Comparing Information Visualization Tools Focusing on the Temporal Dimensions

Eva Wohlfart[1]
Wolfgang Aigner[2], Alessio Bertone[2], Silvia Miksch[2]

[1] Vienna University of Technology, Institute of Software Technology and Interactive Systems (ISIS), Favoritenstrasse 9-11 / 188, A-1040 Vienna, Austria

[2] Danube University Krems, Department of Information and Knowledge Engineering (ike), Dr.-Karl-Dorrek-Strasse 30, A-3500 Krems, Austria

Abstract

Empirical comparisons and categorizations of information visualization tools lack important considerations: the former undervalue the need for a theoretical background, and the latter tend to have too much distance from the user because they do not consider definite user tasks. Therefore, our work combines these approaches and presents the results of both a qualitative evaluation and a recently published categorization. We focus on the visualization of temporal data and reveal that current tools realize only a small part of the visualization possibilities in this field.

Keywords—**categorization, information visualization, evaluation, qualitative**

1 Introduction

Time is an important data dimension with distinct characteristics that is common across many application domains. Approaches to display and interact with temporal data range from linear timeline visualizations to novel ideas employing visual metaphors. However, due to only a few comparison studies available, it is quite cumbersome to assess and select Information Visualization (InfoVis) toolkits. In this paper we use a recently published categorization [8] and a set of user tasks to perform a detailed comparison of several visualization tools, focusing on the temporal aspects of data. We show that these aspects are rarely considered in the design of visualization tools and their functions.

Empirical analyses of InfoVis tools use predefined tasks to assess their usability. Categorizations for InfoVis tools evaluate them by assigning meaningful criteria. Both approaches have a limited view: the former undervalues the need for a theoretical background, and the latter tends to have too much distance from the user because they do not consider definite user tasks. We overcome these problems by combining both approaches and showing the overlapping results. Thus, we aim at an improved view of the examined tools, and contribute to the following aspects:

- we use a novel method to compare InfoVis tools: by combining a theoretical categorization and a qualitative approach we enhance the quality of the comparison.
- we evaluate a recently published categorization and point out possible improvements that should be considered in designing such a categorization.
- by combining a categorization and user tasks we give a detailed overview of the features and shortcomings of current InfoVis tools.

2 Related Work

Considering the evaluation of InfoVis tools we perceive two directions: empirical evaluations with statistical analysis and theoretical approaches which aim to find a general taxonomy. These taxonomies rarely address the specific needs of temporal visualization, e.g. the relationships between temporal primitives. This illustrates the importance of our work that combines both empirical and theoretical approaches to yield an even profounder evaluation of a visualization tool. Related work that led to this concept is presented next.

The temporal visualization framework presented by Chittaro and Combi defines four aspects: time points, time intervals, temporal relations and logical expressions [9]. Our work restructures these aspects by integrating them either in the categorization or in the task definition process. The differentiation between time points and time intervals is reflected in the first criterium *time* of the categorization. Furthermore, the low-level tasks in the practical work deal with the sequence of measured values, which is similar to the aspect of temporal relations.

Wiss and Carr [14] construct a classification framework considering three cognitive aspects: attention, abstraction and affordances. Visual features, like color or shape, help directing the *attention*. *Abstraction* can be seen as information hiding to ease the perception of the visualization, e.g., filtering or grouping. The *affordances* of a visual object are the cues it gives to show what can be done with it, e.g., a button affords pushing. These cognitive aspects are considered in several parts of our work: the categorization analyses if dynamic concepts are used and judges the level of abstraction. In the qualitative evaluation, filtering mechanisms and the visual representation are explored.

Some parallels to the employed categorization are found in the taxonomy by Daassi, Nigay and Fauvet [10]. They define four steps of the visualization process: time, point of view on time, time space and point of view on the time space. In the second step, *point of view on time*, the representation of the time values is chosen, e.g., whether time is considered linear or cyclic. This step is reflected in the categorization criterion *time* that considers the structure of time. At the third step, *time space*, the time values are mapped to a visualization, correspondent to the *representation* criterion.

