

Visualizations at first sight. Do insights require training?

Michael Smuc¹, Eva Mayr¹, Tim Lammarsch², Alessio Bertone², Wolfgang Aigner², Hanna Risku¹ and Silvia Miksch²

¹ Research Center KnowComm, Department for Knowledge and Communication Management, Danube University Krems

² Department of Information and Knowledge Engineering, Danube University Krems
Dr. Karl Dorrek Str. 30, 3500 Krems, Austria

{michael.smuc, eva.mayr, tim.lammarsch, alessio.bertone, wolfgang.aigner, hanna.risku, silvia.miksch }@donau-uni.ac.at

Abstract. Understanding novel visualizations can be a challenge even for experienced users. During iterative usability engineering phases in the DisCō project, visualizations of time-oriented data are explored by domain experts and non-experts. The aim of our study is to analyze the generation of knowledge and understanding by means of visualizations without previous user training. Focusing on applicability in various business domains for personnel planning and time scheduling, we tested mockups of visualizations with a method based on user-reported insights. Results show almost identical behavior of domain experts and non-experts when generating insights into the data from scratch. In the course of working with a visualization, an interchange of insights into the visualization and insights into the data was found.

Keywords: Usability engineering, insight method, Visual Analytics, time-oriented data, participatory usability design

1 Introduction

User studies for the evaluation of complex visualizations with non-experts often start with intense training sessions (e.g., [1][2][3][4]). In this context training focuses not only on the usage of tools but also on the acquisition of expert and domain knowledge [5][6]. For various scientific domains, such as bioinformatics [7], meteorology [8], and mathematics [9], the demand to integrate huge amounts of expert knowledge to obtain deeper insight and understanding of data visualizations seems evident. But this kind of knowledge-based training is not necessarily needed to this extent for other domains. In the business domains of personnel planning and workforce scheduling, complexity arises mainly from the interference of irregular and highly entangled time structures and huge multivariate datasets.

Although most of the central variables sound familiar from experience in daily work life and are comprehensible for most of us (e.g., volume of sales, occupancy rates), there are many differences between experts and novices in the interpretation of

2 Michael Smuc¹, Eva Mayr¹, Tim Lammarsch², Alessio Bertone², Wolfgang Aigner²,
Hanna Risku¹ and Silvia Miksch²

data. Experienced personnel planning consultants are time-structure-experts with deep knowledge and cognitive scripts for data checking and aggregation, a strong background about conventions, policies, and procedures about customer's business and elaborated mental models about economy. All these attributes are necessary for rich data interpretation, problem solving, and the generation of forecasting models. If structures of time are suitably put into pictures, many valuable insights into data can be gathered during explorative analysis.

One decisive point for personnel planning consultants is that results and data insights have to be communicated in most cases to semi- or non-experts for time structures. Usually time for additional explanations about the inner logic of the visualizations is limited during presentations. Although all of us are confronted with various time structures in daily life, such as business days, weeks, or months, the question raises, how fast and accurate can novel visualizations of time-oriented data be comprehended? How intuitive are these new visualizations for non-experts as well as for experts of "temporal analytics" [10]? In this paper we want to address these questions.

This paper is structured as follows: At first we want to give some insights into the challenges of depicting time-oriented data. Because the users' understanding is a crucial factor in the field of data visualization, we conducted a case study that will be described in the subsequent sections.

2 Visualizing Time – A Challenge for Visual Analytics

During the last decade capabilities to both generate and collect data and information have seen an explosive growth. Advances in scientific and business data collection (e.g., from remote sensors or from retail and production devices) as well as in data storage technology enabled us to store high volumes of data and information. However, they clearly overwhelm the traditional manual methods of explorative data analysis. The generation of new methods and tools for more complex explorative data analyses are the subject of the emerging field of "Visual Analytics" [11]. Its basic idea is the integration of the outstanding capabilities of humans in terms of visual information exploration and the enormous processing power of computers to form a powerful knowledge discovery environment.

Exploring trends, patterns, and relationships are particularly important tasks when dealing with time-oriented data and information. Time-oriented data are highly complex in their structure: there are different granularities (e.g., minutes, hours), different forms of divisions (e.g., 60 minutes per hour but 24 hours per day), different calendar systems (e.g., Gregorian or Business calendars), and irregularities like leap seconds or leap years. Moreover, time contains natural, but also social cycles and reoccurrences; for example seasons or holidays.

Within the DisC¹ project we aim to develop novel Visual Analytics methods to visually as well as computationally analyse multivariate, time-oriented data. The

¹ From Latin "discō" (inf. *discere*), which means "I learn".

novel visualizations are developed as a plugin for TIS² (described in [12][13]). In the DisCō project, the development of visualizations is accompanied by usability studies to support the human reasoning process and ensure intuitive human-computer-interaction.

3 Usability Studies

The usability study design is characterised by an early focus on users and their tasks. The different stages of usability evaluation can be described in three phases, namely task and user analysis, iterative process of User Driven Design, and Usability Testing (cp. Table 1).

Table 1. Usability evaluation in three phases

Phase	Methods	Questions & Topics
1. Task and User Analysis	10 qualitative interviews (users and software developers)	<ul style="list-style-type: none">○ Who are the potential users of the system (e.g. educational background)? What kind of expectations do they have?○ What are the typical goals and information needs of users?○ What context do users work in (infrastructure, tools, ...)?
2. Iterative Process of User Driven Design	<ul style="list-style-type: none">○ Mockup testing, think-aloud, interviews○ Usability inspection and heuristic evaluation○ Focus groups in order to provide additional information about preferences	<ul style="list-style-type: none">○ Discussion and optimization of visualizations with potential users, analysis of insight patterns○ Identification of guideline violations; in-depth testing○ Participatory development of interaction design
3. Usability Testing	Summative testing procedures	Validate usability design

The present study relates to the iterative design process (phase 2), where the first mockups of visualizations were tested. In order to present the context of this study, we want to give a brief overview of some results of the already completed user and task analysis (see table 1, phase 1). Further, some specific peculiarities will be discussed with regard to existing findings in literature.

Who are the users of the visualization tools? The participants described their work as business consultants, human resources planners, and controllers. Their tasks are to plan and forecast personnel requirements and to evaluate organisational interventions in this field. In addition, they analyse time-related personnel data for different groups

² [TIS] (Time Intelligence Solutions ®) is a flexible Business Intelligence Software developed by XIMES GmbH. It focuses on time related data with the aim to support personnel planning, controlling, and forecasting. Thereby, TIS facilitates the fast development of company specific solutions and the exchange of methods and ideas in this field.

or locations in conjunction with metadata and various management ratios. The users described the application scenarios of data analysis in different industries, like transport and logistics, service industries, retail, health care, and the public sector.

