RWTH AACHEN UNIVERSITY

Improving the understanding of statistical charts with in situ simulation

Master's Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Verena Kuhr

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof.Dr.Ulrik Schroeder

Registration date: 22.01.2015
Submission date: 31.03.2015

I hereby declare that I have created this work completely on my own and used no other sources or tools than the ones listed, and that I have marked any citations accordingly.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Aachen, March 2015
Verena Kuhr

Contents

List of Figures

List of Tables

Abstract

Statistical analysis plays an important role in evaluating and understanding the results of experiments. It lends evidence to claims made by readers who want to understand the outcome. However, several studies have indicated a lack of adequate statistical knowledge among researchers. Researchers have difficulties statistically analyzing charts of experiments found in papers. The reason for this is an inadequate statistical education and a misunderstanding of rules that can be applied when analyzing charts. As the visualization of data with uncertainty is vital for statistical analysis, but not easy to understand and interpret, several researchers tried to solve this problem. Existing data visualization software requires access to the data being visualized. Though, they are not helpful for the readers who do not have access to the underlying data.

As an attempt to solve these problems, we introduce SimCIs: an interactive visualization tool which simulates possible data underlying confidence interval graphs. The user uploads a paper with confidence interval (from now on abbreviated as CI) charts to SimCIs, to get it analyzed. The complexities of statistical analysis are lessened by embedding knowledge upfront. SimCIs also enables users to interact with visualizations and thereby improves the user's ability to develop interests. According to our user study, we found that SimCIs helps users to be more confident when interpreting results correctly. The system also helps users to perform statistical analysis tasks correctly which they have not done before. In addition, SimCIs shows promise to be used as a learning tool, which can improve users' statistical knowledge.

Überblick

Statistische Analyse spielt eine wichtige Rolle beim Verstehen und Auswerten der Ergebnisse von Experimenten. Sie verleiht den Aussagen von Lesern Glaubwürdigkeit, die das Ergebnis verstehen wollen. Allerdings haben mehrere Studien gezeigt, dass viele Forscher keine adequate statistische Kenntnisse vorweisen können. Forscher haben Schwierigkeiten statistische Graphiken von Experimenten in Artikeln zu analysieren. Der Grund dafür ist eine unzulängliche Ausbildung in Statistik und ein Missverständnis von Regeln, die angewendet werden können um Graphiken zu analysieren. Da die Visualisierung von Daten mit Unbestimmtheit für die statistische Analyse wichtig, aber nicht leicht zu verstehen und zu interpretieren ist, haben mehrere Forscher versucht, dieses Problem zu lösen. Vorhandene Datenvisualisierungssoftware verlangt Zugang zu den Daten, die visualisiert werden. Diese Programme sind jedoch nicht nützlich für die Leser, die keinen Zugang zu den zu Grunde liegenden Daten haben.

Um diese Probleme zu beheben, führen wir eine neue Software mit Namen SimCIs ein: Ein interaktives Visualisierungswerkzeug, dass mögliche Daten simuliert, die den Konfidenzintervallen der Graphiken zugrunde liegen. Der Benutzer lädt einen Artikel mit Graphiken der Konfidenzintervalle in SimCIs rein, um diese analysieren zu lassen. Die Prinzipien der statistischen Analyse werden durch Visualisierungen verdeutlicht und somit verständlich gemacht. SimCIs ermöglicht es den Benutzern ebenfalls mit Visualisierungen zu interagieren, wodurch sie interessierter sind die Analyse zu verstehen. Mithilfe einer Benutzerstudie konnten wir zeigen, dass SimCIs Benutzern hilft überzeugter in ihrer Antwort zu sein, wenn ihre Ergebnisse richtig sind. Das System hilft Benutzern ebenfalls statistische Analyseaufgaben richtig durchzuführen, die sie vorher nicht durchgeführt haben. Außerdem ist es vielversprechend SimCIs als ein Lernwerkzeug zu verwendet, welches die statistischen Kenntnisse von Benutzern verbessern kann.

Acknowledgements

First and foremost, I would like to thank my supervisor Chat Wacharamanotham for his valuable guidance and competent advice throughout this thesis.

I would also like to thank all people at the Media Computing Group and all people who participated in my study for giving me valuable feedback.

Furthermore, I would like to thank Prof. Dr. Jan Borchers, my thesis advisor, and Prof. Dr. Ulrik Schroeder, my second examiner, for their support.

Finally, special thank goes to my family for supporting me during the course of this thesis.

Thank you!

Verena Kuhr

Chapter 1

Introduction

Visualisations are a helpful and powerful way to analyse data (Olston and Mackinlay [2002]). In these visualizations, it is important to show the presence, nature, and degree of uncertainty to the user. According to BIPM et al. [2008], uncertainty is defined as doubt and uncertainty of measurement means the doubt about the validity of the results of a measurement. A common method to show uncertainty is to use error bars. They convey the degree of statistical uncertainty. This visualisation of uncertainty is important because otherwise, the data could be misinterpreted and there is a high possibility that this leads to inaccurate conclusions. It can be as important for judgment as the actual mean values and error rates of different groups (Correll and Gleicher [2014]).

Visualisation of uncertainty is a powerful way to analyse data.

A common way to interpret uncertainty is to do it with the null hypothesis significance testing, called NHST (Cumming and Finch [2005]). However, many who want a reform of statistical practices advocate a change from NHST to CIs. CIs are one way to visualize means with error bars and therefore to visualize uncertainty. According APA *Publication Manual*, CIs are in general the best reporting strategy. These figures can convey an overall pattern of results at a quick glance.

Use confidence intervals to visualize uncertainty.

According to Cumming and Finch [2001], there are four main reasons for using CIs.

Four reasons using CIs.

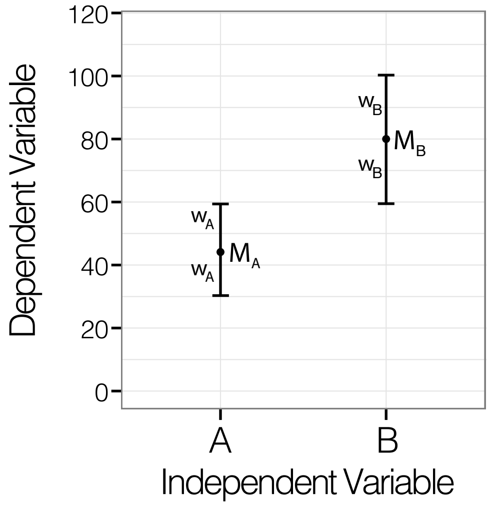


Figure 1.1: Two 95% CIs, A and B with a mean (M), and the interval.

1. CIs give point and interval information that is accessible and comprehensible. Therefore, they support substantive understanding and interpretation.

2. There is a direct link between CIs and the null hypothesis significance testing (NHST). Noting that an interval excludes a value is equivalent to rejecting a hypothesis. This hypothesis asserts that value as true at a significance level. A CI may be regarded as the set of hypothetical population values consistent with the data.

3. CIs are useful in the cumulation of evidence over experiments. They support meta-analysis and meta-analytic thinking, focused on estimation.

4. CIs give information about precision. They can be estimated before conducting an experiment and the width used to guide the choice of design and sample size. After the experiment, they give information about precision that may be more useful and accessible than a statistical power value.

Because of these reasons, we will focus on CIs to visualize uncertainty in this thesis.

A CI, as it can be seen in figure 1.1, is visualized by a mean (M) and an interval with an upper limit and a lower limit centered on the mean (Cumming and Finch [2005]). Each half of the interval from the mean to the limit is labeled with w. The mean is the point estimate of the population mean. The interval estimate indicates the precision, or likely accuracy, of the point estimate. There are different confidence levels of CIs. The 95% CI is the standard confidence level, chosen for the most studies. As the CI can be described as an estimated range of values of a study with a given high probability of covering the true population value, it is essential to be careful with this statement. This carefulness is important whenever probability is mentioned in connection with a CI. Because the probability statement about the upper and lower limit varies from sample to sample, it would be incorrect to state that one interval, derived from specific samples, has a probability of $.95$ of including the population mean μ. This would suggest that μ varies, but μ is fixed, although unknown. According to Cox and Hinkley [1979], a 95% CI can be expressed in terms of samples, where a repeated procedure on multiple samples could result in a CI that would encompass the true population parameter 95% of the time. This CI would differ for each sample.

CIs are an estimated range of values with a given high probability of covering the true population value.

There are only a few steps necessary to get a CI chart visualization from the data. The sample mean M is the mean of all data values of the distribution. The interval can be calculated by extending a distance w on each side of M. "w" is called the margin of error. The detailed calculations can be seen in appendix A.1 "Calculation from different input charts to 95% CI". The lower limit (from now on abbreviated as LL) is calculated by subtracting w from M and the upper limit (from now on abbreviated as UL) is calculated by adding w to M. Therefore the CI is often mentioned to be (LL, UL).

The CI is (LL, UL).

Although using CIs to represent uncertainty has many advantages, there are two main difficulties. Many researchers have important misconceptions about CIs. Belia et al. [2005] made a user study with authors of journals. The results showed that these people had problems in interpreting whether CIs are statistically significant or not. They also

CIs have two difficulties to work on.

had problems to distinguish between standard error bars, which do not include a critical value and CIs (see A.1 "Calculation from different input charts to 95% CI" for more infos). These results indicate that there is a need of better graphical conventions to display interval estimates and to signal more clearly how intervals may be used for interpretations. This result already exposes the second difficulty, that there are only a few accepted guidelines as how CIs should be represented or discussed.

This thesis will focus on a web-based solution to improve uncertainty understanding.

Because there are still problems of misconceptions about CIs, it is important to have a system that helps understanding and interpreting statistical charts. We try to accomplish that by implementing a system, which helps the user to give accurate answers, being confident with their answers, and learning about the general principles of how to interact with CI charts. In this system, it shall be possible for the user to simply upload the corresponding paper with the chart. Then, the user can see different simulated distributions of the CI and interact with different visualizations. As designing visualizations to support decision-making and perform debiasing is not trivial, we will focus on the use cases first, before continuing with the next chapter where related work is studied. Which user groups will benefit from this tool and which purposes it will accomplish, are explained in the following.

The system shall help students and researchers to understand a chart in a paper.

Students as well as researchers have to read a lot of papers when they are writing a thesis for example. Therefore, it is important for them to understand the charts. We can assume that researchers already have statistical knowledge. The statistical knowledge is not that clear for students. According to this article[1], students from Germany learn descriptive statistics, probability calculations, and have basic knowledge in judgment statistics. Students in the United States are able to read bar charts and data tables to analyze data in minimum[2]. According to these information, the system should be usable for people with various knowledge of statistics to understand CI charts or specific problems with

[1] http://stochastik-in-der-schule.de/sisonline/jahrgang26-2006/heft3/2006-3_kaun.pdf

[2] http://www.prb.org/Publications/Lesson-Plans/WorldPopulationDataSheet.aspx

the chart correctly. Therefore, it could also be used as a tool to teach statistics. Teachers could use the system to point out special tasks to do with a specific chart. This could help to deal with the difficulties, mentioned above. Principles that help designing such a system are explained in the second half of 2 "Related work".

The system shall help students and researchers syntax checking their own charts to make them better understandable.

Another possibility would be to use the system to syntax check a chart. This can become interesting for researchers and students who made a study and conducted a chart. As these people might already have a deeper statistical knowledge the focus of the usage of the system is shifted for this task. Now, the system must not help to understand a chart but can used to check whether all important aspects are visualized in the chart and given in the text.

In the next chapter, we will go into detail which previous work already explored the visualization of uncertainty and discuss different visualization tools to improve the understanding and interpretation of statistical charts. Several design principles are given on which the design of this thesis' system, called SimCIs, will be based on. The third chapter 3 "System design - SimCIs" will explain the interaction steps, which are possible with the system, how everything is visualized, and how the final design resulted from several iterations. The following chapter 4 "Evaluation" will evaluate the system to see whether it really helps the user to understand and interpret CI charts. A summary is given at the end in chapter 5 "Summary and future work", which is followed by possible implementations for the future.

Chapter 2

Related work

In this chapter, we will discuss literature concerning how uncertainty is visualized. In the first section, we will focus on static and interactive ways to visualize uncertainty. The exemplary systems for static and interactive visualization described in there are base on general design principles. These principles and existing techniques to augment data visualization are named in section two. Before these visualizations of uncertainty can be used, the system needs to have the underlying data of the chart. According the use cases which are given in the introduction, the extraction of the data directly out of the chart will be necessary for my thesis. There already exist many systems with interactive, automatic, and crowd sourced data extractions. Therefore, these ways will only be described shortly in section three. The main focus of this thesis stays on the visualization, its interactivity and ways to evaluate such systems. To be able to identify and to generate systems which visualize uncertainty in a way that users can understand it correctly, we provide a taxonomy for evaluating visualizations in the last section. Summarizing, a design space shows which areas in interactive visualizations needs further investigation.

The main focus of this thesis stays on the visualization, its interactivity and ways to evaluate such systems.

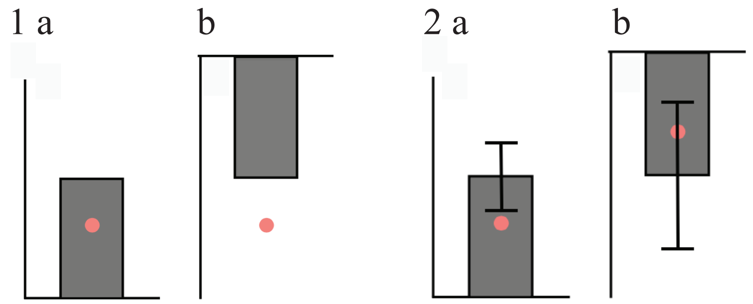


Figure 2.1: Possible test points for (a) a rising graph and for (b) a falling graph with bars centered at zero (adapted from Newman and Scholl [2012]). 2 shows additional error bars, where the test points are located (a) outside the margin of error and (b) inside the margin of error.

2.1 Uncertainty visualization and interpretation

Visualization of the uncertainty to help understanding the data itself can be static and interactive.

When analyzing uncertain, noisy, or incomplete data, measurement error and confidence intervals are as important as the actual mean values (Correll and Gleicher [2014]). Visualization of the uncertainty can help to understand the data itself and therefore also to interpret it. There are two common ways when visualizing uncertainty in confidence intervals. The first section will describe how uncertain data can be presented in a static way. The second section will show previous work which focus on interactive visualizations and animation of the data. These ways can help the viewer to directly read answers to tasks off of the chart (Ferreira et al. [2014]). It should help to read the charts correctly and understand the meaning of the underlying data. However, there are also limitations in previous work which will be outlined and addressed in this chapter to build the base of the design of SimCIs. The main part of the system and the interaction design will then be explained in chapter 3 "System design - SimCIs".

2.1.1 Static visualizations

According to Correll and Gleicher [2014], the most common way to visualize mean and error is to use a bar chart with error bars. However, bar charts suffer from the within-the-bar bias (see Newman and Scholl [2012]). The bias provides a false metaphor of containment, where users think that values are likelier to be true than others, because they are visually arranged in the bar charts. An example, showing this bias, can be seen in figure 2.1 (1). Here, the authors made an experiment showing, that test points originated from within the bar produced greater likelihood ratings than those outside the bar. The test point below the mean was judged to be significant more likely when the bar was rising to zero and contained the point, as it can be seen in figure 2.1 (1 a), than when the bar was falling to zero and not contained the point, as it can be seen in figure 2.1 (1 b). The same results did also occur vice versa.

Commonly used bar charts to visualize uncertainty suffer from the "within-the-bar-bias".

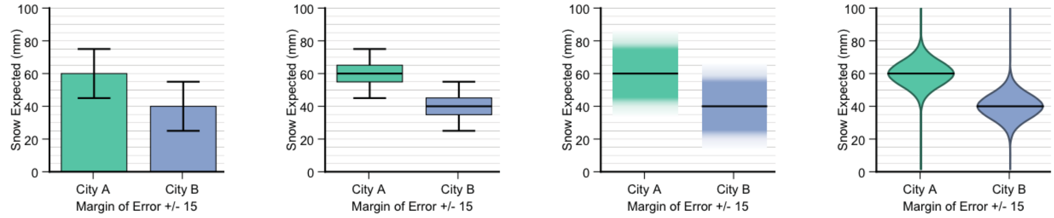
Correll and Gleicher [2014] focused on the problem of binary interpretation, while trying to mitigate the "within-the-bar" bias. The binary interpretation leads the user to only compare between the two states within the margin of error or outside. Correspondent to the authors, this would mean that users would define the point in figure 2.1 (2) a as not inside the margin of error and the point from figure 2.1 (2 b) as inside the margin of error. These problems make it difficult for the viewers to make confident and detailed inferences about the outcome.

These bar charts also suffer from the "binary interpretation".

Scientists worked on alternative representations of uncertainty to deal with the problem of binary interpretations. In accord with the authors, alternative visualizations communicate the implications of mean and error data more effectively to a general audience. Having a careful design of a visualization can convey the general notion of varying levels of errors also for people who have no deep statistical background. To get a design which accomplishes these points, the authors mentioned five goals:

Five goals to design visualizations which deal with the bias and the binary interpretation.

1. The encoding should clearly present the effect size.
2. The encoding should promote the right behavior



(a) Bar chart (b) Modified box plot (c) Gradient plot (d) Violin plot

Figure 2.2: Different visualizations of uncertainty (Correll and Gleicher [2014]).

from the viewer with their confidence in decision.

3. The encoding should proffer the estimation or comparison of statistical inferences that have not been explicitly supplied.

4. The confidence needed to be displayed continuously to avoid the all or nothing binary encodings.

5. The encodings should mitigate known biases in interpretation which requires visually symmetric encodings about the mean.

With these points in mind, the authors compared three different encodings which shall fix the problems that occur while using bar charts.

Modified box plots mitigates the ”within-the-bar-bias”.

Common used **box plots** with error bars only encode the actual distribution of the data and not the distribution of a potential population mean. This distribution then is several analytical steps removed from a confidence interval. Another problem is that large boxes make the viewers underestimate the length of error bars and vice versa Stock and Behrens [1991]). Despite all these problems, box plots have many extensions to show a wide variety of higher order statistics. Correll and Gleicher adapted this encoding to use the positive aspect without dealing with the problems. Therefore, the authors set the whiskers as margins of error and the box includes the area from 25% until 75% of the interval, as it can be seen in figure 2.2. This modification results in additional levels of comparison, having three levels called “outside the error bar”, “inside the error bar”, and “inside the box”. A point inside the box is

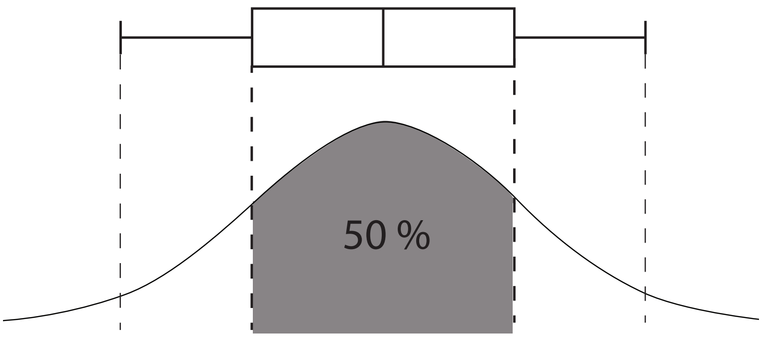


Figure 2.3: Modified box plot with theoretical distribution.

then within a 50% confidence interval, like in the example of figure 2.3. With the visual symmetry around the mean, it mitigates the "within-the-bar" bias. Having no bar anymore, but only symmetry about the mean let the user be aware of a false causality. There is no risk anymore of interpreting values likelier to be true than others only because they are visually arranged in the bar chart and others are not. The next two encodings are generally used for visualizing distributional information and were adapted by the authors to use them for inferential tasks.

In a **gradient plot** transparency is used to encode uncertainty, like shown in figure 2.2. A 95% confidence interval has minimal transparency. Below the lower limit and above the upper limit the transparency then increases until there is a full transparency at the limits of the (fictional) 100% confidence interval. However, since the inverse cumulative probability function decays so rapidly, there is a block of solid color surrounded by "fuzzy" edges. The authors used these imprecise edges to extend the comparison explanation to the statement "if two samples are very statistically similar then their "fuzzy" regions will overlap". Furthermore, this imprecise visualization provokes a willingness to critique which were not encouraged by a precise visualization of error bars. The visual symmetry also mitigates the "within-the-bar" bias as the visualization of the modified box plot does.

Gradient plot mitigates the "within-the-bar-bias" and provokes a willingness to critique.

Violin plot mitigates the ”within-the-bar-bias” and let users read the chart more precisely.

For the **violin plot**, width is used to encode uncertainty (see figure 2.2). Near to the mean, the width is the widest. The width then decreases exponentially with increasing distance to the mean, which visualises the values that become less likely. This results in a smooth, violin-like shape with interior glyphs. Width as a positional encoding of distributional data has a higher precision according the visual channel than color. This benefits in viewer estimation tasks. Furthermore, it encourages the accordance of comparison of values beyond the discrete within/ outside the margin of error judgments. This strong, high fidelity visual encoding let the user read the chart more precisely. As the other two encodings also did, the visual symmetry mitigates the within-the-bar bias.

Gradient plot and violin plot should be preferred when visualizing uncertainty.

All of these alternate encodings have low costs and benefit in performance advantages to a general audience. The experiments from Correll and Gleicher emphasize this statement. Furthermore, visual encodings that overcome the two problems, like the gradient and the violin plot, should be preferred to use instead of bar charts. However, the experiment did not state which encoding is the best replacement. The cultural costs could also be high, because the viewers might prefer suboptimal, but known encodings.
Furthermore, the experiment did not conduct decision-making tasks or a bigger set of encodings like outliers, regression, or multi-way comparisons.

In 3.2.2 “Possible distributions”, we will use violin plots to show skewed data of a CI. This design decision is discussed in chapter 3 “System design - SimCIs”.

2.1.2 Interactive visualizations

As teaching and learning statistics becomes a bigger field in elementary, secondary, and postsecondary education according to Mills, there is a considerable interest to employ effective instructional methods (see Rohatgi [2015a]). To teach these concepts, researchers in general recommend computer simulation methods. However, according to the author, only very little empirical research is done in this field to support this recommendation.

Effectiveness of computer-assisted statistics instruction in learning statistics is given by providing a meaningful average performance advantage.

