Interactive Generation of Narrative Visualizations for Risk Communication

Anna Kleinau

Supervisors: Dr.-Ing. habil. Monique Meuschke
Prof. Dr.-Ing Christian Hansen

Faculty of Computer Science
Otto-von-Guericke-Universität Magdeburg

Master Thesis

ISG September 2023

Acknowledgements

I would like to express my gratitude to Dr. Ing. habil. Monique Meuschke and Professor Bernhard Preim for their continuous support throughout this project. I would also like to thank the narrative visualization group for their valuable input and support. Finally, I would like to extend my thanks to all the participants of my user studies for generously contributing their time to the success of my work.

Abstract

The urgent need for better health communication is especially evident when considering the millions of deaths caused each year worldwide by lifestyle choices and behavioral risk factors. These deaths show that simply researching and understanding the risks caused by these factors is insufficient, as they must also be effectively communicated to the general public.

Narrative visualization can help achieve this by exploring how data can be visualized and incorporated into an engaging story that will capture the interest of the public.

In my thesis, I investigate how risk visualizations can be designed for a general audience, and subsequently, I develop a tool to assist story authors in creating data-based risk visualizations for their data stories.

With the tool I developed, the story author can receive recommendations for risk factors based on their data set. Using methods from visualization and annotation generation, these recommendations are presented as suitable visualizations based on current research in risk communication. The visualizations adapt to the current intention of the story author, whether it is to explore the data set, convince the public to change their behavior, or educating the public about risk factors for a specific disease. Using the tool, the story author can select the most relevant risk factors, customize the visualizations and export them for integration into their data story.

I evaluated my tool with domain experts, as well as the resulting visualizations with the general public. The results demonstrate that the tool is usable and that the visualizations are understandable and engaging for the general public.

By combining the research fields of risk communication, visualization generation and narrative visualization, my work provides a novel approach to support domain experts in communicating risks and risk factors to the general public. With my approach, experts can create annotated, data-based risk visualizations without requiring expertise in risk communication, visualization design or narrative visualization.

Kurzfassung

Die dringende Notwendigkeit für eine bessere Gesundheitskommunikation zeigt sich besonders bei Betrachtung der Millionen von Todesfällen weltweit, die jedes Jahr durch Lebensstil- und Verhaltensrisikofaktoren verursacht werden. Diese Todesfälle zeigen, dass es nicht ausreicht, nur Risiken und Risikofaktoren zu erforschen und zu verstehen, sondern dass sie auch effektiv an die breite Öffentlichkeit kommuniziert werden müssen. Narrative Visualisierung kann dieses Ziel unterstützen, indem sie untersucht, wie Daten visualisiert und in eine fesselnde Geschichte eingebunden werden können, die das Interesse der Öffentlichkeit weckt.

In meiner Abschlussarbeit untersuche ich, wie Risiko-Visualisierungen für ein allgemeines Publikum gestaltet werden können. Ich entwickle anschließend ein Tool, das Story-Autoren dabei unterstützt, datenbasierte Risiko-Visualisierungen für ihre Datenstories zu erstellen.

Mit dem von mir entwickelten Tool kann der Story-Autor Empfehlungen für Risikofaktoren auf Basis seines Datensatzes erhalten. Mithilfe von Methoden aus der Visualisierung und der Annotationsgenerierung werden diese Empfehlungen als geeignete Visualisierungen präsentiert, deren Design auf aktuellen Forschungen zur Risikokommunikation basiert. Die Visualisierungen passen sich der aktuellen Absicht des Story-Autors an, ob es darum geht, den Datensatz zu untersuchen, die Öffentlichkeit davon zu überzeugen, ihr Verhalten zu ändern oder die Öffentlichkeit über Risikofaktoren für eine bestimmte Krankheit aufzuklären. Mit dem Tool kann der Story-Autor die relevantesten Risikofaktoren auswählen, die Visualisierungen anpassen und sie für die Integration in seine Datenstory exportieren.

Ich habe mein Tool mit Domänenexperten, sowie die resultierenden Visualisierungen mit der breiten Öffentlichkeit evaluiert. Die Ergebnisse zeigen, dass das Tool nutzbar ist und dass die Visualisierungen für die breite Öffentlichkeit verständlich und ansprechend sind.

Durch die Kombination der Forschungsfelder Risikokommunikation, Visualisierungsgenerierung und narrative Visualisierung bietet meine Arbeit einen neuartigen Ansatz zur Unterstützung von Domänenexperten bei der Kommunikation von Risiken und Risikofaktoren an die breite Öffentlichkeit. Mit meinem Ansatz können Experten annotierte, datenbasierte Risiko-Visualisierungen erstellen, ohne Expertise in Risikokommunikation, Visualisierungsgestaltung oder narrativer Visualisierung zu benötigen.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In today's world, it becomes more and more obvious that having information available is not enough, it has to be communicated effectively. This is particularly evident in healthcare. Diseases are being unequally distributed through society, stating the question which factors put persons at risk and were we can intervene to prevent diseases. Major health risk factors like smoking or obesity [13] are being preventable through behavior change, making communicating risks to the general population of utmost importance. In 2019, an estimated amount of over 8.7 million deaths worldwide are caused by smoking, 7.9 million by dietary risks, 2.4 million by alcohol use and 800.000 by low physical activity [39].

In Germany, there is an urgent need to improve health communication. An estimated amount of 59% of the population have a low health competency, a percentage even worse than in previous studies [49].

Data and research on the links between risk factors and diseases are available through the field of epidemiology. Epidemiologists study how diseases are distributed in society and analyze patterns in their distribution [46]. Large-scale epidemiological studies, such as the SHIP study in Germany, collect data from representative samples of society, allowing for a multitude of analyses and research [26]. However, this research is highly scientific and complex and needs to be simplified to be presented to a general audience.

Much research is done on how risks can be communicated effectively. Recently, data visualizations have been shown to improve the effectiveness of risk communication [3, 57] but have the potential to cause misunderstandings, be too complex or affected by biases [57]. Additionally, research from the field of narrative visualization can be used to engage the audience by combining visualization with storytelling techniques and interaction [52].

However, educating the public about risks requires a multitude of different abilities. It involves combining knowledge on the domain of interest, such as clinical knowledge, with knowledge about effective risk communication. It also involves knowledge about effective visualization techniques when using data visualizations and the computational knowledge required to analyze data and create data-driven visualizations. However, there is growing research on how visualizations can be automatically created from data. Such visualization generation tools can be generally usable [62] or designed with specific purposes in mind [7, 17, 24].

A visualization generation tool for the purpose of risk communication will not only enable researchers without a computational background to create convincing visualizations, but also support the researcher in creating effective visualizations following current research on risk communication.

1.2 Goals of this Project

With this thesis I aim to develop a visualization generation tool that supports experts in creating effective data visualizations for risk communication. I will consider the following research questions:

- **RQ1:** How can visualizations be designed to effectively communicate risk factors to a general audience?
- **RQ2:** How can annotations enhance those visualizations?
- **RQ3:** How can a tool support experts in creating such visualizations based on a given data set?
- **RQ4:** Are the tool and its resulting visualizations effective in communicating risk factors to a general audience?

The tool will integrate the knowledge and capabilities of multiple research areas.

Knowledge from the field of risk communication and narrative visualization will be used to design effective visualizations to educate a general public (**RQ1**). The growing research on automatic visualization and annotation generation will be used to enhance the visualizations with annotations (**RQ2**) and provide a tool that supports the user in creating a data story from a given data set (**RQ3**). The tool will also use research on risk factor calculation to recommend appropriate risk factors to the user. Knowledge from the fields of risk communication and narrative visualization will be combined to create convincing visualizations based on these risk factors to educate a general public.

The project will also entail two evaluations of the tool (**RQ4**), a qualitative study examining the usability of the tool by experts and a quantitative study evaluating how the generated visualizations are perceived by a general public. The tool and evaluations will especially consider the different intentions a user might have and if the tool supports each of them effectively.

1.3 Structure of the Thesis

After motivating this thesis, I will extensively provide background knowledge on each of the entailed research fields (Section 2.1). I will introduce research on risk calculation, risk communication, narrative visualization, visualization generation and annotation generation. Then, I will discuss related work to this thesis (Section 2.2). In the third chapter, I will present the methodology used to develop the tool (Chapter 3), starting from requirement analysis over design choices, data processing, fact selection, visualization generation and annotation generation. Then I will describe the implementation and the final tool (Chapter 4). Afterwards I will describe the two evaluation studies on how the tool can be used by experts and by a general audience and discuss the results (Chapter 5). I will finish with the conclusion and future work (Chapter 6).

Chapter 2

Literature Review

2.1 Background

I will begin by providing background information on this thesis. Firstly, I will explain the concept of risk factors and their calculation methods. Next, I will discuss the communication of risks to the general public. Then, I will introduce the research field of narrative visualizations. After that, I will explain the process of automatically generating visualizations. Finally, I will discuss the use of annotations in visualizations.

2.1.1 Risk Factor Calculation

The likelihood of an individual person to have a disease can be heightened by different risk factors. The factors can be separated into *environmental influences* like air quality, *predisposition* through, for example, genes or *behavioral characteristics* like smoking and drinking alcohol [46].

Epidemiological Studies

Finding risk factors is an important task in epidemiology. Epidemiology is a research field analyzing how diseases are distributed in a population and their potential causes [46]. To accomplish this task, data on disease and risk factor distributions is collected through different types of epidemiological studies. Cohort studies investigate the effect of potential risk factors by comparing persons with or without a potential risk factor. Case-control studies focus on specific diseases by comparing the prevalence of risk factors in persons affected versus persons not affected by the disease. Cross-sectional studies follow a more general approach by surveying a sample from a specific population at a specific point in time and collecting information on both exposure and disease status [46]. The advantage of cross-sectional studies is that they do not sample persons based on their exposure or disease status. This allows estimations of the prevalences of both diseases and risk factors in society [59], and to analyze which subgroups of the population are affected by the disease or risk factor. Because of the available prevalence estimates, this work focuses on cross-sectional studies. They will not only enable the viewer to access the significance of risk factors, but also their actual impact, stating how many persons are affected by it. However, when using data from cross-sectional studies, potential biases must be considered. For instance, a non-response bias may occur when individuals who responded to the survey exhibit different characteristics from those who did not [59].

Effect Measures

Based on the collected prevalence information, associations between factors and diseases in a data set can be analyzed. For this purpose, various effect measures can be employed. Identifying associations is crucial to determine whether factors are interdependent or if they represent potential risk factors for a disease. In the following, I will present some common measures.

The most common measure of association between continuous variables is the *pearson correlation*. It measures the linear correlation between two variables. The *pearson correlation* is calculated by dividing the covariance (cov) of the two variables by the product of their standard deviations (σ).

$$r_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

For categorical variables, *Cramer's V* is commonly used. It is a measure of association between two categorical variables who can have two or more categories. It is calculated by dividing the chi-squared statistic by the number of observations (n) and the minimum of the number of rows (r) and columns (c) minus one.

$$V = \sqrt{\frac{\chi^2}{n(min(r,c)-1)}}$$

In risk factor research, often the data is processed in the binary formats of *exposed/not exposed* to a risk factor and *diseased/ not diseased* [67]. This creates a 2x2 table of the prevalences of each combination, see Table 2.1. The data is separated into persons who are exposed and diseased (a), not exposed but diseased (b), exposed but not diseased (c), and not exposed and not diseased (d). Based on this table, two common measures are presented as follows.

	exposed to risk factor	not exposed to risk factor
have disease	a	b
do not have disease	c	d

Table 2.1 2x2 table used for effect measures to calculate associations between a risk factor and a disease.