Users reported that they were often confronted with ill-defined problems; although customers are aware of having a problem, they are frequently unable to describe or even to name it. Hence, data analysis has to be mainly exploratory at first. Similar to the workflow in exploratory data analysis in scientific domains (see [14][15][16] and [8]), the workflow in the field of human resources planning lacks linearity. It can rather be described as an iterative or circular process. During their data analysis process, users reported several cases where it was necessary to go back one or more steps and change, normalise, or filter data. On the data level, frequently reported cases are inconsistencies in data like missing values or implausible time shifts and missing consistency with other variables. These validation processes are supported by an extensive toolset and are reported as one of the most time consuming parts of the analysis. On the user level, new insights and deeper understanding of data and their relations, adapted frames of understanding [15][17], and newly developed hypotheses (derived from previous results) often generate the necessity for recalculation.

Testing statistical hypotheses is not of high priority in these domains. Due to narrow time boundaries, results usually have to be gathered within minutes or hours, data analyses that span over weeks or months rarely occur. Visualizations are not seen as a “by-product” of the analysis (as stated by [16]), but as a tool supporting analysis and presentation of results. Users favour visualizations which can be presented to their management board or CEOs with little additional effort of editing or with automatic support for creating new visualizations to illustrate their insights and findings. Therefore, one of the requirements for the software application is intuitive visualization of highly complex multivariate, time-oriented data – so that it can be understood without extensive training lessons.

User requirements, as described above, have to be satisfied. The aim of participatory design is to meet these requirements. To test the support of the human reasoning process, various novel methods have been developed over the last years. Classical benchmarking metrics, like efficiency and efficacy or time and error rate, turned out to be of limited use for the improvement of information visualizations as they are typically task based and used in a highly standardised experimental setting. Therefore, tasks have to be compact and predefined. Definitive, unambiguous, and distinctive answers are forced through the experimental setting and time constraints leave little room for deeper elaboration of the findings [18]. Due to the exploratory nature of knowledge discovery in Visual Analytics, new paradigms [5][19] for testing “beyond time and errors” [20][21] were promoted to fill these explanatory gaps. One of these metrics is the qualitative and quantitative measurement of user reported insights [7], which will be described in the following sections.

3.1 Defining Insights

A review of literature on insights in visualizations from Psychology and Visual Analytics showed that no commonly accepted definition exists [1][7][22][23]. Hartmann [22] sees the problem of the term insight in using “specialized meanings for

[an] ordinary word” (p.242). While in Cognitive Psychology insights are sometimes defined as a sudden realization of a problem’s solution, as an aha-experience [23], in Visual Analytics insights are often defined as discoveries about data [7]. We are not able to solve this ongoing discussion, especially as the granularity of insights is unclear: is a new insight gained when a pattern is detected in visualizations, when scripts for data analysis are identified, or when a mental model about the data is completed? Therefore, we define insight only for the purpose of this study as *the generation of new knowledge by individuals out of visualization for data analysis*. We focus on the finest granularity of insight, a single observation of the visualization.

4 Insight Study

This study addresses the question of whether prior knowledge about a specific visualization is necessary to generate insights. Most studies implemented extensive training [1][3][4][7] prior to the generation of insights or invited participants who were already familiar with this visualization [7][8][9]. By selecting this procedure, researchers implicitly assumed that some knowledge about a specific visualization is necessary. But is this the case?

To shed some light on these questions we conducted a case study. For this purpose we carried out and analyzed interviews about two visualization-mockups.

4.1 Methods and Materials

To provide a better understanding of the applied visualizations, we would like at first to give a brief introduction into Cycleplot and Multiscale visualization techniques. Afterwards, the participants of the case study and the analysis procedures will be described in detail.

4.1.1 Cycleplots

One of the more prominent effects of the structure of time is the fact that special patterns emerge from granularities, like day of week or hour of day. For example, values from one Monday are closer to values from the next Monday than values from the next Friday. The next Monday also has a next Monday that is similar. Values from a Tuesday are of course similar to values from the next Tuesday. A pattern emerges that is seven days long. Patterns like this are called cycles. Another phenomenon that is often found in time-oriented datasets is a trend. A trend is a long-term effect that overlays short term patterns or cycles. If a cycle has very obvious fluctuations, it can be hard to detect a trend that is overlaid by it. Fig. 1 shows a dashed line depicting a cycle and a dotted line depicting a trend. The solid line shows what happens if this cycle and this trend overlap.

Simple attempts on visualizing time-oriented data aggregate to one granularity and show a certain range of data in a single curve. Improvements upon this visualization split the data and show several curves. E.g., seven curves are shown for the days of week, but only one week is shown in total and each curve shows one day straight

from the beginning to the end. These visualizations only show a very general trend. However, it is possible that different trends exist for the several parts of a cycle. E.g., Mondays are getting quieter while Sundays are getting busier. This fact is hidden by the visualizations we have described so far.

Cycle, Trend, and Combination of Both

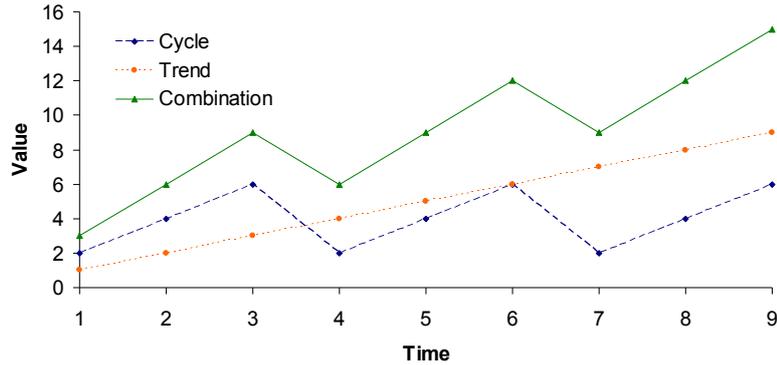


Fig. 1. Example for cycles and trends. The cycle is plotted as a dashed line, the trend is plotted as a dotted line, their sum is plotted as a solid line.

Polzeinsätze an Montagen im August 2005
 Vertikale Striche gruppieren nach Tageszeiten
 Innerhalb jeder Gruppe entsprechen die Datenpunkte den Tagen 8.8., 15.8., 22.8., 29.8.

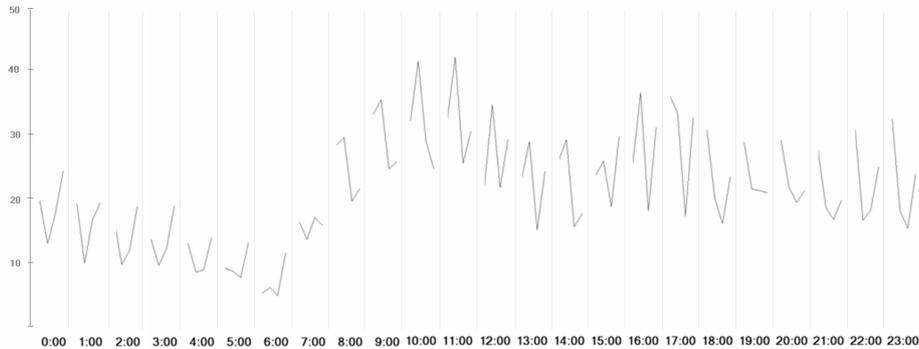


Fig. 2. Cycleplot example, used for mockup testing (text in German). The horizontal axis shows several bins grouped by time of day, each containing the Mondays over the course of one month. The vertical axis shows number of simultaneous police assignments (title translated “Police assignments on Mondays in August 2005”).