In 2003 and in 2007, researchers like Hsu [2003] and Schenker [2007] already found out through meta-analyses, that computer-based tools can be effective in statistics instruction. In 2011, Sosa et alċonducted a study in which they went further, by focusing on aspects of computer based tools that are most closely associated with learning and achievement outcomes (Sosa et al. [2011]). They extend previous work by giving attention to additional attributes, such as different technology types, student engagement, student control over the learning process, and the nature of feedback. These attributes were tested in 45 experimental studies. The results showed, that they account for differences in effectiveness of computer-assisted statistics instruction by providing a meaningful average performance advantage. Further analysis suggested that the effect is larger when having more time for the instructions, graduated students who perform the tasks, and an embedded assessment. However in general, regardless of the features, computer-assisted instructions yields larger effects than non-computer-assisted instructions. According to the authors, research to the role of interactivity, engagement, and feedback getting more important as educators continue work on improving the efficacy of technology-based statistics instruction.

Simulation is beneficial when learning about uncertainty.

Cumming [2012] worked on a solution which focuses on the interactivity and feedback of computer-assisted statistics instruction. This includes some Excel spreadsheet as exploratory software for confidence intervals and his book "Understanding The New Statistics" to understand and see possible ways how to interpret uncertainty. Here, several simulations show how data points are arranged to a specific normal distributed CI, what p value means, and compares two CIs for example. The user also can only start, pause and stop the simulation. All these simulations show samples which are generated randomly, or which are based on numbers inserted by the user at the beginning, like sample size, mean, or standard deviation. According to his changes, the CI is redrawn and simulation is shown. These simulations are provided in 6 Excel spreadsheets with 47 sheets in total to communicate the full meaning of confidence intervals. To fully understand how to interact with these sheets and to get all its information it is necessary to read the book. There are a few spreadsheets from Cumming, where the user can insert the underlying data of such charts. However, it does not provide the possibility to insert a chart. All in all, his work suggests a benefit of using simulation to learn about CIs. In my work, I will go further by making it convenient to create simulations within the context of reading a paper.

Ferreira et al. created a system with 5 task based annotations to understand and interpret uncertainty.

Ferreira et al. [2014] compared five interactive visualizations to understand and interpret uncertainty in bar charts directly. Here, the user has to add the underlying data into the system to get it visualized. These different visualizations of uncertainty, which will be explained later on with figure 2.4 and figure 2.5, shall help the user to read the answers of tasks directly from the chart. The authors chose these tasks to be bar comparison, identification of the probability that a bar represents the minimum and maximum, comparison to a constant, comparison to a range, and identification of items within rankings. At first, the authors set several design goals based on former literature which can be used as guidelines for creating visual annotations:

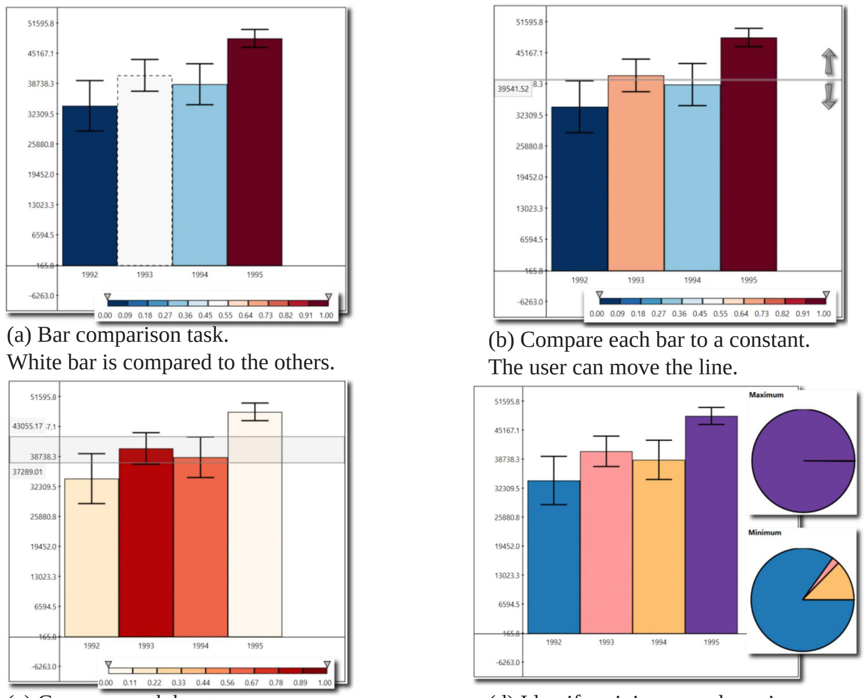
They created guidelines for creating visual annotations.

1. The visualization should be easy to interpret by only adding minimal additional visual complexity.

2. Consistency across tasks is necessary for the user to be able to change between tasks without losing context on the dataset.

3. Spatial stability across sample size shall be comprised, by changing the visualization proportional to the size of the change of the data. This can be done by smoothly animating between visualizations at two successive time intervals.

4. Minimize the visual noise, by displaying some information in one chart type and other information in a different visual representation. According to the authors, having no such additional complexity would cause more confusion and therefore is important to add.



(a) Bar comparison task.
White bar is compared to the others.

(b) Compare each bar to a constant.
The user can move the line.

(c) Compare each bar to a range.
The user can change and move the range.

(d) Identify minimum and maximum.

Figure 2.4: Visual annotation tasks (Ferreira et al. [2014]).

These criteria were fulfilled by applying interactive annotations to base visualizations. Five task-based annotations are given with which the visualization changes directly as the users' action occurs. The annotation techniques are interactive color encoding and pie-chart representation of probability for the comparison tasks. To compare means with uncertainty in a CI chart with the others, the user simply selects that bar. As it can be seen in figure 2.4 (a), that selected bar gets highlighted and the other bars get color encoded to denote the degree of uncertainty. The color encoding works with a divergent color scale which ranges from dark blue, which means "definitely smaller", to dark red, which means "definitely larger". The white color at the center represent an unknown relationship between both charts. The same color encoding is used in the second task, when comparing each bar to a fixed value (see figure 2.4 (b). The user can set a benchmark, visualized by a horizontal line. According to the position of the benchmark to the error bars, the color of a bar is set, as used for error bar com-

Task solving is supported by the system with the annotation techniques "color encoding" and "different visual representations".

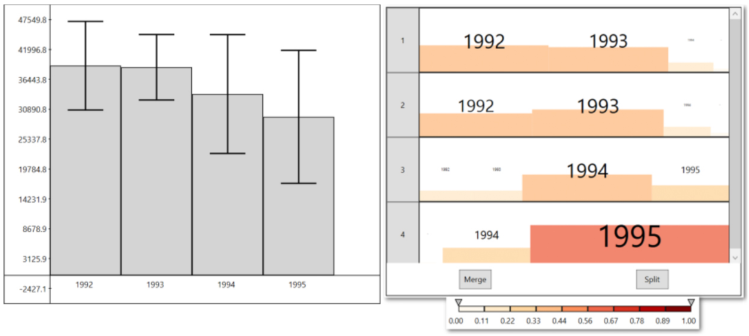


Figure 2.5: Visual annotation task showing standard confidence interval bars (left) and a corresponding ranked list visualization (right) (Ferreira et al. [2014]).

parison. Furthermore, the user can move the benchmark up and down along the y-axis and the change of the bars according the color encoding gets simulated. Like shown in figure 2.4 (c), the bars can also be compared to a range which works the same as the comparison to a benchmark. In this third task, the user can set an upper and a lower limit of a range and the error bars get compared to it. Here, the color encoding is used to show whether it is likely of an error bar to be inside of the range set by the user. Again, the authors encode the relative certainty in a color gradient. As a fourth task, the user can select the option to compare all error bars. This results in two pie charts, one representing the probability of all distributions to be the maximum, also known as highest error bar and one representing the probability of all distributions to be the minimum, also known as lowest error bar (see figure 2.4 (d)). This different representation is used by the authors to avoid the confusion that could occur with a second different bar chart. Because the total probability across all bars must equal 100%, the authors chose the pie chart representation. Here, the color encoding is used differently than in the former tasks. A qualitative color mapping is used to help the user identifying bars and the related regions in the pie charts. When the user wants to see the general rankings, which is the fifth task, he can also select that option. There the ranking can be shown in another representation directly beside the bar chart, as it can be seen in figure 2.5. For each rank, which is shown in separate rows, the user can see the probability of each bar. This probability is shown by the height, width, and color. With increasing width and height, it is more probable for a bar to be in that rank. The color encoding is the same as for the comparison to a range.

The visual annotations meet all guidelines.

With the color encoding and the qualitative color mapping in the annotation tasks, the system fulfills the first guideline by only adding minimal additional visual complexity. Consistency across the tasks, as the second guideline, is ensured by showing the error bar representation all the time. The third guideline, spatial stability across sample size, is given with all annotations, where smooth animation between visualizations at two successive time intervals is given. This animation is used for changing the color of bars when one condition, like the position of the benchmark, changes. However, the spatial stability is not given for the pie chart representation and the ranking. These alternate representations directly appear. Also these different representations help to minimize the visual noise, without the animated transformation from one chart type into another, it can be potentially hard for the user to understand the process of getting different visualizations for the same data. Brosz et al. [2013] focused on this problem. Their results will be shown later on in this section.

The experiment stated that the system needs some training time and is only suitable for experts.

To show that these representations for the annotation task help the user to read the answers to the tasks off the chart, Ferreira et al. [2014] made an experiment. In the experiment, user had to answer 75 questions in random order. The results witnessed the advantage of task-specific annotations which helped the user to make decisions about samples. The users were more confident with their accurate responses than they were without these annotations. However, the experiment could not show that these annotations result in a more accurate decision. Furthermore, the encoding and the usage of the system has firstly to be learned before the user can work with the system. The user still needs some training time to use the system correctly and to know how to read the visualization correctly. The system is designed for expert analysts who are at least familiar with bar charts and confidence intervals and have basic training in statistics. In general, the guidelines for creating visual

Figure 2.6: Transmogrification from a bar chart to a pie chart (Brosz et al. [2013]).

annotations, special and defined annotation tasks, and their annotation techniques result in a system, that helps the user in being more confident with their decisions. Therefore, I also based my concept for creating visual annotations on these guidelines, set special annotation tasks and used different visual representations that shall help the user to understand the CI chart. More information of this process in provided in chapter 3.1 "Interaction walk-through and design rationale" and 3.2 "Visualization of problem cases".

Users are more confident with their decision when using the system.

In general, according to Brosz et al. [2013], data visualizations are necessarily limited to specific representations, which are normally constrained by the format and by the representation choices of the designer. Furthermore, the authors say that no representation can be perfect for all purposes, for all people, or for all data. In the paper of Brosz et al. [2013] a system is presented, where graphic transformation from one shape to another gets animated in real-time to give the user the possibility to chose different visual representations quickly. With this transmogrification, the user is aware of how one dataset in a chart type can be transformed into another form.

Transmogrification explains the process of the animated transition from one chart type into another.

As it can be seen in figure 2.6, a normal bar chart gets modified so that the bars are ordered in a circle to be visualized in a pie chart, for example. Here, the user had the bar chart, which is represented on the right, and chose the circle form. After that, the bar chart visually moved along the link and formed a circle. This animation facilitates seeing the relationship between the original bar chart and the transmogrified space. The user can also see this animated process again in his desired speed by dragging a slider that appears at the bottom of the interface. With this system the user can sketch, transform, and compare visualizations from one or multiple sources.

All these static and interactive visualization examples base on general design principles. They will be discussed in the next section.

2.2 Existing techniques augmenting data visualization

Several techniques help augmenting visualization. According to the visualization type, different design principles can be applied. In the following, we will distinguish between the design principles of static and interactive visualization. Each principle will get assigned a number to refer to the principles later on in chapter 3 "System design - SimCIs" when discussing the design of the system.

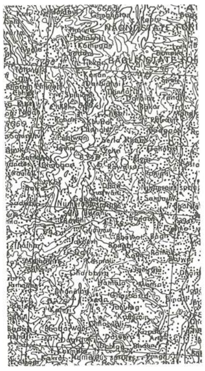
2.2.1 Design principles for a static visualization

Edward Tufte worked on the topic augmenting visual information to show the data. Therefore, he made several guidelines which will be discussed in this section. These guidelines can be used when designing systems.

Layering and separation

Layering and separation reduces the noise and enriches the content.

Layering and separation of the visualization reduces the noise and enriches the content without producing cluttering and confusion (Tufte [1991]). This can reveal the data at several levels of detail and present many numbers. Layering and separation can be achieved by distinction of color, value of the color (brightness), shape, or size for example. Coloring plays an important role to reduce noise and enrich content. Small spots of intense color and saturated color for



(a) No separation and layering with failure to communicate information.

(b) Correct separation and layering.

Figure 2.7: Layering and separation (Tufte [1991]).

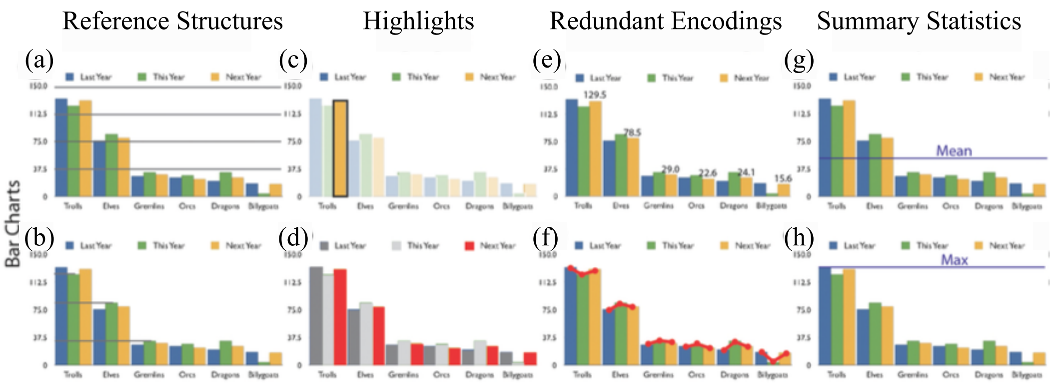


Figure 2.8: Possible types of overlays (Kong and Agrawala [2012]).

carrying information help to highlight important parts and set connections between the different layers. As it can be seen in figure 2.7 (a), having a same visual level with equal values, equal texture, equal color, and nearly equal shape results in an undifferentiated, unlayered surface with jumbled up, blurry, incoherent, chaotic, and unintentional optical art. Information cannot be communicated. In figure 2.7 (b) more detail than the perfect jumble is shown, having separated and layered information.

Overlays

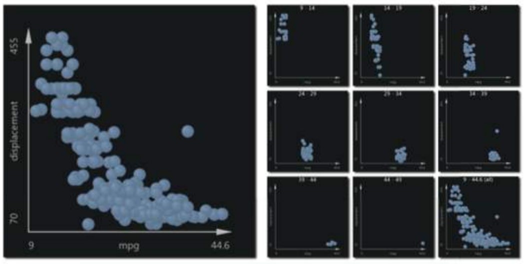
Kong and Agrawala [2012] elaborated these principles by exploring five possible types of overlays that aid chart reading. For this section, we will focus on four types, which are relevant for this work. Regular or special grid-lines like it can be seen in figure 2.8 (a) and 2.8 (b) on main parts of the charts provide reference structures which assist the perceptual process of anchoring and projection. Bartram et al. [2011] recommended an alpha value of 0.2. The resulting grid line is not too obtrusive but not too hard to see when needed. While there is no difference in alpha for black or red colored grids, blue grids need a higher value to be more robust. Highlighting of parts of the chart like shown in figure 2.8 (c) and 2.8 (d) with different texture, border color, or shadows for example focuses the viewer's attention to specific tasks (Kong and Agrawala [2012]). With redundant encodings like data values (see figure 2.8 (e)) on bars the user can more effectively read details. Lines (see figure 2.8 (f)) on the other hand communicates trends better. To save the user from time consuming cognitive functions, summary statistics can be shown, like mean, minimum and maximum, or SD. This can be seen in figure 2.8 (g) and 2.8 (h). To defeat graphical distortion and ambiguity, it is important to not show too many of these overlays. "Above all else show the data" is one of the main principles from Edward Tufte. All these principles are useful for different tasks that are important to understand and interpret CIs. How these principles will be integrated in the design of my system will be explained later on in chapter 3 "System design - SimCIs".

Kong et al. explore five possible types of overlays that aid chart reading.

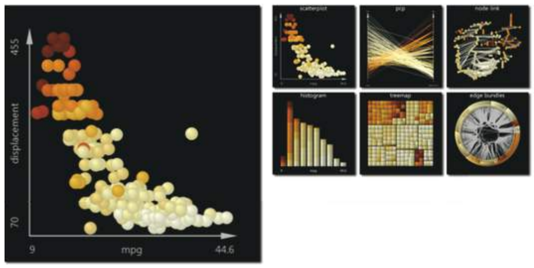
Small multiples

To show changes, alternatives, or differences in data rather than focusing on graphical design, small multiples can be shown the paper of Edward [2001]. Here, the same chart type is shown a lot of times, where each chart represents a different data set. This encourages the eye to compare different pieces of data.

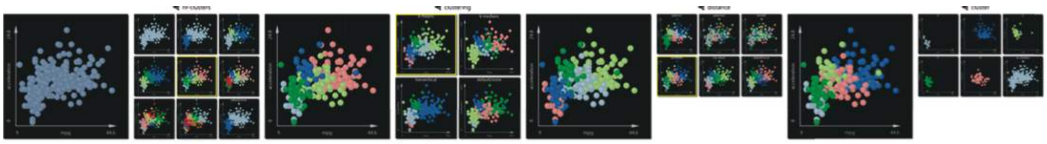
Small multiples help the user to focus on changes, alternatives, or differences.



(a) Filter split on visible (x-axis) attribute mpg.



(b) Mapping split on visualisation type.



(c) Small multiple createn based on visual analytics parameters: number of desired clusters, cluster method, clustering distance metric and finally a split on the cluster themselves.

Figure 2.9: Generation of small multiples (van den Elzen and van Wijk [2013]).

Use small multiples for data analysis.

Based on this principle and large singles, van den Elzen and van Wijk [2013] worked on a new visual exploration method for effective and efficient data analysis. With this system the user shall be able to easily compare the effect of different parameters and a history trail of exploration path. As it is relevant for our work, we will focus on the first point, the comparison. The second point of keeping history is not important for this work at the moment.

There exist four different ways to generate small multiples from a large single.

Comparing the effect of different parameters with small multiples prevent the user in being inefficient or leading to missing interesting features because a parameter value is not inspected. Now, users can compare the separate images and look for patterns, trends, and outliers. The small multiples are generated from a large single in a splitting operation. In this operation one parameter is selected to be varied over the small multiples. The authors defined four different types of parameters for filtering, mapping, visual analytics, and binding.

Filter split according a selected attribute.

For the filter splits, small multiples are created based on a large single and a selected attribute. Different value ranges of the chosen attribute gets then split manually or automatically and divided into different small multiples. An example scenario can be seen in figure 2.9 (a), where the x-axis is chosen as attribute. Each small multiple show the data values in a different x-axis range.

In general the mapping split is used to create small multiples for each visualization type, as it can be seen in figure 2.9 (b). Here, the user has the possibility to explore what visualization type is best for their problem and effortless try different visualization types, or even all at once. However, visual mappings have a variety of parameters like axes, color, and size. These parameters can also result in a generation of alternatives shown as small multiples. The mapping split will also be used for this thesis' system design, to show different distribution types of a given CI. How this is implemented will be shown in chapter 3.1 "Interaction walk-through and design rationale" and 3.2 "Visualization of problem cases".

Mapping split creates small multiples for each visualization type in general.

The visual analytics method, where the small multiples are based on visual analytics parameters, shall enhance the exploration experience. The user has to fulfill several steps one after another in which he firstly sees a number of desired clusters. This can be seen in figure 2.9 (c) on the left. After he chose one cluster, different clustering methods are shown. After that the clustering distance metric is shown in small multiples and at the end a split on the clusters themselves occur.

Visual analytics method enhances exploration experience using cluster parameters.

In binding, advanced split operations are also possible on small multiples, which then is called multiple split. As this is not relevant for this thesis, we will not discuss this further more in detail.

Binding enhances multiple split.

Graphical integrity

In general, it must be visually clear that the graphical design serves a clear purpose and induce the viewer to think about the substance rather than about the design Edward [2001]. To encourage that the graphical display is closely integrated, not only the statistical description of the dataset but also the verbal descriptions of the dataset are impor-

Graphical design must serve a clear purpose.

tant. This can help to avoid distorting what the data have to say.

Tufte derived six main principles resulted from graphical integrity. The following ones from ”The Visual Display of Quantitative Information” (p.77) are relevant for this thesis:

1. ”The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.”
2. ”Clear, detailed, and thorough labeling graphical distortion and ambiguity.”
3. ”Show data variation, not design variation.”
4. ”The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”
5. ”Graphics must not quote data out of context.”

How these aspects get integrated in my system will be shown in chapter 3 “System design - SimCIs”.

2.2.2 Design principles for an interactive visualization

As simulation is an important aspect to teach and learn statistics (see section 2.1.2 “Interactive visualizations”), it is necessary to have design principles of the way visualizations can change because of users input. According to Ware [2004], a good visualization is characterized by finding more detailed data about anything that seems important. This means, that each object shall be capable of displaying more information as needed, disappearing when not needed, and accepting user commands to help with the thinking process. Such an interactive visualization is a process that is made up of a number of interlocking feedback loops that can be categorized into three broad classes.

Manipulation loop

The lowest level of feedback loops is called the manipulation loop. We have a massively parallel processing of visual scene into elements of form, opponent colors, and elements of texture and motion. Here, objects are selected and moved using basic skills of eye-hand coordination. Delay takes an important role as it can disturb the performance of higher level tasks when only a fraction of a second of delay occurred in the interaction cycle. Colin Ware listed eight laws that describe low level control loops. The important laws which are relevant to this thesis are shown in the following. **Choice reaction time** can be modeled with a simple rule called Hick-Hyman law where the reaction time is based on the number of choices C.

Three laws are important to describe control loops: Choice reaction time, hover queries, and vigilance.

$$Reaction\ Time = a + b * log(C)$$

$log(C)$ represents the amount of information processed by a human operator, expressed in bits of information and a and b are empirically determined constants. Another important factor is the speed-accuracy trade-off, where time always suffer from the accuracy and vice versa. **Hover queries** provide extra information of the object. Normally this is done with a delay. When hovering over queries with no delay, dragging over a set of objects can rapidly reveal the data contents and allowing an interactive query rate of several per second. The detection of infrequently appearing targets can get very hard, because people perform poorly on **vigilance tasks**. There are several techniques to improve the performance. Reminders at frequent intervals, can help when there are several different targets that have to be detected. Another way is to make the target perceptually different or distinct from irrelevant information with color, motion, or texture distinction.

