Uncertainty Visualization

Two User Studies

PRAKTIKUM

im Rahmen des Studiums

Visual Computing

eingereicht von

Andreas Roschal, Fabian Schwarzinger
Matrikelnummer 1225600, 1225307

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: Mag. DI Dr. Theresia Gschwandtner

Uncertainty Visualization

Two User Studies

PRACTICAL

in

Visual Computing

by

Andreas Roschal, Fabian Schwarzinger
Registration Number 1225600, 1225307

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: Mag. DI Dr. Theresia Gschwandtner

Abstract

TODO Abstract

Contents

CHAPTER 1

Introduction

Data sets containing information retrieved from the real world usually also contain some amount of uncertainty. This uncertainty is often inherent to the data, for instance because certain measurements can never be exact or because some kind of aggregation is already done when acquiring the data. This is also true for temporal data. Sometimes the exact time of an event is not known (e.g., 'time of the big bang'), is given in an inexact way (e.g., 'since a few hours') or is an imprecise prediction of the future (e.g., 'it will take one or two days'). To incorporate these uncertainties into visual representations and make them visible to the user, several approaches have been proposed [8, 3, 10, 2, 6].

To find out more about the strengths and weaknesses of these techniques and to find out which technique fits certain tasks best, several studies have been conducted. In 2009 Sanyal et al. [12] asked 27 participants to solve four tasks with the help of four commonly used uncertainty visualizations. In 2012 Corell and Gleicher [4] compared four visual encodings of statistical uncertainty in a user study. In 2012 MacEachren et al. [9] conducted two studies, which targeted the intuitiveness of visual encodings and their performance in map reading tasks respectively. In 2015 Gschwandtner et al. [5] compared six visual encodings in a comprehensive user study.

To build upon the results of those studies, we are conducting two additional user studies. The first one (referred to as *Drawing Study*) aims to find out more about the intuitiveness of visual encodings. By asking people to draw visualizations for given tasks, we find out how people think about given problems and what kind of representations they think are most appropriate in those situations. The second study (referred to as *Evaluation Study*) is very similar in its design to the one by Gschwandtner et al. [5]. The difference is, that we do not only compare different visual encodings of uncertainty, but also include a representation in the comparison that completely omits the uncertainty of the underlying data. Through this approach we find out in which situations the visualization of uncertainty adds helpful information and in which situations it is only a counterproductive distraction.

In this report we thoroughly describe the design, execution and results of our user studies. Furthermore, we present some of the most relevant related work that has been done, which can be found in Chapter 2. In Chapter 3 we explain the design of our studies. This chapter is split into two main parts - the first is about the *Evaluation Study* and the second part regards the *Drawing Study*. In the following Chapter 4 the results are presented, which are discussed in detail in Chapter 5. We conclude this report with a summary of our approach and its most important findings in Chapter 6.

CHAPTER 2

Related Work

Since we are designing and conducting two user studies, other similar studies are of great interest to us. Through those existing works, we can learn more about the state of the art of study design and evaluation.

Obviously, the user study that guides our work the most is the one by Gschwandtner et al. [5], as our aim is to build up on this work. This study compares six different techniques for the visualization of uncertainty in the temporal domain. To determine which technique works best for certain tasks, five different types of tasks were designed. The first type is about finding out how users interpret the different visualization techniques. In the second type of tasks the users are asked to read the boundaries of uncertainty intervals from the visualization. The third type is about determining the extent of an uncertainty interval. In the fourth type of tasks, the users have to gauge certain probabilities using the visualization and the last type of tasks asks the users for their opinion about the visualization.

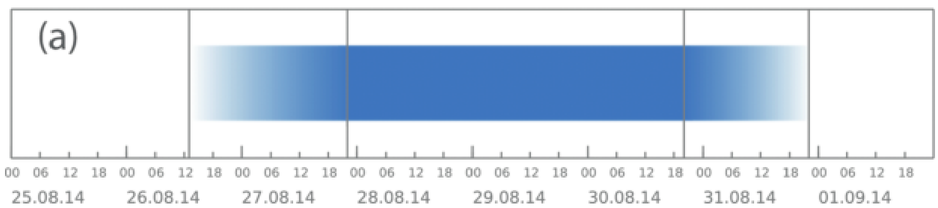


Figure 2.1: *A gradient plot shows the certain parts of an interval as a solid color, while the uncertain parts are represented by a color gradient. [5]*