Cycleplots have been introduced by Cleveland [24][25] as a means to detect multiple trends in substructures of time. When generating a Cycleplot, the user has to choose a granularity that is used to group the data into bins. The important point is that this granularity should not be the coarsest one but the one that contains the most prominent cycle. Inside each bin, the whole selected range is plotted, but filtered for the value of the granularity that corresponds to the bin. Finer granularities than the

one used for the grouping are usually aggregated to an average value. Fig. 2 shows an example. The number of simultaneous assignments of a police department has been plotted over the course of a month. These assignments mostly result from emergency calls, so it is difficult to plan them in advance. Experts in shift planning have the task to find out how many assignments are most likely to take place. In this example, the data has been filtered to show only assignments on Mondays. It is aggregated to use hours as the smallest granularity. Every hour of the day is shown in a separate bin. Note that the course of the line inside a bin does not reflect the minutes of that hour, but instead shows the average value of that hour of the day for several different Mondays. Trends are visible, but they are different for each hour. This visualization is well-suited for classification done by the user.

The Cycleplot performs very well in the detection of trends and some other aspects. However, for a data set the size of our example, it needs aggregation or filtering. When showing all available data points at a lower level than days, the optical clutter becomes too strong. We will now present a visualization that can show more details.

4.1.2 Multiscale Visualizations

To deal with a huge amount of time-oriented data to be displayed, several interaction mechanisms are usually employed: *Scrolling* is used to show only a part of the dataset. There is no loss of detail by employing scrolling, but it is difficult to overlook the data or compare values that are temporally apart from each other. *Filtering* (e.g., showing only Mondays) is very similar to scrolling. *Aggregating* (e.g., calculating the mean over a certain amount of time) is effective for gaining overview, but it hides a great deal of information.

To reduce the need for scrolling or aggregation, it is necessary to show as much data as possible in a limited space. This can be done by encoding the position on the time axis using two dimensions. In these visualizations, it is necessary to find a different way to encode the value. One way is using different colors for different values. The position within a two-dimensional grid is determined by one time granularity for the position along the horizontal axis (e.g., day of week) and the value of another time granularity for the position on the vertical axis (e.g., week of month). A well-known example for this arrangement is a calendar sheet, with the day of the week on the horizontal axis and the week of month on the vertical axis. While a calendar plots the number of the day of month in different colors inside each square, a matrix-based visualization usually fills the whole block with a color that represents a data value. If the size of such a block is shrunk down to a single pixel, we also refer to them pixel-based visualizations.

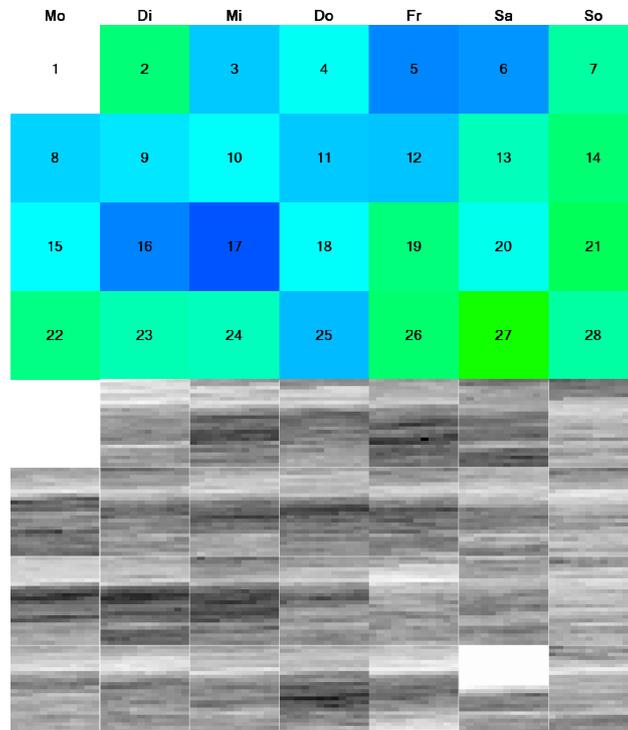


Fig. 3. Multiscale visualization example. Each block shows one day. Inside each block of the finer scale display (bottom), rows represent hours and each pixel in a row a 5-minute-interval.

Multiscale visualizations have been developed by Shimabukuro et al. [26] in a work mainly focused on the combination of time-oriented and space-oriented data where overview and detail are combined in one visualization. Although the work of Shimabukuro et al. has a strong focus on space-oriented data, the Multiscale visualization is solely for time-oriented data. A range of time (e.g., a month) is divided in a matrix-based visualization (e.g., a calendar sheet). In a parallel view, a second version of the visualization is shown. In that view, each block is further divided (e.g., in hours and minutes) using again a two-dimensional matrix. Therefore, each larger block contains another matrix-based visualization. The visualization makes strong use of the structure of time, allowing the combination of four granularities and their data content. Unfortunately, the original paper from Shimabukuro et al. does not stress the granularity aspect further as it is focused in a different direction. For us, it is important to stress that any combination of granularities can be used in any order.

The example in Fig. 3 shows the same dataset of police assignments as the Cycleplot example (described above). However, the finer granularity consists of 5-minute-blocks. The fact that even at this time raster, a whole month can be displayed (without filtering for Mondays) shows how large datasets can be displayed in full detail using the Multiscale visualization (this visualization consists of 7803 separate

data points). We have used a visual mapping of data value to color (hue) on the coarse level and a visual mapping of data value to gray scale on the fine level.

We have increased the flexibility of the Multiscale visualization by allowing the mapping of any granularity on any of the four nested axes. We can combine both views within one visualization by mapping the average value at the coarse level to hue and the average value of the fine level to gray scale (see Fig. 4). This variant consumes less space and combines overview and detail more integrally, freeing the user from the need of switching the focus back and forth.

