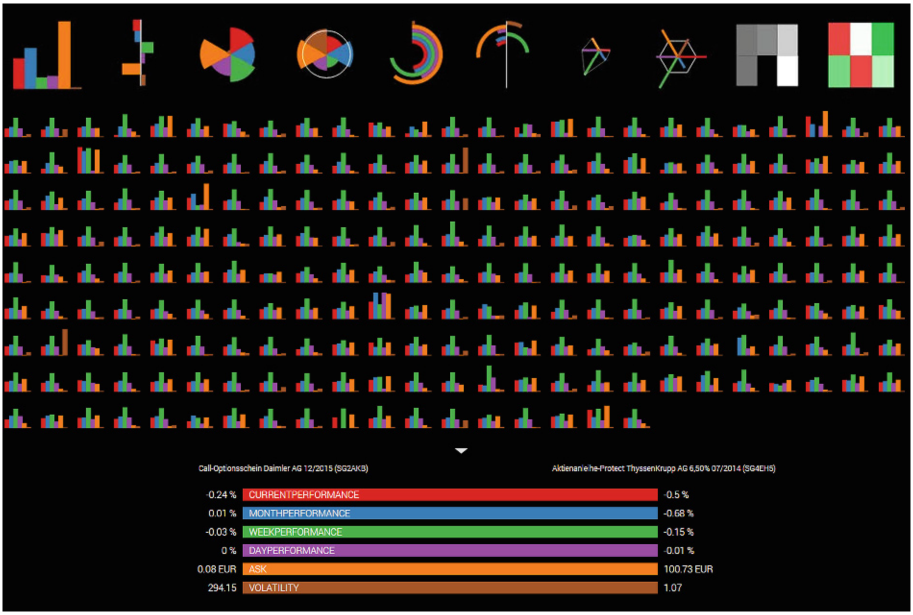

Fig. 3. Different variants of simple glyphs visualized with GlyphViz based on a finance dataset

this decision is guided by the data, which will lend itself to one visual form or a combination of several visual forms. In this design step, many intermediate visualizations have to be designed to end up with the final result. In every iteration, it has to be evaluated, whether the selected basic building blocks of the visualization works for the information design. It is hard to sketch out shape or forms without using real data. Hence, it is very important to walk through this ViDaLa step with real data. To address this problem, a software tool is used that is based on the concept of glyphs, a common form where a data set is depicted by a collection of visual representations [16]. Different properties of the data are depicted as different visual variables of a glyph. To address this challenge of determine the suitability of various visualization dependent on the dataset the GlyphViz was developed.

With GlyphViz different forms can be examined by a fast and flexible creation as well as adaption of glyphs (Fig. 3). Additionally, the tool provides sort functions according to similarity at a pre-selected data. This interactive tool supports decision making by depicting different representations of the data. These decisions should be well-founded due to design and aesthetics, but also based on the dataset and the story of the visualization.

3.5 Creating an Interaction Graph

Interactive data visualizations enable users to focus on interesting parts and details, to customize the content and even the graphical form, and to explore large amounts of data. Therefore, the change in the presentation of a data set fundamentally extends visualization capabilities. Interaction methods involve either how the data interacts with itself or how users can interact and control the data representation. In software, it is easy to provide many options to the user. Yet, it is more difficult, and more important to figure out an efficient and effective interaction in information visualization that is most relevant to the majority of the tasks executed by the majority of users. For instance, highlighting elements and showing details on demand are interactions that are useful for almost all data visualizations. Furthermore, assembling multiple representations in multiple views [17] and coordinating them clarifies different aspects of the data set at the same time. These are only some tangible quality standards for interactions in visualization as described in [4, 18, 19]. The benefit of such a configurable visualization is to display different structurally similar data sets that the user can interactively change. Additionally, when interactions are used in such a manner, an interactive visualization can make a much larger data set accessible than a comparable static graphic.

A software tool to support this ViDaLa step is still under development, yet focused on user-centered interaction within views, linking views as well as easily implement essential transitions and animations. Intended interactions will follow the guidelines of Shneiderman [4], the taxonomy of tools that support visualizations [20] and the seven general categories of interaction techniques described by Yi et al. [3]. Commercial tools such as Spotfire [21] and Tableau [22] provide examples of such interactions in visualization specification by drag-and-drop operations.

3.6 Composition and Assembling

The final stage of an interactive visualization is developed throughout this stage, but it is not only about the design. During this stage, the determined visual representations and the defined interactions are interweaved. During this implementation step, the designer builds with the assistance of the developer an understandable visual representation of the information and its hierarchy. It is an iterative process to realize the narrative of the visualization. In its first iteration it is not the final result, but a design to prove against the storyboard as well as visual elements and interactions. In the following iterations, improvements on the structure as well as refinements of the implemented strategy will take place. The ability for a designer to iterate based on the existing tools is the key to the semi-automatic information design. This might include the addition or removal of features, modification of the visual design, or the improvement of interactions.