The actual study was conducted with 73 participants, which all were bachelor students in computer science. The students were recruited from a course in information design and visualization, which implies a certain knowledge about this topic. To automatically track relevant data, such as completion time and accuracy, during the study sessions, the EvalBench software library

[1] was utilized. This library was designed especially for the evaluation of visualization. To analyze the results Gschwandtner et al. ran an analysis of variance(ANOVA) for each task and subtask and backed up their results with a non-parametric Kruskal-Wallis test. Their analysis showed, that the technique *ambiguation*, which can be seen in Figure 2.2, works best for tasks in which the user has to judge the exact duration and bounds of an uncertainty interval. If the user has to determine certain probabilities within the uncertainty interval, *gradient plots* (see Figure 2.1) work best.

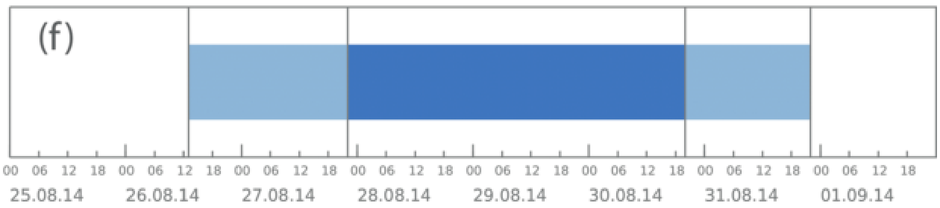


Figure 2.2: *This technique is called ambiguation and shows the uncertain parts of a time interval in a lighter color than the certain part, which is represented by a solid color. [5]*

In our *drawing study* our main focus lies on gauging how people think about certain situations and what kind of visualizations they associate with them. The goal is to find out what is intuitive for most people. These insights are valuable for the design of new visualizations, especially those aimed at non-expert users. MacEachren et al. [9] also tried to find out more about intuitive design of visualizations through user studies. Similar to us they conducted two separate user studies, which also have similar goals to ours. Their first study compares many different sets of symbols for the visualization of uncertainty, to find out which are most intuitive to people. Every set consists of three symbols, which encode high, medium and low uncertainty of 3 different kinds (accuracy, precision and trustworthiness), for 3 data domains (spatial, temporal and attribute). Some example sets can be seen in Figure 2.3. In total 102 symbol sets were rated by 31 undergraduate students for their intuitiveness on a scale from 1 to 7. After this first series of tests, the most unintuitive symbol sets were filtered out, which left 76 sets over. Those were again rated by 72 participants with a background in GIScience.

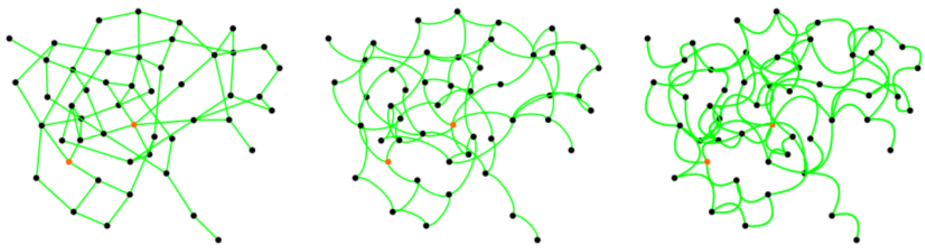
Figure 2.3: *Every column shows a set of three icons which represent high, medium or low uncertainty respectively. [9]*

After this first study about intuitiveness, the 20 highest rated symbols for every combination of uncertainty type and data domain were compared in a second experiment. The goal of this subsequent study is to compare the selected visualizations' performance, so the combined results of both studies yield the best visualizations for a given task, which is intuitive and efficient at the same time. To compare the chosen symbols, two quadratic matrices with 9 symbols each were visualized side by side. The participants were asked to answer the question which of the two matrices featured a lower overall certainty, based on the presented symbols.