Polizeieinsätze August 2005, Tagesdurchschnitt und 5-Minuten-Raster

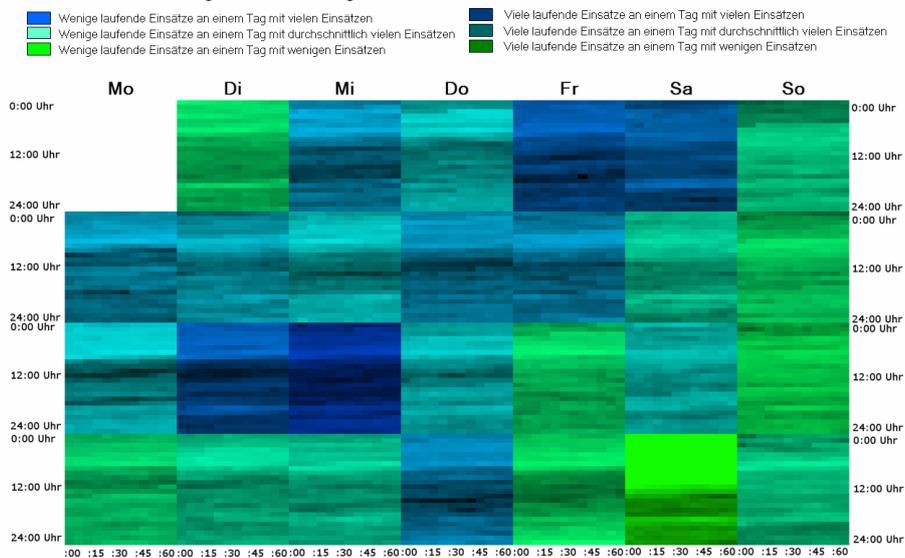


Fig. 4. Multiscale visualization example, used for mockup testing (text in German). Each block shows one day. Inside each block, rows represent hours and each pixel in a row a 5-minute-interval (title translated: “Police assignments in August 2005, daily mean and 5-minute-raster”).

Both examples reveal the following phenomenon: Tuesday 16th and Wednesday the 17th have a hue that shows very high activity of the police. While the hue of Monday 15th suggests that there were less assignments on that day, a look at the value, showing the finer granularity, reveals that the high number of assignments already started on Monday morning. The quiet night from Sunday to Monday hides that fact if the data is aggregated to days, as in our initial example. The Multiscale visualization can reveal the true circumstances.

It is possible to make comparisons along any of the granularities when reading a Multiscale visualization. Inside each block, we can read along an hour or compare hours to each other. Between blocks, we can try and detect trends over the course of the week, or you can compare a day of the week to the same day one week later. In the example, Sundays are very similar too each other, but there are very different kinds of Saturdays.

These two novel visualization techniques – the Cycleplot and Multiscale visualization – were tested using the procedure described in the following section.

4.1.3 Procedure

To test the visualizations (Fig. 2 & 4), participants were asked about their insights into the visualizations using the think-aloud method. The two visualizations were presented as printed out mockups, without trainings and further explanations. When a participant was not able to generate any more insights into a visualization after repeated requests, a standardized short explanation of the visualization (cp. 4.1.1 and 4.1.2) was provided. After this explanation, we asked the participant again to generate insights and suggest further improvements. To reduce the interviewer’s workload, audio protocols were recorded.

4.1.4 Analyses

Prior to the analyses we transcribed the audio protocols from all interviews. Then, we segmented the transcripts into smallest meaningful units (see [27]) and created a time-stamp for each unit from the audio protocols.

As first step in the analysis of the interviews we coded each transcript segment as an insight or not an insight: we coded an insight when new knowledge was generated by the participant. We then grouped the insights according to the knowledge generated: Segments referring to the same knowledge were coded as one insight with the time-stamp of the first reference.

For the categorization of insights we followed the procedure of North and Saraiya [7][18]. We coded insights with respect to the following characteristics:

- *Observation*: Occurrence of insights for each visualization
- *Time*: When was the insight generated (e.g. before or after explanation)?
- *Correctness*: Was the insight correct?
- *Category*: Which types of insights were observed?

A first analysis of the insights showed two main categories of insights: Insights into the visualization and insights into the data the visualization displays. At first sight, this categorization seems to be different to existing insight studies, but according to North [18] insight “capture[s] the intuitive notion of a visualization’s purpose” (p. 6). This definition compasses both categories of insights used in our study.

Visualization insights can be further divided into insights what the visualization displays (“how-insights”), insights how the visualization has to be read (“meta-insights”), and insights how the visualization can be improved (“improvement-insights”; see Table 2 for examples of insights in each category).

In addition, we coded data insights according to the purpose of the two visualizations: Cycleplot enables the differentiation of trends and cycles. Therefore, we coded data insights into the Cycleplot as either trend or cycle. Multiscale displays both overview and detail in one integrated view. Therefore, we coded data insights into the Multiscale as either overview or detail (see Table 2).

Additionally, we coded whether prior knowledge was integrated by the participant for generating the insight and what knowledge was used (see Table 2 for an example). To visualize the uptake of prior knowledge or insights we built an uptake graph [28].

Table 2. Examples for insight categories.

Visualization Insights	How-insight	“The more green the less assignments, the more blue the more assignments.”
	Meta-insight	“Okay, first I’m looking at the days, if I can detect any patterns.”
Data Insights	Improvement-insight	“It would be good to be able to filter out one day.”
	Cycleplot: Cycle	“Starting in the morning it rises to a peak around 10, 11 am. Then it calms down at noon with a second peak around 4, 5 pm. and falls down again.”
	Cycleplot: Trend	“The first Monday is high, descending on the second, and rising again on the third and forth.”
	Multiscale: Overview	“Sundays are rather low, on average.”
	Multiscale: Detail	“Especially at noon it’s higher than before or after noon. It’s always darkest then.”
Integration of Prior Knowledge		“It decreases until 6 in the morning, to a minimum. I assume this is due to [...], to my knowledge, change of shift.”

4.1.5 Case Description

The recordings of three participants were analyzed for our study. Two of them were experts in time-scheduling and personnel planning/controlling (abbreviated as P.Y. or participant Y., P.X or participant X. in the following sections), with an experience of 5-10 years as consultants in various business domains (service industries, retail, production, health care and public sector). We interviewed also one expert (P.Z. or participant Z.) in data visualization with focus on dynamic organizational mapping and Social Network Analysis. He has 3 years experience in research and development in the academic field and consulting.

The cases were selected carefully on the basis of previous interviews in the phase of task and user analysis. The selected participants were those who showed the most demonstrative behavior when analyzing data. We are aware that three cases will not suffice to produce generalizable results. Still, according to Holzinger [29], at least three participants are necessary for think-aloud-analysis. Also, it has to be pointed out that these results are preliminary, since three other interviews are currently under transcription (see discussion). So far we have not found any indications in the additional interviews that argue against the results described below.

4.2 Results

Observation. On average, each participant generated 34.67 data insights overall. Comparing data insights for each visualization, the Multiscale ($M = 19.67$) produced more data insights than the Cycleplot ($M = 15.00$). Still, if we relate the number of data points for each visualization to the number of insights produced, relatively few insights were produced using the Multiscale. In contrast, the amount of insights into the Cycleplot was very high using this relative point of view.

Time. Participants had insights into the visualizations straightforward from the beginning; the first insights started between 13 and 79 seconds after instruction. The time course of insights for both visualizations per participant can be seen in Fig. 5.