The task definition process was guided by existing empirical approaches. In general, the composition of user task sets follows common rules: single and multiple variable ranges are covered, and relationships between the attributes have to be discovered [12], [13]. More advanced tasks consider the comparison of multiple criteria and the detection of trends.

3 Comparison Setting

Before evaluating the InfoVis tools, some aspects had to be taken into consideration. The adequate tools and the best suiting categorization had to be chosen. Also, a set of tasks had to be found that corresponds as much as possible to real-life visualization requirements. The following sections describe these parameters.

3.1 Tool Selection

In this work, we focus on two-dimensional InfoVis tools to increase the meaningfulness of the evaluation because of their direct comparability. Furthermore, tools adopting a third dimension often appear more complicated to the user and are therefore used less. The three-dimensional tool cviz [2] was not compared to the others because the data import did not work and the contacted support did not answer. We selected the four tools Spotfire 2.0 [4], Tableau 3.0 [5], Xmdv Tool 7.0 [7] and ILOG Discovery Preview Version [3] to provide a profound overview of the currently available InfoVis tools. They use a linear structure for the visualization. We did not choose InfoVis tools using other temporal structures because they are only available for specialized tasks and not for any desired data. Tableau, Spotfire and ILOG Discovery are produced and promoted by a corporation, whereas the Xmdv Tool is distributed free-of-charge and developed by persons all over the world.

3.2 Categorization

In our work we aim at a detailed comparison of InfoVis tools visualizing temporal data. A lot of categorizations have been published that deal with InfoVis tools, but only a few of them focus on time-dependent data. We are confident that the degree of adaption to the type of data influences the quality of the categorization. Therefore, we chose the categorization by Aigner et al. [8] that builds upon three criteria: time, data and representation. Moreover, a lot of ideas that are mentioned in related research are considered in it. In the following, we present the parts of the categorization in detail.

The time axis can be made up of time points or intervals. The structure of the time axis can be linear, cyclic or branching. Considering the criterion data, it distinguishes spatial and abstract data with no inherent spatial structure. It makes a difference if each temporal primitive is associated with a single data value (univariate data) or if multiple values have to be considered. Visualizing the raw data is useful in many scenarios, but complex data sets require data abstractions tailored to the user's needs. Static representations show still images of points in time. Dynamic representations use the physical dimension time to visualize the information. The presentation space itself can be 2- or 3-dimensional.

3.3 Task Selection

For the practical part, a data set containing pollution data measured in Great Britain was used [6]. We enhance the task categorization provided by Alan M. Eachren [11] and divide the user tasks into two categories: basic tasks handling only one or two variables at a time and advanced tasks dealing with more complex problems, like the determination of trends. Thereby, we achieve a controlled qualitative evaluation that still focuses on the needs in real life. In the following, a list of the defined tasks is shown. Table 1 shows the connection of the defined tasks and the categorization.

Basic Tasks

Which temporal range is covered by the Ozone measurements in Plymouth?

Was a value of 3 of Nitric Oxide ever measured together with a value of 27 of Nitrogen Dioxide?

Which year had the highest ozone pollution in London?

How much did the Sulphur Dioxide pollution change between January and July 2003?

Advanced Tasks

Are there pollutants having high values at specific seasons?

Are there any correlations between pollutants?

Has the general pollution increased in the last five years?

3.4 Qualitative Evaluation

The previously defined tasks were accomplished by a visualization expert with every tool and the results compared. The following paragraph describes the accomplishment of two example tasks.

One of the basic tasks dealt with the co-occurence of two events: "Was a value of 3 of Nitric Oxide ever measured together with a value of 27 of Nitrogen Dioxide?"