The iteration-based implementation of the visualization is supported by our Visual Toolbox. The focus of this tool is to exploit the power and flexibility of the JavaScript visualization library D3 [23] in the user-centered information design of complex and multidimensional datasets. The information design technology is used to provide powerful information visualizations based on the process model of ViDaLa. The Visual Toolbox is an attempt to build re-usable visualization and interaction components for D3 without taking away the power of the library itself based on building blocks of interactive visualizations. The toolbox is a very recent collection of components, with the goal of keeping these components very customizable, staying away from standard chart solutions. It provides an editor based on graphical elements to fulfil the realization step and is currently maintained by a team of frontend software engineers at

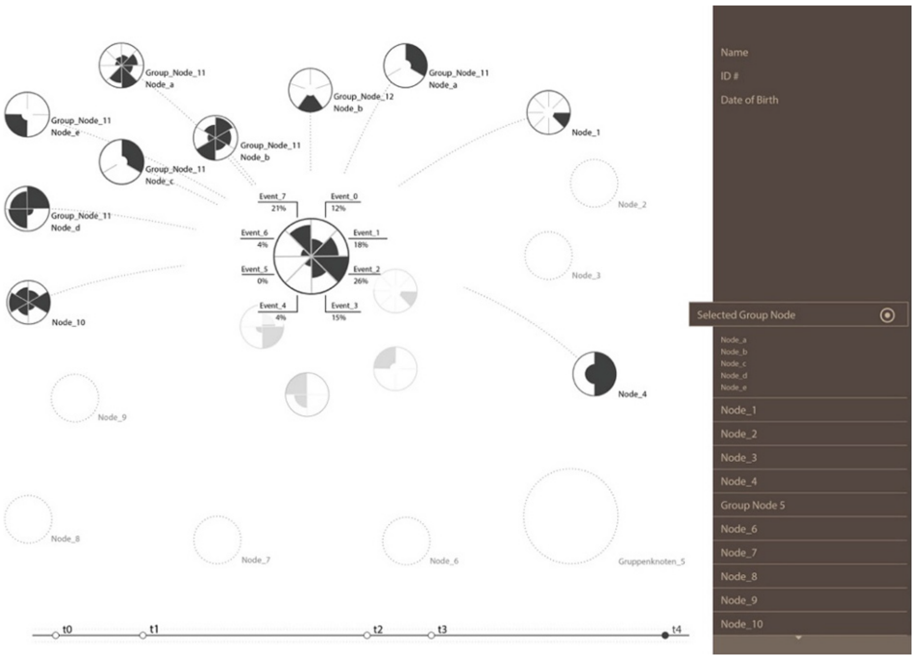

Fig. 4. Information design of the graph-based visualization of a patient model that depicts a detailed node (descriptions and probability of events) in the center of the graph and its relations to other nodes and group nodes as well as a menu on the right side to navigate within the network.

Gesellschaft für Technische Visualistik. The Visual Toolbox harnesses the power of visual intelligence to handle complex and multidimensional data and enables the user to shape them into an interactive visualization.

4 Case Study

To evaluate the ViDaLa design process, digital medical records of patients in a hospital setting were identified as a complex and multidimensional dataset (Fig. 4). The digital patient record consists of a large number of isolated data entries that are provided by various health information systems and medical devices. In order to transfer the digital patient and decision process model into real clinical applications, the ViDaLa process was applied and the digital medical records of patients were visualized in a graph based-visualization. In the context of interactive representations of clinical processes and information management, a user-intended categorization of information architecture was developed. It was decided to visualize digital patient records as a network visualization. A circle glyph combined with a pie chart is used to visualize data of the patient in a network graph to present dependencies, connections and hierarchies.

5 Conclusion

With this paper, we illustrated the importance of visualizing information in business units as well as in research and development. The presented ViDaLa approach constitutes a design process for interactive visualizations using a process model that is based on the analysis of given datasets and analysis of a given problem domain.

The presented tool chain is a consistent set of individual tools that are used to perform information design tasks to create an interactive visualization. In general, all of the ViDaLa tools are executed, however, not necessarily executed consecutively. Moreover, the Visual Toolbox contains a visualization library that consists of graphical elements that can be composed and arranged according to the goals of the visualization. The approach presented in this paper allows developers that are not specifically trained in design and information visualization to create visualizations, which assist users in exploring data in order to gain insights and information. Applying the ViDaLa process model, the development time due to a model-based approach can be reduced, while retaining customization of complex visualizations.