The relative risk, also called *Prevalence Rate Ratio (PRR)*, compares the proportion of *diseased* persons in the *exposed* group with the proportion in the *not exposed* group. A higher PRR indicates a higher risk of the disease for the exposed group, compared to the not exposed group.

$$PRR = \frac{a/(a+c)}{b/(b+d)}$$

The *Prevalence Odds Ratio (POR)* compares the proportion of persons who either are *exposed and diseased* or *not exposed and not diseased* with the rest. A higher POR indicates a higher association between the risk factor and the disease.

$$POR = \frac{a*d}{b*c}$$

Zocchetti et al. argue that the choice of risk measure should depend on the given task, with both measures being similar when the disease is rare (prevalence < 0.1) but with discrepancies at higher disease prevalences [67].

Dealing with Multiple Risk Factors

When considering multiple risk factors together, for example, to investigate interactions between them, one commonly used method is logistic regression [46].

Regression analysis requires the use of continuous variables. Standardization can be used to make the model weights associated with each variable more interpretable. However, categorical variables can be adapted for use in regression models through methods like one-hot encoding. In this method, each class of a categorical variable is encoded as a separate binary variable stating if the given data point is part of that class or not [43]. Data imputation methods may be needed to fill in missing data in the data set [1]. Additionally, the variables used as input for regression models are assumed to be independent. If this is not the case, feature selection methods can be used to select a subset of the variables. For example, in stepwise feature selection, variables are added one by one to the model to ensure diversity [15]. Such approaches can either be hypothesis-based, with an expert already knowing which variables to include, or hypothesis-free without prior knowledge based only on the data [28]. Visual methods, like the 3D Regression Heat Map by Klemm et al. [28], can be used to support the user in this task.

Logistic regression, as a subtype of regression, is used for analyzing multiple independent variables in respect to one binary outcome. This is accomplished by using the sigmoid function, that maps the output to a value between 0 and 1. The output can then be used for classification by setting a threshold like 0.5 to separate predictions into two classes.

The trained regression models can either be used by directly considering the weights of the model, which provide information about the importance of the different variables. Alternatively, the model can be used for predictions of the outcome. In this case, the goodness of fit of the model can be measured by metrics like the accuracy or F score. The accuracy of a model directly states the percentage of cases predicted correctly. The F score additionally considers if a prediction is a false positive, true positive, false negative or true negative, making sure that the model does not only predict the majority class but creates good predictions for cases of all classes.

Predictive and Explanatory Variables

The presented effect measures provide information about associations and correlations between the factors and disease. However, it is important to consider that correlation does not equal causation. This distinction and its influence on epidemiological research is discussed in-depth by Schooling and Jones [51]. The term risk factor is used for both predictive variables, like symptoms of a disease predicting the disease, as well as for explanatory variables like the causes of a disease, with some risk factors being both.

Both variable types are relevant. However, they are useful for different tasks and models. Schooling and Jones differentiate between two tasks. The task of prediction aims at creating models good for identifying subgroups of a population at an elevated risk of a disease. These subgroups might be suitable candidates for targeted interventions. The goal here is to find factors that, when combined, can explain the greatest amount of variance in the available data set. The task of explanation, however, aims at identifying the causes of a disease. This requires much more in-depth research and focused studies avoiding a number of biases like selection bias or confounding bias. The selection bias appears when there is a bias in the selected participants of a study. The confounding bias happens when a confounder is ignored. A confounder is a hidden factor that influences both the risk factor as well as the disease, explaining their association.

2.1.2 Risk Communication

Risk communication is an interdisciplinary field asking the question of how risks can be effectively communicated to a given audience, often either patients or the general public. It is defined by Powell and Leiss [44] as the “process of communicating responsibly and effectively about the risk factors associated with industrial technologies, natural hazards, and human activities”.

Leiss [31] presents a compelling argument for why simply researching and gaining knowledge about risk factors is not sufficient, and why effective risk communication is necessary. Risk communication can reduce fears about risks by setting them into context as well as increase awareness of risks and therefore enforce necessary behavior change. In that, communicating risks improves public dialogue and risk management, as well as leads to better priority setting. When risk management strategies are created, communication is needed to build trust in them.

As in all communication, when wanting to communicate risks, all parts of the process have to be considered. Balog-Way et al. [4] differentiates between messengers, message attributes and the audience. Messengers are the indiviuals or organizations who are aiming to communicate a certain risk. They influence the risk communication process by the trustworthiness and fairness associated with them. Message attributes, the aspects of the message itself, have to be considered under a series of lenses like framing effects, where small changes have a big impact on the conveyed message [23]. Additionally, the messenger has to consider the effects of choice of words and created feelings. Balog-Way et al. [4] highlights the importance of scientifically evaluating risk messages to avoid unintended effects. Lastly, the audience of risk communication can be influenced by various factors, such as geography, culture, and identity, in how they perceive messages.

A challenge in risk communication is that more and more the “language of risks” is used in public places, making the handling of the gap between how risk assessment experts think about risks and how the public thinks about them of great importance. Leiss [31] describes how the public perception of risks differs from that of experts in the field. Contrary to the expert view of risks as quantitative probabilities, he argues that the public thinks often in the binary of something being safe or not. Additionally, for the public, other factors are important like if they are voluntarily exposed to a risk or not and if the risk is familiar, in which case it is tolerated to a much higher degree. One example of this is the perception of the risks of car crashes and risks associated with radiation.

Health Risk Communication

The focus of this thesis will be on health risk communication. Specifically for the communication of risks in healthcare, numerous research has been done, although most often in the context of comparing risks associated with different treatments, not in communicating risks to a general audience.

Extensive guidelines are presented by Trevena et al. [57] on how risks should be communicated. Despite their focus on patient decision aids, their results are easily generalizable to other cases of health risk communication too. Important best practices include considering framing effects like the effect of relative and absolute risk presentations, the importance of context and how uncertainty can be conveyed. They emphasize the consideration of graph literacy and numeracy of the targeted audience, as well as consistency across visualizations. Considering the format used for probabilities, they emphasize keeping the denominator consistent instead of using "1 in X" formats. Research on the most effective format is still inconclusive, with some studies showing that natural frequencies are easier for statistical calculations than percentages [21] and others showing that for treatment risk and benefit communication percentages are preferred [61].

Schmälzle et al. [50] review research on risk perception from the perspective of psychophysiology. They highlight the importance of intuitive and affective processes in risk perception and decision making. For example, they describe the phenomenon of unrealistic optimism, describing that persons tend to compare themselves against others in a too favorable way. They explain that such biases and aspects of risk perception have to be kept in mind, and campaigns should focus on delivering a clear message with clear behavior recommendations.

Health Risk Visualization

In this section, I will elaborate on prior work on how risk communication can be improved by using adequate visualization techniques.

Ancker et al. [3] provides an extensive review of research on how graphs for risk communication should be designed. They differentiate between the goals of accuracy, behavior change and likeability. Each of these goals requires different, sometimes contradicting design choices. For example for accuracy, graphs showing nominator as well as denominator in a part-to-whole relationship are considered useful. In contrast, for the goal of behavior change the numerator should be emphasized to increase the perceived risk. For likeability, simple graphs are preferred. However, when participants used their preferred format it did not lead to an increase in performance.

Additionally, Ancker et al. [3] reviewed research of different graph formats. Some key insights are that the often used survival or mortality curves for risk visualizations display detailed information but are considered very complex, leading to many individuals struggling to understand them or use them correctly. Pictographs are recommended as a very effective method for risk communication and risk ladders as a means to place unfamiliar risks in context.

Adding on to that, Hawley et a. [20] performed a study to compare different graph formats for treatment risk communication. They compared basic and modified versions of bar charts, pie

Figure 2.1 Pictographs used by Hawley et al. [20] in their study on graph formats for risk communication.

charts, pictographs and tables. They differentiate between verbatim knowledge as knowledge about precise information like numbers or differences between numbers, and gist knowledge as more general knowledge about the essential points of a graph. Study results rated tables as best for verbatim knowledge and pie charts best for gist knowledge. However, considering overall performance, pictographs are recommended by the authors with good results in both categories. The pictographs used in their study are shown in Figure 2.1. As participants with higher numeracy generally performed good across all formats, they highlight that visualizations should focus on low numeracy users who are most impacted by the choice of graph format.

Research on using interaction in risk visualizations is conflicting. Whilst it may increase a patient's active processing of the risk information, Zikmund et al. [66] provide an example of how interaction may instead lead to worse knowledge and participation. Hakone et al. [19] replaced their interactive graphs with static graphs as users often did not use the interactivity or found it confusing. However, the interactive and narrative tool of Bissett et al. [6] received positive feedback from many participants, who viewed it as more engaging than other similar tools.

2.1.3 Narrative Visualization

When designing visualizations for the general public, user engagement, clear messaging and ease of understanding are of special importance. The research field of narrative visualization combines techniques from visualization with storytelling to create visualizations and data stories that are engaging, easily understandable and memorable [52]. Segel and Heer [52] introduce the emerging field of narrative visualization and present a foundational design space for such visualizations. Their definition of genres of narrative visualizations shows the versatility of narrative visualization, from typical slideshow formats used for narrated presentations to interactive data stories enabling free exploration. The focus on storytelling leads to a special consideration of the importance of consistency and connection between visualizations to support the viewer in following the story.

The goal of visualization is often considered to be generating or spreading insights [8]. As Hamming (1973) already stated, "the purpose of visualization is insight, not pictures". Visualizations should therefore be designed as a means to a goal, not as an end in itself. Insight is defined by Saraiya et al. [47] as "an individual observation about the data by the participant, a unit of discovery". Depending on the type of narrative visualization, data stories can be either designed to convey a specific insight or to enable the user to discover insights themselves. The presented tool in this work will enable both, allowing experts to use visualizations to gain insights and to then share these insights with a general audience.

Especially when designing visualizations to convey a specific insight, every aspect of the visualization has to be carefully considered. Hullman and Diakopoulos [23] use the term "Rhetoric" to refer to processes representing an intended meaning in a (narrative) visualization. Their interpretations are shaped by the viewer and viewing context of the visualization. This definition highlights the importance of studying the intended user group to effectively convey a message. Sometimes even small changes in a data story can have a big impact on the resulting conveyed message. These framing effects require special attention when designing narrative visualizations [23]. The work of Hullman and Diakopoulos [23] explores and describes rhetoric techniques for information visualizations and how they frame the interpretation of the visualization. They categorize rhetoric techniques into the following system:

- Information Access Rhetoric: Techniques that change which data is accessible to the viewer, for example by filtering or aggregating the data or omitting outliers.
- Provenance Rhetoric: Techniques that provide information about the data source or data uncertainty, for example citations, error bars or information about the creator of the visualization.
- Mapping Rhetoric: Techniques changing how the data is mapped to visual elements, for example by using visual metaphors, contrasts or redundancy.
- Linguistic-based Rhetoric: Techniques that use wording and font style to influence the interpretation of the visualization, for example by using textual highlights or rhetorical questions.

- Procedural Rhetoric: Techniques that use defaults and interaction to anchor and guide the user, for example through suggestions and search bars.

These rhetoric techniques can be used on the different editorial layers of a visualization distinguished by Hullman and Diakopoulos [23], meaning that framing and influencing of a message can happen on all levels of the visualization generation process. The editorial layers are:

- Data layer: The data itself, its preprocessing and manipulation.
- Visual Representation Layer: The mapping of the data to visual elements.
- Annotation layer: The addition of textual and graphical annotations to the visualization.
- Interactivity layer: The addition of interaction to the visualization.

In this work, rhetoric techniques will be used to adapt the visualizations to the intention of the user. They will also be considered to support the expert in effectively finding risk factors to include in their data story.