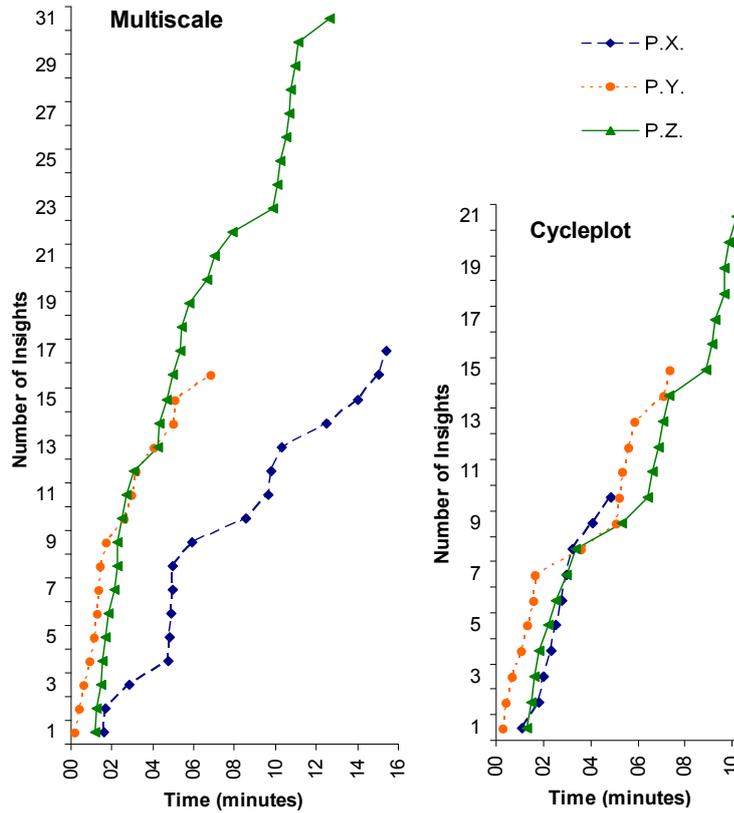


Fig. 5. Sequence of insights for Multiscale (left) and Cycleplot (right) per participant over time (P.X. = participant X.; P.Y. = participant Y.; P.Z. = participant Z.)

Participant Z. generated most insights into both visualizations, but many of them after the explanations. If the amount of insights before the explanation is compared to the amount of insights afterwards, it can be seen that the expert participants X. and Y. generated most insights before the explanation (64 and 90 %), whereas participant Z. generated most insights after the explanation (62 %). Further analyses were conducted to see whether there are qualitative differences between participant Z.’s insights on the one hand and X. and Y.’s insights on the other hand.

Correctness. Overall, almost no incorrect insights were observed. It was audible sometimes that a wrong conception of the visualization was corrected, but the final insight was correct. For example, “I missed that: It’s not one week, but four weeks.”

Category. First, data and visualization insights for each visualization technique were differentiated. A similar amount of data and visualization insights was observed for each technique (cf. Table 3).

Table 3. Data and visualization insights into the two visualizations (left two columns) and time of first insight (right)

	Multiscale Insights		Cycleplot Insights		First Insight	
	Visualization	Data	Visualization	Data	Visualization	Data
P.X.	9	9	5	5	81 sec.	203 sec.
P.Y.	7	9	9	6	15 sec.	53 sec.
P.Z.	18	13	13	8	74 sec.	112 sec.

Most visualization insights focused on how the visualization can be improved ($M = 5.33$ for the Cycleplot; $M = 5.67$ for the Multiscale). The Multiscale needed more insights how the visualization works ($M = 4.33$) than the Cycleplot ($M = 1.33$). On the contrary, more meta-insights were gained into the Cycleplot ($M = 2.67$) than into the Multiscale ($M = 1.33$). Comparing the three participants, X. and Z. generated more improvement insights, whereas Y. focused more on how- and meta-insights.

Regarding data insights into the Cycleplot, only participant Z. generated data insights with respect to categories, cycles and trends. Participant X. generated only insights with respect to cycles, whereas participant Y. generated only insights with respect to trends and once remarked that “a daily cycle can also be seen”.

Data insights into the Multiscale were more evenly distributed: Every participant gained insights on an overview as well as on a detail level. Again, participant X. focused on one level: He generated eight overview-insights and only one detail-insight.

Integration of prior knowledge or prior insights. The most important observation – though intuitively logical – is that insights into the visualization always precede insights into data. On average, the first insight into the visualization was after 57 seconds, into the data after 122 seconds (cp. Table 3, right column). As can be seen in the following uptake graphs (see Fig. 6 and Fig. 7), many data insights build upon a preceding insight, how the visualization works.

This is very clear for participants Y. and Z., but participant X. seemingly did not verbalize every visualization insight during think-aloud procedure – many of his insights require other insights that were not explicated and therefore have to be inferred by the researcher. We therefore focus in the following paragraphs on participants Y. and Z., who reported their insights more accurately and completely. We are going to analyze their uptake of prior insights and prior knowledge in the Cycleplot first and then present their insights into the Multiscale.

Fig. 6 shows the uptake graph of both participants for the Cycleplot (see Fig. 2): In uptake graphs, dependencies of events (here insights) on two or more different levels can be visualized in chronological order. The right column displays the integration of prior knowledge P. Visualization insights are viewed as H (how-insight), M (meta-insight), and I (improvement-insights) in the middle column. The left column shows data insights on cycles C, trends T, and other O. Visualization and data insights are numbered according to the time point when they were gained.

P.Y. is the only participant who fully verbalized all visualization insights, how the Cycleplot works (H1 to H3). However, in his analysis of data, he had mainly insights into trends (T1 to T4, building on H3), but did not use the insights H1 and H2 for interpretation of the data.