Task	Lev. of Abstraction		# Variables		Temp. Prim.	
	Data	Abstract	Univ.	Multiv.	Points	Interval
Which temporal range is covered by the Ozone measurements in Plymouth?	x		x			x
Was a value of 3 of Nitric Oxide ever measured together with a value of 27 of Nitrogen Dioxide?	x			x	x	
Which year had the highest Ozone pollution in London?		x	x			x
How much did the Sulphur Dioxide pollution change between January and July 2003?		x	x			x
Are there pollutants having high values at specific seasons?	x	x	x		x	x
Are there any correlations between pollutants?	x			x	x	
Has the general pollution increased in the last five years?		x	x	x		x

Table 1: Connection of Categorization and User Tasks. *Our work combines both a categorization and a set of user tasks. The table shows the user tasks and their corresponding categorization categories, which are important for accomplishing the task themselves*

With Tableau, the drag-and-drop technique and the so-called shelves (see also Figure 1) are used to display the result: Tableau adjusts the time axes dynamically to show the relevant data points. With the scatterplot in Spotfire, a wrong result was retrieved first due to overlapping points. Then the visualization was split to single days and the correct result was displayed. Using the Xmdv Tool, the user brushes along the parallel coordinate axes to show the right values. Because of the slowed down performance of ILOG Discovery, this task was not carried out with this tool.

One of the advanced tasks was "Has the general pollution increased in the last five years?". Herein, the user has to analyze the whole dataset and draw conclusions about the general trend. Dragging the year, region and measure values variables to the shelves of Tableau creates a bar visualization of the pollution over the last five years. It indicates that populous regions are most polluted, as London and Birmingham in the southern center of Great Britain, mostly due to high Ozone and Nitrogen Dioxide values. Plymouth in the southwest is the least polluted city. A bar visualization is used in Spotfire as well. The Xmdv tool lacked aggregation functions to solve this task, and ILOG Discovery's performance problems also did not allow the completion.

3.4.1 Tool Overview

Summing up the experiences of the practical work our impression is that Spotfire and Tableau offer more functions than the other tools. The interaction process employed by Tableau allows access to all important functions and is easily learned, because it is based on the drag-and-drop interface metaphor. Tableau's user interface provides excellent feedback about which visualization parameters are associated with the variables displayed. Spotfires implementation of this aspect raised some problems: sometimes irritating results were produced because a filter limited the data range but the user had to scroll to see the setting of this filter. But Spotfire offers slightly more functions than Tableau, e.g., more aggregation functions and different kinds of filters.

ILOG Discovery and the Xmdv Tool provide several visualizations for the display of data. The user chooses the visualization he/she wants to use and works with it. Simple interaction mechanisms are offered as well, range filters by ILOG Discovery and brushing by the Xmdv Tool. But the facilities for changing the visualization, e.g. the color of items, are relatively small compared to Tableau and Spotfire. ILOG Discovery had serious problems handling the size of the data set, and the Xmdv Tool lacked the aggregation of variable values. Therefore, these tools give a first overview of a data set, but for in-depth analysis other tools, like Tableau or Spotfire, are more suitable.

3.4.2 Handling of Temporal Information

ILOG Discovery and Xmdv Tool treated the data like any other sequential input and positioned the data points along an axis. Spotfire and Tableau recognize the date format correctly and combined temporal values inside an interval by a user-specified aggregation function. The control of the granularity of the time axis is realized quite similar in Tableau and Spotfire. Using Spotfire, the default granularity "year" can be changed by clicking on the axis that represents the time. Tableau employs the well-known metaphor of plus and minus signs used in directory trees to adjust the temporal granularity. Figure 2 illustrates this concept.

3.5 Categorization

After the practical part including some initial training and the task accomplishment we apply the categorization to the InfoVis tools. The tool features are collected and summarized in Table 2.