Narrative Visualization in Medicine

Meuschke et al. [35] provides a first proof-of-concept how narrative visualization can be used in medicine to inform a general audience about diseases. Their data story on liver cancer is shown in Figure 2.2. Concepts of narrative visualization are adapted for communication of medical topics and main challenges presented. Their work focuses on individuals interested in medical topics, but without medical expertise. This user group is interested in topics such as the effects of lifestyle on diseases and disease risk, as well as education about diseases and treatment options. They highlight the importance of easy-to-understand medical visualizations, simplifying medical concepts and dealing with problems of occlusion and noise in medical imaging typically directed at experts. To capture the attention of the audience, visual data representations are combined with engaging illustrations and highlighting techniques. The research of Meuschke et al. aims to enable researchers without a creative background to improve their science communication using the tools of narrative visualization. This goal is also shared by this thesis.

Further aspects researched in the field of narrative visualization in medicine are the effect of human protagonists in data stories [38], as well as the effect of different narrative genres [37]. However, research on narrative visualization in medicine is still in its infancy. Therefore, the main contribution of Meuscke et al. is a research agenda for narrative medical visualization. The agenda includes open questions like how narrative medical visualizations should be designed

Figure 2.2 Data story educating a public audiene about liver cancer by Meuschke et al. [35]. A first application example of how narrative visualization can be applied to medicine.

for patients and experts, as well as how they should be evaluated. Their first point, however, is the need for better authoring tools for narrative medical visualizations. This work aims to contribute to this goal by providing a tool for experts to communicate risks and risk factors to a general audience.

2.1.4 Visualization Generation

This thesis will built upon prior work in visualization generation to create a visualization generation tool for risk communication.

The foundational work of the research area of visualization generation is the work by Mackinlay [33]. The presented visualization tool is designed to be application-independent in order to automatically create visualizations for a number of different use cases. The paper presents how given information can be transformed into visualizations. Visualizations are created using a composition algebra and graphical languages. Then the created visualizations are scored by how close they fit the given information. Such visualization tools found great

Figure 2.3 Visualization recommendations generated by the visualization generation software *Voyager* by Wongsuphasawat et al. [62].

commercial success, with a popular tool being *Tableau*[1], a software to automatically generate visualizations given a set of data variables.

However, these tools still require the user to specify the fact that should be visualized. Further work explored not only how visualizations can be generated, but also how facts can be automatically extracted from data and then visualized. The additional fact extraction makes such tools well-suited for initial data exploration. A noteworthy software implementing this is *Voyager* by Wongsuphasawat et al. [62]. Their interface presenting visualization recommendations is shown in Figure 2.3.

Such visualization generation systems became widely successful, with two commercial examples being the "Show me" feature, a group of UI commands to automatically create presentations from data, implemented in *Tableau* [34] and *QuickInsights*, a tool for finding, scoring and visualizing interesting patterns, implemented in *Microsoft Power BI* [14]. For academic work, Zhu et al. [65] present an extensive overview of research on automatic visualization and annotation.

[1] https://www.tableau.com

Architecture of Visualization Generation Tools

Visualization generation tools can be structured in multiple ways. The structure used by Wongsuphasawat et al. [62] in *Voyager* uses a visualization browser for the user to interact with and select visualizations. Visualizations are created through the process of first a recommendation engine creating interesting visualization specifications, which are passed through to a compiler and subsequent renderer creating the visualizations from those specifications.

The structure used by Wang et al. [60] in their tool *DataShot* consists of fact extraction, fact composition and visual synthesis. First facts are formulated and scored in the fact extraction phase. Then, facts are grouped into topics in the fact composition phase, with a set of important and diverse facts being selected for the fact sheet. Lastly in the visual synthesis phase, facts are visualized, styled and a layout is created.

Both structures generally consist of elements to generate visualizations, score, and visualize them. Depending on the workflow of the specific tool, they are adapted to include more or less interactivity.

Visualization Generation

Commonly formal languages are used to quickly describe and create a number of visualizations. Mackinlay [33] defined multiple graphical languages for different parts of a visualization, for example languages for positioning of the elements, types of marks and types of style elements like colors and size. Using a compositional algebra, these languages are combined to create a complete visualization specification.

A different technique is used by Wongsuphasawat et al. [62]. They are using a descriptive logic to create an exhaustive list of all possible visualizations based on specific rules. They are then ranking them following the guidelines of Cleveland and McGill [10] and Mackinlay [33].

Scoring Functions

Mackinlay [33] presents a scoring method to evaluate how well a visualization displays specific information. The method ranks different visualizations based on two criteria. Effectiveness describes if exactly the given information is included in the visualization. Expressiveness measures how well the visualization conforms to the human visual system and chosen medium, following the design principles outlined by Cleveland and McGill [10].

For also scoring the facts themselves to decide which information should be visualized, further criteria were needed. Ding et al. [14] score facts based on an insight's significance and impact. Significance states how noteworthy an insight is based on statistical measures, whilst impact states how important the insight is in the context of the whole data set. For example, strong correlations are considered very significant, but only impactful if the correlated data is of large scale or an important part of the data set. They also eliminate “Easily Inferable Insights”, defined as insights that are based on functional dependencies between variables and too trivial to be of any interest for the user. An example of this would be a correlation between age and birth year.

Wang et al. [60] also use impact and significance measures, but they aim to avoid selecting too many similar facts in their visualizations. Users are likely more interested in a diverse range of information. Therefore, they do not use the individual scores directly. Instead, they select visualizations using a density-based top-n algorithm.

User Interaction

Finally, the user must be able to interact with the tool to select visualizations and influence recommendations. Various metaphors and user interfaces have been proposed to facilitate intuitive use of visualization generation software.

Perry et al. [41] use a card deck metaphor to add or remove visualizations from the current selection like someone would add or remove cards from a card deck. Additionally, brushing and linking techniques are used to guide the user by highlighting selected values or data points in all other visualizations.

Wongsuphasawat et al. [63] present different techniques to allow the user to interactively steer recommendations. Using *wildcards*, the user can specify some parts of a visualization manually and let the algorithm automatically recommend solutions for the rest. Using *related views*, recommendations are displayed based on the current context.

Cui et al. [12] display recommendations as a feed, integrating the recommendations as a virtual assistant recommending them in a dialogical manner.

2.1.5 Annotation Generation

To support the interpretation of the final visualizations by the general public I decided to include annotations. Annotations are part of the broader category of graphical overlays, which is defined by Kong and Agrawala [30] as “visual elements layered onto charts to facilitate a larger set of chart reading tasks”. Such overlays can be reference structures like grid lines, redundant encodings like labels for visual elements or, as in this case, annotations. In a later work on the related concept of visual cues, Kong et al.[29] outline the conflict between creating visualizations with detailed information and not overloading it. Visual cues and graphical overlays are a way to resolve this conflict by guiding the attention of the user through the visualization. Furthermore, visual cues can organize information and relate elements of the visualization to each other.

Annotations have a long history in medical visualizations due to the labeling of anatomical structures. Annotations typically used here are *internal labels* which are directly added in close proximity of the associated structure, and *external labels* positioned outside of the model and connected via a line with the structure. Whilst *internal labels* are easier associated with their structure, *external labels* are easier to read and do not pose occlusion problems. They are also better suited for longer texts [40]. When combined with interaction [30], the amount of visual clutter caused by annotations can be greatly reduced, for example by just showing the labels of the selected elements.

However, Kong et al. [29] outline cases in which visual cues can be detrimental, for example when the visualization is already too overloaded with information and annotations would just increase the visual clutter. Additionally, annotations are not suitable in all contexts.

Visual cues are especially helpful to attract and orient a general audience, as they can highlight and direct attention to the most compelling insights [45]. Annotations can directly state the insights in natural language format next to the visualizations containing them, which will help especially users with low graph literacy to understand the visualization.

Annotations can also set visualizations into context with the rest of the data story or document. In their work, Ren et al. [45] analyses the use of annotations for data-driven storytelling. The use of annotations for this purpose was also observed by Segel and Heer [52] where annotations were used in interactive slideshows to convey a narrative.

Automatic Annotation Generation

The potential of annotations to guide attention, convey insights and help with storytelling gave rise to research on how annotations can be automatically generated for (narrative) visualizations.

Kandogan [27] provides an approach of how to automatically annotate point-based visualizations. He argues that annotations should be used to provide additional information instead of redundantly encoding what the user already is able to see themselves. His generation of annotations was implemented using templates to generate appropriate wording. His work also featured interaction with the annotations, for example to highlight or save them.

The difference between annotations that redundantly encode information already present and annotations that provide additional information is also described by Gao et al. [17]. They distinguish between *observational annotations* describing what is already visible in the graph and *additive annotations* adding new information. Most researchers in the area favor *additive annotations* [17, 24] as they do not replace the users' ability but will provide new information. However, research like the work of Wang et al. [60] using textual descriptions to describe visualizations, illustrate the potential of *observational annotations* in helping users to understand the visualization. This is especially the case when the visualizations are unfamiliar or complex [55]. *Observational annotations* can be generated by using statistical functions combined with natural language templates to display their results [55].

Annotations are especially useful in research on how visualizations can be generated for specific domains. For example, Hullman et al. [24] annotates stock visualizations with relevant news articles and Gao et al. [17] adopts this approach of using news articles to annotate maps.

2.2 Related Work

Research related to my work can be found in multiple areas. I will first introduce different approaches to risk communication using visualizations. Then I will present tools for narrative visualizations and visualization generation. Finally, I will discuss other visual analytics tools for epidemiological data.

2.2.1 Visualization for Public Health Risk Communication

Visualizations and data stories are used in a variety of contexts to aid in (health) risk communication. Research in the field has examined different aspects of how risks can be effectively communicated.

One approach to risk communication is to translate abstract risks into real-world visualizations. In the context of medical images, risk communication based on realistic images can be achieved through medical imaging technologies. Hollands et al. [22] review such approaches that use imaging technologies to show individuals their own medical data. Although these approaches are found to be effective, they are not easily accessible to a general public due to the need for medical imaging technologies. In my work, I will instead use data of epidemiological studies. This data is not as personalized but provides insights into risk factors.

Other approaches to risk communication research how engaging the audience may improve communication. In the context of communicating health risks, Crovato et al. [11] created and evaluated the use of a serious game to increase adolescents' awareness of food safety. Their game featured a detective story investigating a case of food poisoning caused by raw milk consumption. A similar approach is taken by Bissett et al. [6] who built a visualization tool to increase awareness of the risks of alcohol consumption. They feature comic-style illustrations for an entertaining and engaging experience. In contrast, my work will focus on visualizations directly derived from data. This will enable my tool to be used for a variety of different use cases and data sets.

Few systems use visual analytics to communicate risks. One such system is *PregnancyLine* by Li et al. [32], which supports pregnant persons in assessing potential risks during their pregnancy and communicating with their doctors. The system uses multiple interactive views designed for persons without medical expertise. These views include statistical graphs interpreting exam results as normal or abnormal, timelines of the pregnancy, and more intuitive visual metaphors of ultrasound scan reports. In their work, persons from the general public interact directly with the visual analytics tool. My work will instead use visual analysis to support the expert in creating visualizations, that are shown to the public afterwards. This enables the expert to include their expertise in the process and will not require the general public to learn how to use a visual analytics tool.

2.2.2 Tools for Narrative Visualizations

With my work I want to support experts in creating data stories for risk communication. Tools for such narrative visualizations are sparse. Amini et al. [2] developed their software *DataClips* to support users in creating narrative data videos. The software provides a libary of visualizations and animations that can be used by the story author to create a data story.

Satyanarayan and Heer [48] present *Elipsis*, a tool that can be used to generate a variety of scenes and narrative structures for data stories. Like my work, they use a formal language to describe the stories. They also provide a user interface to interactively create stories. No programming knowledge is required, as all user input is automatically translated into the formal language.

Both are general-purpose tools for manual story generation. My tool will further support the user through recommendations and is more domain-specific. It will also not support users in the whole story creation process, but only generate visualizations and instead feature easy export functions to be used in cooperation with such tools.