These concepts are applied in the system SimCIs, which will be explained later on in chapter 3.1 “Interaction walkthrough and design rationale” when talking about interactive refinement. The hovering technique will be used to display more information when hovering over the chart or the small multiples. To provide vigilance during the whole process in the thesis’ system, outlier visualization is provided. The speed-accuracy trade-off will get important in chapter 4 "Evaluation", when comparing the results from the user study.

Exploration and navigation loop

The law "focus, context, and scale" solves the focus-context problem using four different techniques.

The intermediate level of feedback loops is called exploration and navigation loop. In this level, five laws help the analysts to find their way in large visual data space and to try better understanding or perceiving a problem. **Locomotion** and **viewpoint control**, as well as **frames of references and map orientations** shall help to navigate in a 3D environment that shall reflect navigation in the real world. Therefore, the first three laws are not relevant here. The fourth law is **focus, context, and scale**. During the exploration process, the user will face the focus-context problem, where he has to find a detail in a larger context. In this thesis, the context is provided by the uncertainty chart and the focus is represented by the small multiples, which show different distributions. This will be explained in detail in chapter 3.1 "Interaction walk-through and design rationale". In general, same interactive techniques can be applied to solve the focus-context problem. These four different visualization techniques are distortion, rapid zooming, elision, and multiple windows.

Multiple windows having one overview window and several detailed windows which are connected via links.

The technique, called *multiple windows*, is interesting for this thesis, as it comprises one window showing an overview and several others showing expanded details (see figure 2.10). As this technique is very similar to the small multiples idea, explained in section 2.2.1 "Design principles for a static visualization", this technique will be used in the thesis' system SimCIs. How it is implemented will be explained later on in chapter 3.1 "Interaction walk-through and design rationale" and 3.2 "Visualization of problem cases". This technique solves the problem of visually disconnected windows from the overview window. Lines are added that connect the boundaries of the detailed window with the boundaries of the overview window. This technique is better than the other techniques, because it is non distorting and shows focus and context simultaneously.

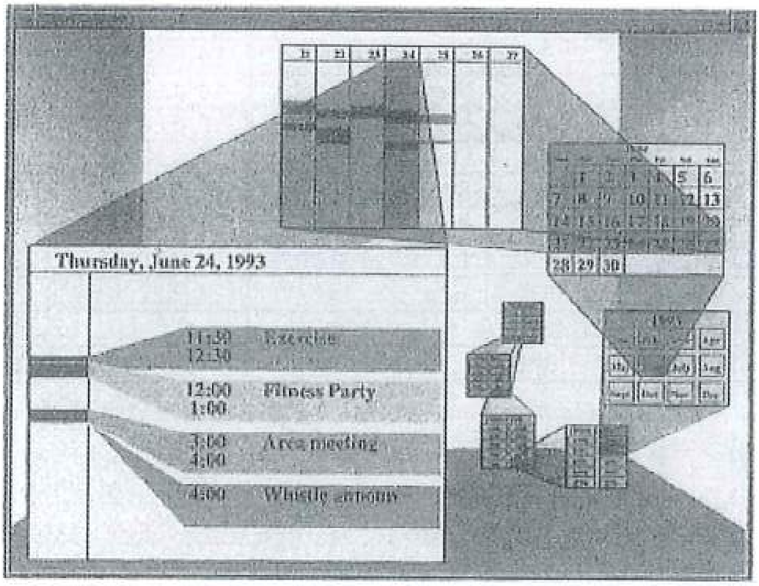
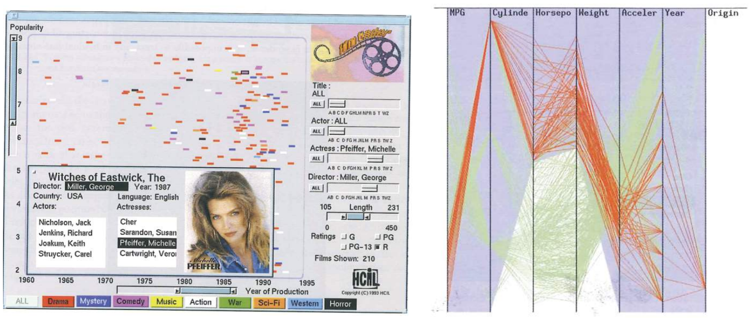


Figure 2.10: Multiple windows technique on spiral calendar. Information in one window is linked to its context within another by a connecting transparent overlay (Ware [2004]).



(a) Dynamic query sliders (b) Brushing objects

Figure 2.11: Rapid interaction techniques to have a fluid mapping between the data and its visual representation (Ware [2004]).

Rapid interaction with data enabled with dynamic queries interface and brushing.

The fifth law in the exploration and navigation loop is the **rapid interaction with data**. It is important to have a fluid and dynamic mapping between the data and its visual representation. With the principle of transparency, the user is able to apply intellect directly to the task so that the tool itself seems to disappear. To achieve this sense of control, the responsiveness of the computer system is the key psychological variable. This interactive data mapping is the process of adjusting the function that maps the data variables to the display variables. There exists several techniques to display this interactive data. Ware [2004] mentioned two mainly used techniques. With the *dynamic queries interface*, the range of data values that are visible and mapped to the display variable are limited, when the data set is very large and complex. This results in data range sliders which isolate and visualize a subset of the data when adjusted, like shown in figure 2.11 (a). With *brushing*, subsets of data elements can get highlighted interactively in a complex representation. Figure 2.11 (b) shows a parallel-coordinate plot, where each data dimension is represented by a vertical line. The user can interactively select a set of objects by dragging the cursor across them. They then get highlighted by using a different color representation. When having a data object which appears in several views, selecting it in one view will also highlight it in the other views. This also enables visual linking of components of an object.

In this thesis a form of brushing will be used. When the user highlights information of the chart in the text, different representation (small multiples) that fits to this information gets highlighted. More information and the realization will be given in chapter 3.1 "Interaction walk-through and design rationale".

Problem-solving loop

In five levels, requirements are formulated and separated into parts to finally solve the task.

In the highest level of feedback loops, called problem-solving loop, the analysts form hypotheses about data and refine them through an augmented visualization process. Here, thinking can be augmented by visual queries on visualizations of data. This level is segmented into five levels. In the highest and initial level, called **problem solving strategy**, a problem context and its provisional steps for solving it are settled up. This includes the formulation of a set of requirements. As the system of this thesis shall also be usable by non-experts in statistics, which might not have the knowledge of which requirements are important, these steps are taken over by the system itself. More details are given in chapter 3.1.3 "Interactive refinement by selecting information from the text". The next level, called **visual query construction**, focuses on the formulation of parts of the problem to afford a visual solution. This can include to read the data values out of the graph. Therefore, the graphical data extraction is an important part during this process. Different ways of graphical data instructions will be explained in the following section 2.3 "Graphical data extraction". The **pattern-finding loop** starts with the search for elementary visual patterns important to the task. Two or three simple solutions or one complex solution can be stored in the visual working memory. One solution must be retained while alternate solutions are found. This can be supported by the system when giving the user the possibility to highlight a potential solution. We will use that for the interaction design, which is described in chapter 3.1 "Interaction walk-through and design rationale". The last two levels, the **eye movement control loop** and the **intrasaccadic image-scanning loop** are not relevant for this thesis.

Implications

According to (Ware [2004], the model with its three levels presented above has three main implications for data display systems:

1. To allow the user to use the advantages of the pattern-finding capabilities of the middle stage of visual processing, the data should be represented in a way so that informative patterns are easy to perceive.

2. To help the user think about the problem and not the interface, the cognitive impact of the interface should be minimized.

3. For low-cost, rapid information seeking, the interface should be optimized.

The complexity of visual query patterns result from the expert level of the system users.

There exist a lot of visual query patterns with respect to graph examination. For example, trend estimation, correlation identification, outlier detection and characterization, or identification of structural patterns. To perform a visual query rapidly and with a low error rate, the query should consist of a simple pattern that can be held in the working memory. For expert users, the query patterns can have greater complexity. However also for experts, some patterns are easier to understand than others because of the laws of pre-attentive processing and elementary pattern perception. When a system designer starts with the visualization of the system, he has to be aware of the expertness of the people who will use the system at the end. In general, he has to use simple informative patterns to represent data (implication 1) so that the users can use the advantage of pattern-finding capabilities. When he designs the system for experts, the patterns can get more complex. This user awareness combined with the complexity of patterns of the interface is also important so that the user can focus on the problem instead of the interface (implication 2) and to ensure rapid information seeking (implication 3).

These basic rules are important to have in mind when constructing a system with interactive visualizations. All users will try to manipulate objects, explore the data space, and solve a task in a way, similar to the rules explained above. The implications, based on these steps give initial considerations which design decisions are necessary to do. To support the user in understanding and interpreting the visualization, the system should support these three levels of feedback loops.

2.3 Graphical data extraction

As we could see in section 2.1.2 "Interactive visualizations", there already exist several ways to understand and interpret CIs based on detailed underlying data. However,

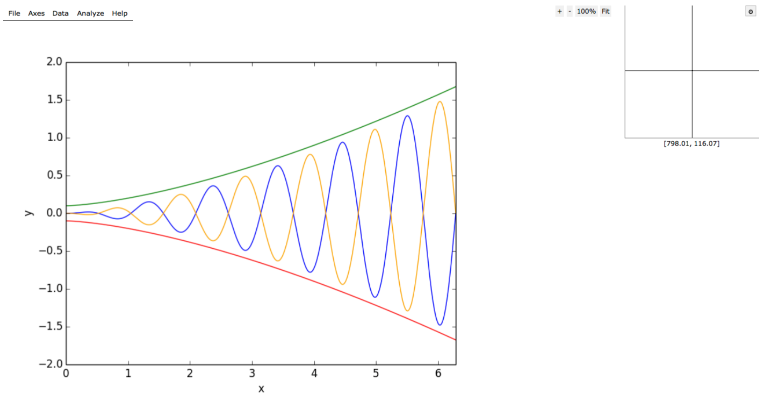


Figure 2.12: WebPlotDigitizer to interactively extract underlying data and set additional data points (Rohatgi [2015b]).

it is often hard and time-consuming to get this data, as we already discussed. Often, referencing text in the paper includes relevant facts which are not directly seen in the chart (see paper of Choudhury et al. [2013]). Therefore, graphical data extraction of the CI chart to get important data values is a relevant part for my work. This will be the base to label the graph with its main values and to simulate data which then both helps to understand and interpret CIs. As there already exist several systems that automatically extract data out of the chart and the text, this topic will be out of the scope of this thesis. However in the following, three general ways extracting data from general graphics will be explained.

2.3.1 Interactive data extraction

One way to get the underlying data out of a chart is the combination of user interaction and computation. One exemplary system is WebPlotDigitizer (see Rohatgi [2015b]) shown in figure 2.12. Here the user can upload the chart in a web browser and simply redraw the graph by drawing the x- and y-axis and labeling them with the according values and numbers. With automatic curve extraction algorithms, rapid extraction of a large number of data points.

Systems like WebPlotDigitizer help users to interactively extract the underlying data.

The user can also add data points by clicking in the chart. These data points can also be rearranged by the user. As an analysis task, the user can choose between getting an angle or a distance between two points. Therefore, he has to select two points for the distance and three points for an angle. Then, the system calculates the distance or the angle according the previously set axis. This technique for data extraction is very important for more complex charts like stacked or grouped bar charts, where marks, axes, and data values cannot yet extracted automatically. However, the users action can vary greatly from chart type to chart type. Therefore, it is easier and more time consuming to use automatic data extraction whenever possible.

As we have seen in this section, interactive data extraction and analysis is already combined with partly integrated automatic data extraction. How automatic data extraction can look like will be explained with some exemplary systems in the next section.

2.3.2 Automatic data extraction

Systems like ReVision extract the underlying data to redesign different chart types.

The paper of Savva et al. [2011] presents a system, called ReVision, that automatically redesigns visualizations of a given bitmap image for an improved graphical perception. At first, the chart type is identified by computer vision and machine learning techniques. The steps explained in the following are visualized in figure 2.13.

As first step, the image is classified.

The image classification process is divided into seven steps. In the first step, the image gets normalized. At next, square patches get extracted and patch standardization is applied. Path clustering then results in centroid patches, that correspond to the most frequently occurring patch types, called codebook patches. They capture frequently occurring graphical marks like lines, points, corners, arcs, and gradients. In the patch response computation, for each extracted patch the nearest codebook patch is determined by Euclidean distance. Then the feature vector formulation starts, where the dimensions of the codebook patch response map get reduced by dividing the image into quad-

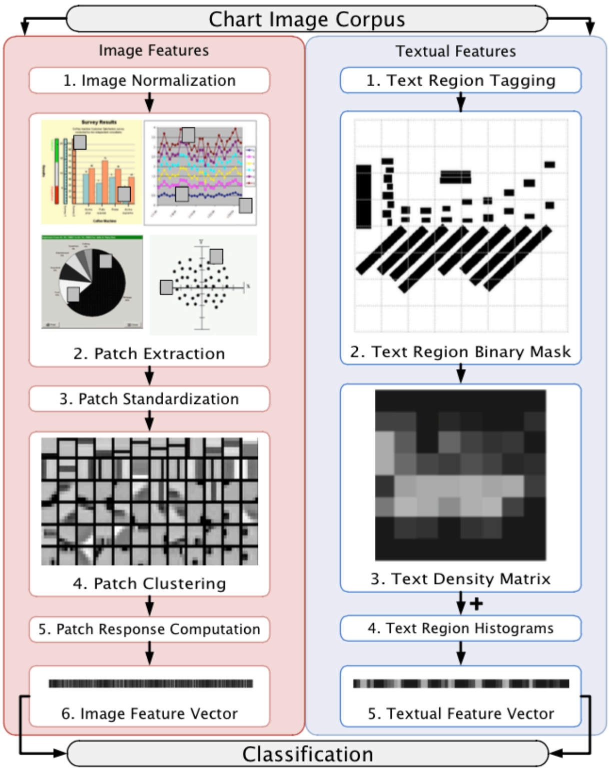


Figure 2.13: Chart image and text classification (Savva et al. [2011]).

rants and summing the activations for each codebook patch in a given quadrant. In the last step, the feature vectors get classified using Support Vector Machines with quadratic kernel function. Beside the image classification to determine the category of a chart image, also the text classification is important.

As second step, the text gets classified.

As text regions in chart images often correlate with the visualization type, it can further improve the classification accuracy. Therefore, the authors designed a tagging interface to annotate chart images with position, size, angular orientation, and content of text regions. The tool extracts the text image region and performs OCR with the Tesseract open source OCR engine. Then the text region binary mark is constructed which indicates which pixel belongs to text regions. By subdividing the mask into 8 x 8 blogs, the text density matrix is constructed. After that, the values

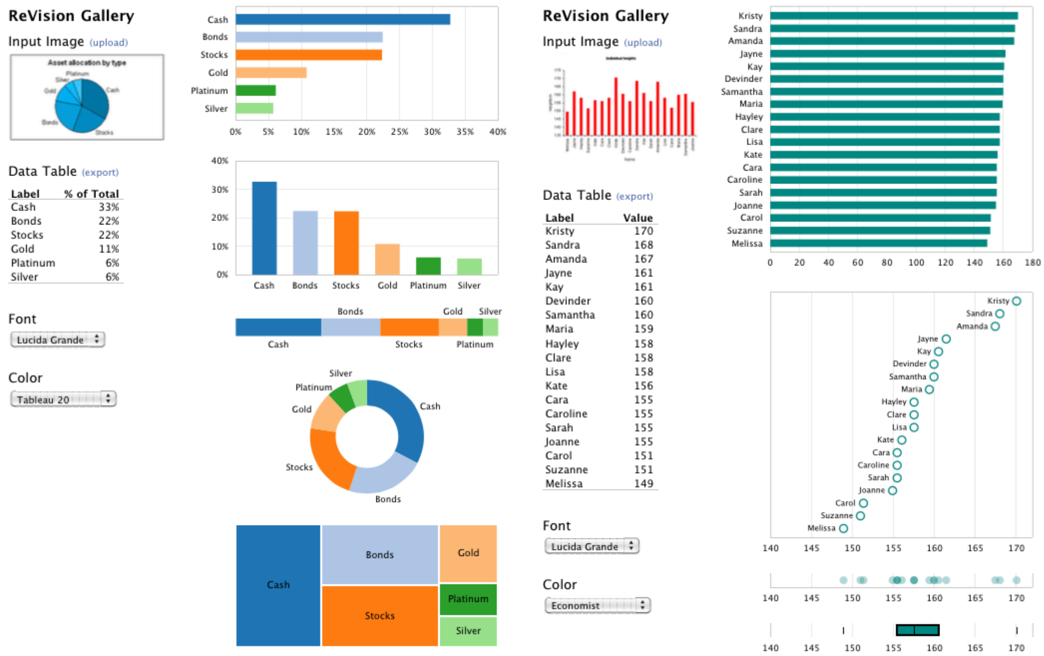


Figure 10: ReVision Design Galleries. Given an extracted data table, the gallery presents a variety of chart types, sorted by proposed perceptual effectiveness rankings [20]. Users can also select and compare color schemes and typefaces.

Figure 2.14: ReVision system with two input charts and their redesigned charts (Savva et al. [2011]).

get linearized and normalized 10-bin histograms of the distributions of text region length, width, center position and pairwise orientation and distance between text region centers get computed. These histograms then get concatenated with the text density vectors to create a final textual feature vector.

96% classification accuracy.

With these steps, the chart gets 96% accurately classified across area graphs, bar graphs, curve plots, maps, pareto charts, radar plots, scatter plots, tables, and venn diagrams. Afters these steps of classification, the graphical marks get extracted to infer the underlying data.

The data extraction phase results in a table with an ID and a data tuple and the redesigned visualizations appear.

In the data extraction phase, the geometry of the extracted marks and the text region tags from the classification stage result in a table containing an ID and a data tuple for each mark. Then, an interactive gallery of redesigned charts appear, based on perceptual design principles. The system chooses different visualizations depending on the input chart type and extracted data. For pie charts (see figure 2.14, left), among others bar charts are generated to to support part-to-part comparison. With bar charts as input (see figure 2.14, right), the system generates a sorted bar chart and a labeled dot plot to support comparison of individual values, and small dot and box plots to enable assessment of the overall distribution. According to the authors, this interface helps the users to view alternative chart designs and to retarget content to different visual styles. The user can change mark types, colors, or fonts.

As the different visualizations help the user to get different insight, these visualization types will also be included in the system of this thesis. How this will be implemented in the system will be explained in detail in chapter 3 "System design - SimCIs".

The user gets different insight with different visualizations.

In Choudhury et al. [2013], not only the charts but also the text in the paper belonging to that chart can be extracted. With a Java based PDF processing library PDFBox, text and raster images together with their IDs get extracted from PDF files. Vector graphics cannot be extracted with this system. After that, line classification in the text is used to find the caption based on the ID and to extract the caption. Figures and caption are matched by extracting the figure number beneath the image and searching for that in text. Several researchers worked on automatic data extraction like Gao et al. [2012], or the authors of datathief[1].

Other systems also extract the text in the paper automatically.

2.3.3 Crowdsourced data extraction

Another possibility to extract data is to provide crowdsourcing when selecting parts of the text to highlight corresponding parts of the chart. Linking a chart to its corresponding text can be helpful because there might be additional information in the text which help to further understand the graph. The linking helps the user to better understand the relation between the text and the chart. Kong et al. [2014] provide a mechanism to extract references be-

Users highlight text and corresponding parts in the chart get highlighted due to crowdsourced linking.

[1] http://www.datathief.org

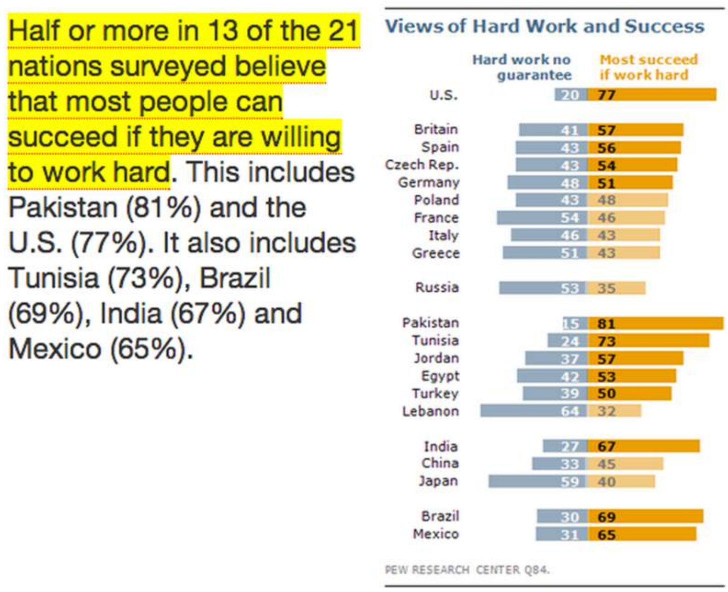


Figure 2.15: Highlighted text (yellow background) refers to 13 bar segments highlighted in the chart (dark orange) (Kong et al. [2014]).

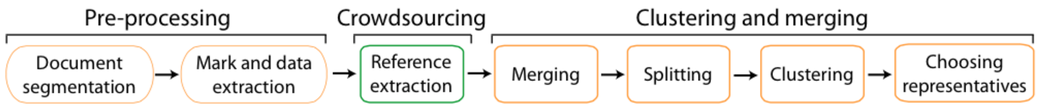


Figure 2.16: Three stages of extraction (Kong et al. [2014]).

tween text and charts with a crowdsourcing pipeline. The result then looks like shown in figure 2.15. The selected text stays highlighted and the corresponding information in the chart is highlighted, shown in darker orange. This reference extraction is done in three stages (see figure 2.16). In the first stage, called preprocessing stage, segments input documents into paragraph-chart pairs and extracting the marks and data tables from the charts. The segmentation is important to have micro tasks that workers can complete quickly in good faith. In this paper, the text splitting and chart pairing is done manually, but can also be enabled to crowdsourcing. The mark and data extraction is already done automatically, using the same procedure as Savva et al. [2011], which already explained in section 2.3.2 "Automatic data extraction". When the user now highlights a part of the text, the reference is extracted via crowdsourcing (stage 2) and the according part in the chart gets marked (stage 3). This idea of extracting references between text and charts, which results in highlighting of a chart when the user selects parts of the text, is also a relevant part for my thesis. I will adapt this idea by filtering a set of different charts (small multiples) according the text that is selected by the user. As text can provide additional information for a chart, it can nail down different possible visualizations of a problem which helps the user to understand the chart. How this will affect my thesis in detail will be discussed in chapter 3.1.3 "Interactive refinement by selecting information from the text".