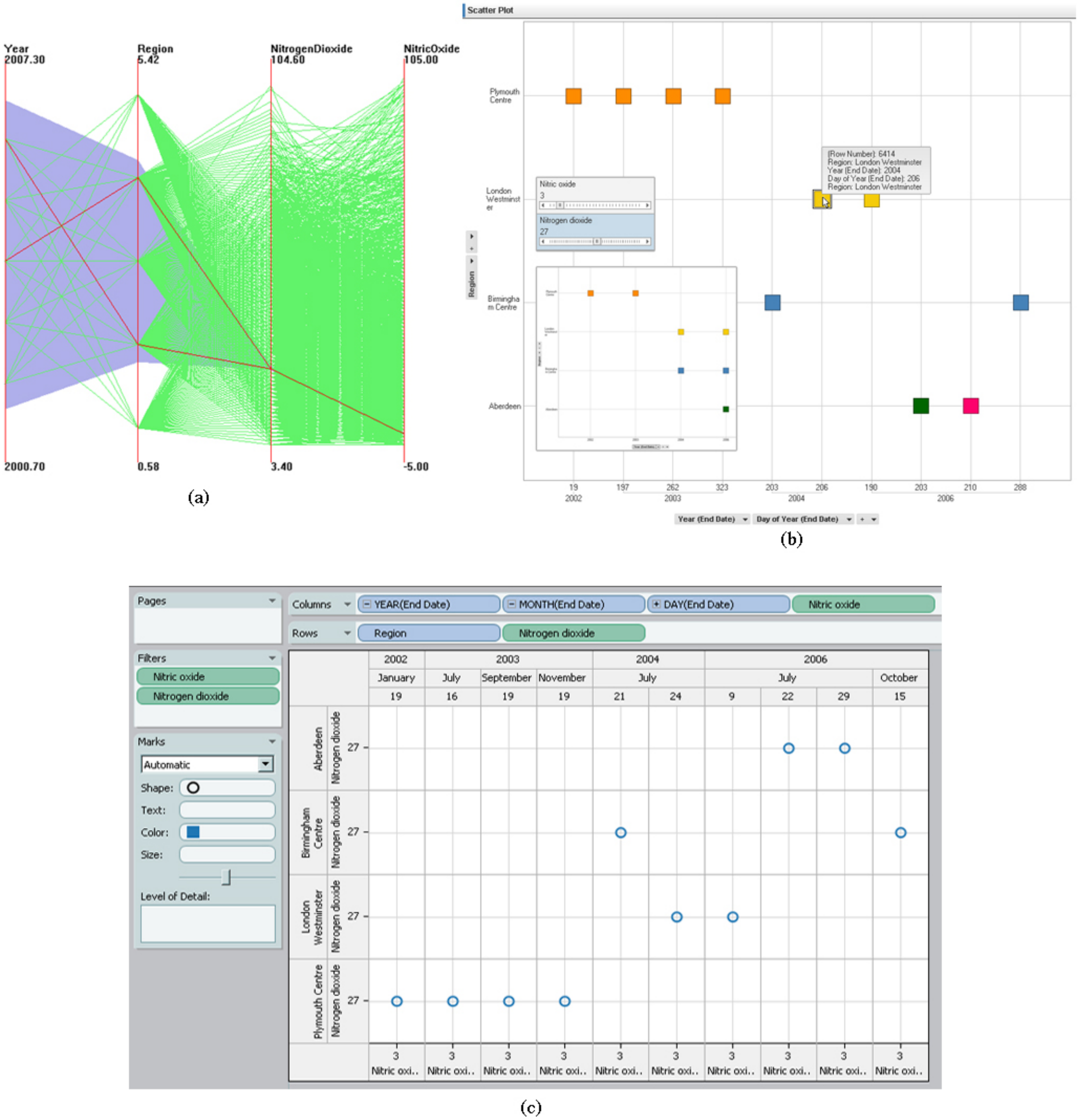

Figure 1: Task Comparison. *The three tools Xmdv Tool, Spotfire and Tableau show corresponding Nitric Oxide values of 3 and Nitrogen Dioxide values of 27. In this case, the correct result is ten times. (a) By brushing along the parallel coordinates visualization of the Xmdv Tool the matching points are revealed by red lines, as shown on the top left. The wrong result of two lines may result in missing data points during import or overlapping data points. (b) Spotfire visualizes the same task. In the small window, the matching points in time are displayed without splitting up the visualization to single days. Therefore, only 7 points seem to meet the condition. After refining the visualization, the correct 10 points are shown. (c) The two fields at the top of Tableau's user interface, the so-called shelves, represent the variables visualized. The user drags the variables there and the visualization is generated. Tableau then adjusts it to show only the relevant data points. The user sees at first glance on which date and in which region a data point was measured.*

3.5.1 Evaluation of the Categorization Scheme

The categorization gives a good overview to guide an initial judgement of visualization tools. In general, the application of the categorization was done without problems.