2.2.3 Visualization and Annotation Generation

In this work, I use tools for generating visualizations to facilitate risk communication. These tools automatically generate visualizations for a given context. They often use annotations to support the generated visualizations.

Wongsuphasawat et al. [62] introduced the general-purpose visualization generation tool *Voyager* that automatically generated visualization recommendations based on a tabular data set. The tool suggests visualizations by first creating univariate data visualizations and then gradually adding dimensions. In my work, I will use a similar architecture that separates the user interface, recommendation generation and rendering. However, my approach will be

Figure 2.4 Fact sheets generated by the visualization generation software *DataShot* by Wang et al. [60].

tailored to risk visualizations, providing clear guidance on which visualizations to use based on current research in risk communication. I will also automatically annotate the visualizations.

Wang et al. [60] presents *DataShot*, a tool to automatically generate a fact sheet from a given data set. Their visualizations are easy to understand, visually pleasing and are annotated with a textual description. Figure 2.4 shows some of the generated fact sheets. The work promotes a new approach of user interaction by first automatically providing the user with a first draft that can then be manually refined through interaction. In my work, I will use the same approach, as well as have a similar focus on easily understandable visualizations. I also adapted templates for textual descriptions and annotations to provide additional information. However, my visualizations will be designed for a specific purpose and will not have a specific presentation format. Instead, I will provide export options to create the final data story in another tool.

Shi et al. [54] created a system called *Calliope* that, similarly to *DataShot*, starts with an input spreadsheet and selects facts, but instead of creating a fact sheet, it creates a data story. The additional consideration of a narrative incorporates, like my work, narrative techniques into visualization generation. However, my work will focus on risk communication and does not fully automate the story generation but instead lets the user choose and incorporate their intentions and domain knowledge.

Further visualization generation tools were designed for specific use cases and data sets. Hullman et al. [24] presents a tool, the *Contextifier*, that automatically generates stock visualizations for news articles. The visualizations are created without any human intervention by semantically analyzing the text and connecting it to stock data and other news articles. An

Figure 2.5 Stock visualization generated by the *Contextifier* software by Hullman et al. [24].

example visualization is presented in Figure 2.5. The visualizations are automatically annotated with the headlines of related news articles to provide context information. Like the *Contextifier*, my tool will automatically generate visualizations for a given context and annotate them with related information. However, my tool is designed in a more interactive way letting the story author choose and customize the visualizations. I will also use further information entailed in their dataset for my annotations, instead of using headlines of external news articles.

Gao et al. [17] uses a similar approach as Hullman et al. [24] to create interactive, annotated maps from news. Again, semantic information from the news article is use to find relevant places and news articles. This work introduces interactivity in their visualizations, whilst in my tool all visualizations will be static. I adopted the use of additive annotations from this work and from Hullman et al. [24] to provide information in the annotation not available in the original visualization. Opposed to their tools which will first use additive annotations before using observational ones, I will use observational annotations first and then use additive annotations. I will explain my reasoning for this in Section 3.6.

Bryan et al. [7] also adopted a similar approach to Hullman et al. [24] to automatically create visualizations for temporal data. The visualizations are designed using the principles of narrative visualization. Annotations are again used to provide context. The user can interactively adapt the visualization before being provided with a static export. Like this work, I will provide the user with automatically generated initial templates that can be interactively adapted. I also provide a static export, as most currently used tools do not support interactive visualizations.

However, my work will not contain a single central visualization. Instead, my tool will be used to create a series of multiple images closer to a slideshow format.

2.2.4 Visual Analytics for Epidemiological Data

From an expert's perspective, this work aims to create a visual analytics tool for epidemiological data.

Chishtie et al. [9] provide a review of current research in the area. Among the 55 articles they examined, most focused on infectious diseases or analysis of medical records. However, nine papers focused on the closely related area of population health monitoring.

A tool that might also be used for risk factor analysis is the tool by Benis and Hoshen [5]. They analyze epidemiological data over time to identify clusters and profiles. My work will not consider temporal data but focus on cross-sectional data.

Another noteworthy tool was created by Shaban-Nejad et al. [53] to integrate population health data. Among other things, their tool helps public health practitioners in monitoring diseases and risk factors by integrating knowledge-based causal graphs with data-based correlation analysis. Risk factors can also be monitored and the effect of interventions tracked. In contrast, my work derives risk factors from data but enables the domain expert to integrate their causal knowledge by choosing and customizing the risk factors that will be included in their final presentation. My work also focuses on educating the public, instead of providing decision support for practitioners.

Multiple further visual analytics tools and frameworks were designed for specific use cases and data sets. Garcia-Marti et al. [18] use visual analytics as part of a process to identify risk factors for tick bites. Their visualizations included heatmaps and geospatial graphs. Yu et al. [64] used visual analytics to analyze multiple data sets on risk factors for respiratory diseases. In contrast, my work will be able to generate potential risk factor visualizations for a variety of use cases and data sets.

Chapter 3

Methodology

In this chapter, I will describe the methodology used to develop the presented tool. I will start by describing the requirements for the design of the tool, followed by the design choices made. I will then describe the individual parts of the tool starting with the data processing, followed by the fact selection, visualization generation and annotation generation.

3.1 Requirement Analysis

To derive the requirements for the tool, I will first characterize the user groups and then derive requirements from these characterizations.

3.1.1 User Characterization

This work targets two separate user groups, domain experts and the general public. Domain experts will be enabled by the tool to create data-based visualizations for their presentations or data stories. These data stories are then intended to be presented to the general public, requiring to consider this user group. I will proceed to characterize both user groups in more detail.

Domain Experts

Domain experts posess the necessary knowledge to identify and interpret risk factors. This entails the statistical knowledge required to calculate risk factors from data, considering confounding factors and potential problems like insufficient data sizes. This allows them to interpret visualizations of risk and contextualize them appropriately. In addition, experts who will use the tool with their own data set will possess the required data set knowledge to interpret the variables and domain knowledge to interpret the computed results. They will have a background understanding of the research and current knowledge in the area and potential open questions on the common risk factors for their disease or hazard. Furthermore, they will be familiar which risk factors are genetic or lifestyle-related, enabling them to choose the most relevant risk factors for their target audience.

As the authors of the final data story they will be responsible for the presented visualizations and will have to integrate them with the remaining data story.

General Public

The final visualizations will be targeted to the general public. This user group must be engaged and motivated to read the data story or pay attention to a presentation, as they most likely have no personal involvement with the topic. As medical expertise cannot be assumed, they will not know medical terms or domain knowledge and will not be familiar with the data set. The general public is a broad and therefore very diverse user group, resulting in viewers with different levels of statistical expertise, as well as visual and numerical literacy. They will also have different backgrounds and contexts which will influence their interpretation of the visualizations. Furthermore, their relations to the risk and risk factors will differ, with different levels of previous knowledge, contact with the disease or hazard and different risk factors applying to them.

3.1.2 Requirements

I formulated requirements based on these user characterizations, research on risk communication and discussions with visualization experts. Considering the story authors, I identified the following requirements for the tool.

- R1: usable for different cross-sectional data sets: The tool has to be generic enough to be able to be used by different experts with different epidemiological and cross-sectional data sets, risks and risk factors.

- R2: data security: The tool should consider the security of the data sets that may contain sensitive information or have restricted access.

- R3: recommendation of important risk factors: The tool should support experts in selecting the most relevant risk factors for their target audience by providing scoring metrics indicating the significance of each factor.

- R4: support in effective risk communication and visualization: The tool should support experts in communicating the risk factors according to the intention of their data story by providing adequate visualizations and annotations following current research in risk communication.

- R5: high customizability of the visualizations: The tool should provide visualizations that are highly customizable to enable experts to create visualizations that fit the design of their data story.
- R6: easy export of visualizations: As the visualizations will be integrated into external data stories, the visualizations should be easily exportable using common formats like png or pdf.

For the resulting visualizations directed at the general public, the following requirements arise.

- R7: intuitivity: The visualizations should be intuitively understandable for a general public.
- R8: engagement: The visualizations should motivate viewers to engage with them.
- R9: accuracy: The visualizations should have an adequate accuracy.
- R10: motivation for behavior change: If the expert intends to motivate the public for behavior change, the visualizations should support that.

3.2 Design Choices

Following the requirement analysis, I derived the following design choices.

- Web-based application: The tool will be web-based to enable easy access without the need for installation. By not using a backend server, all calculations using the data of the user will be done on their local computer, ensuring data security (R2).
- Statistical Methods: The tool will support experts in selecting appropriate risk factors (R3) by providing statistical methods to identify the most relevant risk factors. This will include correlations with the risk, as well as the consideration of confounding factors and issues like small data set sizes.
- Dashboard Metaphor: The tool will use a dashboard metaphor to group the visualizations into the ones selected by the expert and the ones displayed as recommendations. This enables the expert to export the specific set of visualizations relevant to them (R6).

- Template-then-adapt approach: The tool will support experts in risk communication by providing templates considering state-of-the-art research on risk visualization. These templates will then be highly adaptable to the individual data story of the expert (R5).
- Intentions: The tool will support different intentions catering to different needs of the expert by adapting the visualizations accordingly. This will enable the expert to use the tool for multiple purposes, like the exploration of risk factors, as well as to motivate behavior change (R10), or the education of a general public.
- Simplification: The visualizations are kept simple in order to make them more intuitive (R7). This results in creating simple visualizations for each fact and risk factor, for eaxample by using common graph types like bar charts and pie charts.

3.3 Data Processing

I will now go through the individual parts of the tool, starting with the data processing. When the data set is uploaded, there are multiple computational tasks that need to be performed. In this section I will describe the initial pipeline for first creating groups and bins for each factor, selecting from them the risk groups of each factor, determining similarity with other factors and finally selecting and scoring potential risk factors.

For the scope of this work the data set is assumed to be cross-sectional, meaning that prevalences can be inferred and time will not be considered.

3.3.1 Creating Groups or Bins

For visualizations as well as for the computation of correlations with categorical variables, the values of each factor in the data set are sorted into groups or bins. As there is no prior knowledge which columns are categorical and which are continuous, the following assumptions are introduced.

1. Factors with less or equal than ten unique values are considered to be categorical. The unique values are directly used as groups. This threshold is chosen to prevent the creation of too many groups, which would make the visualizations too complex.
2. All remaining factors with more than five unique numerical values are considered to be continuous. The threshold is chosen to make sure that the factor is most likely numerical. Their values are binned following the binning algorithm detailed below.

3. All other factors are assumed to be too specific for each individual to be used for population-based calculations. They are therefore ignored. This includes for example participant IDs or textual comments.

The binning of continuous variables is done with the goal of creating four equal-sized bins. However, the bin size and boundaries of each bin are made to be a multiple of 5 to make the bins more readable. Additionally, bins with less than 5% of all participants or an absolute value of less than ten participants at the beginning and end of the distribution are merged with their neighbors to reduce the number of bins. The resulting higher bin sizes also enable more accurate and statistically significant computations.

Automatic computation of bins is unaware of any domain standards how a variable should be binned. Therefore, I decided to give the user full customization over these bins and enable them to freely split, merge or change them.

Missing values are handled by creating a separate bin for them. As the amount of missing values in the data set is probably not very relevant for the general public and may introduce confusion, I decided to automatically exclude them. However, the user has the option to include them if they choose to do so.

3.3.2 Selecting Risk Groups

The next step of the initial data processing is the selection of risk groups. Risk groups are used to separate a factor's bins into two binary categories of persons *in the risk group* and *not in the risk group*. This binary classification can then be used to compute conventional risk measures for scoring of factors like the prevalence odds ratio and relative risk, as well as for annotations and to compare factors to each other in contextual visualizations.

Risk groups are classified by a simple heuristic, as there is no knowledge available beforehand on which percentages are considered high risk by the expert. First, the mean of the highest and lowest likelihood of the bins is calculated. Then, all bins below are considered *not in the risk group*, and bins above considered *in the risk group*. The user is able to interactively change which bins should be part of the risk group and which not.