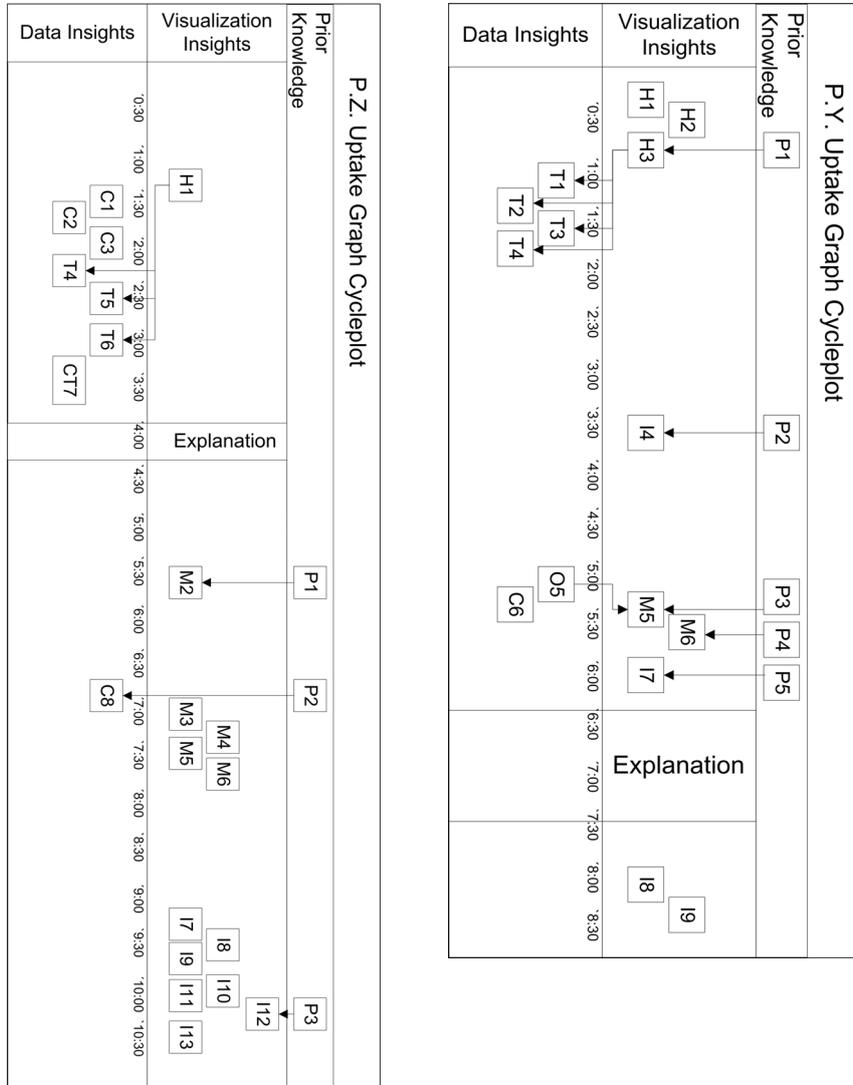


Fig. 6. Uptake graph for participants Z. (left) and Z. (right) for the Cycleplot.

Maybe the integration of prior knowledge (P1) and the identification of the practical value of these insights let him focus only on trend insights, neglecting cycle insights. In addition to the trend insights (T1 to T4) he extracted a further data insight that led to an insight of a further use of this visualization, namely a combined view of means and standard deviations (M5) that can also be used to identify outliers (M6). All other visualization insights (4 to 9) focus on improvement possibilities (14, 17 to 19). After the explanation, P.Y. did not generate any further data insights.

Visualizations at first sight.
Do insights require training? 15

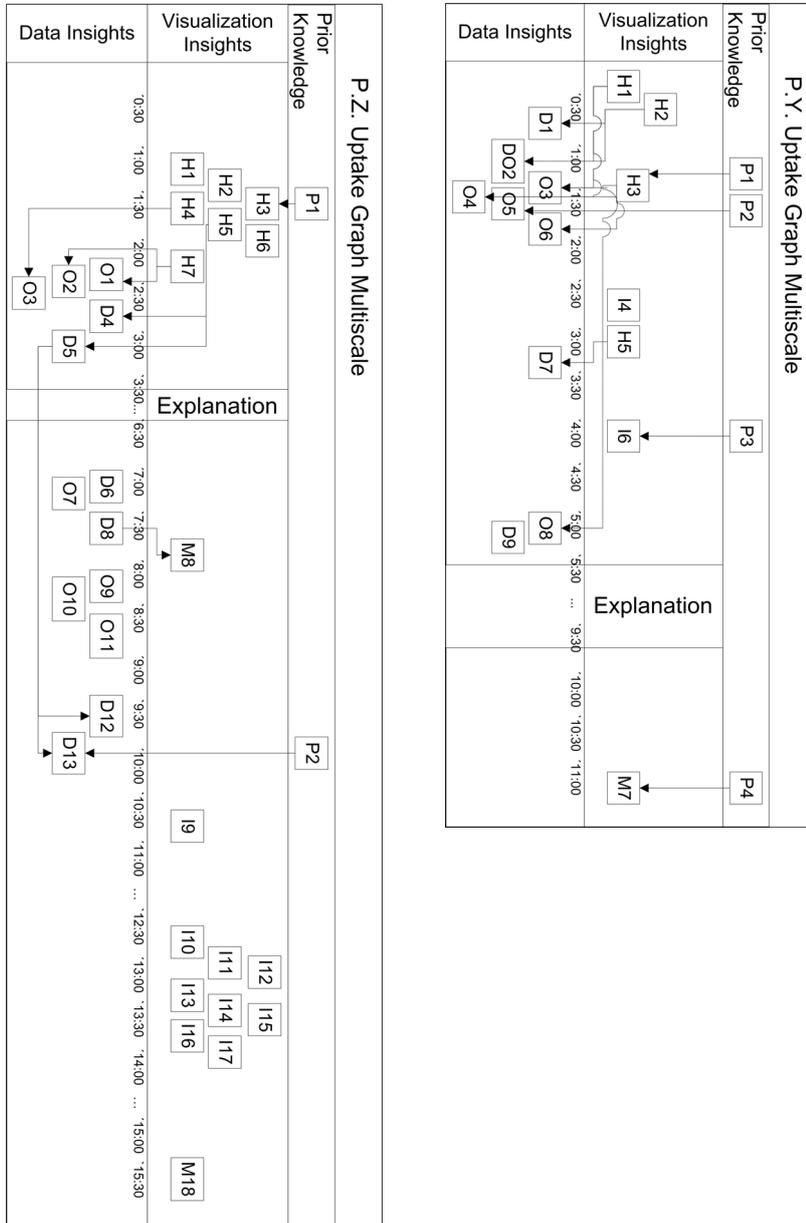


Fig. 7. Uptake graph for participants Y. (upper) and Z. (lower) for the Multiscale

What can be seen from P.Y.'s uptake graph is a high use of prior knowledge for the interpretation of the visualization; P.Y.'s cognitive scripts and knowledge about time

data influenced to a high degree how he extracted information from this visualization. In contrast to P.X., this was especially true for improvement insights into the visualization.

Comparing all participants' insights into the Cycleplot, P.Z. generated most insights – into the data and into the visualization. Still, he seemingly did not verbalize all visualization insights, as the first data insights identifying cycles (C1 to C3) do not build upon a visualization insight. In contrast the trend insights (T4 to T6) build upon an explicated insight how the visualization works (H1). All other visualization insights were either meta-insights (M2 to M6) or insights, how the visualization can be improved (I7 to I13). As was the case for the two experts, also P.Z. did not generate many data insights after the explanation. The only data insight he generated afterwards (C8) interestingly builds upon a data insight P.Z. had into the Multiscale visualization (both built upon the same data set!). This integration of prior insights into the data shows that P.Z. built a mental model on the data set across the two visualizations.

Fig. 7 shows the uptake graph of both participants for the Multiscale (see Fig. 4): it is built analogous to the one for the Cycleplot. The data insights in the first column are coded as overview O or detail D.

Participant Y.'s uptake of prior knowledge and prior insights during the exploration of the Multiscale is highly complex: nearly all his data insights build upon prior visualization insights. Very interesting is his data insight O5 where he uses prior knowledge to interpret his insights into the data. As for the Cycleplot, he does not produce any data insights after the explanation.