ViDaLa aims to make complex and multidimensional data more focused and more understandable for users via visualizations tools and visual analytics methods. The overall goal is to place a strong emphasis on quality over quantity, especially in the big data world. Visualizing data in an effective, creative way will provide more relevant, understandable information. ViDaLa will help to simplify the design process necessary to create visualizations that focus on the value for the end user.

6 Future Work

With the ViDaLa model as a tool-supported information design process, open research questions about the effective and efficient perception of information can be pursued with great ease. This semi-automatic design process offers possibilities to broaden the range of information design in the context of analytics and business intelligence. With the foundation of the current process model, the aim is to identify more criteria, methods, and techniques that can be integrated into the ViDaLa tools of the trade. Therefore, design and development of a concrete interface can be enhanced and the number of design iterations can be reduced. In addition, the software tools supporting the ViDaLa process model are subject to constant further development and maintenance. Ideally, the tools from specification to implementation form a toolchain that can be executed consecutively, by making the output of each tool or design step the input or starting environment for the next one.

References

1. Chen, C.: Top 10 unsolved information visualization problems. IEEE Comput. Graph. Appl. **25**, 12–16 (2005)
2. Heer, J., Agrawala, M.: Software design patterns for information visualization. IEEE Trans. Vis. Comput. Graph. 12(5), 853–860 (2006) doi:http://doi.ieeecomputersociety.org/10.1109/TVCG.2006.178
3. Yi, J.S., Youn, K., Stasko, J.T., Jacko, J.A.: Toward a deeper understanding of the role of interaction in information visualization. IEEE Trans. Vis. Comput. Graph., 13, pp. 1224–1231 (2007) doi:10.1109/TVCG.2007.70515
4. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings of IEEE Symposium on Visual Language, pp. 336–343. IEEE Computer Society Press (1996)
5. Borchers, J.O.: Interaction design patterns: twelve theses, pp. 1–6. Workshop Pattern Lang. Interact. Des. ACM Press, Hague, Netherlands (2000)
6. Lau, A., Vande Moere, A.: Towards a model of information aesthetics in information visualization. In: 11th International Conference on Information Visualization 2007 IV07, pp. 87–92. IEEE (2007)
7. Lang, A.: Aesthetics in information visualization. Trends Information Visualization, 8 (2009)
8. Kosara, R.: Visualization criticism-the missing link between information visualization and art. In: 11th International Conference on Information Visualization 2007 IV07, pp. 631–636. IEEE (2007)
9. Fry, B.J.: Computational information design. Massachusetts Institute of Technology (2004)
10. Munzner, T.: A nested model for visualization design and validation. IEEE Trans. Vis. Comput. Graph. **15**, 921–928 (2009)
11. Satyanarayan, A., Wongsuphasawat, K., Heer, J.: Declarative interaction design for data visualization. In: Proceedings of 27th Annual ACM Symposium User Interface Software Technology, pp. 669–678. ACM (2014)
12. García, M., Harmsen, B.: Qlikview 11 for Developers. Packt Publishing Ltd, Birmingham (2012)

13. Fry, B.: Visualizing Data, 1st edn. O'Reilly Media, Sebastopol (2008)
14. Keck, M., Kammer, D., Iwan, R., et al.: DelViz: exploration of tagged information visualizations. In: Interaktion und Visualisierung im Daten-Web (2011)
15. Ware, C.: Information Visualization. Perception for Design, 2. Auflage. Elsevier Ltd, Amsterdam (2004)
16. de Almeida Madeira Clemente, M., Keck, M., Groh, R.: TagStar: a glyph-based interface for indexing and visual analysis. In: Proceedings of 2014 International Working Conference on Advanced Visual Interfaces, pp. 357–358. ACM (2014)
17. North, C.L., Shneiderman, B.: A Taxonomy of Multiple Window Coordinations. Department of Computer Science, University of Maryland, USA (1997)
18. Shneiderman, B.: Designing the User Interface. Strategies for Effective Human-Computer Interaction, 4th edn. Addison-Wesley, Boston (2004)
19. Tidwell, J.: Interaction design patterns. In: Conference on Pattern Language Programming (1998)
20. Heer, J., Shneiderman, B.: Interactive dynamics for visual analysis. Queue **10**, 30 (2012)
21. Spotfire. http://spotfire.tibco.com/de/. Accessed 12 Feb 2016
22. Tableau. http://www.tableau.com
23. Bostock, M., Ogievetsky, V., Heer, J.: D3: Data-Driven Documents. IEEE Trans. Vis. Comp. Graph. Proc. InfoVis **17**, 2301–2309 (2011)