As seen in these works, automatic and semi-automatic data extraction from charts are already explored in the literature. Therefore, this thesis focuses on the interaction with the chart in order to understand a specific uncertain chart and assumes that the data will be extracted automatically.

2.4 How to evaluate a system with visualization

Checklist for creating visualizations of uncertainty.

Heer and Shneiderman [2012] made a taxonomy of interactive dynamics for visual analysis which can be used to compare and to detect missing features of the system that support basic tasks. It can be used as checklist of items to consider when creating a new analysis tool for developers. The points which are relevant for the visualization of uncertainty are:

1. Visualization
2. Filter
3. Select
4. Navigation
5. Coordinated Views
6. Organized visualization

7. Recording
8. Annotate
9. Share results
10. Guiding

The taxonomy can be divided in three main parts. These parts will be explained in the following, by naming the important aspects which is needed in each system. Furthermore, examples are given of how a realization could look like. As these concepts are one of the main parts for the design of my system, it will be explained in detail when talking about the design process in chapter 3 "System design - SimCIs".

2.4.1 Data and view specification

When the user shall be able to explore large data sets with varied data types, a flexible visual analysis tool must provide controls for the user to specify the data and views of interest. These controls important for visualizations of uncertainty comprise visualization and filtering.

Visualize to choose which data can be shown.

When **visualize (heuristic 1)** data by choosing visual encodings, the user can actively choose which data can be shown and how. Therefore, the user shall have the possibility in a program to choose between multiple charts types and their data variables, like the size and color or visualized marks. For this thesis, a system will be designed which allows the user to see the data in different chart types, like in pie charts for the probability distribution as it was already done by Ferreira et al. [2014].

Filter to focus.

To focus on relevant items, the user shall be able to **filter (heuristic 2)** out data. The user has the possibility to shift his focus among different data subsets by using radio boxes or checkboxes for categorical and ordinal data. Another way is to filter with sliders, which is also possible with temporal data. The system from this thesis will give the user the possibility to filter out small multiples by highlighting additional information in the text of the paper.

2.4.2 View manipulation

The user should be able to manipulate the view, like selecting, navigating, coordinating, or organizing, in order to understand and interpret the visualized data.

The user shall be able to **select (heuristic 3)** items in order to highlight, filter, or manipulate them. This can be done with mouse hovering, mouse click, or region selections. Ferreira et al. [2014] used the selection to highlight bars according a special benchmark or range. This thesis focuses on hovering over the small multiples in order to manipulate the large single.

Select to highlight, filter, or manipulate.

Navigation (heuristic 4) is necessary to examine high-level patterns and low-level detail. One common pattern here is the idea to firstly show an overview, then zoom and filter and at last show details-on-demand. Different kinds of navigation are already explained and the implementation in the thesis' system is given in section 2.2.2 "Exploration and navigation loop" when talking about focus, context and scale. Here it is also shown that overview and details can be shown simultaneously.

Navigate to examine high-level patterns and low-level detail.

Coordinated views (heuristic 5) are beneficial for linked, multi-dimensional exploration. This helps to see data from different perspectives. To compare several view, the small multiples approach from Edward Tufte is generally used as well as in this thesis' system. Here, several visualizations in the same type and with the same measures and scales are placed in spatial proximity.

Coordinated views to see data from different perspectives.

When having multiple views and workspaces, the collections of visualizations have to be **organized (heuristic 6)** in a way. In general all windows shall be able to be opened, closed, maximized and to lay out different components by the user. Having a large display, all information can be seen at once, which should result in a way so that the control panel with sliders, check boxes, radio buttons, and a search box could be on the far right. With a details-on-demand window and annotation box at the bottom. Typical systems allows analysts to add single views, more advanced

Organize views for individual use.

systems allow analysts to use an automated support where multiple windows are opened and closed in a group. These views are then automatically resized and set in spatial proximity. This thesis' system provides some kind of automatic visualization organization of different distributions visualized in small multiples. How this is implemented can be seen in section 3.2 "Visualization of problem cases".

2.4.3 Process and provenance

The process of iterative data exploration and interpretation is an important part in visual analysis and consists of recording, annotating, sharing, and guiding.

Record to refine and review history.

Recording (heuristic 7) the actions and insights helps to review and refine the history of work. At a minimum, this interaction history shall provide basic undo and redo support. Here, low-level input like mouse and keyboard events are easy to capture, while high-level semantic actions become more valuable. The representation of these historical analytic actions can include chronological timeline-like views as well as sequential comic-strip-like views. The sequential views show the steps taken in visual analysis. With the visual history, several interactions are possible. Prior analysis states can be revisited and incomplete explorations can be resumed. For this thesis, the system provides simple undo support when filtering the small multiples.

Annotate to document individual observations.

When the user is able to **annotate (heuristic 8)** patterns, outliers, and views of interest, he can document his observations, questions, and hypotheses. One possibility is to allow textual annotations of states within the visual history. In more expressive annotations, direct interactions with the view can be done, like the selection techniques that are discussed earlier in section 2.4.2 "View manipulation". Freeform graphical annotations can be used to highlight parts of a charts, by using circles or arrows. They allow high degree of expression, but they have no explicit connection to its underlying data. Because they are not data aware, they can become meaningless when using filtering or sorting operations. Annotations can be made data aware when realizing them with selections. This also enables the analyst to search for all comments or visualizations that refer to that data item. Furthermore, these annotations are machine readable, which makes it possible to export selected data and to identify data subsets of high interest. In the paper of Kong and Agrawala [2012] , the techniques for text or free hand annotations to allow the user to create arrows and other reference marks are discussed. However, systems which already focus on visualizing uncertainty, like the ones from Ferreira et al. [2014] or Cumming [2012], do not support annotation techniques. As this thesis' system will also not support this technique, this is an important aspect for future work.

Share to collaborate work or support a process.

Recording and annotation helps the user to **share (heuristic 9)** his results and insights easily with colleagues, to collaborate the work of multiple groups, or to support a process. At minimum, it should be possible to export views or data subsets. It is also important to export the settings of the control panels, so that others can see the same visualization. There are two main possibilities to collaborate the visualization. One of it is sharing via application bookmarking. The system should be able to model and export its internal state. This enables the user to take up the exploration where the former collaborator left of. This is also useful for the collaborators as it provides a common ground for discussion. The other way to share the visualization is to publish it. With this possibility, visualization dashboards can be shared as interactive web pages with a subset of functionalities for some follow-up analysis. It is also important to have in mind which context of use is given for sharing when designing the system. Are the collaborators working synchronously or asynchronously? And are they co-located or distributed?

Guide to create reusable work flows and get all information the system provides.

Guiding (heuristic 10) is important for experts to create reusable work flows, which can be used by less knowledgeable team members. Non experts can use a system directly when having special guiding support which helps them to see all information the system provides. According the authors, more research is needed to identify these effective visual-analytics processes. However, there are different forms of narrative visualizations where interactive graphics are structured to tell stories of the data. A step by step manual through a linear narrative visualization guides the user with the help of supporting text and annotations. At a story's conclusion, interactive control should be provided for further exploration. These narrative structures communicate key observations from the data and provide a tacit tutorial of available interactions by animating each component along with the story. When the user has already seen demonstrations of the interactive controls, the system opens up for free form exploration. This demonstrated how guided analytics can be used to disseminate data-driven stories to a general audience.

2.5 Design space of the interaction with uncertain data visualizations

The visualization of spread, combined with existing techniques to augmenting data visualization, and the taxonomy of interactive dynamics for visual analysis result in the following design space in table 2.1 "Design space of the interaction with uncertain data visualizations.". This design space for interactive visualizations sums this chapter up and shows us the open aspects which need further investigation.

[2]Symbols:
⧫: Cumming [2012] ■: Ferreira et al. [2014] ○: Thesis' system

Table 2.1: Design space of the interaction with uncertain data visualizations.

Interaction			[2] **Visualization of data** Single point (1D)	Interval (2D)
Specification	Visualize	Redundant encodings (labels, lines, etc.)	⧫ ○	⧫ ○
		Reference structures (grid lines)		○
		Vigilance	⧫ ○	⧫ ○
		Transmogrification		
		Layering and separation (coloring)	⧫ ○	■ ⧫ ○
		Small multiples	○	○
	Filter	Dynamic queries (sliders)	⧫	⧫
	Data	Simulate Data	⧫ ○	⧫ ○
		Data extraction		■
Manipulation	Select	Hover Queries	⧫ ○	⧫ ○
		Brushing (highlighting)		■

Table continued on next page.

Interaction			Visualization of data	
			Single point (1D)	Interval (2D)
	Navigate	Distortion technique		
		Rapid zooming technique		
		Elision technique		
		Multiple Windows	⧫ ○	■ ⧫ ○
	Coordinate	Variation only in data	⧫ ○	⧫ ○
	Organize	Manually	⧫	■
		Automatically	○	○
Process and Provenance	Record			
	Annotate	Not data aware		
		Data aware		
	Share		⧫	⧫
	Guide	Linear process	(⧫)	(⧫)
		Flexible process	○	■ ○
	Summarize		(⧫)	(⧫)

Continued table from previous page.

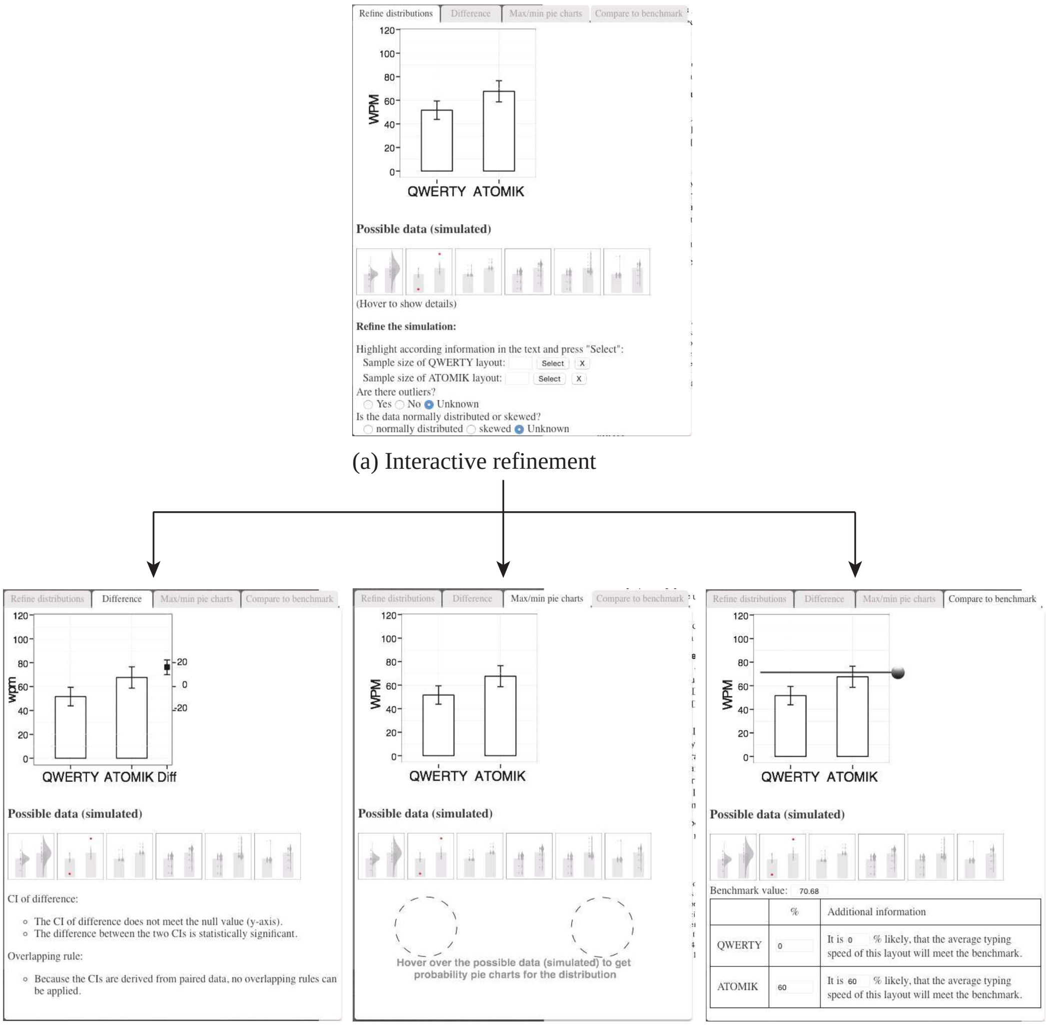
Chapter 3

System design - SimCIs

During this chapter the design of the system SimCIs will be explained. The whole interaction process and the design rationale will be explained at first in section 3.1 "Interaction walk-through and design rationale". In section 3.2 "Visualization of problem cases", all parts will be explained in detail, focusing on the visualization. The mathematical detail is also mentioned here. The first two section will also go into the design principles discussed in chapter 2 "Related work" and appropriate adaption get explained. In the last section 3.3 "Design iterations", the whole iteration of the system design from paper prototypes until the web based system will be shown.

3.1 Interaction walk-through and design rationale

When the user has uploaded the paper with the chart that has to be interpreted, he sees an adapted version of the chart showing a 95% confident interval. How this conversion occurred will be explained later on in section 3.2.1 "Possible input chart types". This updated version of the chart is seen in a panel with four tabs for different interaction purposes as it can be seen in figure 3.1. What interaction can be done in the different tabs will be explained in



(a) Interactive refinement

(b) CI of difference (c) Pie charts (d) Benchmark

Figure 3.1: Overview of the interaction UI of SimCIs.

the following.

3.1.1 Interact with CI chart

Different interaction purposes with the chart in four tabs.

In the first tab (see figure 3.1 (a)), the user is able to switch between original chart and converted chart to see what changed. Therefore, a button is given which shows the original chart as long as the mouse is hovering over the button. This prevents the user from interacting with the original chart what shall not be possible. Furthermore, the hovering enables a faster switch which helps the user easier compare the changes from one to the other chart. There are two types of overlays, mentioned in 2.2.1 “Overlays”, which are used here and in the other interactive tabs. When hovering over the chart, redundant encodings are given in form of labels showing the mean, upper limit, and lower limit of the error bars. These hover queries (see 2.2.2 “Manipulation loop”) showing extra information of the chart helps the user to more effectively read details. When switching to the next tab (see figure 3.1 (b)), an additional CI of difference is shown in the chart. Which information is visualized here will be explained later in section 3.2.5 “CI of difference and the overlap rule”. The third tab (see figure 3.1 (c)) shows two pie charts for each distribution with the probability for each bar in the chart to be the maximum and minimum distribution. The last tab (see figure 3.1 (d)) shows a benchmark, which the user can drag vertically along the chart to see how likely it is that the CIs meet the benchmark. The benchmark can be used as reference structure in a special grid line form, because the value is shown all the time. They assist the perceptual process of anchoring and projection. From the 95% CI chart, small multiples get created, which are directly below the chart in each tab.

3.1.2 Interact with alternate explanation from simulated data

Small multiples show possible distributions of the CIs.

The problem of the CI chart is, that the user does not know on which data points this chart is based on. As the user only uploads the paper, no underlying data is given. However, the distribution of the results can influence the message, a CI chart conveys. Small multiples, as already shown in chapter 2.2.1 “Small multiples”, are there to show differences, changes, or alternatives. In this thesis, they will be used to show how distributions could look like which will result in the CIs represented in the chart. The different possible distributions get simulated with R[1]. The essential code $rnbinom(N, mean, .5)$ calculates the Negative

[1] http://www.r-project.org

Binomial. The Negative Binomial as well as the Binomial counts the samples. However we chose the Negative Binomial, which counts failures until a fixed number of successes. As the Binomial would only count a fixed number of trials, we will always have the same number of successes with the Negative Binomial.

Hover to see additional information of the distribution.

The small multiples, the large single, and the paper are visible all the time, which is comparable to the multiple windows exploration technique (see 2.2.2 "Exploration and navigation loop"). The user can see the context all the time in form of the paper and the large single representation on all paper pages represented as overlays. On the same time, he is able to focus on different distributions which are represented in the small multiples. With the navigation taps in the interaction panel (see figure 3.1), the user can also focus on each interaction purpose, while having the context of the chart and different possible distributions. When the user inspects the small multiples and hovers over them, a bigger version of them will appear as overlay at the place of the large single. Again, this technique called hover queries, mentioned in section 2.2.2 "Manipulation loop", helps the user to get additional information. The user can see the small multiple in larger form with additional textual information of the distribution. In the third tab, when hovering over the small multiples, also pie charts representing the probability of this distribution in the chart to represent the maximum or the minimum error bar are shown. This concept was already provided in 2.1.2 "Interactive visualizations" by Ferreira et al. [2014] for CI charts in general. Here, as there is no underlying data to base on, the probability pie charts are represented for each possible distribution. Which distributions are shown, how they are ordered, and how the pie charts result will be explained in section 3.2 "Visualization of problem cases". However, the focus get shifted to the interactive refinement of the small multiples at first.

3.1.3 Interactive refinement by selecting information from the text

Reduce number of small multiples by adding additional information from the text.

As several small multiples with different possible distributions are represented, the user shall be able to reduce the number of small multiples by adding additional information from the text of the paper. This mapping process can be compared to the mapping technique explained in 2.2.1 "Small multiples". The user can see three different highlighting tasks. The tasks are highlighting the sample size from all error bars, check text for outliers, and check text according the distribution of all error bars. The user can select the sample size in the text and press a button to apply this information. The text stays highlighted, the information appears next to the task, and the small multiples get actualized. For the other tasks the user can press different radio buttons. For the outlier task he can check whether there exist outliers, whether there are no outliers, or whether this information is unknown. In the distribution task, the user can select whether the data points are mentioned to be normally distributed or skewed. The default is set to unknown, as it is done for the outlier task. According to these information, possible distribution get sorted out when their visualization does not fit to the known information anymore. When some small multiples gets sorted out, the other small multiples automatically get reorganized so that there is no empty space between them. Therefore heuristic 6 (see 2.4.2 "View manipulation"), the organization of views, is applied in this system. The mapping of information in the text with the chart reduces the number of small multiples and therefore helps the user to get a more accurate insight of the possible distribution of the chart. Furthermore, these tasks help the user with the problem-solving strategy and provides possible solutions for the pattern finding loop in form of the reduced number of small multiples (see section 2.2.2 "Problem-solving loop"). As the reduced number of different distributions highlights possible data, this whole process can also be seen as a form of brushing from section 2.2.2 "Exploration and navigation loop". It is also possible for the user to undo each of these steps. This might become necessary when the user highlighted wrong information or when he wants to see all different distributions again. As these steps might take more time than simply checking the chart in the paper, the user can get more accurate insights of this chart. This called speed-accuracy trade-off (see section 2.2.2 “Manipulation loop”, choice reaction time) will be tested later on in chapter 4 “Evaluation”.

After this section described how the user used the system and the design rationale, the following section will now focus on the visualization of all parts. The order in which the visualizations are explained is based on the order of interaction steps shown above.

3.2 Visualization of problem cases

As Hao et al. [2005] mentioned, the position and size of displays show how important it is for the user. Top left displays are of higher importance as well as bigger sized displays. This information was included in the design of the system so that the user can focus on understanding and interpreting the chart. Therefore, the context in form of the panel with the 95% CI chart and small multiples with different possible distributions is shown in the top left corner, as it should grab the users attention from the beginning on. The paper is shown in the middle but more in the background.

3.2.1 Possible input chart types

Original charts can have SD, SE, or CIs in general.

As we already discussed in the introduction, there exists different ways to present uncertainty. The common interval based ways are standard deviation (SD), standard error bars (SE), and confidence intervals (CI). Because confidence intervals are the best way to represent uncertainty Cumming [2012], the user shall be able to insert any type. This chart then gets transformed to a CI chart, if it is not already one, before next steps of reverse engineering are realized. How the transformation is calculated is provided in appendix A.1 “Calculation from different input charts to

95% CI”. The CI chart is then drawn with R[2].

Switching between original chart and 95% CI chart to show the difference.

Former prototypes showed the whole process of transformation (see section 3.3.1 “Paper prototype”). However, the final design only provide a comparison of original chart and 95% CI chart which is similar to the Magic Lenses concept to keep it simple and comprehensible for the user. The concept called Magic Lenses (see Bier et al. [1993]) provides a detail and context technique which enables the user to see the details in changes, while being aware of the context of chart interpretation. Like looking through a lens, a difference is provided between the original and the CI chart. As the CI chart is drawn with R, both charts can have different styles, like text font, background color, etc. But the main difference between the charts will be the interval length of the error bars.
After having the final CI chart, the next step is to show possible distributions in a scatter plot like chart.

3.2.2 Possible distributions

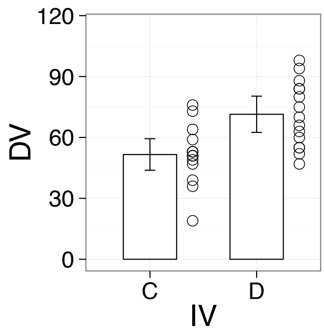
Data of CI charts can be normally distributed, skewed, and can have outliers.

In general, it is assumed that a CI is normally distributed Cumming and Finch [2001]. However, as CIs often present data that results from experiments, there can also be other kinds of distribution. There can exists outliers, or the distribution is non central which means that the data is skewed in one direction. These possible distributions can affect the way the chart can be interpreted. An example of how the distribution can affect the interpretation is given in figure 3.2. A user could assume that there is no big difference between the CIs when looking at the normal distribution shown in figure 3.2 (a). However, when the user can see the skewed data with outliers like in figure 3.2 (b), he might think about the influence of outliers and skewness in general for a distribution. Because of a few outliers, the most data points are on different positions.

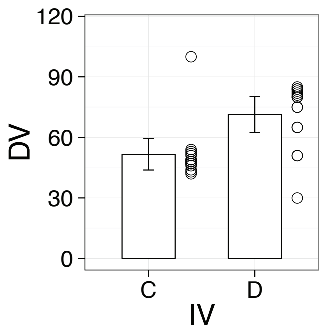
These different distributions result in the same CI.