Participant Z's uptake graph can be segmented in four very clear exploration phases: During the first phase, he gains insights how the visualization works (H1 to H7). In the second phase he uses this knowledge to interpret the data, first on an overview (O1 to O3), then on a detail level (D4 to D5). P.Z. did not fully understand the visualization, he did not decode the 5-minutes-raster. It was explained to him, but did not lead to any data insights. Still, after the explanation he generated further data insights (D6 to D13), suggested improvements for the visualization (I9 to I17), and reflected it on a meta-level (M18). An interesting insight was D8, where he reflected his analysis behavior on a meta-level (M8).

Comparing both participants, Y.'s uptake is much more complex and his insights more often build upon prior knowledge. The time sequence of Z.'s uptakes is more clearly structured in phases. Looking across both visualizations, "individual styles" of exploration can be identified: P.Y. uses his prior knowledge to interpret both visualizations and reports his insights in full. P.Z. has similar exploration phases in both graphs and analyses the visualizations in a more step-wise fashion. What can be observed for both participants is that the Multiscale requires more how-insights into the visualization before data insights are gained than the Cycleplot.

4.3 Discussion

Overall, our analyses show that participants generated insights into visualizations from scratch, without training and even without expert-knowledge in the domain of

time-oriented data analysis. To gain insight into the data, they first had to gain insight into the visualization on which their data insights could build upon. However, they did not need a full understanding of the visualization prior to data insights; they only had to gain some relevant visualization insights first. One of our participants did not obtain a full understanding of the Cycleplot visualization at all, but he still gained insights into the data. In the course of interpreting a visualization, insights into the visualization and insights into the data interchange and partly build upon each other.

The data insights gained at first sight of a visualization were not – as one might expect – superficial, but most of them were rather complex: Henry and Fekete [30] suggest to categorize data extractions in dependence of the needed degree of interpretation as highly complex (e.g., correlations, patterns, trends), medium complex (e.g., groups, outliers), or little complex (e.g., single values). Following this categorization for the data insights into the Cycleplot, 79 % are highly complex and 11 % medium complex. The data insights into the Multiscale are slightly less complex, but still 42 % are highly complex and 52 % are medium complex. So overall, the complexity of the data insights is rather high, even though participants had only mockups for interpretation.

The high quality of the insights is also reflected by their correctness: No incorrect insights were observed. An open question is whether no incorrect mental models about the data existed, whether these models were corrected too fast to be verbalized, or whether these models existed but were not verbalized.

Even though all participants generated a high amount of insights from scratch, the insights of the experts highly depended on their prior knowledge. Still, it seems that their analysis scripts were not fully activated: An important step of data analysis, the plausibility checks and validation of the data set, was not conducted: half a day missing in the month view was not mentioned by any participant.

The prototypical progress of insights over time for all participants can be divided in three phases: first, they generated (at least one) insight, how the visualization works. Second, they generated data insights that often built upon these first visualization insights. In the end, they suggested improvements for the visualization that would let them generate further data insights. The last phase might be artificial, as participants were instructed to suggest improvements and additional interactive features. However, some of these insights were generated before this instruction, suggesting that phase three also exists without instruction during mockup testing.

Participants were instructed to generate insights into two visualizations, the Multiscale and the Cycleplot. Slightly more insights were gained into the Multiscale. Interestingly, all participants had an audible aha-experience once during the interview, when they figured out that the Multiscale visualization contains a monthly calendar view and not only a week. Therefore, the calendar-metaphor seems to suit this visualization well and leads to fast understanding when recognized. As intended, insights into the Multiscale were both on the level of details and of the overview.

In contrast, for the Cycleplot not all participants generated insights on both intended levels, namely cycles and trends: only one participant generated both kinds of insight, one participant only gained cycle insights, and one participant decisively focused only on trend insights. These results suggest that this visualization is more difficult to decode than the Multiscale. This might be due to the fact that no everyday

metaphor, like the calendar, exists for the Cycleplot. Therefore, more effort is needed to create an “ad-hoc-metaphor on the fly”. These results show a need for improvement of the visualization or of the caption so that everyday metaphors can be applied, trends and cycles are highlighted alike, and both insights are supported.

During the testing, participants generated many suggestions for improvements: the most important suggestions were derived from problems encountered with interpreting the mockup of the visualization. Additional improvements were suggested, which are mostly related to participants’ analysis scripts and prior knowledge of interactive feature.

- Individual selection of coloring (Multiscale): One participant had problems to differentiate between the green and blue colors used.
- Addition of parallel curves / metadata (Cycleplot): All participants asked for the possibility to display other curves parallel to the Cycleplot to identify correlations.
- Selection of data displayed (Multiscale, Cycleplot): Changing the initial selection, but also adding and filtering of data.
- Details onClick (Multiscale, Cycleplot): Being able to see the real value onMouseover or onClick.
- Additional statistics (Multiscale, Cycleplot): Displaying mean values for days, month, etc.; for the Multiscale it was also suggested to display the “mean color” (e.g. for weekdays) or the mean daily course.
- Zooming, scrolling, annotation (Multiscale, Cycleplot)

5 Conclusions

Many other studies [1][4][7][19] have used insight-methodology to measure the productivity related to visualizations. In contrast, in our study, insight evaluation techniques were used to analyze and improve two visualization mockups and to gain deeper knowledge about the workflow of exploratory analyses.

To stay close to the typical process of analysing time-oriented data, we asked experts in personnel planning to participate in this study. As we are confronted with time in our everyday life, we also asked a non-expert to participate in this study to compare novice and expert users. Overall, we did see little differences between experts and non-experts: both were able to generate insights into the visualization and the data from scratch in similar time and quality. However, the experts more often used their prior knowledge to interpret data. On the other hand, it seems that existing cognitive scripts also hindered alternative analysis processes. Both experts were unable to extract different data insights from the Cycleplot (cycles and trends), whereas our novice user was able to do so. Preliminary results from other interviews with novices currently under transcription show similar capabilities of non-experts. Other explanations for this difference could be a higher motivation of the novice user or that our experts – although familiar with time-oriented data – are lacking background knowledge and expertise in analysis of data from the police sector. This is also reflected by rare interpretations of gained data insights. Further studies within

this project will account for this fact by a more task-oriented design, which will allow each user to use data of his/her own domain.

Another limitation of our study is the open design of the instruction: participants were not given any problems to solve or tasks to complete. Rather, they were asked to tell whatever they see in the data. This led to a very explorative analysis behaviour, might have impeded the activation of existing analysis scripts (like checking for missing data), and provided a wide space for interindividual differences. Therefore, results should be generalized only tentatively beyond our cases.

Even though we found many differences between participants, all participants were able to produce insights from scratch. This indicates that no instruction is needed, but that these visualizations can be understood with only little effort at first sight. Therefore, these two visualizations are also suited to visually present results to an auditorium unfamiliar with them, as they are rather self-explaining or easy to explain. Our analyses further show that no complete understanding of visualizations is needed to obtain insights, but already partial understanding is sufficient.