Walny et al. [13] conducted a study with the goal of providing deeper insights into the way people think about and use visualizations to communicate their ideas. This study is relevant to our work, because it features a similar approach as our *drawing study*. A total number of 69 researchers were observed using whiteboards during brainstorming, thinking, communication and similar actions. Whiteboards were chosen as a visualization medium, because they support a variety of thinking tasks, like personal and collaborative cognition, group meetings and planning. The results of the study feature interesting insights, such as different uses of emphasis techniques and the usage of ellipses as a focus and context technique. Our *drawing study* aims to provide similar insights through a similar approach, by also observing users in their creation of visualizations and reviewing those drawings.

In another study of greater exploratory nature, Walny et al. [14] asked 22 participants (mostly computer science students) to sketch visualizations of a given dataset. The data was provided in a table format and was about appropriateness ratings of certain behavior in given situations. The student's task was to create visualizations to find interesting patterns in the data and articulate them in a post-sketching questionnaire. The results were analyzed through multiple coding passes, which showed that, even though 9 out of 22 participants claimed to have no experience in visualization, most of the sketched representations could be classified as known types. As already stated, the study is of exploratory nature and therefore it does not answer many questions, but rather raises interesting questions and gives direction to future research.

There are user studies in the domain of information visualization which have the goal of determining if the visualization of a certain kind of information or a certain way of visualizing it is feasible or not. Our *Evaluation Study* is one of those studies, since it aims to answer the question if it is advisable to visualize temporal uncertainties or if this information is not used in decision making and should therefore be omitted. Another similar study by Xu et al. [15] is concerned with the feasibility of curved lines in graph layouts. In the study, graphs were visualized with either straight edges or three different kinds of curved edges and users were asked to perform certain tasks with a given graph. The completion time of those tasks and whether the final user decision was correct or incorrect, served as objective measures to rate the different graph layouts. Furthermore, the users were asked to give their personal opinion on which graphs they prefer and find visually more pleasing. An example graph in the three different layouts can be seen in Figure 2.4.



(a) Straight edge (STL) (b) Slightly curved edge (SCL) (c) Heavily curved edge (HCL)

Figure 2.4: *The same graph is drawn in three different curve layouts. (a) shows the graph drawn with straight edges, while (b) and (c) use slightly and heavily curved edges respectively. [15]*

Robertson et al. [11] conducted a study to get insights about the feasibility of animations in trend analysis visualizations. To answer the question if animation helps in the comprehension of visualized trends and in the completion of corresponding tasks, users were provided with one dynamic and two static data representations and asked to perform tasks with the presented data. One of the static representations showed the change within the data in traces of the changing data points, as can be seen in Figure 2.5. The tasks were either questions regarding the presented data or some kind of analysis task. During every study session, the completion time and the accuracy of the given answers were automatically recorded. To evaluate the results of the study, four hypothesis were formulated and tested for support within the resulting data.

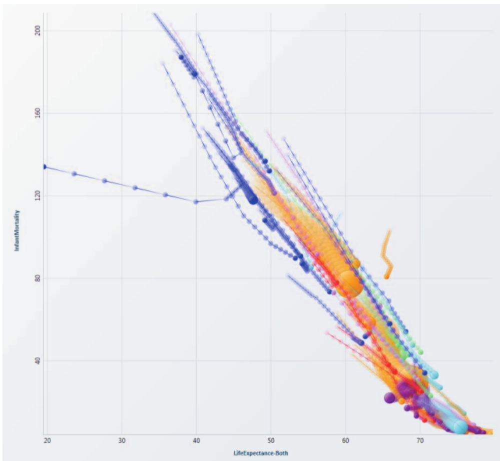


Figure 2.5: *The visualized data points are changing over time, which is statically represented by their traces. Traces are generated by drawing each point in its different stages over time and connecting these stages with lines. [11]*

CHAPTER 3

Method

In this chapter the study designs of both user studies are presented. Both studies have the aim of unraveling insights about the visualization of temporal uncertainty, but their approaches are wildly different. This is due to the varying research questions they try to answer. For this reason the two designs are presented in the following two separate sections.