Now, it is visually clearer that there is a significant difference between the CIs. However, all different kinds of distribution result in the same confidence interval with the same

[2] http://www.r-project.org



(a) Normal distribution



(b) Skewed distribution with outliers

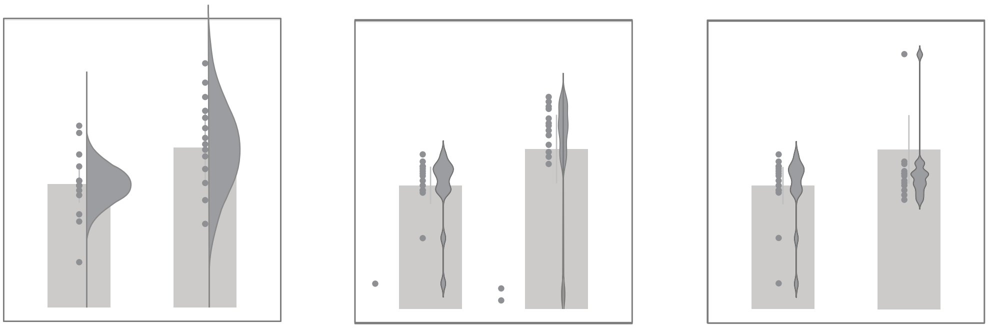
Figure 3.2: Different distributions effect the interpretation of CIs.

error bar. This means, that the error bars are not interesting when comparing different distributions in small multiples. These information get relevant when starting the layering and separation of small multiples. As it was used in the example in section 2.2.1 "Layering and separation", coloring and degrees of transparency will be used to highlight different interesting aspects of each distribution. Therefore, these visual marks can be grayed out to fade it out the spotlight and the visual marks of the distribution can be highlighted. In the following, we will discuss all three possibilities and derive a visualization that let the user easily differentiate these kinds. Because the user uploaded only the paper with the chart, there is no underlying data. Therefore, the different distributions result from simulated data using R[3].
As the data points are simulated to be normal distributed, having outliers, and being skewed, there is no further classification needed by the software to recognize these distributions.

Normal distribution visualized with normal curve.

For **normally distributed** data, a lot of data points from an experiment are located around the mean. Less data points are located away from the mean near the upper and lower limits of the CI. Having a normal distributed CI resulted from several experiments, the distribution of sampled means are normally distributed. This can be visualized with a normal curve that corresponds to the probabil-

[3] http://www.r-project.org



(a) Normal distribution (b) Skewed in same direction (c) Skewed in different direction

Figure 3.3: Visualization of different distributions resulting in the same CI.

ity density function of a t-distribution. This violin plot like representation, based on the work of Correll and Gleicher [2014], consists of the same visual metaphor that values become less likely the bigger the distance from the mean is. Figure 3.3 (a) shows exemplary how the visualization can look like. The normal curve is a bit transparent so that the user has no problems to the the bar behind the curve.

Skewed data points are visualized with an adapted violin plot.

Data points of a CI can also be **skewed**. This is visually represented by a large number of data points which are positioned on one side of the interval and only a few data points located on the other side. Here the violin plot from Correll and Gleicher [2014] is taken and adapted with scaled kernel density estimation. This highlights the skewness of data points, as areas of the violin plot are bigger when more data points are in that area. The resulting design can be seen in figure 3.3 (b) for data points of two CIs skewed in the same way and in figure 3.3 (c) for differently skewed data points. In the system, there are four skewed charts; both are skewed down, both are skewed up, first chart is skewed up and the other is skewed down, and vice versa.

There is a high possibility for the distributions to have outliers, or data points with a certain outlier factor. However, the outliers are not highlighted in the normal and skewed distributions, as the focus shall stay on the distribution. More information according to the outliers and their outlier factor are discussed in the next section.

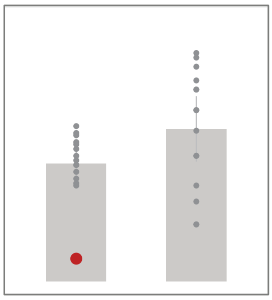


Figure 3.4: Chart showing a distribution with an outlier (left) and a normal distribution (right).

3.2.3 Local outlier factor

Outliers visualized with red bigger data points.

When doing an experiment, it is normal that outliers can occur. These points are visually easy to see as they are more distant to the other points. Because an outlier changes the resulting confidence interval dramatically, it is important to highlight these points. As we learned from Edward Tufte in section 2.2.1 “Layering and separation”, small spots of intense and saturated color for carrying information help to highlight important parts. Therefore, outliers are marked as red, bigger dots (see figure 3.4). This reminds the user of the possibility of outliers which are mentioned in section 2.2.2 “Manipulation loop”, when talking about vigilance. How the degree of outlierness will result in the data point size will be explained in the following.

Local outlier factor to calculate size of data points.

For many knowledge discovery in database (KDD) applications, finding outliers or other rare instances can be more interesting than finding common patterns (see Breunig et al. [2000]). According to the authors, it is more meaningful to assign the degree of an outlier to each object instead giving them a binary property. This degree, also called local outlier factor (LOF), shows how isolated the object is according to the surrounding neighborhood. This factor can be used to find considerable outliers in real-world datasets. To calculate the LOF of a given dataset D, some steps have to be done which are given in A.2 “Calculation of the local outlier factor”. As previous work studied outlier detection in the field of statistics, we will also use this mechanism in our system. It is acceptable to only use

these basic steps to calculate the local outlier factor, because for CIs, the data points only differ in y-coordinates, what is a relatively simple setup. However, this approach is necessary to use, because the sample size can get very high.
Because outliers have bigger impact on the resulting CI than a data point near to the mean, this point should be highlighted by having a bigger size. Using the outlier factor to resize the data points in a scatter plot however needs additional work. According to Kriegel et al. [2011], the simplest way to translate an outlier score S to a normalized scale is to apply a linear transformation. With this translation, the minimum occurring score of a dataset is mapped to 0 and the maximum occurring score to 1. The resulting formula looks as follows:
$Norm_S^{linear}(p) := (S(p) - S_{min})/(S_{max} - S_{min})$.
Here, this normalized score is then multiplied with the original size of the data point and added to that former value. The corresponding R code is shown in appendix B.1 "Generate the local outlier factor".

The normal distribution, the skewed distributions and a distribution with an outlier result in the small multiples that are ordered horizontally below the large single. As it is assumed that the data of a CI is distributed normally in general, it is positioned at first.

3.2.4 Pie chart representation with minimum/ maximum probability distribution

Bootstrap to re-sample simulated distributions and get probability pie charts.

The pie charts shall help the users to see the probability of each confidence interval in a chart to be the minimum and maximum error bar. Using the bootstrap technique, the CIs get resampled 10000 times. Therefore, the number of repetitions is the same as for the resampling method from Ferreira et al. [2014]. Each time it is checked which mean is bigger and which is smaller than the other. At the end, the comparison is taken to make one pie chart for the maximum probability and one for the minimum probability. The underlying code is provided in appendix B.2 "Bootstrapping to generate probability pie charts".
These two pie charts get calculated for all different distribu-

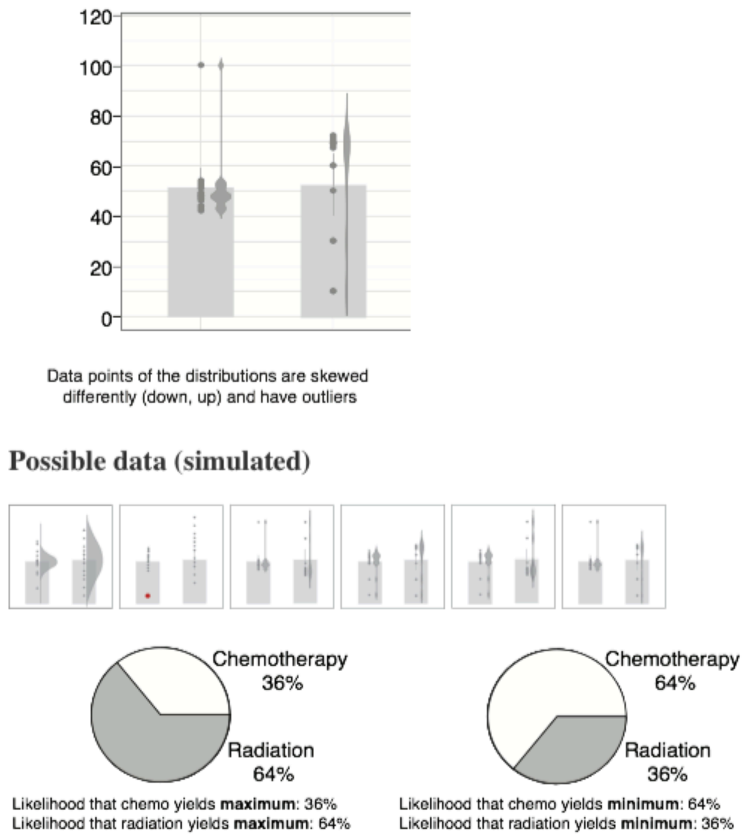


Figure 3.5: Two pie charts represent the probability for the CIs Chemotherapy (left) and Radiation therapy (right) to be the maximum and minimum CI.

tions which were mentioned in section 3.2.2 "Possible distributions". For each chart there are two pie charts at the end. These two pie charts are represented below the small multiples in the third tab, when the user hovers over one small multiple.

One exemplary setup can be seen in figure 3.5. On the top, an exemplary distribution can be seen. According these data points, bootstrapping is applied and the pie charts get constructed. On the bottom, the user can the see probabilities for each CI to be the maximum or minimum error bar. According the concrete example, shown in figure 3.5, 36% of bootstrapped distribution means from Chemotherapy were higher than the means from the Radiation therapy. 64% of the bootstrapped distribution means from the

Radiation therapy were higher than the means from the Chemotherapy. Therefore, it is more probable that the Radiation therapy will yield the maximum.

3.2.5 CI of difference and the overlap rule

This section focuses on the direct comparison of the two CIs, independent from different possible distributions. A CI of difference is provided and an overlap is shown for independent data in the second tab. When calculating the CI of difference between two CIs, it is important to know whether the data came from two independent groups or from one group doing a pre- and a posttest. Whether the data is paired or independent is given in the paper in general. More information about the concrete calculation can be found in appendix A.3 "Calculating the CI of difference". For independent data, the CI of difference is always larger than each of the original intervals, as the sampling error in the difference is a compounding of sample error from each of the two independent means (see Cumming and Finch [2005]). For paired data, the CI of difference is always smaller than each of the original intervals, as the margin of error is based on the SE of the differences, which reflects a high correlation. As the CI of difference is always lower than the original CI, based on the y-axis, it gets a separate scale. When the CI of difference meets 0 on the y-axis, the general assumption is, that the difference is not significant. The more distance between the CI of difference and the null value, the the more significant is the difference between both CIs.

CI of difference is shown which indicates whether the difference between two CIs is significant different or not.

When having CIs from independent groups, the overlap rule can also be applied to see whether the difference is significant or not. Therefore, this information is provided additionally to the CI of difference for independent data. For the software only a few calculation steps are necessary to get the proportional overlap. At first, the average margin of error has to be calculated. It is the sum of the margin of errors from the original CIs divided by two. Then, the overlap is calculated by subtracting the lowest upper limit from the highest lower limit. To get the proportion over-

For independent data, the overlap rule can be applied. The proportion overlap is visualized.

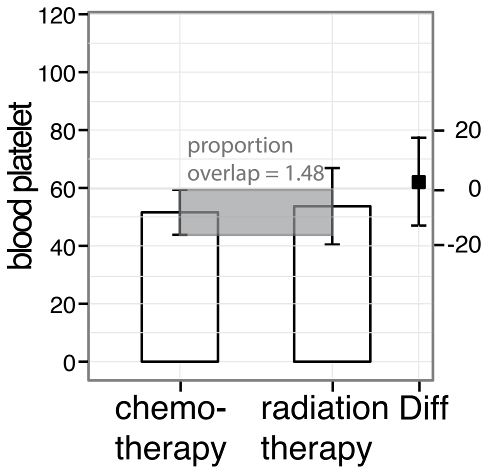


Figure 3.6: The CI of difference is represented with a separate y-axis and because the chart results from independent data, the overlap in shown.

lap, which is relevant for the overlap rule, the calculated overlap is divided by the average margin of error. Every value of the proportion overlap smaller than zero get set to zero as this value is only relevant to show how big an overlap is. When the proportion overlap is about 0.50 or less, the p value is equal less $.05$ which means that the difference is significant. This means visually, that the overlap of the CIs is not more than half the average margin of error (represented visually in the CI of difference). The smaller the proportion overlap is the more significant is the difference. One exemplary setup can be seen in figure 3.6. Here, the overlap is very big and has a proportion overlap of 1.48. According the overlap rule, which can also be applied visually, this difference between the CIs is not significant. The CI of difference is shown on the right side with a separate y-axis. It shows the same result as the overlap rule. Because this CI does meet the null value, the CIs are not significant different.

Visual annotation guidelines of Ferreira et al. are met.

All in all, this section explained how the principles, explained in chapter 2 "Related work" are used to create a system which is visual clear and provides mechanisms to help the user in understanding and interpreting uncertainty in form of confidence intervals. The concept for creating visual annotations is based on the guidelines of Ferreira et al. [2014], which are written down in section 2.1.2 “Interactive visualizations”. Special annotation tasks and different visual representations are accompanied with text to be easy to interpret and only adding minimal additional visual complexity. The consistency across the tasks is ensured by showing important information, like different distributions, a benchmark, or CI of difference, always in the space of the large single. Here, the dataset is always represented in the same visual representation. The only change in the visual representation is given with the use of pie charts to show the probability of error bars to be the maximum and minimum bar. However, this accomplishes the guideline of minimizing visual noise, by displaying different information in different visual representations. This also ensures the spatial stability across the sample size. How these visualizations changed from the first approach to the final design will be explained in the next section.

3.3 Design iterations

In the following, illustrations of the different iteration steps are shown. It starts with a paper prototype in section 3.3.1 “Paper prototype”, continues with two iterations on a software prototype in section 3.3.2 “Software prototype”, and ends in a web based prototype in section 3.3.3 “Web-based prototype”, which shall show how a working system could look like. In section 3.3.4 “Iteration in design”, a table summarizes the changes from one iteration to the next one with the reason why these changes were made.

3.3.1 Paper prototype

Paper prototype focused on whole process without going into detail.

Figure 3.7 shows general transformation steps from SD to SE to CI chart, a transformation from a pie chart to the scatter plot, the transformation from the scatter plot to a histogram, and after the user selected whether the data is paired or independent the CI of difference arises.
The process from the underlying data to an SD chart is visualized by firstly showing all data points which then merge

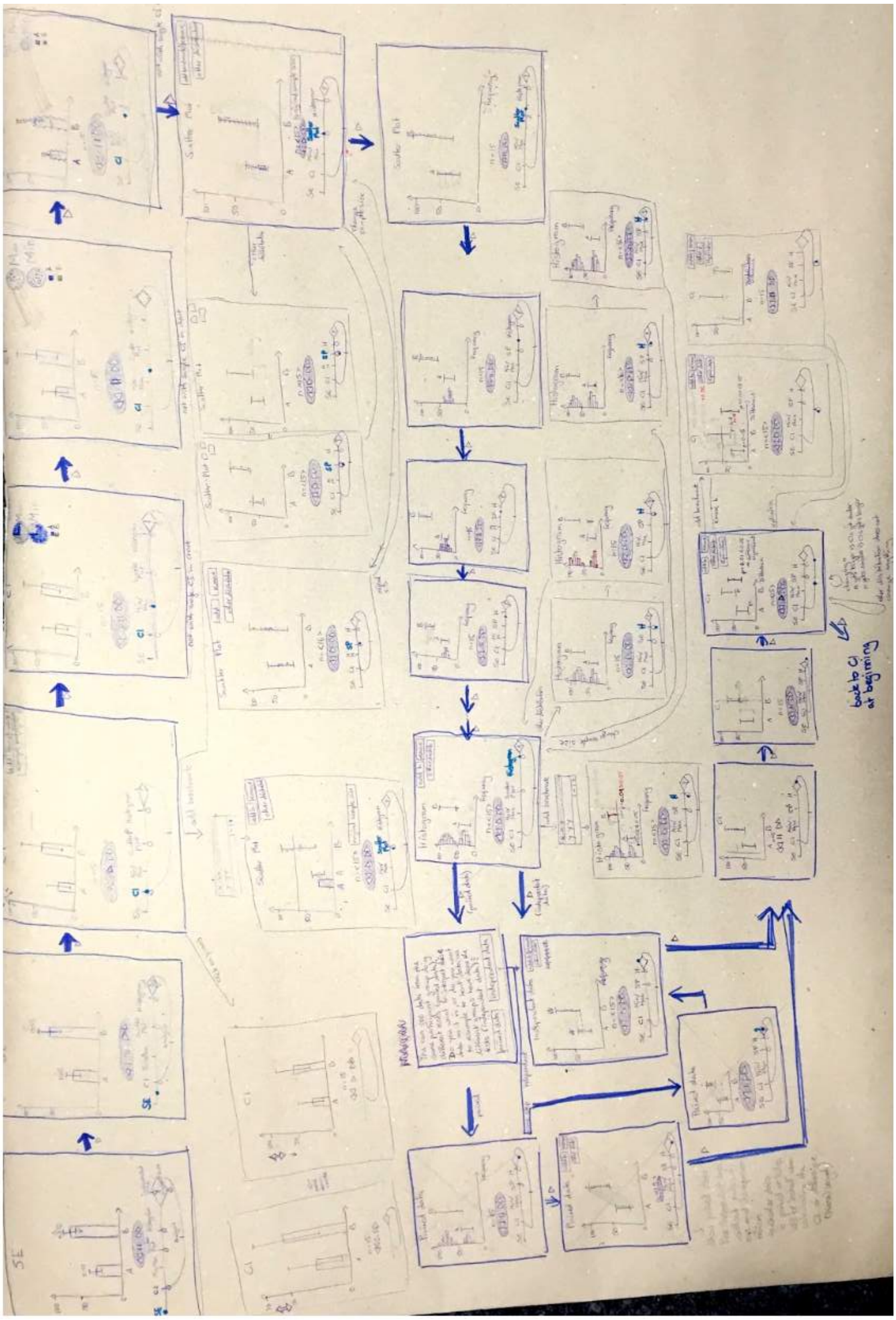


Figure 3.7: Paper Prototype.

to the mean and the variance. Furthermore, the formula is shown, which provides the user with relevant information to reproduce this process on his own.
For the visualization from SD to SE, an icon representing the number of users who did the experiment pass each interval of a chart. When passing the interval, the interval gets shortened to its SE value. As it is done for the previous transformation, the formula is shown.
During the transformation process from SE to CI, a "95% CI" label passes the intervals, which then enlarge again according the right factor.
At the end, special gridlines at the UL and LL as well as at the mean of each CI appear, having an alpha value of 0.2. On these gridlines the exact y-axis value is given so that the user gets assisted in the perceptual process of anchoring and projection.
These transformation steps can begin at each point which guarantees the user a correct CI chart at the end, no matter whether a bar chart with underlying data, a SD chart or an SE bar chart was inserted to the system. During all these steps a navigation bar indicates on which step the simulation is for that moment.

3.3.2 Software prototype

First Keynote prototype

On the first iteration of the software prototype, all visualizations and transformation were done in more detail. In figure 3.8, the transformation steps to a CI chart are shown. Figure 3.9 shows the transformation steps from a scatter plot to the pie charts. Here, a form of leverage ("Discovering Statistics Using SPSS") is used, which is replaced by LOF later on. Figure 3.10 shows the generation of small multiples. The visualization of a normal distributed scatter plot, a scatter plot with outliers, and skewed data points, as well as the order of the small multiples were still not clear at this point of time. The skewness curve did not change the width and the height according the data points. At last there is the transformation from a scatter plot to a histogram shown in figure 3.11. Each data point moves to the

First software prototype focused on the transformation of the visual representations.

y-axis and based on that the normal curve is drawn.

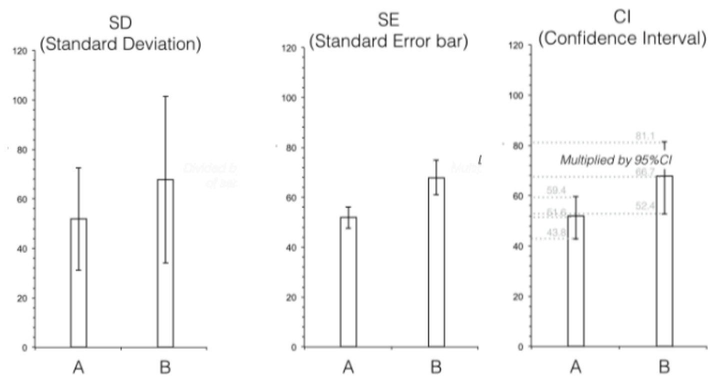


Figure 3.8: First Keynote prototype illustrating the transformation from SD to SE to CI chart.

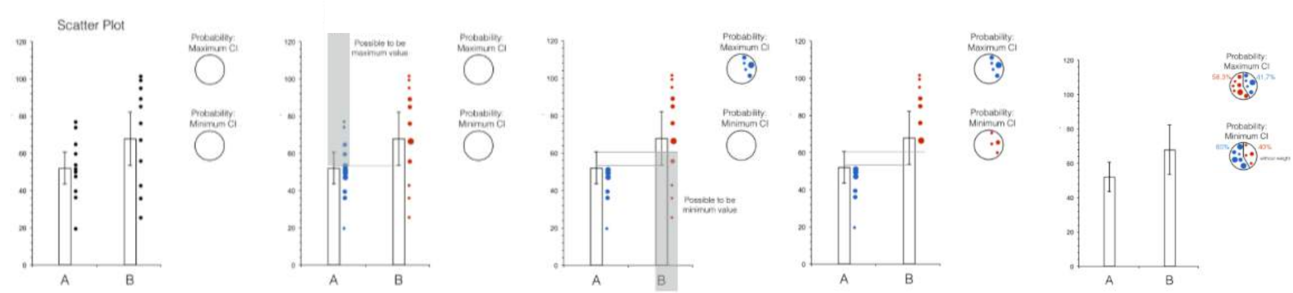


Figure 3.9: First Keynote prototype illustrating the transformation from scatter plot to a pie chart.