Considering the criterion level of abstraction, one has to choose if a tool shows raw data or provides abstractions, e.g. calculated aggregation values. Here the number and character of abstraction methods available are important to compare visualization tools adequately. Therefore, a further differentiation of this criterion would be an enhancement to the categorization.

The ability to interact with visualizations fundamentally influences the users process of exploration. Interaction methods essentially contribute to the usability of a tool. Thus, the addition of a fourth category considering the factor interaction would improve the categorization as well.

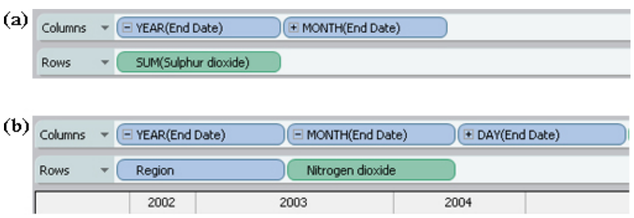

Figure 2: Granularity Adjustment. *Using Tableau, the user adjusts the granularity of the time axis by clicking on plus and minus signs at the left of the time interval icons. Figure (a) shows the shelves before splitting up to days. By clicking on the plus sign of the month symbol the new icon representing days is added to the shelf, as shown in Figure (b). The visualization's detail is changed accordingly to show the exact measured values of each day.*

3.6 Results

All examined InfoVis tools consider time as a linear dimension. This sequential view of time is reflected in the common visualizations available: parallel coordinates, scatter plots and simple bar or line plots.

Aggregation functions that enable the user to compare collected data in different intervals are the most common support to ease the exploration of large data sets. The Xmdv Tool does not have the possibility to aggregate data values but handles large data sets with hierarchical clustering: cluttered visualizations are clarified by combining multiple similar visualization elements to one element.

A smooth input process without any changes to the input file was not available for any of the examined tools. The real-life data set used contained missing values which caused errors during read-in. Some tools, namely the Xmdv Tool and ILOG Discovery, were not able to recognize a date format. Input file formats were quite diverse as well: besides the common csv format, other proprietary formats were demanded that required adding some lines to the beginning of the file.

Large data sets containing multiple variables and a lot of data points pose a big problem to several of the examined visualization tools, e.g., ILOG Discovery. For example, the data set used in this work had more than 9000 lines and caused problems ranging from slowed down user interaction to complete crash of the tool. However, Tableau allowed the user to access these data by saving the project after reading in the data and therefore using some kind of caching.

3.7 Possible Improvements of InfoVis Tools

The conversion of the data set to different file formats made the read-in quite cumbersome. Therefore, a standard for formatting input files is needed, especially for the format of temporal data, which was recognized by only two tools. The complying with such a standard would greatly improve the input data handling.

Some of the examined tools had essential problems dealing with the size of the used data set. The judgement of a visualization tool not only includes the kinds of data and formats supported, but also its performance abilities. Therefore, it should either support real-life data without problems, as in the case of Tableau and Spotfire, or specify the allowed data set size in their documentation if they do not support large data sets.

Some of the above mentioned problems is solved by the recently available Time Intelligence Solution (TIS) [1], which supports the analysis of time-oriented data. On the one hand, TIS uses a standard format as input format to allow flexibility and reuse of different operators and some simple visualization methods. On the other hand, TIS is able to handle a huge amount of input data with good performance, but needs to be improved for the visualization and interaction capabilities.

4 Conclusions

The InfoVis tools compared have quite different features and allow more or less user interactions. Once the range of functions increases, the simplicity and clarity of what can be done and how becomes more and more important. The structuring of user interfaces of InfoVis tools is definitely an important field for further research.

The specific temporal information contained in the data was correctly recognized by only two visualization tools. This illustrates how rarely time-oriented data are considered when designing a visualization tool. Due to the frequent relation of scientific data to the dimension time awareness for the specific needs of temporal data has to be raised.