3.3.3 Finding Similar Variables

Multiple parts of the tool require finding similar factors to a given factor. Similarity computations are used to find potential risk factors, find similar factors already inside the dashboard or similar factors in general. Similarity is evaluated using pearson correlation for two continuous variables or Cramer's V when at least one categorical variable is involved. When the computation of Cramer's V involves a continuous variable, its bins are used.

3.3.4 Finding Potential Risk Factors

In this subsection, I want to describe how factors are scored and recommended as potential risk factors to the user.

Before describing my calculation, I first want to address the distinction discussed by Schooling and Jones [51] between predictive and explanatory risk factors. In my work, the term "risk factor" is used in a predictive sense as it is based on associations and does not imply a causal relation. However, the task of the tool is not to find vulnerable subgroups in the data, but to educate the public about health risks. This requires simplification, the most obvious being that the visualizations for the general public are being kept simple by only considering one factor at a time and not their interplay. While I will still train a model that takes into account this interplay, my goal doing that is to recommend a more diverse set of risk factors to the expert rather than creating perfect predictions. Although it would be interesting for the user to find causal explanatory relationships in the data, it is neither possible with the available data nor within the scope of this work. Fortunately, an expert will create the final data story and can assess which variables only have a correlation and which may be causal, making it unnecessary to exert the required effort.

For recommendation of risk factors, the user can choose between four scoring metrics that I selected based on common effect measures for associations between factors. For more details on their calculation, the reader is referred to Section 2.1.1.

- The relative risk score sorts factors by the relative risk calculated by using the classification of persons as part/ not part of the risk group.
- The odds ratio score sorts factors by the prevalence odds ratio calculated by using the classification of persons as part/ not part of the risk group.
- The correlation score sorts factors by their correlation with the outcome.
- The regression score sorts factors by the goodness of fit of a regression model to predict the outcome.

I decided to use the regression score as the default because of the following advantage. Regression analysis can take into account not only the current factor, but also any factors already selected in the dashboard. This enables the score to not only recommend factors considering how strongly they are associated with the outcome, but also considering that new factors should be diverse and different from the already selected ones. This prevents the user from receiving only recommendations for similar factors that may all be based on the same confounder, like weight, BMI and waist circumference.

Regression Score calculation

Because of its importance I want to go into more detail how the regression score is calculated.

As regression requires significantly more computational power than the other methods, it is only performed on the most promising subset of factors. The subset is determined by a minimum correlation boundary. Only factors with a correlation above this boundary are considered for the regression score. The boundary is set to include approximately 20 factors by setting the boundary to the correlation value of the factor with the 20th highest correlation. However, the boundary is set to be at least 0.05 and at most 0.9 to prevent factors from being included that are too weakly correlated and factors from being excluded that are very strongly correlated.

To perform regression on categorical factors, they are encoded using one-hot encoding. One-hot encoding also allows the encoding of missing values, which are encoded by setting all one-hot encoded values to zero. For continuous factors, missing values are replaced by the mean of the factor. Whilst more complex handling of missing values may lead to better results, this simple approach is resource-efficient and sufficient for the purpose of a regression model only used for recommendations that must not be giving the best predictions.

To handle the binary outcome variables, logistic regression is used [46]. As standard parameters, a train-to-test ratio of 0.9 to 0.1 is chosen, together with a batch size of 10. Learning rate and epochs are adapted to the individual data set by decreasing the learning rate and epoch number the bigger the data set is. This is done to increase performance on bigger data sets that may otherwise take a long time to compute. Additionally, the user is able to change these parameters.

How is regression now used to recommend risk factors? The first decision that had to be made was if one model is trained on all factors and its weights used for recommendation, or if individual models are trained for each factor. Training only one model will directly recommend a diverse set of factors. However, regression models struggle with multicollinearity, which is the case when factors used for prediction are highly correlating. In this case, the effect of each individual factor is harder to predict and the calculated weights more uncertain. This may lead to inaccurate assumptions about the underlying relationships between predictor variables and the risk that is predicted [36]. As I cannot assume that the data sets entered into my tool are preprocessed to avoid strongly correlating factors, multicollinearity is likely to occur. I also cannot know which of the highly correlating factors is the most important one to the user and their data story, making it unfavorable to automatically preselect factors.

By training individual models, this problem is avoided. However, this approach will recommend similar factors because the individual models will not be aware of which factors are highly correlating. To solve this problem, the user is included in the process. Each time the user adds a factor from the recommendations to their dashboard, this factor is used as an interaction term in the regression models for the new recommendations. This results in a stepwise feature selection with the user as the selector. New scores are then calculated by sorting the factors by the improvement in the goodness of fit of the model trained with the new factor and all already selected ones, compared to a model trained only on the selected ones. That way, the new recommendations will be diverse and enabling the user to consider interaction terms when selecting factors. Of course, the user can choose if factors should not be considered as interaction terms.

To measure the goodness of fit of the logistic regression model, which is then used as the factor's score, I considered multiple measures. All measures are computed on a separate test set not used for training. The first measure I considered was the accuracy of the model, which is the percentage of correctly predicted outcomes. However, given the nature of cross-sectional studies, the data is probably very imbalanced. This may result in models neglecting the minority class as high accuracy can be achieved by just predicting the more common class. To prevent this, I considered the F score, which considers all classes by calculating the harmonic mean of precision and recall [56]. However, my decision previously to train separate models for each factor in practice often led to factors not having a strong enough influence to change the prediction, despite having a positive effect. Therefore, I decided to not consider the final class prediction, but the predicted sigmoid value of the regression model. The sigmoid function used in logistic regression transforms the output of the model to a prediction value between 0 and 1. For the final class prediction this value is then rounded to 0 or 1. By using the sigmoid value directly I can consider, even if the class prediction stayed the same, if the models prediction was closer to the true value. To calculate the goodness of fit, I therefore calculate the mean error on the sigmoid values of each data set item of each class and then, to achieve equal weighting of both classes, take the mean of the two errors.

3.4 Fact Selection

In this section, I will describe the process of selecting the facts that are later visualized and presented to the user. To select facts, I considered which information might be interesting to the expert user and their audience, as well as the narrative structure of their possible data story. I found three basic components: General information, individual risk factors, and contextual information.

3.4.1 General Information

First of all the user will want to tell the audience some general information about the risk or hazard itself. Most information about the risk, e.g. in case of a disease, its symptoms and therapy, will have to be provided by the story author themselves. However, the tool can provide prevalence information calculated from the data set. This information can be used to inform the viewer on how common the risk is.

When using data-based visualizations, the user may want to present information about the data set itself to their audience. Whilst again information like the data source and time of collection must be provided by the user, the tool can provide information about the number of participants in the data set. This information can be used to give the audience an idea of how representative the data set is.

3.4.2 Individual Risk Factors

Next, the tool will provide information about the individual risk factors. To select facts for an individual risk factor I considered which information is most important to assess the likelihood and severity of a possible risk factor.

For this I was inspired by the fact scoring methods of Ding et al. [14] and Wang et al. [60] separating each fact's or pattern's relevance into its *significance* and *impact*. The *significance* of a fact considered their statistical relevance, which is based on how likely it is that the observed pattern is not due to chance or how strong the observed pattern is. The *impact* of a fact considered its practical relevance, which is based on how generally applicable the fact is, with higher scores for facts concerning bigger subgroups of the data set.

Adapting these concepts to risk factors, *significance* describes the strength of the correlation between a risk factor and the outcome, whilst *impact* describes how many persons are affected by an increased risk. Unlike Ding et al. [14] and Wang et al. [60] I decided to encode *significance* and *impact* information as two separate facts to keep each fact and the resulting visualizations as simple as possible.

Additionally to this vital information, another fact that the user may be interested in and that is available from the data set is the similarity of a factor with other factors. This information is presented only on demand to keep the initial template presented to the user as simple as possible. The calculation of similar factors is described in Section 3.3.3.

Facts not included are information about the regression used to calculate the risk factor scores. As this information may be relevant to the expert it is included in the user interface itself. However, for a general audience, this information is too technical and complex and would only distract from the data story.

3.4.3 Contextual Information

Contextual information is important to understand the relevance of risk factors and a risk itself. Setting a risk in relation to other risks can help the audience to more accurately assess the danger and likelihood of a risk. This can be done, for example, by using risk ladders or scales that order risks by their respective likelihoods, as described by Ancker et al. [3]. However, even if the provided data set includes information about other risks, the tool will not be able to identify this information as the user only provided information on which factor is the outcome. Therefore, such contextual information will not be included in the tool.

Nevertheless the tool is able to compare risk factors to each other. Instead of just providing a simple list of all risk factors, I wanted to create facts that provide more information to set the factors into context. These facts should be easy to understand by a general public and relate the risk factors to each other in a meaningful way.

The most straightforward comparison involves assessing the risk groups of each risk factor based on their individual probabilities of absolute risk. This gives a short overview of which groups of persons are most at risk. For example, persons with a BMI of over 40 may have a 33% risk, compared to persons with an age over 50 who have an 18% risk of diabetes.

In addition to these absolute risk values, the second comparison examines the relative increase in risk associated with each factor. This is accomplished by calculating how many times higher the risk is for the risk group compared to persons not in the risk group. For example, persons with a BMI of over 40 may have a 2.7x higher risk of diabetes than others, compared to persons with an age over 50 who have a 4.5x higher risk than others.

Lastly, I wanted to include information on which factors are the most important when assessing the risk. For this, I compare each factors influence on the regression model by taking their weight, or in case of one-hot encoding, the maximum weight used in the model. This fact is also very useful for the expert user to assess how the risk factors interact with each other in the model and as interaction terms.

I also considered multiple facts that I did not include in the final visualization. First of all, a comparison of the risk difference would state by how many percent points the risk in the risk group is higher than for the rest. However, percentages in this visualization are used differently than in the rest of the visualizations by stating not the percentage, but percentage point difference. As percentages are used throughout the tool, I did not want to confuse the viewer. Additionally, the viewer might interpret this visualization wrongly by assuming that the percentage states by which percentage the risk increases, therefore interpreting 20% not as a difference of 20 percentage points, but as a risk increase of 20%. I also considered including the risk ratio and relative risk, as they are common in the medical field. However, they are not intuitive to a general public and would require a lot of additional explanation.

3.4.4 Final Fact Selection

Based on the considerations above, I decided to include the following facts in the tool:

- General Information
  - Risk Prevalence
  - Number of participants in the data set
- Individual Risk Factors
  - Significance
  - Impact
  - Similarity
- Comparison of Risk Factors
  - Absolute risk
  - Relative risk increase
  - Influence in regression model

Looking at this list, I have a small number of possible facts. This enables me to encode each fact manually in the tool and consider the best way to visualize and annotate them individually. The small number of facts also makes further fact scoring algorithms unnecessary as they can all be comfortably presented to the user.

3.5 Visualization Generation

Before diving into the individual visualizations, I want to go into some general considerations when designing the visualizations. What I describe here is the iterative development of concrete visualizations. As this tool will be usable for a variety of use cases and persons, these designs will be fully customizeable. Following the template-based design methodology, the visualizations described here will be the templates from which the user can start to create their individual visualizations. The goal of these templates is to give the user the best start that can be automatically generated, with elements and designs chosen based on current research on risk communication and risk visualization. Furthermore, the visualizations were crafted through iterative development, involving regular presentations to various individuals in order to identify potential misunderstandings and challenges at an early stage. Subsequent enhancements were made based on this feedback. This process held particular significance in determining the most appropriate textual descriptions to accompany the visual representations.