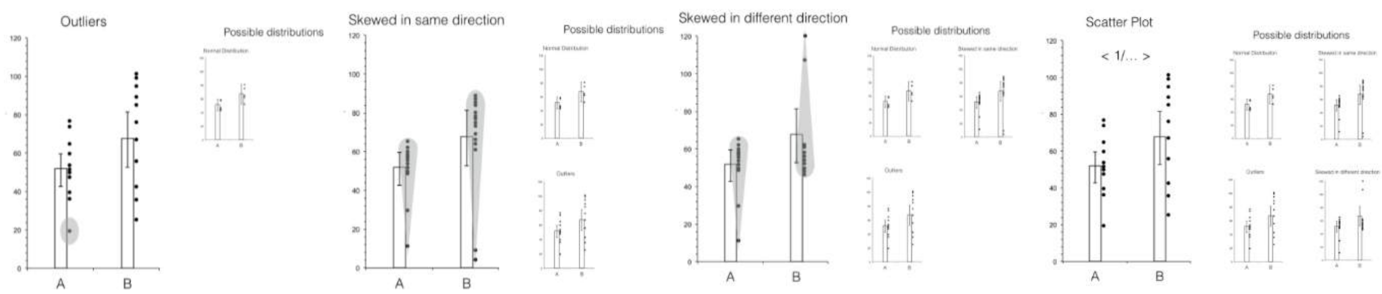


Figure 3.10: First Keynote prototype illustrating the generation of small multiples.

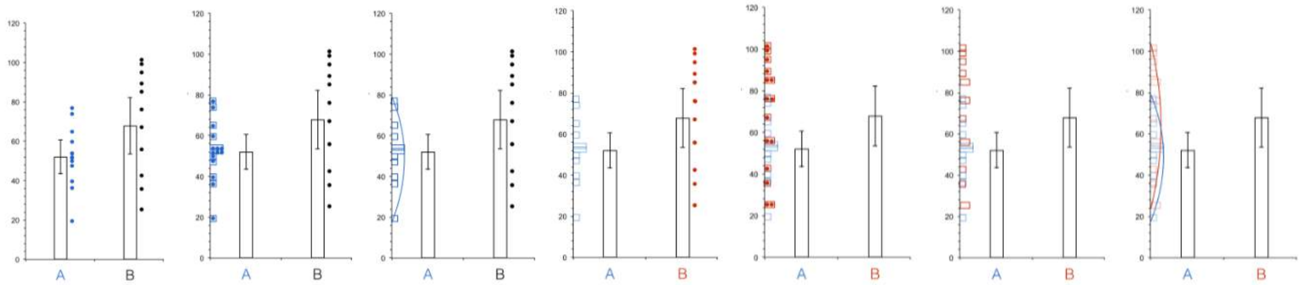


Figure 3.11: First Keynote prototype illustrating the transformation from scatter plot to histogram.

Second Keynote prototype

The second iteration only focused on changes. This is the first prototype, where the paper is also visible in the screen (see figure 3.12 (a)). Furthermore, the navigation bar got simplified (see figure 3.12 (b)) and the focus was set on the user interaction to filter out small multiples according additional information from the text (see figure 3.12 (c)).

Second software prototype focused on visualization of the system design and the interactive refinement.

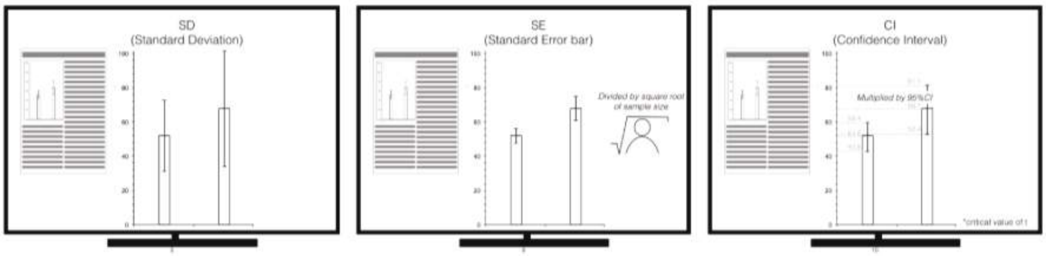
3.3.3 Web-based prototype

The web-based prototypes already were implemented. They focused on the working user interaction with the system.

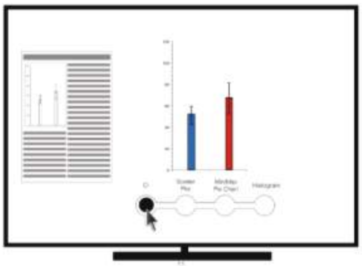
First web-based prototype

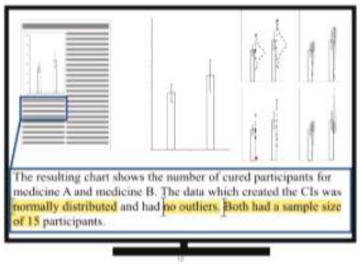
The first web-based design had an overall appearance as it can be seen in figure 3.13. The text is now centered, the small multiple is on the left to grab the user attention, and the small multiples are on the right in thumbnail form. The user does not see the transformation from the original chart to the 95% CI chart anymore. He can switch between these two representation by hovering over the button in the top left corner.
The benchmark visualization is added (see figure 3.14 (a))

First iteration with large single and tasks on the left, paper in the middle, and small multiples on the right.



(a) Transformation from an SD to an SE to a CI chart.





(b) Navigation bar. (c) Text highlighting to filter small multiples.

Figure 3.12: Second Keynote prototype based on changes.

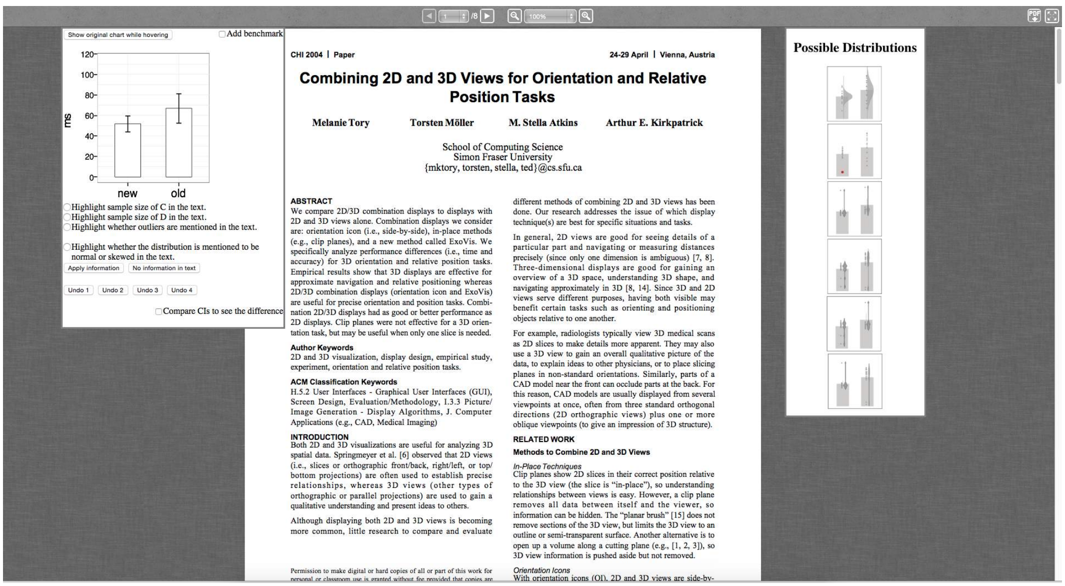
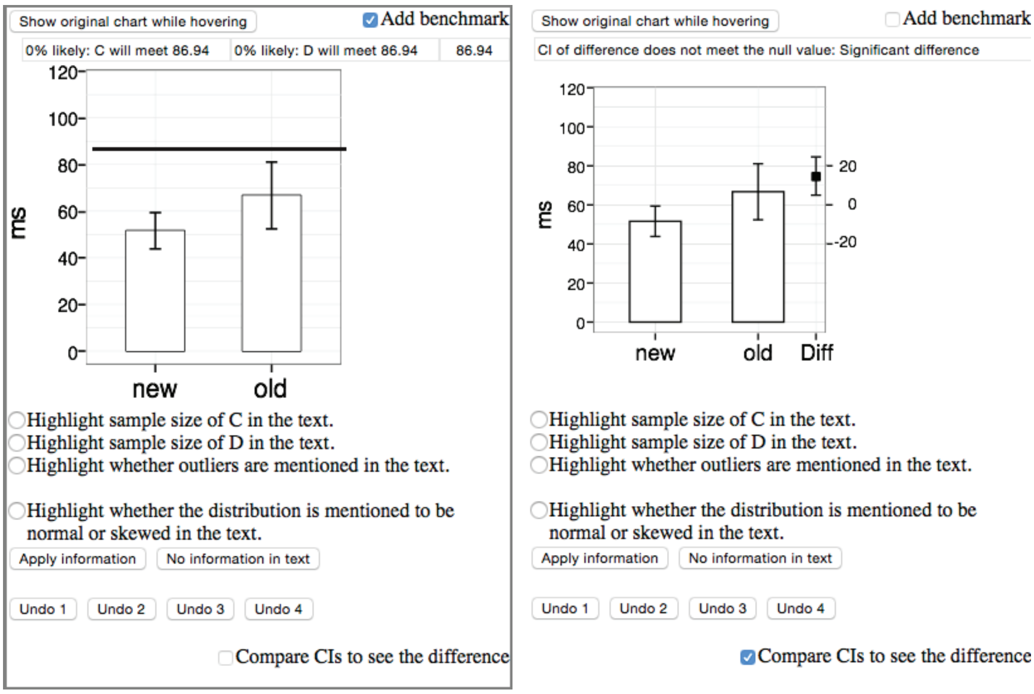


Figure 3.13: First web-based prototype: Overall appearance.



(a) Benchmark (b) CI of difference

Figure 3.14: First web-based prototype: Interaction.

by clicking on a check box on the top. Additional information about the likelihood of the CIs to meet the benchmark are given.
By clicking on the check box in the right bottom corner, the CI of differences can be seen in the large single. This can be seen in figure 3.14 (b), where also additional information of the meaning of the CI of difference is given.

Second web-based prototype

In the second iteration a panel with several tabs was added so that the user can focus on one task at time. As it can be seen in figure 3.15, the small multiples are directly beneath the large single to provide a visual connection. When the user now refines the small multiples according the information in the text, he can see the results directly near to the interaction space. This can be seen in figure 3.16 (a). In the

Second iteration with panel, small multiples, and tabs for each task on the left and paper in the middle.

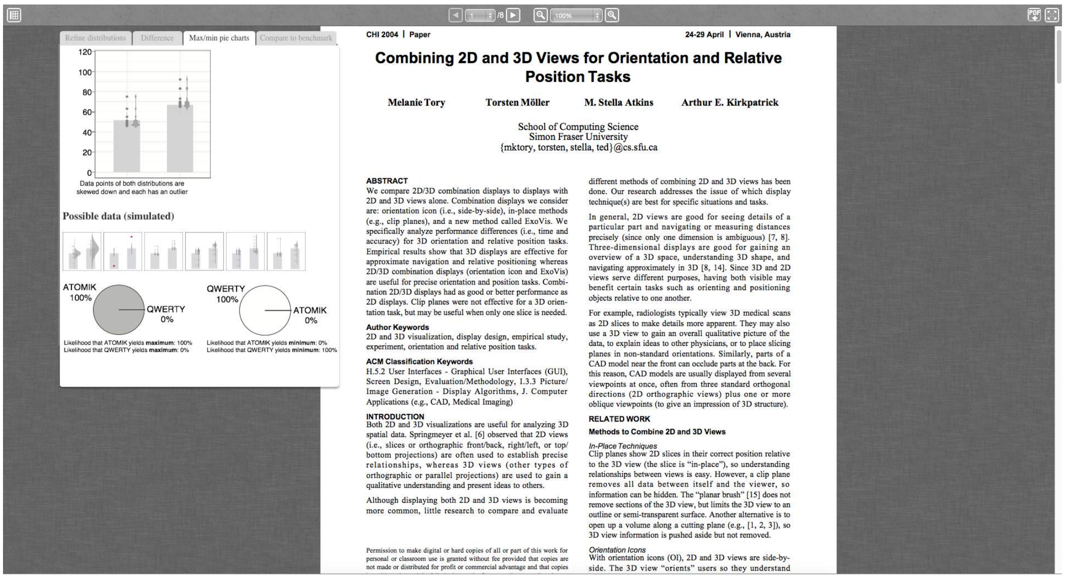
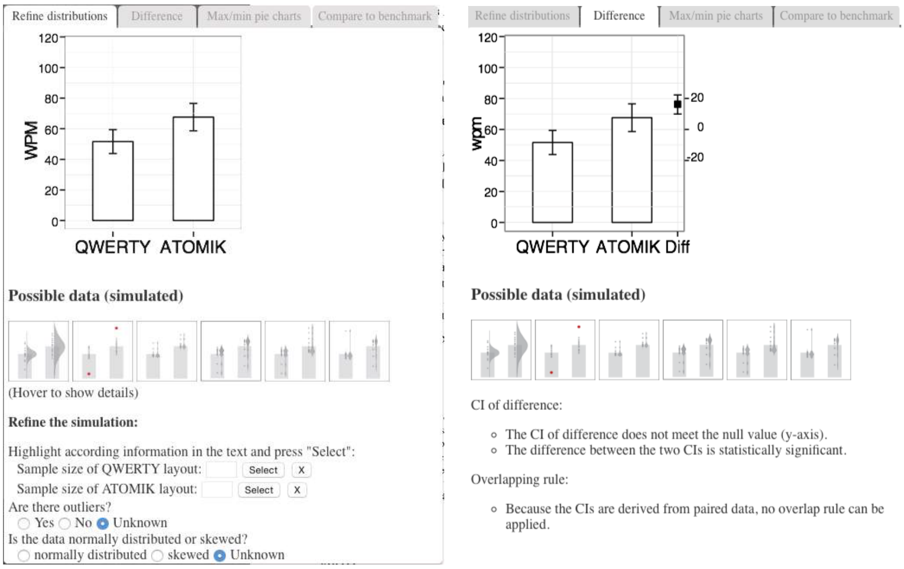


Figure 3.15: Second web-based prototype: Overall appearance.

tabs, the CI of difference (see figure 3.16 (b)), the pie charts (see figure 3.17 (a)), and the benchmark (see figure 3.17 (b)) can be seen separately. Additional information are given below the visualizations that shall help the user to understand and interpret the visualizations correctly.

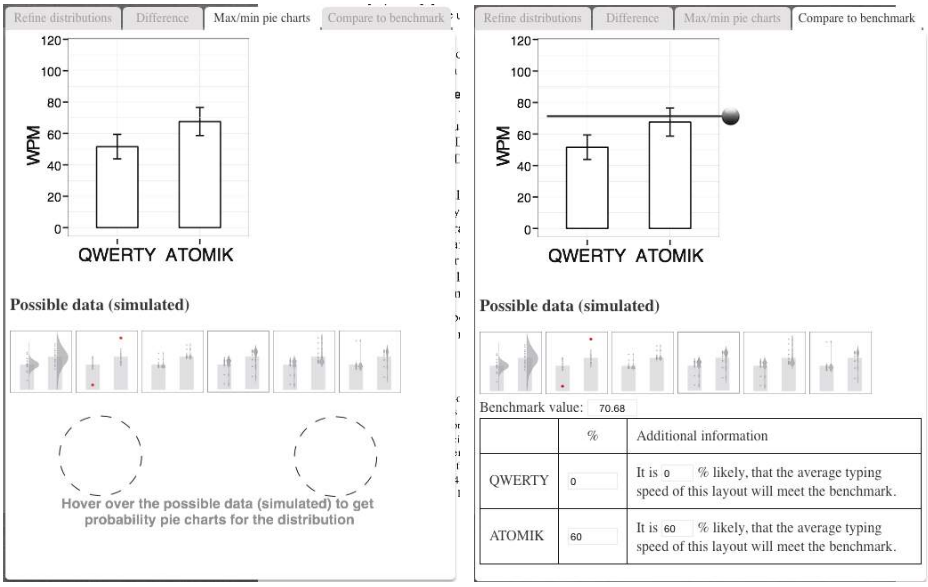
3.3.4 Iteration in design

The following table 3.1 “Changes during iterations.” presents all changes that were made when an iteration from one to the next prototype occurred. The reasons, why these changes were made are explained in the last column.
In chapter 4.4.6 “Final changes”, final changes will be mentioned, that are done after the user study.



(a) Interactive refinement (b) CI of difference

Figure 3.16: Second web-based prototype: Interactive refinement (left) and CI of difference (right).



(a) Pie charts (b) Benchmark

Figure 3.17: Second web-based prototype: Pie charts for probability distribution (left) and benchmark comparison (right).

Table 3.1: Changes during iterations.

Paper prototype → 1st Keynote prototype

Changes	Reasons for changes
Add formulas for transformation SE to SD to CI.	Get more precise charts which is necessary for later system.
Transformation from scatter plot to pie chart (not other direction anymore).	Pie charts get constructed from data points in scatter plot (reverse engineering).
Add several scatter plots with different distributions (normal dist., outliers, one example skewed in different/ same direction).	Shows different possibilities of how the data points for the CI could look like.
Histogram is fully constructed firstly for one CI, secondly for the next CI.	Viewer pays full attention to one CI.
User must not decide, whether it is paired or independent data anymore.	User with less understanding of statistics or CIs does not have any problem in this case.

1st Keynote prototype → 2nd Keynote prototype

Changes	Reasons for changes
Corresponding paper is always seen in screen.	User has a reference and can see corresponding text. Therefore, he is able to apply more information to the chart.
Highlighting corresponding parts in the text narrows down the number of possible distributions.	User only sees plausible distributions which helps him to understand the CIs.
Refined and shortened navigation bar.	Easier for the user to read and to follow.

2nd Keynote prototype → 1st web-based prototype

Changes	Reasons for changes
Histogram transformation deleted.	This transformation should only show the distribution of the data points. As this is now done with different forms of the violin plot, this transformation is not necessary anymore.
Red dots for outliers.	Tufte [1991]: Small spots of intense and saturated color for carrying information help to highlight important parts.
Violin plot for normal distribution.	Correll and Gleicher [2014]: Visual metaphor that values become less likely the bigger the distance from the mean is.
Violin plot with scaled kernel density estimation for skewed distributions.	Shows detailed distribution of all data points. The bigger the violin plot is, the more data points are in that area.
Large single and tasks on the left, paper in the middle and small multiples on the right.	Have small multiples in thumbnail form without hiding parts of the paper.

1st web-based prototype → 2nd web-based prototype

Reasons for changes	
Panel with several tabs, each with a separate interaction task.	This separation shall help the user to focus on one task at time.
Small multiples are located below the large single.	Results of interactive refinement can be directly seen above. No visual separation anymore helps the user to link the action of refinement with the small multiples.
Radio buttons for outlier detection and distribution.	Faster interaction.

Chapter 4

Evaluation

In the last chapter, the interaction, the visualization, as well as several iterations were explained. Now, the system will be evaluated to see whether it is really an improvement in understanding and interpreting statistical charts, compared to the interpretation, the user would make when reading the paper with the chart. In section 4.1 "Research question", the focus will be on the concrete research questions, that will be answered with this evaluation. The whole experiment design and procedure will be explained in the subsequent section 4.2 "Design and procedure". At first, the tasks will be named that the user will do during the study. Then, the experimental design with dependent and independent variables will be described before entering the procedure of the whole experiment. The next section 4.3 "Participants", shortly goes into detail, which people participated in the study. The last section shows the results of the user study. The research questions are answered and quantitative results are mentioned. At the end, a final table is given with final changes. The problems which resulted in the study are given as well as an explanation why the final changes will solve these problems.

4.1 Research question

As discussed in chapter 1 "Introduction", CIs are hard to understand and to interpret; even for experts. As the system SimCIs shall improve the understanding and interpretation of statistical charts with in situ simulation the following research questions arise:

Does the interactive visualization of SimCIs help the user...

1. ... reaching the right decision?
2. ... feeling confident in his decision?
3. ... learning how to interpret CIs?

To answer the *first research question*, the answers of the participants will be compared with the correct solution. This results in a test score which are given in percent. 0% means that the answer for this data analysis question is incorrect and 100% means that the answer is correct. This score is needed for some data analysis questions which do not have a binary answering scheme. This will be explained in more detail in section 4.2.1 "Experiment data analysis question".

The *second research question* will be rated with a confidence score ranging from 0 (not confident) to 5 (very confident). This shall indicate whether the chosen simulations and visualizations really helped the user to understand the answer and being confident or not.

The *third research question* can be answered with the quantitative results as well as compared test scores and confidence scores. Here, the procedure of the experiment is important. More information will be given in section 4.2.4 "Procedure".

4.2 Design and procedure

In this section, the whole process of the design will be explained, which yield to results that answer the research questions.

Table 4.1: Tasks from Amar et al. [2005] (A) and Ferreira et al. [2014] (F). Bold ones will be used for this experiment.

Tasks	Example	Source
Retrieve Value	Name concrete values like UL, LL, or mean	A
Filter	Compare CIs to a constant	A, F
Compute Derived Value	Name mean of difference	A
Find Extreme	Identify minimum/ maximum	A, F
Sort	Rank CIs	A, F
Determine Range	Compare CIs to a range	A, F
Characterize Distribution	Distribution of data points in small multiples	A
Find Anomalies	Detect outliers	A
Cluster	Find a cluster of data values	A
Correlate	Find correlation between CIs	A
Compare	Compare CIs to each other	F

4.2.1 Experiment data analysis question

Several data analysis question are mentioned that base on uncertainty and information visualization.

To be able to answer the research questions, the users have to answer several data analysis questions based on uncertainty and information visualization. Amar et al. [2005] already provide ten low level components of analytic activity for information visualizations. They reach from simpler value checking tasks to smaller analytical comparison tasks. Ferreira et al. [2014] took these tasks as a base and formulated five tasks which focus on uncertainty visualizations. These tasks can be seen in table 4.1 "Tasks from Amar et al. [2005] (A) and Ferreira et al. [2014] (F). Bold ones will

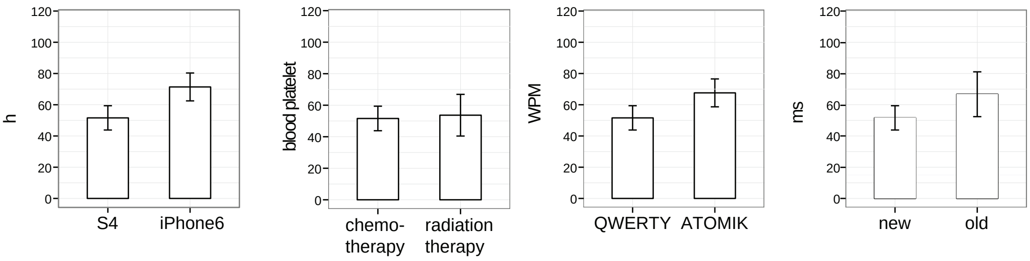


Figure 4.1: Four different charts which were used in the user study.

be used for this experiment.".