3.1 Design of the *Evaluation Study*

Our *Evaluation Study* is building on the work of Gschwandtner et al. [5]. Gschwandtner et al. [5] compare different visual encodings for temporal uncertainties in order to find out which representations work best for which kinds of tasks. Our goal, on the other hand, is to evaluate for which kinds of tasks it actually makes sense to additionally visualize the information about the temporal uncertainty at all, i.e. in which cases does the user benefit from it and in which cases is the visualized uncertainty more of a distraction and actually not supporting the user in fulfilling his or her tasks. In order to examine those proposed deliberations, we designed our *Evaluation Study* and implemented it using the EvalBench framework. Figure 3.1 shows an example of the user interface of our study implementation inside EvalBench.

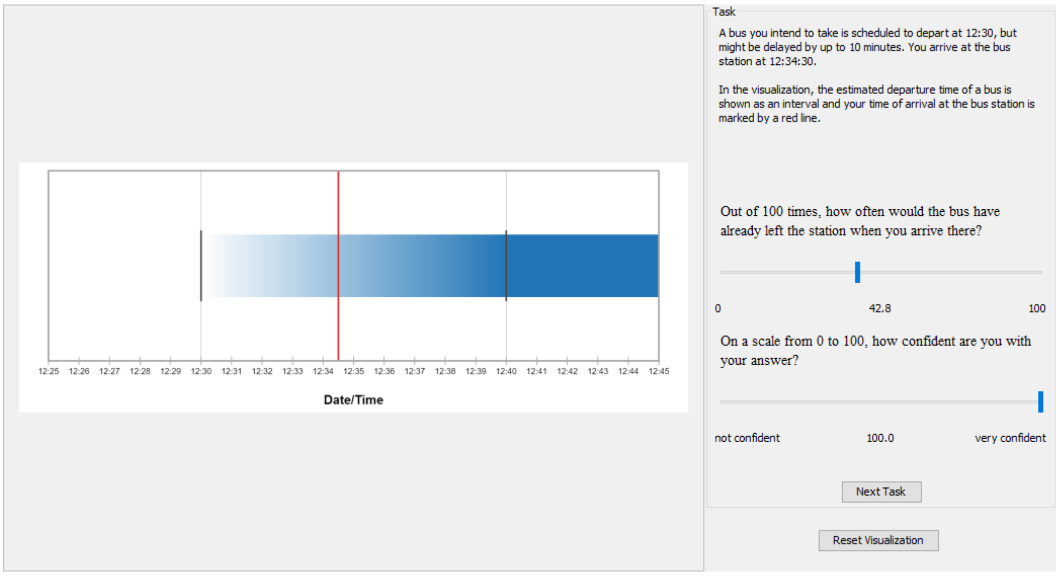


Figure 3.1: *A screenshot of our study implementation using EvalBench. On the left side the task supporting visualization is depicted. On the right side the task description and inputs for answers are displayed.*

The study was conducted on 3 different user groups. We applied a within-subject-design on the tasks and a between-subject-design on the visualization type. Hence, all participants had to solve the same tasks, but depending on the assigned user group, the participants got different kinds of visualizations, supporting them in fulfilling their tasks. The first user group got Gradient Plots, the second user group got Ambiguation Plots, and the third user group only got visualized mean values, so no visual information about the temporal uncertainty was given at all. However, textual information of the uncertainty intervals, was always provided in the task description for all user groups.

We defined four different task types representing typical questions which might be asked when it comes to temporal uncertainties. Therefore, our study consists of four sequential sessions, covering a wide range of possible tasks. For the first task type, the uncertainty interval of the start- or finish-time of some uncertain time event is given. Furthermore there is a specified point in time, usually lying somewhere inside the uncertainty interval. The participants now have to estimate the probability of the time event having already started or ended at this specified point in time. Figure 3.2 shows how tasks of the first session look like for all user groups and how the answer is selected by the participant. Additionally to the task related question, we are also asking for the participant's confidence in his or her given answer.

For the second task type, there are always two parallel uncertainty intervals showing the uncertain finish-times of two possible time events. Again, there is a specified point in time. In these tasks, the participants have to compare the probabilities of both time events at the specified point in time and decide for which uncertainty interval the probability is higher. Figure 3.3 shows how tasks of the second session look like and how the answer is selected.