Despite these limitations, the method used appeared to be helpful for our purpose, as this study focuses on formative rather than summative testing of visualizations. The insights gained by our participants can thereby be used to further improve and develop visualizations of time-oriented data. Insight studies, especially when focusing on insights into both the visualization and the data, turned out to be a promising methodology for participatory design of visualizations.

Acknowledgments. The DisCō project (project number: 813388) is supported by the program “FIT-IT Visual Computing” of the Federal Ministry of Transport, Innovation and Technology, Austria. Thanks to our project partner XIMES GmbH for valuable feedback and help in finding participants for our study.

6 References

1. Rester, M., Pohl, M., Wiltner, S., Hinum, K., Miksch, S., Popow, C., Ohmann, S.: Evaluating an InfoVis Technique Using Insight Reports. In: Proc. 11th International Conference Information Visualization (IV 2007), pp. 693--700. IEEE Computer Society, Washington, DC (2007)
2. Perer, A., Shneiderman, B.: Systematic Yet Flexible Discovery: Guiding Domain Experts through Exploratory Data Analysis. In: Intelligent User Interfaces. pp. 109--118. ACM: New York (2008)
3. Robinson, A.C., Chen, J., Lengerich, E.J., Meyer, H.G., MacEachren, A.M.: Combining Usability Techniques to Design Geovisualization Tools for Epidemiology. *CaGIS* 32(4). 243--255 (2005)
4. Saraiya, P.B.: Insight-Based Studies for Pathway and Microarray Visualization Tools. Doctoral Thesis. Virginia Polytechnic Institute and State University (2006)
5. Shneiderman, B., Plaisant, C.: Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies. BELIV06. ACM, New York (2006)
6. Andrews, K.: Evaluating Information Visualisations. BELIV06. ACM, New York (2006)
7. Sarayaia, P.B., North, C., Duca, K.: An Insight-Based Methodology for Evaluating Bioinformatics Visualizations. *IEEE Trans. Vis. Comput. Graph.* 11(4), 443--456 (2005)

20 Michael Smuc¹, Eva Mayr¹, Tim Lammarsch², Alessio Bertone², Wolfgang Aigner²,
Hanna Riskul¹ and Silvia Miksch²

8. Trafton, J.G., Kirschenbaum, S.S., Tsui, T.L., Miyamoto, R.T., Ballas, J.A., Raymond, P.D.: Turning Pictures into Numbers: Extracting and Generating Information from Complex Visualizations. *Int. J. Hum-Comput. St.* 53(5), 827--850 (2000)
9. Zazkis, R., Dubinsky, E., Dautermann, J.: Coordinating Visual and Analytic Strategies: A Study of Students Understanding of the Group D4. *J. Res. Math. Educ.* 27, 435--437 (1996)
10. Kuchar, O.A., Hoeft, T.J., Havre, S., Perrine, K.A.: Graphically Speaking - Isn't It About Time? *IEEE Comput. Graph.* 26(1), 80--83 (2006)
11. Thomas, J.J., Cook, K.A.: Illuminating the Path: The Research and Development Agenda for Visual Analytics. IEEE CS Press, <http://nvac.pnl.gov/agenda.stm> (2005)
12. Gaertner, J.: Evidenzbasierte Personal-Bedarfsbestimmung - Unterstuetzung durch Time Intelligence Solutions [TIS] <http://www.gor-hcm.de/htdocs/Folien/2008/Gaertner.pdf> (2008)
13. Gaertner, J.: Time Intelligence - Bedarfsanalysen, Forecasting, Arbeitszeitcontrolling. In: Fank, M., Scherf, B. (eds.) *Handbuch Personaleinsatzplanung*, pp. 173--185. Datakontext, Frechen (2005)
14. Shneiderman, B.: Inventing Discovery Tools: Combining Information Visualization with Data Mining. *Information Visualization* 1(1), 5--12 (2002)
15. Pirolli, P., Card, S.: The Sensemaking Process and Leverage Points for Analyst Technology as Identified through Cognitive Task Analysis. In: *Proc. of 2005 Int. Conf. on Intelligence Analysis*, http://vadl.cc.gatech.edu/documents/2_card-sensemaking.pdf (2005)
16. Springmeyer, R.R., Blattner, M.M., Max, N.L.: A Characterization of the Scientific Data Analysis Process. In: *Visualization '92*, pp. 235--242. IEEE Press, New York (1992)
17. Russell, D.M., Stefik, M.J., Pirolli, P., Card, S.K.: The Cost Structure of Sensemaking. In: *Conf. on Human Factors in Computing Systems*, pp. 269--276. ACM, New York (1993)
18. North, C.: Toward Measuring Visualization Insight. *IEEE CG&A* 26(3), 6--9 (2006)
19. Plaisant, C., Fekete, J., Grinstein, G.: Promoting Insight-Based Evaluation of Visualizations: From Contest to Benchmark Repository. *IEEE Trans. Vis. Comput. Graph.* 14, 120--134 (2008)
20. Bertini, E., Plaisant, C., Santucci, G.: BELIV'06: Beyond Time and Errors; Novel Evaluation Methods for Information Visualization. *Interactions* 14(3), 59--60 (2007)
21. Ellis, G., Dix, A.: An Explorative Analysis of User Evaluation Studies in Information Visualisation. BELIV06. ACM, New York (2006)
22. Hartmann, G.W.: The concept and criteria of insight. *Psychol. Rev.* 38, 242--252 (1931)
23. Jones, G.: Testing Two Cognitive Theories of Insight. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1017--1027 (2003)
24. Cleveland, W.S.: *The Elements of Graphing Data*. Hobart Press, Summit, NJ (1994)
25. Cleveland, W.S.: *Visualizing Data*. Hobart Press, Summit, NJ (1994)
26. Shimabukuro, M., Flores, E., Oliveira, M.D., Levkowitz, H.: Coordinated Views to Assist Exploration of Spatio-Temporal Data: A Case Study. In: *2nd Int. Conf. on Coordinated and Multiple Views in Exploratory Visualization*, pp. 107--117, IEEE Press, New York (2004)
27. Ratwani, R.M., Trafton, J.G., Boehm-Davis, D.A.: Thinking Graphically: Extracting Local and Global Information. *CogSci 2003*. Erlbaum, Boston, MA (2003)
28. Suthers, D., Dwyer, N., Vatrappu, R., Medina, R.: An Abstract Transcript Notation for Analyzing Interactional Construction of Meaning in Online Learning. *System Sciences, 2007. HICSS 2007*, p. 4c. Computer Society Press, Hawaii (2007)
29. Holzinger, A.: Usability Engineering for Software Developers. *Communications of the ACM*, 48(1), 71--74 (2005)
30. Henry, N., Fekete, J.: Evaluating Visual Table Data Understanding. BELIV06.. ACM, New York (2006)