The tool has been structured to offer users a variety of simple visualizations, with each individual fact presented through its own visualization, as opposed to a single intricate and convoluted visualization encompassing all the facts collectively. This was already argued for by Tufte [58] in 1990 and is supported by Ancker et al. [3] who observed that simple graphs increased likeability and are better understood, especially by persons with low graph literacy. In their study on different risk graphs, persons liked simple bar graphs most and already had first problems in accurately understanding slightly more complex visualizations like survival curves. Therefore, I restrained from designing completely new visualizations and focused on well known ones.

A crucial part of designing these visualizions turned out to be the phrasing of titles and axis labels, as a quick understanding of the graphs is important. Presenting the first prototypes to individuals revealed the importance of including all relevant information in the title, even if it may already be part of the user interface, as the interface will not be visible to the final viewers. This includes information on the current factor and the outcome. I considered formulating the title as a question, but this further lengthened the title and did not seem to improve intuitivity. The final titles are kept short and simple, but clearly state what the graph is about. This style is inspired by the work of Wang et al. [60] who also use short, descriptive titles for their visualizations, often omitting any more axis titles or legends. Like Wang et al., I also use colors-coding of words in the title to parts of the visualization in an attempt to further improve intuitivity. As all texts are based on general templates, there may be grammatical errors in the final texts. Therefore the user can easily change these texts to fix any errors or improve the texts to their liking. When adapting the texts, the user is able to use some pre-defined variables for the outcome label, the label of the current column or what each row of the data set stands for (like people or patients). This allows the user to easily adapt the texts to their needs without having to manually change the texts for each visualization, keeping consistency throughout the visualizations. The user is also able to change the color of each text fragment or make it bold or italic.

When combining multiple graphs their interplay has to be considered. Wang et al. [60] here proposed the concepts of *Inter-Consistency* and *Intra-Diversity*. In the context of fact sheet design where multiple graphs on different facts are combined, they proposed that similar facts should be visualized in the same, consistent manner. However, following *intra-diversity*, if facts are different (apart from their type), they should be visualized differently to create a more diverse fact sheet.

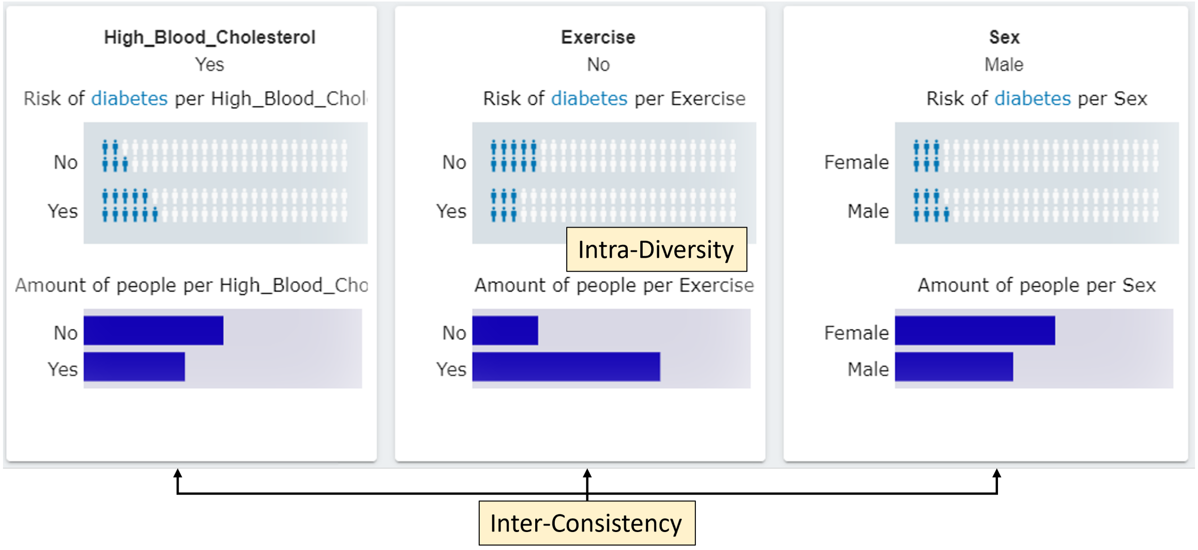


Figure 3.1 Multiple visualizations together are designed following the principle of *Inter-Consistency* to keep visualizations similar between risk factors, and *Intra-Diversity* to create for each risk factor a diverse set of visualizations.

In the context of this tool, I considered a similar mix of keeping a consistent scheme across visualizations but providing some diversity to increase engagement. Across all variables, the same graph designs are used for the same facts. This *inter-Consistency* makes it easy to switch between variables as they are all visualized in the same way. However, when looking at one variable, there is a diversity in graph visualizations created for its facts. See Figure 3.1 for the applied concept. Additionally, to keep consistency across all visualizations, a global color scheme is used, as well as for each factor the same bins and labels in all its facts.

All visualizations are designed as static visualizations. This is done to simplify the export and import of the visualizations into other tools. Research on the effect of interactivity is still sparse and mixed [57], with first evidence in risk communication showing that the visual appeal may come at a price of worse understanding [66].

3.5.1 Visualization Templates

In this subsection, I will go through the templates for the individual facts presented in the last chapter.

Impact Graphs

The impact graphs should effortlessly convey to the viewer an understanding of the distribution of individuals from the dataset across various bins corresponding to the current factor. As this is

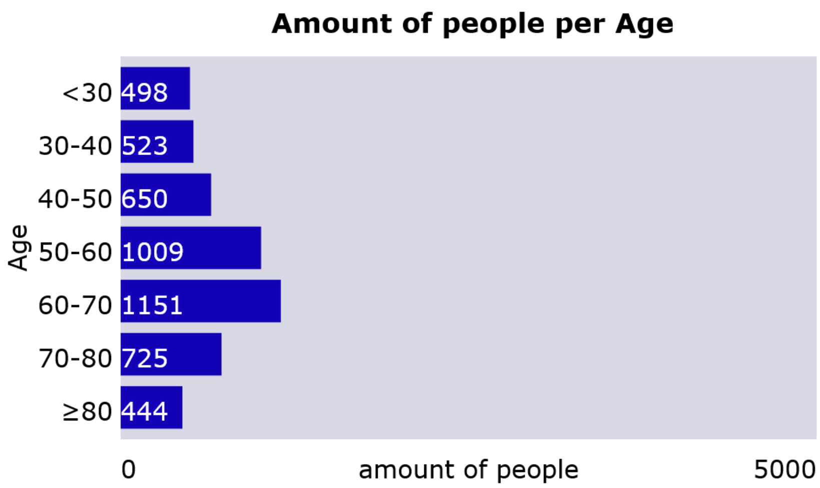


Figure 3.2 The standard template of impact graphs consists of labeled bar charts titled "Amount of people per [factor]."

very simple numerical information I decided to use bar graphs. They have a high likeability due to their simplicity [3] as well as a high accuracy due to their mapping of continuous values to the length of a bar, which is one of the best forms for humans to accurately display information [10].

The most complex part of designing this visualizitation was to come up with understandable wording. Starting from the simple title of "#people per option" first iterations with persons soon revealed the importance of clearly stating which column the graph belongs to (this information was previously only part of the user interface) and using descriptive titles that prevent confusions with other graphs. After multiple iterations the final title is "Amount of people per [factor]". The word "people" here is automatically exchanged depending on how the user defines rows should be called, for example to "patients" or "participants". The word "Amount" is generally usable for absolute values, percentages and proportions. The word "per" is technical, but is chosen as it is short focusing the user's attention on the variables named in the title.

The final visualization can be seen in Figure 3.2. Alternative graph types available to the user are pictographs, pie graphs, and text descriptions. Especially pie graphs are an effective alternative used in the intentions later. Pie graphs significantly decrease accuracy as humans are worse in judging degrees than lengths [10]. However, they are simple and liked by the public, as well as great to visualize the part-to-whole relationship described here. This is researched by Hawley et al. [20] who observed that whilst pie graphs were bad for accurate, verbatim knowledge, they were the best option for overview, gist knowledge in treatment risk communication and therefore probably also for general risk communication.

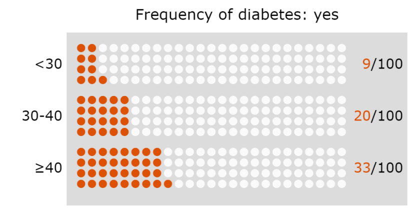
Significance Graphs

The second of the two main facts for each potential risk factor is its significance. This fact requires the visualization of the percentage of persons per group or bin of the factor that is affected by the outcome. In other words, how high the risk is per group. This information could easily be visualized again as a bar chart, taking advantage of how well-known this graph type is and its accuracy. However, for the following reasons I decided to use pictographs instead. First of all, following the principle of intra-diversity, using another graph type will increase the interestingness and engagement of the graphs. Secondly, pictographs are recommended by Hawley et al. [20] as the overall best format for treatment risk communication and looking at their arguments they are also very well suited for more general risk communication as in this work. Their study revealed pictographs as good in both verbatim and gist knowledge. This makes sense as their mapping of numbers to ordered icons still lets the human perception use length or area information for accurate assessment, but the icons provide an intuitive understanding of the subjects of the data as, for example, humans, and the part-to-whole relationship.

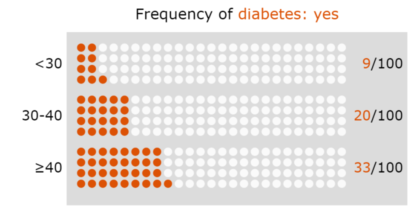
As this graph visualizes part-to-whole relationships I considered how those should be best described in textual form for the labels of the individual groups. Options were using percentages, using natural frequencies with fixed nominator, often called “1 in X” (e.g. 1 in 10 and 1 in 20) and using natural frequencies with fixed denominator (e.g. 5 out of 10 and 8 out of 10). I found clear research against the use of “1 in X” as it resulted in very low accuracy when comparing numbers [3, 57]. If percentages or natural frequencies with consistent denominator are better depends more on the given task [57]. However, as the pictographs already implicitly used natural frequencies with consistent denominators, I decided to use them with the same denominator for the labels as well. I only display the percentages to expert users who may take advantage of the exact numbers and are more familiar with percentages.

Given the extensive research conducted, I had anticipated that the pictographs would be readily comprehensible. However, I realised through numerous iterations of presenting those visualizations to individuals, that the understanding of the pictographs depended heavily on details like the used textual descriptions and labels. Because of this I want to use this graph to provide a more detailed example of how I iteratively improved the visualizations presented in this tool, see Table 3.1 . For simplicity, I reduced the process to six distinct phases.

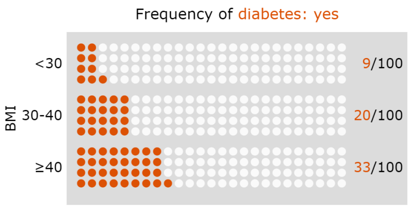
Pictographs can be designed with or without showing the denominator as differently-colored icons. As a standard practice, I chose to display the denominator, as strongly recommended [57]. This approach places the presented numbers into perspective and prevents the phenomenon of denominator neglect bias. This bias can make differences between numbers appear more significant than they actually are, especially when they are not considered alongside their



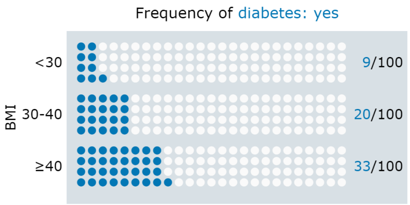
First draft. I created a pictograph with 100 icons, labels of each group and a first title stating “Frequency of diabetes: yes”. I also added the value of each group as annotation.



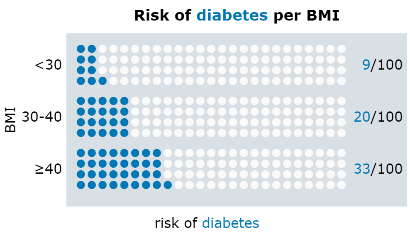
Added color-coding. Color-Coding the title increases intuitivity by creating a link between the colored icons in the pictograph and the risk they are visualizing that is stated in the title.