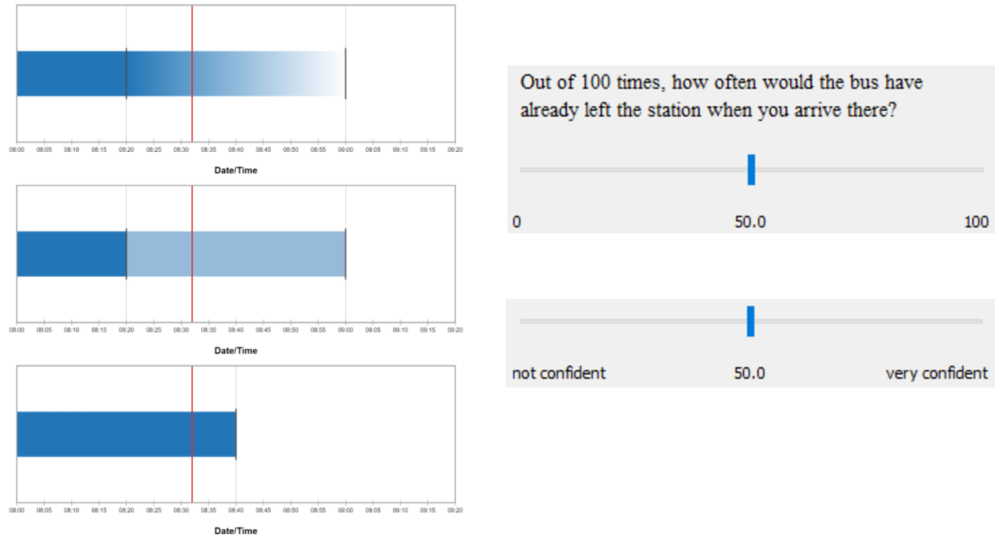


Figure 3.2: *On the left side the different task supporting visualizations for the first session are shown - from top to bottom: Gradient Plots, Ambiguation Plots, mean values. The specified point in time is marked by a red line in all plots. On the right side, there is an extract of the user interface showing the input fields for giving answers. We are not asking for a percentage value, but for some natural count instead, like suggested by Hullman [7].*

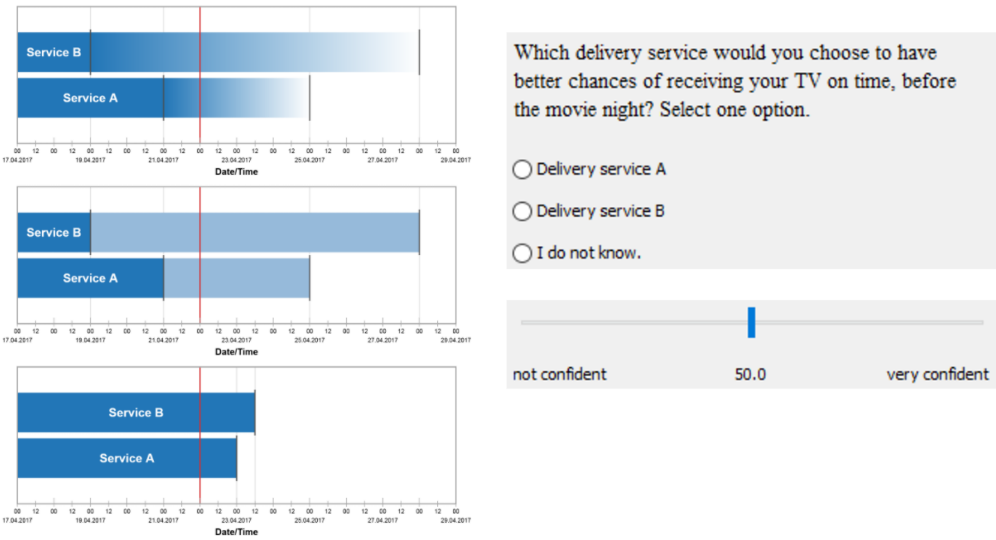


Figure 3.3: *On the left side the different task supporting visualizations for the second session are shown - from top to bottom: Gradient Plots, Ambiguation Plots, mean values. The specified point in time is marked by a red line in all plots. On the right side, there is an extract of the user interface showing the input fields for giving answers. The answer, i.e. the time event with the estimated higher probability, is selected from a set of radio buttons.*

For the third session, tasks are similar to the second session. Again, there are two parallel uncertainty intervals showing uncertain finish-times. However, this time there is no specified point in time, but instead the participants are asked to estimate which event will finish sooner on average. Figure 3.4 shows how tasks of the third session look like and how the answer is selected.