Data analysis questions in the user study.

The bold tasks show the ones which will be used as data analysis questions for this evaluation. The task to retrieve a value is provided by the system, but as this task of reading labels is very easy, it will not be tested. As other tasks are very familiar with tasks that will be used in the evaluation, like determining range and filter (determining constant), only one representative is chosen. For a CI chart, the user had to answer the following data analysis questions:

1. Is there a significant difference between the both CIs?
2. How is the probability for both CIs to be the maximum distribution?
3. How likely is it for both CIs to meet the benchmark?

4.2.2 Experiment design

Within-subject design testing four charts with paper and the system.

We use the within-subject design with the analysis method as the independent variable. Each participant analyzed two charts on paper and two with SimCIs. For each chart, he answered three data analysis questions, in order to see, whether SimCIs complements traditional CI interpretation on the paper itself. In total, each participant answered 12 data analysis questions.

4.2.3 Stimuli

The participants got four different CI charts.

The participants got four different charts like it can be seen in figure 4.1. This figure shows the order in which the participants saw the charts. The leftmost chart shows a positive gab between the intervals, the next has an overlap of the smaller CI, the third one has an overlap of less than the half of the smallest MOE, and the last one has an overlap of more than half of the smallest MOE. The first two charts were derived from independent data and the other two charts from paired data. In general, there are two difficulty levels in the data analysis questions which arise from the chart representation. A paper was changed according to these four different charts with its textual information to provide comparable papers with the same difficulty level. The according text, which the users got with these charts are given in appendix C "Additional information of the user study".

Firstly, CI charts from independent data are tested using the paper and then the system.

After the participant answered the question of influencing CI factors, he got a paper with the chart and several data analysis questions to answer (see final data analysis questions in C "Additional information of the user study"). For each answer, the user should also rate how confident he was with his answer. This should help to see how well he already understands CIs and how well he can interpret the uncertainty. After he finished answering these data analysis questions with the paper, he got a new chart and similar data analysis questions which he should solve using the system SimCIs. As he had to do with the paper, he answered similar data analysis questions and rate the confidence. This should help to see whether the user feels more confident using the system and whether the user feels more confident when giving correct answers. These first two charts showed CIs, which are raised from independent data. They are simpler, as overlapping rules can be applied to rate the difference between the CIs in a chart for example.

Then, CI charts from paired data are tested using the system and then the paper.

After the participant used the system the first time, he was asked again which factors influence CIs. This shall help to see whether the different small multiples, and the tasks during the interactive refinement are understood from the participant and therefore in his mind, also when not interacting with the system. The next chart, which should also be analyzed using the system again, showed CIs arose from paired data. As the user could also read from the system, the overlapping rules cannot be applied here and the user has to focus on the CI of difference instead. Again, he had to answer the same kind of questions and rate his confidence. The fourth chart showing two CIs based on paired data was only in paper form. Answering these data analysis questions and the rating the confidence can give insights of what the user learned from the system and whether he could apply it.

4.2.4 Procedure

The participants had to answer general questions at the beginning and at the end of the study.

After the user signed a consent form, he was asked which factors influence CIs. This should give a first impression of the basic statistical knowledge of CIs. Then the participants answered all data analysis questions using the paper or the system. The order of the charts with their data analysis questions stayed the same and therefore, the same charts were used for paper analysis and the same for software analysis. This was only possible, because all charts had comparable difficulty levels. At the end, the participants had to fill out a short questionnaire to check their statistical knowledge. This questionnaire was given at the end to prevent the user from taking these information when answering the question which factors influence CIs and when answering the data analysis questions of the first chart, only with the help of the paper. I refined this procedure by pilot studies with three participants. In the user study, each session took 45 minutes on average.

4.3 Participants

Nine users between 22 and 26 participated in the study.

The participants were students between the age of 22 and 36. The mean age was 27. Nine users in total participated on the study. All of them were students in computer science and machine engineering. Only 50% of the participants took basic statistic curses, or curses which included statistics.

The participants did not work with statistics regularly.

All participants did not work with statistical charts that often. They only created 2 charts and interpreted 7 charts on average during the last six month. None of them worked with skewed data and they only interpreted 6 CIs on average and created 0 of them. They also only interpreted 8 bar charts and created 1 on average. These results indicate, that the participants are not that into statistics topics and therefore do not count as experts. This was the user group I expected to use my system in oder to see whether it also helps people with little statistical knowledge and people, who are not that used to interpret these kind of uncertainty charts.

4.4 Results

At first, 100% of the participants read the text before checking the data analysis questions. Using the system, they firstly explored all tabs, hovered over the small multiple distributions and over the chart, and checked the interactions possible in each tab. After that, they started answering the questions.

All participants read the text first and then explored the system.

The participants needed 45 minutes for the whole procedure. 36 minutes of that time was spend on the chart data analysis questions. In general, it took them less time to interpret the charts using the paper (14 minutes) than using the system (22 minutes). As the user needed time to inspect the system the first time, it took them 15 minutes on average to answer these data analysis questions. When the participants interpreted the third chart with the system, it took them only 6 minutes on average, which is the same result as for the last data analysis question with the paper. Whether the time changed something on the accuracy of the answers or the confidence with their answers will be explained in the following.

In general, the participants spend more time on the data analysis questions when using the system.

Overall, the participants got a test score of 83.2% using the paper and 93.3% using the system. The confidence answering the data analysis questions was 3.2 using the paper and 4.1 using the system. These numbers range from 1 (not confident) to 5 (very confident) as they were asked to answer in a Likert scale. Concrete values of the test scores and the

Using the system results in 11.8% higher test score and 0.9 higher confidence level.

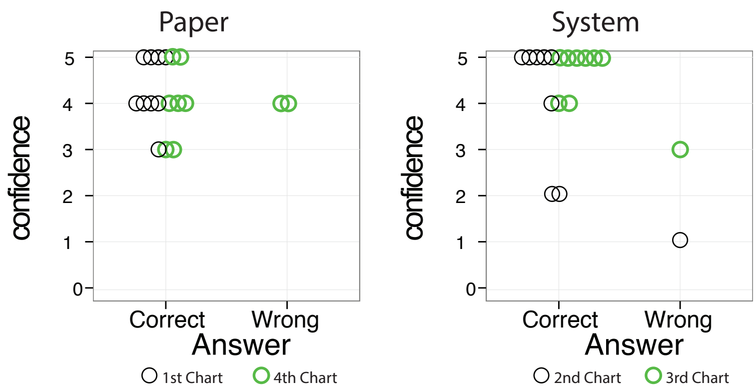


Figure 4.2: Answers and the according confidence for the first data analysis question. Left: Answers using paper; Right: Answers using system

confidence levels for different data analysis questions will be explained in the following.

4.4.1 Question 1 - Significant difference

Answer significant difference with the overlap rule or by calculating the CI of difference.

The first data analysis question compared the two CIs to see whether there is a significant difference or not. For the paper there was the possibility to use the overlap rule for independent data, and to calculate the CI of difference for paired data and independent data. No information about these possibilities were given so that they could only do that when they have former knowledge. The system offered these solutions in the second tab.

The system and the paper interaction got the same test score and confidence level.

Both, paper and SimCIs yielded the same accuracy: 88.9% of the answers were correct. With the system they had the same test score. The users also had the same confidence level of 4.2. However, the participants tend to be more confident when using the system (figure 4.2 (right)). This indicates the positive impact of the system. The outliers in the paper condition as well as in the system condition in figure 4.2 resulted from different participants. As each outlier in the system condition resulted from another chart, they have no deeper meaning. In the paper condition, the out-

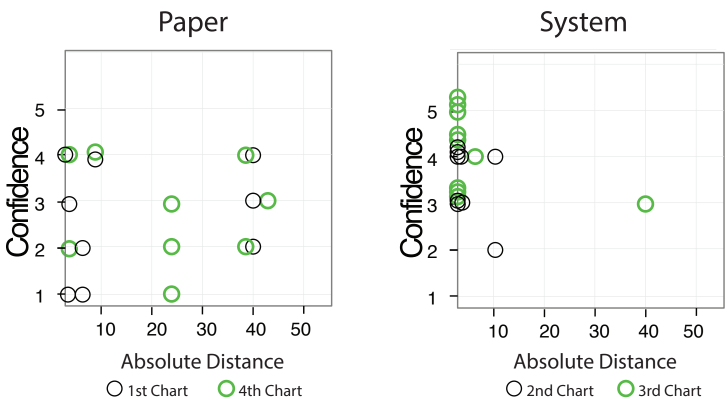


Figure 4.3: Absolute distance of the answers to the correct answer and the according confidence for the second data analysis question. Left: Answers using paper; Right: Answers using system

liers came from the same chart. This correlates with the underlying data and the knowledge of the users. The last chart arose from paired data, where no overlap rule can be applied, but the users made their decision because of the overlap rule. Therefore, it might be good to improve the system in a way, so that the user is aware of not applying the overlapping rule for paired data.

4.4.2 Question 2 - Proportion to be maximum distribution

Indicate proportion to be maximum distribution by first data analysis question and using bootstrapping.

In the second data analysis question, the participants had to name the proportion of each possible distribution to be the maximum one. As this results from bootstrapping, it is hard to guess without the system. In general, the user could know, that when the difference is significant (data analysis question 1), that the proportion is nearly to 0%/100%. With the system, the user can read the answer from the pie charts in the third tab. Here, it is important, that the user refined the possible charts with information in the text, to have less possible distribution and therefore a more accurate value.

The system got a higher test score and a higher confidence level.

Using the paper, the participants achieved a test score of

87.8%. With the system, the test score was 96.2%. The test score was calculated by subtracting the mean of the absolute difference from 100% total accuracy. The mean of absolute difference was calculated for the paper and the system condition. The mean of difference for each answer is given in figure 4.3. On the left the difference for the answers using the paper are shown, on the right, the difference of the answers using the system are provided. Black and green circles show which answers belong to which charts. The mean of difference is calculated with the following formula: $\frac{\sum_{j \in Users} \sum_{i \in Charts} |a_{i,j} - k_{i,j}|}{||Users|| \times ||Charts||,}$ where $a_{i,j}$ are the answers from the user j to the question in chart i and $k_{i,j}$ is the correct answer of the chart i. As it can be seen here, the resulting average difference using the system is smaller than using the paper. Therefore, the system helps the user to give more accurate answers. Furthermore, the participants were more confident using the system (3.7) than using the paper (2.7). In the chart on the left, showing the results for the interaction with the paper, all answers were spread in distance to the correct answer and confidence. This indicates, that there is no correlation between the accuracy and the confidence. As it can be seen for the interaction with the system, people were more accurate and more confident. When being less accurate, the users weren't that confident. This indicates a correlation of accuracy and confidence when interacting with the system. All data points are relatively near to each other, but two outliers can be seen when using the system. These two outliers came from the same person. He had problems to find the right visualization to answer this task. Therefore, they were less accurate and also less confident. For more information about this problem, see 4.4.4 "Qualitative results". 22.2% of the participants did not refine the possible distributions and were less accurate. This explains the little spread of data points, as it can be seen on the confidence level 4. However, these results came from different participants.

4.4.3 Question 3 - Compare to a benchmark

Compare benchmark to CI by measuring the proportion between the mean and the limit where the benchmark is set.

In the last data analysis question, the user had compare the CIs to a benchmark. When using the paper the user can

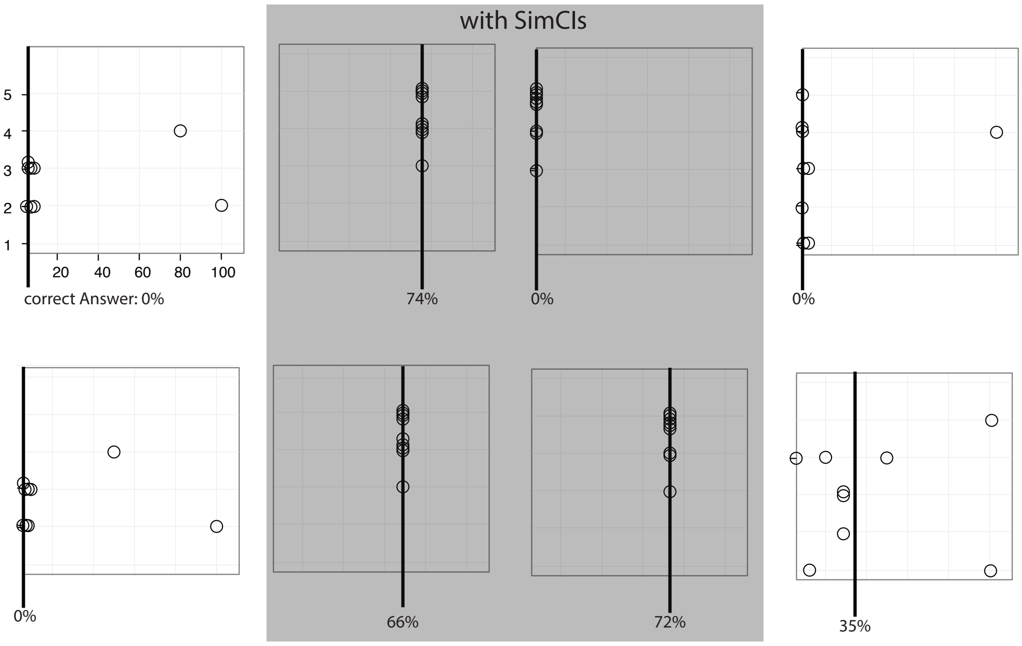


Figure 4.4: Answers compared to the correct answer and the according confidence for the third data analysis question. Left and right: Answers using paper; Middle: Answers using system

compare the benchmark with the confidence interval. If the benchmark does not meet the CI at all, the result will be 0. When the benchmark lies within the confidence interval, the exact position between the mean and the limit has to be measured. The values range from 100% at the mean to 0% next to the limits of the interval. As this data analysis question is possible to solve in the paper condition, the accuracy can be higher when using the system. With the system, the user can move the benchmark to the exact position and read the answers directly from the chart.

Participants achieved the highest difference with this data analysis question using the system.

While the participants achieved a test score of 79.8% using the paper and 100% using the system. This is the highest difference compared to all data analysis questions with the highest test score. Also the confidence is the highest using the system with a confidence level of 4.4. The confidence by answering the data analysis question without the system achieved only a level of 2.8. The answers compared to the correct result and the confidence with the answer for all charts are shown in figure 4.4. The results of the first chart are shown on the left, and the results of the last chart are shown on the right. The top row shows the results from the left CI in each chart, and the bottom row shows the results from the right CI in each chart. Although the confidence varied using the system, all users answered this question correctly. For the interaction with the paper, no correlation can be seen between the accuracy and confidence, like it was for the second data analysis question.

Outliers indicate a learning effect.

For the charts, which were interpreted using the paper, the outliers give some insight. Each result of the first chart shows two outliers. These outliers resulted from two participants. However, using the the system helped both participants to understand the principle of answering this question correctly. Therefore, one participant answered the question correctly for the last chart, where he had to interact with the paper again. The other participant showed incorrect results again. However, this did not result from an incorrect understanding of the principle to answer this question. Here, the user falsely understood the question itself, which result in outliers for the last chart. According to the way he understood the question, the answer was correct, so that these outliers do not need to be regarded. The other outlier for the last chart occurred, because one participant did not fully understand the principle to answer the question. He thought that it is 100% likely that the benchmark will meet the CI independent from the actual position of the benchmark in the interval. The spread of answers for the second CI of the last chart result from the difficulty level of the question. As the user had a benchmark which met one CI in the last chart, he had to calculate the percentage according to the small chart on a paper.

4.4.4 Qualitative results

78% of the participants had improved knowledge of factors that influence CIs after using the system.

In general, the participants learned more about CIs during the study. After they used the system the first time, 78% of the participants were more precise in their answers which factors can influence CIs. Because of the system they also mentioned other aspects, they did not thought of before. Only 22% of the participants answered the same factors at the beginning and at the end. These participants had already a deeper knowledge in statistics and therefore gave more precise answers. As the answers of all participants were differently at the beginning and ranged from "variance" to "population", 100% of the participants mentioned different ways how the distribution could look like after using the system. They named outliers, normal distributions and skewness. When they hadn't mentioned the impact of sample size at the beginning, they did it after answering the system data analysis questions.

Difficulties appeared during the text highlighting in interactive refinement and answering of the second data analysis question.

During the refinement tasks, the participants tried to select the sample size once and highlight them both by clicking on both buttons one after another. This didn't work, as the selection got removed after the text part was highlighted. Therefore, this part could get further improvement. Another difficulty in the interaction with the system was to interact with the pie charts. When the users first read the second data analysis question, where they have to give the proportion of both conditions to be the maximum distribution, they weren't sure how to answer the question. Firstly, most of the participants tried to answer the question with the benchmark tab. Therefore they moved the benchmark, to get probable values. As this did not work for them they needed some time to find the correct visualization. 77.8% of the participants used the pie charts after some time of searching. However, one participants did not connect this tab to the data analysis question. After the study he answered, that he did not thought of getting concrete values for all of the data analysis questions and that he thought to give that information from the context of the given information in the other tab. These participants could not connect the information given in that tab with the question. They were not that confident with their answer, which indicates that the system supports less confidence when the answer is wrong. Though, this pie chart representation should need further investigation. How this could be done is explained in 5.2 "Future work".

The participants liked to work with the system and would use it again.

All in all, the users had fun during the interaction with the system. They were more encouraged to understand each data analysis question and give an accurate answer. When interacting with the paper after using the system, 100% of the participants mentioned that they would like to have the system again to be more accurate in answering the data analysis questions.

4.4.5 Limitation

Only a small number of users participated in the study.

The biggest limitation in this user study is the small number of participants. As statistics are hard to understand and seem to be difficulty for most people, it was hard to motivate users to participate in the study. As this limited number of participants was clear after scheduling the study, there was also the need of a fixed setup in experiment design (see 4.2.2 “Experiment design”). Therefore, the order of paper and system interaction, the order of the charts, and the order of the questions were fixed. This should help to get comparable results also with a small number of participants. However, this results in no counterbalancing.

Potential bias with the questions of statistical knowledge.

Another limitation could be found in the way how the statistical knowledge of the participants was questioned. According to the information of the questions, the users did not worked with a lot of statistical charts and only a few participants took basic statistical courses. Therefore, it could be assumed that there is only little statistical knowledge. However, some participants showed deeper knowledge when interacting with the charts, as they could explain their answers with the knowledge of the overlap rule. This indicates, that there should have been more questions about the statistical knowledge.

Data analysis question might be too easy.

As the participants had this deeper knowledge, the data analysis questions might have been too easy. Especially the first data analysis question achieved high accuracy results by the paper interaction. There was no improvement in accuracy for this analysis question using the system.

4.4.6 Final changes

According to the notes the user made, the following table of problems and changes is created (see table 4.2 “Final

Table 4.2: Final changes of the design.

Problem	Changes	Reason why it solves problem
It is hard to get precise benchmark.	Benchmark can be typed in text field.	When the user can also type a number in the text field, he can get the precise values he wants.
Purpose of ”clear” button in the refinement tab wasn’t that clear.	Changed from ”X” to ”Clear”.	As interaction with the select button was no problem for the users, the same should work for the clear button.

changes of the design.”). All of these changes are implemented and in the final system. Other things that need further investigation will be explained in detail in chapter 5.2 “Future work”.

Chapter 5

Summary and future work

In the last chapters, SimCIs was designed based on the knowledge and insights from previous work. To evaluate the systems impact on understanding and interpreting CIs, we made a user study with nine participants. The next two sections will summarize the findings while evaluating the system as well as open tasks that could be done as future work.

5.1 Summary and contributions

Whether SimCIs improves the understanding and interpretation of statistical charts was tested.

In this thesis, we wanted to develop a system, which improves the understanding and interpretation of statistical charts with in situ simulation. Therefore, the system goal was to help the user in being more accurate and confident in their interpretation and therefore in their answers to analytic data questions. To be confident in their answers, the system should help the users to understand the principles of interpretation and therefore learn more about statistical analysis. Data points are simulated and different interactive visualizations were applied to accomplish these goals. To investigate whether the system meets these points, we made an experiment.

Users were more accurate and more confident using the system.

As the results indicated, people were more confident with correct answers when using the system than only having a paper to analyze the CI chart. Therefore, the first two goals (accuracy and confidence) are met. For the first data analysis question, checking whether CIs are significant different, there was no higher accuracy using the system. In general, the users gave correct answers when interacting with the paper as well as interacting with the system. However, having a more complex analytic question like the second data analysis question, where the user has to get the proportion of both CIs to be the maximum distribution, shows the great impact of the system. Here, the users were more accurate using the system and even more confident. When answering this data analysis question only having the paper, the users answered more inaccurate and there was no correlation between the accuracy and their confidence. Similar results can be seen for the third data analysis question, when comparing CIs to a benchmark. As the users were 100% accurate using the system and their confidence was higher, the accuracy as well as the confidence was more mixed when answering the question only with the paper. For this question, a learning effect could be seen. Therefore, the system also tends to teach some principles for statistical analysis. However, as this result was only shown for the third question, we only partly accomplished that goal.

As they all learned more about the meaning of CIs and and had fun using the system, the users would like to use the system again, when having to analyze CI charts. However, there were some aspects of the system, which could have been better implemented. During the evaluation, points arose, which could be added to the system for greater impact. Furthermore, the evaluation had some limitation which could be addressed in future work. These things and more aspects for future work will be explained in the following.

5.2 Future work

According to the results of the study and the user comments, there are three domains where future work could be interesting.

5.2.1 Improving the system to meet all goals

As the evaluation of the system indicated, there is further improvement necessary to meet all goals (to help the user in being more accurate and confident in their interpretation and therefore in their answers and to learn the principles of statistical analysis). Especially the third goal, learning the principles of statistical analysis to know how to interpret CIs, needs further investigation. One possibility would be to make UI improvements, as it will discussed in the next section. However, also different methods could be added to help improving that goal. For example, transmogrification (see 2.1.2 "Interactive visualizations") could be added in all visualizations. This could help to understand how one visualization results from another one and therefore, it is especially interesting for the pie chart representation.

Methods like transmogrification could improve the third goal.