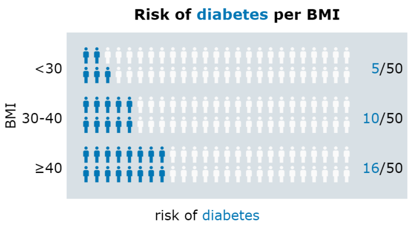
Added variable name. User feedback soon showed a need to show the name of the current variable clearly in the visualization. Before, the column was only stated in the interface, not the visualization itself. The variable name is added as an axis title on the left side.



Improved design. Visually engaging graphs increase user engagement and likeability. The background color is automatically adjusted to the foreground color. Different color schemes are made available.



Improved wording. This was the longest phase with the most user input. Choosing the right wording is crucial for the understanding of the visualization. The final title is “Risk of diabetes per BMI” with an added axis description of “risk of diabetes”



Added human icons and decreased denominator. Instead of the former circle, a human icon is more intuitive and engaging. However, it also produces more visual clutter which is more problematic the higher the icon count is. Therefore, I decided to decrease the denominator to 50.

Table 3.1 Development of the significance pictograph using the example of diabetes and the risk factor BMI.

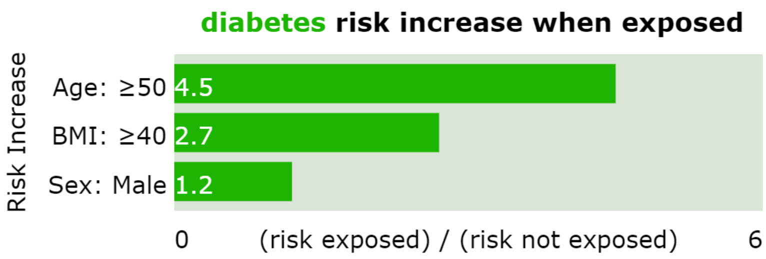


Figure 3.3 This context visualization compares risk factors by how much they increase the risk of outcome. A bar chart is chosen with one bar per factor.

denominator. For instance, consider the distinction between a risk of 3 or 4 out of 10,000. However, when looking at very small risks or wanting to increase the perceived risk, hiding the denominator might be a valuable option. Therefore, I implemented this as one of the intentions and as an optional setting.

As mentioned before, bar charts are a valuable alternative visualization for this fact. Additionally, a text visualization is provided to the user and a visualization using multiple pie charts. Pie charts are again included for their popularity and effectiveness for gist knowledge and the visualization of part-to-whole relationships. I did not include visualizations that do not use binning or only use the risk groups. However, this could be added in future work.

Context Graphs

Lastly, I wanted to consider multiple risk factors together. The basic visualization types are bar charts and pictographs because of their advantages outlined in this section. Bar charts are also recommended by Ancker et al. [3] for risks in context. Here again, the wording played a crucial role for users and a general audience to understand the visualized information.

To visualize the risk increase, I decided to use bar charts as they are the most accurate visualization type and there is no part-to-whole relationship to be visualized. I used as title the wording “diabetes risk increase when exposed” to avoid confusion between the risk groups of the factor and the standard groups. The axis describes the calculation as “(risk exposed)/(risk not exposed)”. The final visualization can be seen in Figure 3.3.

To visualize the absolute risk, which is given in percentages, I used a pictograph in line with the significance visualizations. This familiarity should aid in graph reading. The title uses the phrasing “Risk of diabetes when exposed”, again using the term “exposed” instead of “risk groups” to avoid confusion with the risks of the standard groups or bins of the factor. The final visualization can be seen in Figure 3.4.

The importance of each factor on the regression model is visualized using a bar chart, as again no part-to-whole relationship is visualized. The title uses the phrasing “influence in

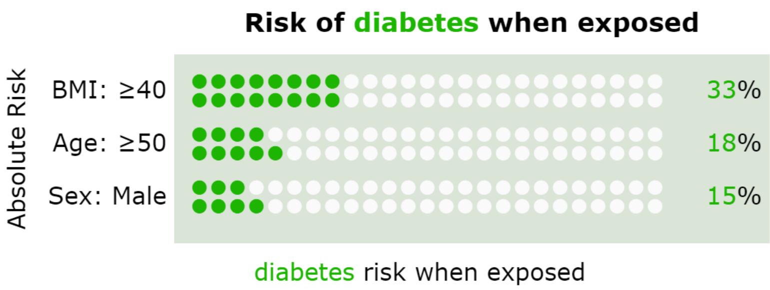


Figure 3.4 This context visualization compares risk factors by the absolute risk of persons in the risk group. A pictograph is chosen for consistency with the significance visualizations.

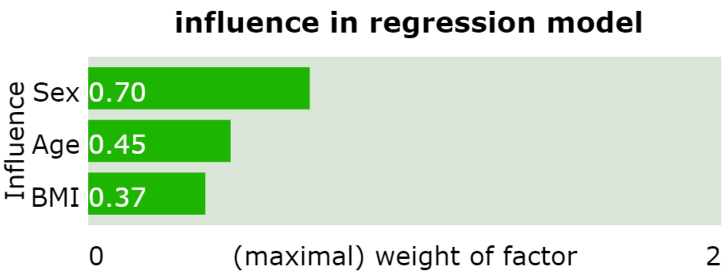


Figure 3.5 This context visualization compares risk factors their influence in the regression model. A bar chart is chosen with one bar per factor.

regression model" instead of using the technical term "weight" to choose a wording that is easily understandable by a general audience. The term weight is just used in the axis description to provide an understanding for the expert user and persons with a technical background. Opposed to the other two context visualizations, this graph compares the variables, not their risk groups, and therefore only the variable names are stated as labels. The final visualization can be seen in Figure 3.5.

Additional Graphs

I will now shortly explain the visualizations for the remaining facts and graphs.

Starting with facts containing general information, the first fact describes the outcome prevalence. When considering how to visualize information on the risk prevalence as part of the general information category, I realized that this required the same graph type as the impact graphs of the potential risk factors. In consequence, to keep consistency, for risk prevalence the same graph is used.

The other general information fact provided is the number of participants. I decided to visualize the number of participants simply as a textual description as this is a single number without any context or anchors that could be used for visualizations.

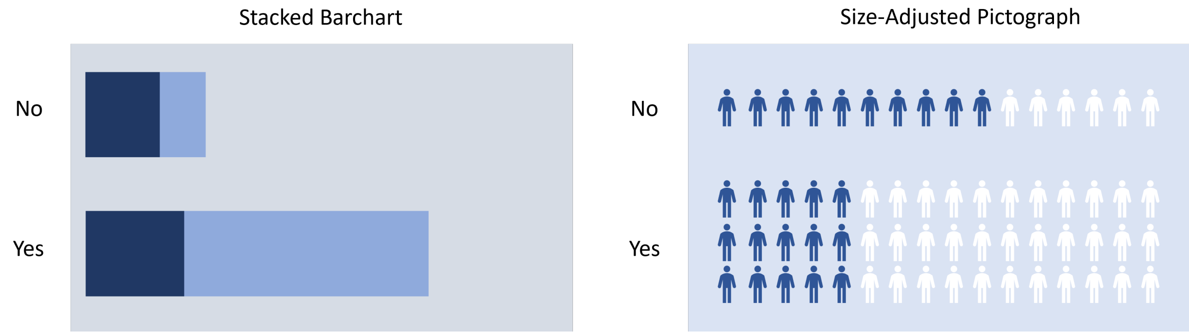


Figure 3.6 Visualizations combining significance and impact facts were considered but not included. This entails stacked bar charts (left) and size-adjusted pictographs (right).

Another important aspect to consider for risk factors is their similarity. I decided to visualize correlations of one risk factor with others simply as bar graph to take advantage of their accuracy.

Finally, for every risk factor, I have also included the option to input personalized text descriptions if the user wishes to provide supplementary information.

Not Included Visualizations

I considered, but decided against more complex visualizations combining significance and impact. First of all, as described at the beginning of this section, multiple simple visualizations are more effective than one complex one. Secondly, even as optional additional visualizations, the following visualizations turned out to be too hard to read and understand.

The first considered visualization were stacked bar charts that color-code the impact bars with the amounts of persons per outcome risk, see Figure 3.6, left. This visualization was hard to read as the user had to manually compare multiple differently sized areas. Especially for groups with small numbers of participants and thus, a small bar, the individual groups would hardly be recognizable.

The second considered visualization adapted the significance pictograph to have the icon number proportional to the amount of persons in the group, see Figure 3.6, right. Even when this was uniformly scaled down, groups varied in participants between numbers like 3 and 500 which would have simply not been readable anymore.

3.5.2 Intention

As mentioned previously, the best visualization always depends on the given context or task. I tried to consider that in my work by adapting the standard templates of the visualizations to three different intentions. My chosen intentions are *Explore*, *Convince* and *Educate*.

The background for this is the review of Ancker et al. [3] comparing graphs for risk communication for the three outcomes of accuracy, behavior change and likeability. They found that the best graph design indeed depends on the given goal, with, for example, pictographs with many icons being better for accuracy but less liked by users than pictographs with fewer icons. Another study by Hawley et al. [20] looked at the outcomes of verbatim knowledge, gist knowledge, patient ratings and chosen treatment. They also found differences like pie charts being very liked and better for gist knowledge but worse for verbatim knowledge. Additionally, Ancker et al. [3] also found that results are impacted by the user's education, culture, numeracy and graph literacy.

For this work, I have drawn from their results to settle on the three intentions *Explore*, *Convince* and *Educate*. For the following reasons I decided against using their outcome types directly. First of all, in real applications all outcome types are necessary to varying degrees. In our case, the expert wants to portray accurate information and optionally influence behavior but still keep the users engaged by creating graphs with high likeability. Also, goals vary between the different target groups of the tool, with the expert wanting accuracy but the general audience requiring likeability. Secondly, the expert using the tool may not be an expert in risk communication and therefore be biased themselves by liking graphs more that score high on likeability without being aware how they might decrease accuracy. So, instead of focusing on individual measurements, my goals are inspired by possible intentions of the expert researcher themselves. First the expert user will want to explore the data set themselves, which will be supported by the *Explore* intention. Secondly, the expert may want to simply educate or change the behavior of the general audience to varying degrees. With both goals often standing in conflict to each other, I decided to separate them into the *Educate* and *Convince* intentions.

To implement the intentions, I mainly changed the graph type, context visibility and unit as there was sufficient research supporting these decisions. With the intentions targeting multiple goals, some design decisions are based on my hypotheses and have to be evaluated in my evaluation and in future work.

Explore

The *Explore* intention is targeted to the needs of the expert user themselves. They will mainly require high accuracy to make educated decisions. Behavior change is not necessary as that is not part of their task. Likeability is important to consider as the expert user will be influenced by that, but not as necessary as for a general audience as the expert will have motivation to use the tool regardless. Additionally, intuitivity is not as essential as the expert has more time to get used to the visualizations.

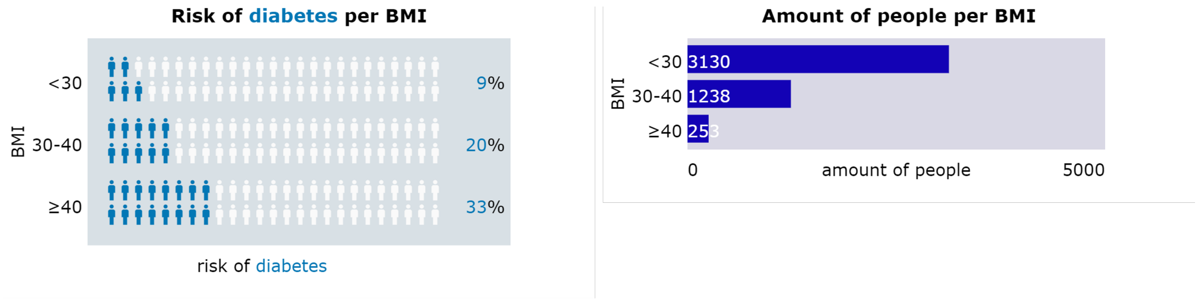


Figure 3.7 The *Explore* intention uses pictographs with contexts for significance facts and bar charts with absolute values for impact facts.