In the fourth session, tasks revolve around overlapping uncertainty intervals of two successive event. So there is some time event with an uncertain finish-time and some time event with an uncertain start-time and those time events are overlapping to some extent. Here the participants have to estimate the probability of the overlap. Figure 3.5 shows how tasks of the fourth session look like and how the answer is selected.

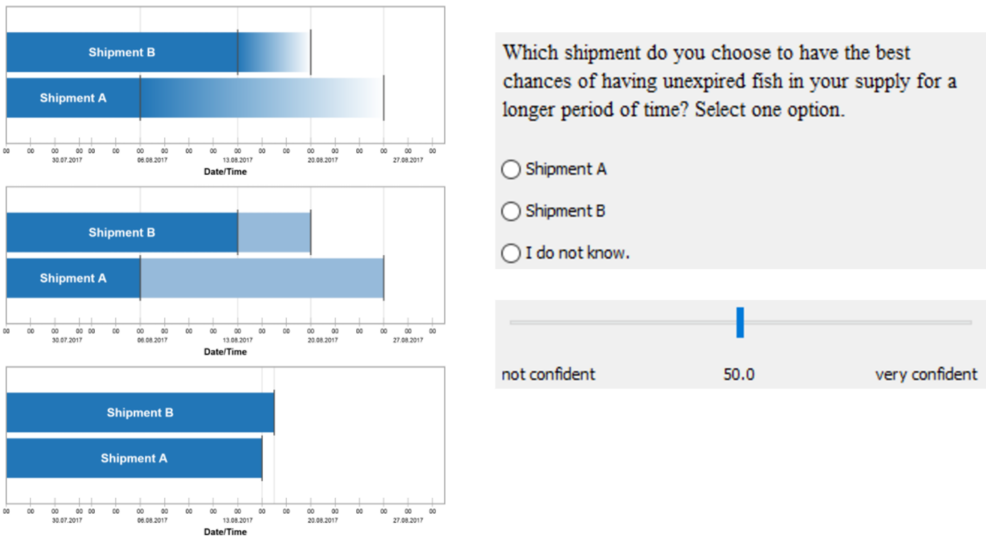


Figure 3.4: *On the left side the different task supporting visualizations for the third session are shown - from top to bottom: Gradient Plots, Ambiguation Plots, mean values. On the right side, there is an extract of the user interface showing the input fields for giving answers. The answer, i.e. the time event which is estimated to end sooner, is selected from a set of radio buttons.*

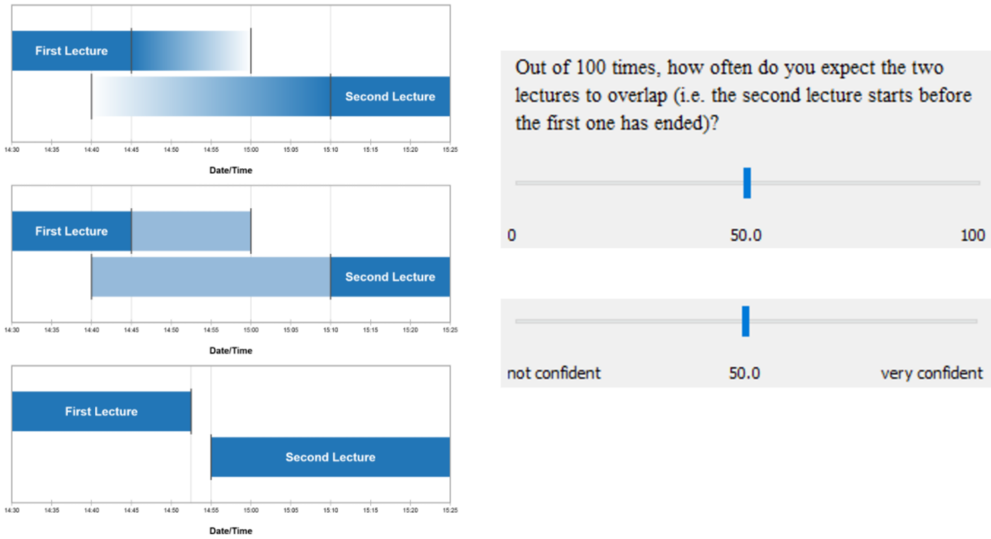


Figure 3.5: *On the left side the different task supporting visualizations for the fourth session are shown - from top to bottom: Gradient Plots, Ambiguation Plots, mean values. On the right side, there is an extract of the user interface showing the input fields for giving answers. Like in the first session, we are again not asking for a percentage value, but for some natural count instead, like suggested by Hullman [7].*