5.2.2 Improvements of the UI

As it was easy for the users to understand the principle of comparing the CI to a benchmark, the other data analysis questions were more difficult for them. The following bullet points will give a short insight of which changes can help to improve the UI:

Bullet points provide ideas of UI improvements.

- Textual refinement
  - Add sliders to set the sample size. This helps the user to understand the impact of the sample size for the CIs.
- CI of difference
  - Information button could show the formula to calculate the CI of difference and therefore improve the understanding of statistical analysis.

- Proportion of CIs in pie charts
  - Information button provides information about bootstrapping. Understanding this principle could help the user with the second and third goal.
  - Change naming of the tab and the explanation text so that the user connects this visualization with the second task.

Add possibility to annotate and download visualizations.

One thing that wasn't integrated in the system at all was to provide techniques for textual or free hand annotations. As mentioned in section 2.4.3 "Process and provenance", these annotations allow the user to create arrows and other reference marks. When the user could then download all visualizations with the according annotations, he could share these results with others to discuss it or revise his findings when looking at it again after some time.

5.2.3 Further user studies

Further studies to test changes in UI and compare SimCIs with other systems.

Future studies could check, whether the improvements in the UI helps to accomplish the goals. Furthermore, the effect of SimCIs could be compared against other existing software, which helps to interpret CIs. Here, one group could test SimCIs and another group could test a software, like the one conducted by Ferreira et al. [2014]. After the each group used one system, it could be checked which system leads to more accurate results and confidence in their answers. It would also be interesting to see which system is easier to use and help to understand statistical analysis in general. Testing the system against a statistical book, like the one from Cumming [2012], could also give insights that could improve the understanding of principles of statistical analysis.

Appendix A

Mathematics behind the system

This appendix shows the underlying calculations of the system in detail.

A.1 Calculation from different input charts to 95% CI

In the following sections the calculations from the underlying data to an SD, to an SE, to the final 95% CI are shown. These steps are also necessary, when calculating the 95% CI from data in general.

A.1.1 Transformation from the underlying data to SD

When having the underlying data, an SD can easily be calculated. The mean can be calculated with the following formula:
$mean = \sum(\frac{datavalues}{samplesize})$
The variance can be calculated with the formula
$variance = \frac{\sum(datavalues - mean)^2}{(samplesize - 1)}$.

The square root of the variance then results in the standard deviation. When only having a bar chart, this deviation could be added to get a SD chart.

A.1.2 Transformation from SD to SE

The next step is the transformation from an SD chart to an SE chart. To get this chart, the sample size has to be applied to the SD interval with the following formula:
$SE = \frac{SD}{\sqrt{samplesize}}$
The sample size is given in the underlying data, or, if the underlying data is not available, can be provided in the paper. However, this means that the sample size must be provided. Otherwise this transformation is not possible.

A.1.3 Transformation from SE to 95% CI

As the last step there is the transformation from SE to CI chart. Therefore, the SE value has to be multiplied with the critical value of t. For 95% CI, which is the normally used CI, and for a sample size equal or larger 10, this critical value is 1.96. This results in the following formula:
$CI = 1.96 \times SE$

A.2 Calculation of the local outlier factor

The following steps have to be done to get the outlier factor of each data point in a distribution.

1. Calculate distances $d(p, o)$ between each two objects p and o in D.
2. Set k-distance of object p.
   The positive integer k is defined as the distance between p and an object o in D such that:
   for at least k objects $o'D/\{p\}$ it holds that $d(p, o') \leq$

$d(p,o)$
for at most $k-1$ objects $o'D/\{p\}$ it holds that $d(p,o') <$ $d(p,o)$

3. Calculate distance $k - dist(p)$ from p to k^{th} nearest neighbor.

4. Get k-distance neighborhood of object p.
Here, $N_k(p)$ contains all objects which are in a k-distance of p such that
$N_k(p) = \{q \in D/\{p\} | d(p,q) \leq k - dist(p)\}$

5. Calculate reachability distance $reach - distk(p,o)$ of an object p with respect to o with
$reach - dist_k(p,o) = max\{k - dist(o), d(p,o)\}$

6. Calculate local reachability density of an object p with
$lrd_{MinPts}(p) = \frac{|N_{MinPts}(p)|}{\sum_{\sigma \in N_{MinPts}(p)} reach-dist_{MinPts}(p,o)}$,
where MinPts specifies a minimum number of objects and is a parameter to define the density

7. Calculate local outlier factor of an object p:
$LOF_{MinPts}(p) = \sum_{\sigma \in N_{MinPts}(p)} (\frac{lrd_{MinPts}(o)}{|N_{MinPts}(p)|}$
$= \sum_{\sigma \in N_{MinPts}(p)} (lrd_k(o) \times \sum_{\sigma \in N_{MinPts}(p)} (reach - dist_k(o,p))$

A.3 Calculating the CI of difference

When calculating the CI of difference it is important to know whether the underlying data is paired or independent (see Cumming and Finch [2005]). As paired data results from one user group who does a pre- and posttest for example, independent data arises from two different groups doing the same experiment with one or several different conditions to compare them later on.
For independent and paired data, the mean M_d is calculated by subtracting the second mean from the first mean. The margin of error MOE is calculated differently.
For independent data, MOE_D is calculated by using the sample sizes N_1 and N_2 of both CIs as well as the samples

s_1 and s_2 as an estimate of the population:
$MOE_D = t_{.95}(N_1 + N_2 - 2) \times s \times (1N_1) + (1N_2)$
, where $s : ((N_1 - 1) \times s_1^2 + (N_2 - 1) \times s_2^2)(N_1 + N_2 - 2)$
The value of $t_{.95}$ depends on the degree of freedom and can be read from a table like in this link[1]. The CI of difference is larger than either of the original intervals, because the sampling error in the difference is a compounding of sampling error from each of the independent means. It results in the following formula for the interval estimate of the difference between the population means (B-A):
$M_d - MOE_D, M_d + MOE_D$
For paired data, the MOE_d is calculated based on the SE of the differences. This means that the MOE of difference can be calculated as it is described for CIs from independent data, but have to be divided through 1.96 at the end.

[1] http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf

Appendix B

Code fragments in R

B.1 Generate the local outlier factor

This section focuses on the generation[1] of the local outlier score, its normalization, and the calculation of the size of the data points which will be drawn in the chart.

```
library(DMwR)
library(reshape2)
library(ggplot2)

a.data <- c(10, 43, 51, 45, 61, 65, 60, 58, 42, 55, 48, 68, 51, 53, 64)
d.data <- c(50, 90, 80, 70, 60, 50, 40, 30, 20, 10, 65, 90, 55, 60, 35)

# Calculate LOF for each distribution
a.data.lof <- lofactor(a.data, 2) # 2 is the kth-distances to be
                                  # used to calculate LOFs
d.data.lof <- lofactor(d.data, 2)

# Normalisation for each distribution
range01 <- function(x){(x-min(x))/(max(x)-min(x))}

a.data.lof <- range01(a.data.lof)
d.data.lof <- range01(d.data.lof)

# calculate size if data points (normal size is 3)
```

[1] http://www.inside-r.org/packages/cran/DMwR/docs/lofactor

```
a.pointsize <- matrix(0, NROW(a.data))
d.pointsize <- matrix(0, NROW(d.data))

for (i in 1:NROW(a.data)){
  a.pointsize[i] = 3+(a.data.lof[i]*3)
}

for (i in 1:NROW(d.data)){
  d.pointsize[i] = 3+(d.data.lof[i]*3)
}
```

B.2 Bootstrapping to generate probability pie charts

This section shows code in R which bootstraps two distributions and visualizes the outcome in two pie charts showing the maximum probability and the minimum probability.

```
boot <- rep(0,10000)
  a.data <- c(10, 43, 51, 45, 61, 65, 60, 58, 42,
    55, 48, 68, 51, 53, 64)
  d.data <- c(50, 90, 80, 70, 60, 50, 40, 30, 20,
    10, 65, 90, 55, 60, 35)
  aHigher <- 0
  dHigher <- 0
     for(i in 1:10000)
     {
     a <- sample(a.data,replace=T)
     d <- sample(d.data,replace=T)
     if(mean(a)>mean(d)){
        aHigher <- aHigher + 1
      }
      else if(mean(d)>mean(a)){
        dHigher <- dHigher + 1
      }
     }

# Draw maximum CI probability pie chart
slices <- c(aHigher, dHigher)
lbls <- c("C", "D")
```

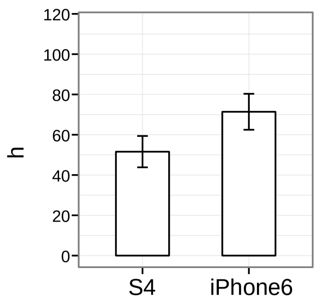
```
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices,labels = lbls, col= c("white", "gray"),
main="Maximum CI probability")

# Draw minimum CI probability pie chart
slices <- c(dHigher, aHigher)
lbls <- c("C", "D")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices,labels = lbls, col= c("white", "gray"),
main="Minimum CI probability")
```

Appendix C

Additional information of the user study

The following figures show the information provided in the paper for each chart as well as the final questions. On the left of each figure, the chart is shown, in the middle the corresponding text, and on the right the specific questions for all charts. Figure C.1, figure C.2, figure C.3, and figure C.4 are shown in the order as they were given to the participants.



Experiment
The companies of Android Galaxy S4 Note and iPhone 6 promote the long battery life of their new models.
In the following experiment, the battery life of the Android Galaxy S4 Note was compared to the iPhone 6.
Both phones have an app installed to measure the battery life and log the measurements.

Participants
30 volunteers were recruited from various levels of different companies. All participants are in the same occupation.
Participants had varied levels of sophistication but worked the same amount of time. Mean age was 34.

Overall Procedure
The participants were randomly assigned to one of the models. 15 participants in each group tested the smartphone on five weekdays. 7 males and 8 females participated in each group.
During the time, they had to perform same tasks in the evening.
The order of the tasks was counterbalanced.

RESULTS
We analyzed our results using the t test.
The results are shown in Figure 4. Error bars represent 95% confidence interval.
The chart shows the battery life in hours.
The data are normally distributed. There are no outliers.

Chart 1 Participant's ID:
Date: Static or interactive:

1. Question: From this data, do you think that the battery life of iPhone 6 is notably better than the S4?

   ❍ Yes ❍ No

Rate how confident you are with your answer

completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

2. Question: How likely is it for the participants to have a better battery life with model S4? And how likely is it with iPhone6?

   S4: __________% iPhone 6: __________%

   Rate how confident you are with your answer

   completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

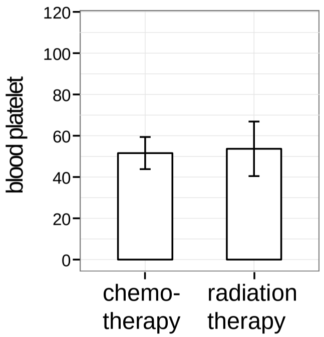
3. Question: How likely is it for S4 and iPhone 6 have battery life at **30 hours** on average? Try to answer as precise as possible by using the graph. (0%: very unlikely – 100%: very likely)

   S4: __________% iPhone 6: __________%

   Rate how confident you are with your answer

   completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

Figure C.1: Text in paper and questions of first chart.



Experiment
Chermotherapy treatment and radiation therapy are effective ways treating leukemia. To see which therapy is more effective in treating leukemia, a clinical trial was conducted.

Participants
30 volunteers were recruited, having various levels of illness severity. Some already had other tratments which did not work and they had to discontinue medications. Other participants started their the treatment during this study. Mean age was 42.

Overall Procedure
The participants were randomly assigned to one of the models.
The study with 15 participants in each group was conducted over one year.
During this time, every two weeks a blood sample was taken from the participants as well as before and after the treatment session. 7 males and 8 females participated in each group.

RESULTS
We analyzed our results using the t test.
The results can be seen in Figure 4. Error bars represent 95% confidence interval.
The vertical axis shows the blood platelet count in 10^3/microliter.
The higher the count is, the better the treatment worked.
By way of comparison, the standard count range from 150 to 400*10^3/microliter blood platelet.
the data are normally distributed without any outliers.

Chart 2
Date:

Participant's ID:
Static or interactive:

1. Question: From this data, do you think that chemotherapy treatment is more effective than radiation therapy?

❍ Yes ❍ No

Rate how confident you are with your answer

completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

2. Question: How likely would it be for the participants to have more blood platelets with the chemotherapy treatment? And how likely would it be with the radiation therapy?

chemotherapy treatment: __________% radiation therapy: __________%

Rate how confident you are with your answer

completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

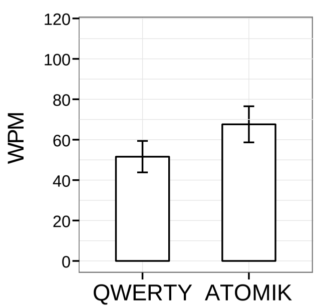
3. Question: How likely is it that each of the treatments will result in blood platelet count of 50? Try to answer as precise as possible by using the graph. (0%: very unlikely – 100%: very likely)

chemotherapy treatment: __________% radiation therapy: __________%

Rate how confident you are with your answer

completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

Figure C.2: Text in paper and questions of second chart.



Experiment
The new keyboard layout ATOMIK shall help to increase the typing speed. To see whether this layout results in a notable increase of words typed per minute (wpm), the new layout was compared to the QWERTY layout.
The typing speed (wpm) and the text is logged to measure the error rate afterwards.

Participants
15 volunteers were recruited from various levels of of professional typists as well as novice and intermediate typists.
Participants had varied levels of experience in typing. Mean age was 27.

Overall Procedure
The participants had to read several sentences one after another and then had to memorize them while typing. The within-subject design was chosen to get comparable results. The order, with which layout the participants worked first, was counterbalanced.
Six men and nine women participated in each group with changing order of sentences.
Participants completed all tasks with one layout (including follow-up questionnaires) before beginning the second layout. Experimental sessions lasted approximately 30 minutes.

RESULTS
We analyzed our results with the paired t test.
The results can be seen in Figure 4. Error bars represent the 95% confidence interval. The vertical axis shows the number of words per minute. Although the data are normally distributed, there are some outliers.

Chart 3
Date:

Participant's ID:
Static or interactive:

1. Question: From this data, do you think that the evidence indicates that the ATOMIK layout yields faster typing speed/ higher wpm?

   ❍ Yes ❍ No

Rate how confident you are with your answer

completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

2. Question: How likely is it for the participants to have a better WPM score with the ATOMIK layout? And how likely is it with the QWERTY layout?

   ATOMIK layout: __________% QWERTY layout: __________%

   Rate how confident you are with your answer

   completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

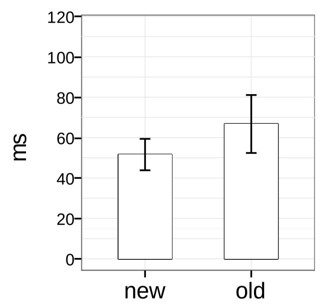
3. Question: The average typing speed is **48 WPM**. How likely is it for each keyboard layout to yield this speed? Try to answer as precise as possible by using the graph. (0%: very unlikely – 100%: very likely)

   ATOMIK layout: __________% QWERTY layout: __________%

   Rate how confident you are with your answer

   completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

Figure C.3: Text in paper and questions of third chart.



Experiment
In the new series of Opel, the speed acceleration was changed. To see whether this change results in quicker acceleration, the new model was tested against its older model.
Some electronics were built into the car to measure the speed acceleration and log the measurements.

Participants
15 volunteers were recruited from various levels of the professional race drivers and alumnus of the race driver education from LechneRacing.
Participants had varied levels of experience with races.
Mean age was 27.

Overall Procedure
All racers drove the same race course with both models.
The within-subject design was chosen to get comparable results. The ordner, with which model the racers drove first, was counterbalanced.
12 males and 3 females participated in this study.

Participants completed all tasks with one car (including follow-up questionnaires) before beginning the second task
Experimental sessions lasted approximately 45 minutes.

RESULTS
We analyzed our results with the paired t test.
The results are shown in Figure 4. The Error bars represent 95% confidence interval. The vertical axis shows the ms to reach 100 km/h.

Chart 4 | Participant's ID:
Date: | Static or interactive:

1. Question: From this data, do you think that this the new model is a considerable improvement?

   ❍ Yes ❍ No

   Rate how confident you are with your answer

   completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

2. Question: How likely is it for race drivers to achieve a better speed acceleration with the new model? And how likely is it with the old model?

   New: __________% Old: __________%

   Rate how confident you are with your answer

   completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

3. Question: Suppose an ideal value of the speed acceleration is 75 ms. For each model, how likely is it for the acceleration to be at 75 ms? Try to answer as precise as possible by using the graph. (0%: very unlikely – 100%: very likely)

   New: __________% Old: __________%

   Rate how confident you are with your answer

   completely uncertain ❍ ❍ ❍ ❍ ❍ completely certain

Figure C.4: Text in paper and questions of fourth chart.

Bibliography

R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 111–117, Oct 2005. doi: 10.1109/INFVIS.2005.1532136.

L. Bartram, B. Cheung, and M.C. Stone. The effect of colour and transparency on the perception of overlaid grids. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):1942–1948, Dec 2011. ISSN 1077-2626. doi: 10.1109/TVCG.2011.242.

Sarah Belia, Fiona Fidler, Jennifer Williams, and Geoff Cumming. Researchers misunderstand confidence intervals and standard error bars. *Psychological methods*, 10(4): 389, 2005.

Eric A Bier, Maureen C Stone, Ken Pier, William Buxton, and Tony D DeRose. Toolglass and magic lenses: the see-through interface. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 73–80. ACM, 1993.

IEC BIPM, ILAC IFCC, and IUPAC ISO. Iupap and oiml 2008. *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement*, 2008.

Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *ACM sigmod record*, volume 29, pages 93–104. ACM, 2000.

John Brosz, Miguel A. Nacenta, Richard Pusch, Sheelagh Carpendale, and Christophe Hurter. Transmogrification: Causal manipulation of visualizations. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 97–106, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2268-3. doi: 10.1145/2501988.2502046. URL `http://doi.acm.org/10.1145/2501988.2502046`.

Sagnik Ray Choudhury, Prasenjit Mitra, Andi Kirk, Silvia Szep, Donald Pellegrino, Sue Jones, and C Lee Giles. Figure metadata extraction from digital documents. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 135–139. IEEE, 2013.

M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *Visualization and Computer Graphics, IEEE Transactions on*, 20 (12):2142–2151, Dec 2014. ISSN 1077-2626. doi: 10.1109/TVCG.2014.2346298.

David Roxbee Cox and David Victor Hinkley. *Theoretical statistics*. CRC Press, 1979.

Geoff Cumming. *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge, 2012.

Geoff Cumming and Sue Finch. A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, 61(4):532–574, 2001.

Geoff Cumming and Sue Finch. Inference by eye: confidence intervals and how to read pictures of data. *American Psychologist*, 60(2):170, 2005.

Tufte Edward. The visual display of quantitative information. *Graphics Press, Cheshire, USA,*, 4(5):6, 2001.

Nivan Ferreira, Danyel Fisher, and Arnd Christian Konig. Sample-oriented task-driven visualizations: Allowing users to make better, more confident decisions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 571–580, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557131. URL `http://doi.acm.org/10.1145/2556288.2557131`.

Jinglun Gao, Yin Zhou, and Kenneth E Barner. View: Visual information extraction widget for improving chart images accessibility. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2865–2868. IEEE, 2012.

M.C. Hao, U. Dayal, D.A Keim, and T. Schreck. Importance-driven visualization layouts for large time series data. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 203–210, Oct 2005. doi: 10.1109/INFVIS.2005.1532148.

Jeffrey Heer and Ben Shneiderman. Interactive dynamics for visual analysis. *Queue*, 10(2):30, 2012.

Y. Hsu. The effectiveness of computer-assisted instruction in statistics education: A meta-analysis. *Doctoral dissertation*, 2003.

N. Kong and M. Agrawala. Graphical overlays: Using layered elements to aid chart reading. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2631–2638, Dec 2012. ISSN 1077-2626. doi: 10.1109/TVCG.2012.229.

Nicholas Kong, Marti A Hearst, and Maneesh Agrawala. Extracting references between text and charts via crowdsourcing. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 31–40. ACM, 2014.

Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. Interpreting and unifying outlier scores. 2011.

GeorgeE. Newman and BrianJ. Scholl. Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic Bulletin & Review*, 19(4), 2012. ISSN 1069-9384. doi: 10.3758/s13423-012-0247-5. URL `http://dx.doi.org/10.3758/s13423-012-0247-5`.

Chris Olston and Jock Mackinlay. Visualizing data with bounded uncertainty. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pages 37–40. IEEE, 2002.

Ankit Rohatgi. Webplotdigitizer. *Journal of Statistics Education*, 10(1):1–20, 2015a.

Ankit Rohatgi. Webplotdigitizer v 3.6. *URL http://arohatgi.info/WebPlotDigitizer*, 2015b.

Manolis Savva, Nicholas Kong, Arti Chhajta, Li Fei-Fei, Maneesh Agrawala, and Jeffrey Heer. Revision: Automated classification, analysis and redesign of chart images. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 393–402. ACM, 2011.

J.D. Schenker. The effectiveness of technology use in statistics instruction in higher education: A meta-analysis using hierarchical linear modeling. *Doctoral dissertation*, 2007.

Giovanni W. Sosa, Dale E. Berger, Amanda T. Saw, and Justin C. Mary. Effectiveness of computer-assisted instruction in statistics: A meta-analysis. *Review of Educational Research*, 81(1):97–128, 2011. doi: 10.3102/0034654310378174. URL `http://rer.sagepub.com/content/81/1/97.abstract`.

William A. Stock and John T. Behrens. Box, line, and midgap plots: Effects of display characteristics on the accuracy and bias of estimates of whisker length. *Journal of Educational and Behavioral Statistics*, 16(1):1–20, 1991. doi: 10.3102/10769986016001001. URL `http://jeb.sagepub.com/content/16/1/1.abstract`.

Edward R. Tufte. *ENVISIONING INFORMATION.*, volume 68. John Wiley & Sons, Ltd, 1991.

Stef van den Elzen and Jarke J. van Wijk. Small Multiples, Large Singles: A New Approach for Visual Data Exploration. *Computer Graphics Forum*, 32(3):191–200, 2013. doi: 10.1111/cgf.12106. URL `http://diglib.eg.org/EG/CGF/volume32/issue3/v32i3pp191-200.pdf`.

Colin Ware. Information visualisation. *Perception for Design*, 2004.

Index

Typeset March 30, 2015