All tools position temporal events along a linear axis. Especially when dealing with large data sets functionalities for the discovery of periodicity would effectively improve the usability. The adoption and implementation of periodic

	Categorization	**Tableau**	**Spotfire**	**ILOG Discovery**	**Xmdv Tool**	Practical Counterpart
Repres.	*Time dependency*	Static	Static	Static	Static, Dynamic	Static visualization Use of animations
	Dimensionality	2D	2D	2D	2D	Two-dimensional visualization
Data	*Level of Abstraction*	Data, Abstract	Data, Abstract	Data, Abstract	Data, Abstract	Raw data visualization e.g., Aggregation functions
	#Variables	Univariate, Multivariate	Univariate, Multivariate	Univariate, Multivariate	Univariate, Multivariate	Display single variables Show variables in relation
	Frame of Reference	Abstract Spatial	Abstract, Spatial	Abstract	Abstract	Abstract/Spatial data can be explored.
Time	*Structure of time*	Linear	Linear	Linear	Linear	Linearly structured visualization
	Temporal Primitives	Time points Time intervals	Time points Time intervals	Time points	Time points	Show single points Visualize intervals

Table 2: Categorization. *The most functions and features are clearly offered by Spotfire and Tableau. They allow the display of both time points and intervals and also display spatial information. All examined tools position temporal events along a linear time axis and do not offer any features that would help the user to discover branching or cyclic temporal structures. Dynamic visualizations can be generated when using the Xmdv Tool.*

data seems to be generally underrepresented and remains as a field for future research.

Acknowledgements

This work was partly supported by the program "FIT-IT Visual Computing" of the Federal Ministry of Transport, Innovation and Technology, Austria. Project number: 813388. For more information about this work, see the Master's Thesis [15].

References

All web references have been accessed on March 20th, 2008.

[1] Time Intelligence Solution, `http://wiki.ximes.com`.

[2] CViz, `http://www.alphaworks.ibm.com/formula/CViz`.

[3] ILOG Discovery, `http://www2.ilog.com/preview/Discovery/index.html`.

[4] Spotfire, `http://www.tibco.com/software/business_intelligence/`.

[5] Tableau, `http://www.tableausoftware.com/`.

[6] UK National Air Quality Archive, `http://www.airquality.co.uk/archive/data_and_statistics_home.php`.

[7] Xmdv Tool, `http://davis.wpi.edu/xmdv/`.

[8] Wolfgang Aigner, Silvia Miksch, Wolfgang Müller, Heidrun Schumann, and Christian Tominski.*Visualizing Time-Oriented Data - A Systematic View*.In Computers and Graphics, volume 31, pages 401–409. Elsevier, 2007.

[9] Luca Chittaro and Carlo Combi.*Representation of Temporal Intervals and Relations: Information Visualization Aspects and their Evaluation*.In Proceedings of the Eigth International Symposium on Temporal Representation and Reasoning, pages 13–20. IEEE Computer Society Press, 2001.

[10] Chaouki Daassi, Laurence Nigay, and Marie-Christine Fauvet.*A Taxonomy of Temporal Data Visualization Techniques*.Information-Interaction-Intelligence, 5:41–63, 2006.

[11] Alan M. Eachren.How Maps Work: Representation, Visualization, and Design.Guilford Press, 1995.

[12] Qing Li and Chris North.*Empirical Comparison of Dynamic Query Sliders and Brushing Histograms*.In Proceedings of the IEEE Symposium on Information Visualization, pages 147–153. IEEE Computer Society Press, 2003.

[13] E. Morse and M. Lewis.*Evaluating Visualizations: Using a Taxonomic Guide*.International Journal of Human Computer Studies, 53:637–662, 2000.

[14] Ulrika Wiss and David Carr.*A Cognitive Classification Framework for 3-Dimensional Information Visualization*.Technical report, Lulea University of Technology, 1998.

[15] Eva Wohlfart.*A Detailed Comparison of Information Visualization Tools Using a Reference Data Set*.Master's thesis, Technical University of Vienna, 2007.`http://ieg.ifs.tuwien.ac.at/projects/infovis-compare/`.