3.2 Design of the *Drawing Study*

The *Drawing Study* is designed to be of exploratory nature. This means that the research question it aims to answer is not as concrete as for instance the one of our *Evaluation Study*. The goal is to gain insights into the intuitiveness of visual encodings and to find out how people would visualize temporal uncertainty by themselves. This information could consequently be used in the design of novel visualizations, aimed at expert and especially non-expert users.

To gain these insights, we describe predefined scenarios, which encompass some kind of temporal uncertainty, to our study participants and ask them to draw a visualization sketch that intuitively represents this given scenario. Furthermore, the participants are always provided with a certain task a hypothetical user should be able to efficiently solve given an implementation of the sketched design.

To elicit the desired sketches of temporal visualizations from our study participants, we have to ask the right questions and also have to pay close attention to ask them in the right way. This means that it is imperative to not suggest any possible answers or solution approaches while communicating the task that should be solved, because this would greatly affect their given answers [7]. For this reason we try our best to make the given scenario and the task as clear as possible to our participant without suggesting anything that would help in the solution of the task.

CHAPTER 4

Results

TODO

4.1 Results of the *Evaluation Study*

TODO

4.2 Results of the *Drawing Study*

TODO

CHAPTER 5

Discussion

TODO discussion

CHAPTER 6

Conclusion

TODO conclusion

TODO sort out Bibliograpy (delete unused papers)

Bibliography

[1] Aigner, W., Hoffmann, S., and Rind, A. (2013). Evalbench: a software library for visualization evaluation. In *Computer Graphics Forum*, volume 32, pages 41–50. Wiley Online Library.

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

[3] Chittaro, L. and Combi, C. (2001). Visual definition of temporal clinical abstractions: A user interface based on novel metaphors. In *Conference on Artificial Intelligence in Medicine in Europe*, pages 227–230. Springer.

[4] Correll, M. and Gleicher, M. (2014). Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE transactions on visualization and computer graphics*, 20(12):2142–2151.

[5] Gschwandtner, T., Bögl, M., Federico, P., and Miksch, S. (2016). Visual encodings of temporal uncertainty: A comparative user study. *IEEE transactions on visualization and computer graphics*, 22(1):539–548.

[6] Harris, R. L. (2000). *Information graphics: A comprehensive illustrated reference*. Oxford University Press.

[7] Hullman, J. (2016). Why evaluating uncertainty visualization is error prone. In *Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 143–151. ACM.

[8] Kosara, R. and Miksch, S. (2001). Metaphors of movement: a visualization and user interface for time-oriented, skeletal plans. *Artificial intelligence in medicine*, 22(2):111–131.

[9] MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., and Gahegan, M. (2012). Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2496–2505.

[10] Messner, P. (2000). *Time shapes: a visualization for temporal uncertainty in planning*. Citeseer.

[11] Robertson, G., Fernandez, R., Fisher, D., Lee, B., and Stasko, J. (2008). Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1325–1332.

[12] Sanyal, J., Zhang, S., Bhattacharya, G., Amburn, P., and Moorhead, R. (2009). A user study to compare four uncertainty visualization methods for 1d and 2d datasets. *IEEE transactions on visualization and computer graphics*, 15(6):1209–1218.

[13] Walny, J., Carpendale, S., Riche, N. H., Venolia, G., and Fawcett, P. (2011). Visual thinking in action: Visualizations as used on whiteboards. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2508–2517.

[14] Walny, J., Huron, S., and Carpendale, S. (2015). An exploratory study of data sketching for visual representation. In *Computer Graphics Forum*, volume 34, pages 231–240. Wiley Online Library.

[15] Xu, K., Rooney, C., Passmore, P., Ham, D.-H., and Nguyen, P. H. (2012). A user study on curved edges in graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2449–2456.