For the highest accuracy, the context is shown in all visualizations and the labels are given as percentages. Percentages are used as it is assumed that the expert has enough time to understand the pictograph without the additional hint of using natural frequencies. Percentages are then the better option as they may have an advantage when comparing chances [57] and experts are probably more used to them. The impact graph is shown as a bar for high verbatim knowledge, with the exact numbers of persons per bin as in the data set. This enables the expert to evaluate closely if the number of persons in a bin is high enough to trust further analysis based on them. The final images are visible in Figure 3.7.

Convince

The *Convince* intention is targeted to a general audience with the goal of changing their behavior. To be effective, this additionally requires high likeability for a general audience to stay engaged. However, information should still be accurate to not mislead the audience.

The most important change here is to hide the context information so that differences between groups appear bigger. In the significance pictographs, the denominator is still shown in the labels but only the nominator is displayed using the icons. This results in only the icons representing persons who are affected by the outcome being shown, creating space for the shown icons to appear bigger. To make this possible, natural frequencies are used for the labels, which are, next to percentages, another great option for general audiences [57]. The impact graph is also scaled up to make the longest bar almost fill the entire visualization width. This approach further amplifies the visible distinctions between the groups, thereby accentuating the differences. The final images are visible in Figure 3.8.

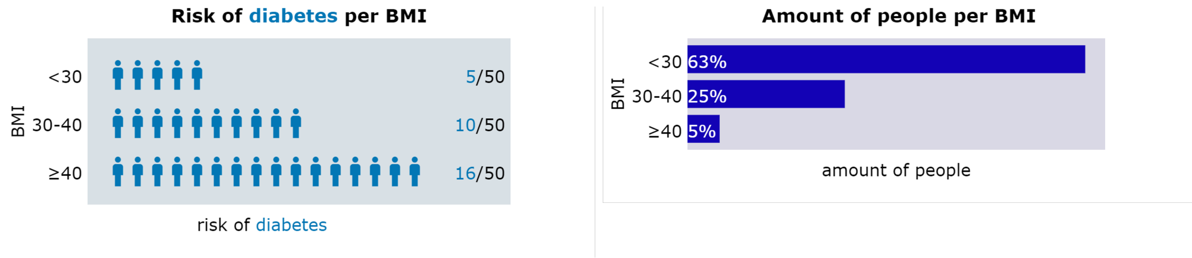


Figure 3.8 The *Convince* intention uses pictographs without context for significance facts and bar charts with percentages for impact facts.

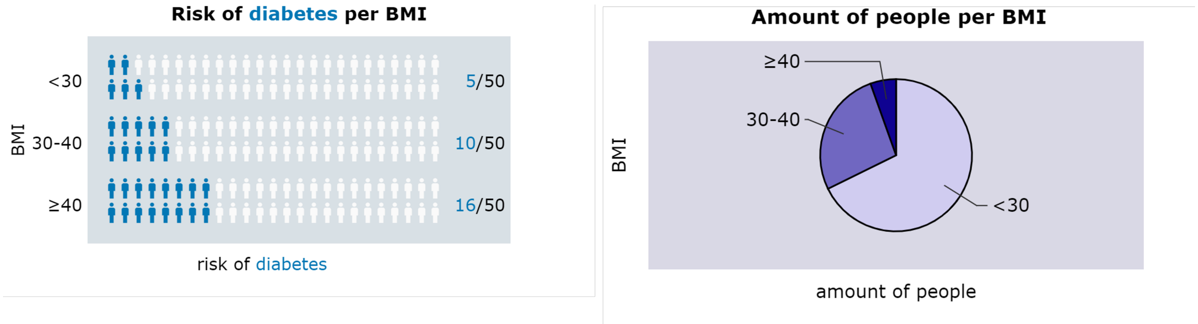


Figure 3.9 The *Educate* intention uses pictographs with contexts for significance facts and pie charts impact facts.

Educate

The *Educate* intention is targeted at a general audience to educate them about a risk. The focus here is on intuitivity and likeability to keep the audience engaged, as well as gist knowledge to convey the main messages.

The context is shown for a more accurate interpretation of the knowledge. Natural frequencies are shown to improve the intuitivity of the graph and numbers. The most noteworthy change in this intention is the use of pie graphs for the impact visualization. Pie graphs are chosen as they are very well liked by the public and provide strong gist knowledge [20]. They also provide an intuitive understanding of the part-to-whole relationship. This comes at the price of being not as accurate in verbatim knowledge as other graphs. The final images are visible in Figure 3.9.

3.5.3 Styling

The styling of the visualizations is important for multiple reasons. Having a consistent style reduces the effort needed by the viewer to switch between and compare risk factors [25]. It also enables customizability to the user's data story and preferences. I also expect a well designed style to improve likeability and therefore engagement of the general audience.

The starting point of my styling options was the fact sheet styling by Wang et al. [60] using seven color schemes and easy customization.

I also integrated multiple predefined color schemes commonly used for visualizations, like the inbuild color schemes of d3[1]. However, they were not designed for coherent styling but for maximum differentiability of colors. Therefore, I decided to additionally provide the user with the option to choose a color scheme based on analogous color schemes [42]. These schemes use colors next to each other on the color wheel and therefore provide a more coherent look. They have the same brightness and saturation and only vary in hue to create consistency and improve readability in the graph. The user inputs a starting color and uses a slider to choose the spread, defining how similiar or different the other colors should be chosen. Then, a color scheme of neighboring colors is created automatically, see Figure 3.10, left. This not only enables the easy creation of different styles but also enables the user to closely match the color scheme to their data story, for example by using one color as their starting color. Of course, the user can customize the pre-defined color schemes as well as the analogous color schemes afterwards by changing each color individually.

The background color of the visualizations is also customizable. Standard options include colors and hues that should work with a variety of different foreground colors, like grey and white. I also wanted to enable the user to set colored backgrounds, without the color interfering with the foreground colors. As the foreground colors inevitably vary, I decided to include a feature that automatically picks a fitting background color for the foreground color used in the graph. This is accomplished by keeping the hue fixed but reducing saturation and increasing brightness. The user is also able to choose a custom background color. For the final choices, see Figure 3.10, middle.

Additionally, the user is able to choose a font color and font family of their liking for the text visualizations, again with pre-defined options and a fully custom option, see Figure 3.10, right.

3.6 Annotation Generation

This section will focus on how the visualizations are improved by automatically adding annotations. I will consider two types of annotations. First of all, a summary annotation can combine and sum up the most relevant information portrayed in the visualizations of each risk factor as a separate text description. Secondly, the individual visualizations can be enhanced with annotations. Both will be discussed separately in the following subsections.

[1] https://github.com/d3/d3-scale-chromatic

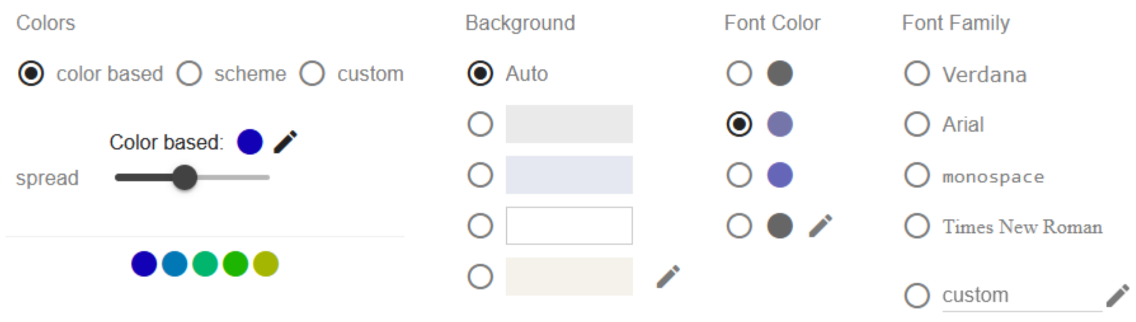


Figure 3.10 The design settings allow the creation of a color scheme, as well as customization of the background, font color and font family.

3.6.1 Summary Annotation

In this work, each possible risk factor is visualized using at least two visualizations, one for the significance of the risk factor and one for its impact. In order to support the user in understanding the visualizations and how they play together, a separate summary annotation is generated. The annotation is generated based on a template that is filled with the individual risk factor information and can be adapted by the user. To design this template, I had to decide which information is the most relevant for the user. As this work is about risk factors, an important information is which group is at risk. However, the risk of the risk group only makes sense in relation to the risk of the rest of the population. I wanted to stay close to the visualizations in order to aid in their readability, so I decided to approximate the risk of the rest of the population with the risk of the most common bin. This also helps to connect the significance visualization with the impact visualization stating how common each group is. In the case of there being no significant differences in risk likelihood, I just state the most likely bin and that there are no significant differences. An example of such a summary annotation with the accompanying visualizations is shown in Figure 3.11.

3.6.2 Graph Annotations

For the annotations of individual graphs, I first want to go into some general considerations and then look at each graph type separately.

There are many different visual cues like arrows, lines or shapes [29, 30]. In this work, I only consider textual annotations and use lines as visual cues to connect the annotations to the corresponding bins of the visualization. Further forms of visual cues might be considered in future work. I decided to use external annotations added on the right side of the visualization because they are easier to implement as they do not interfere with the visualization itself and

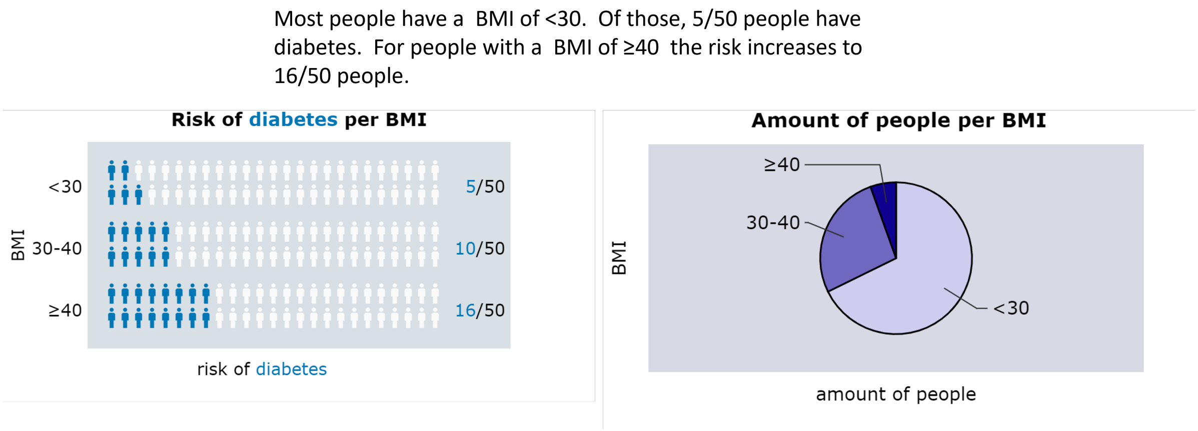


Figure 3.11 Summary annotation with accompanying visualizations.

are generally usable. However, a study by Kong et al. [29] showed that internal cues are more effective than external visual cues. Therefore, in future work, internal cues should be considered.

Annotations can be used for multiple different purposes. In this case, they can provide additional information not shown in the current visualization (additive annotations), highlight observations in the current visualization (observational annotations), describe which groups are currently considered risk groups or be used as narrative techniques. As all of these functions are useful depending on the task of the user, I wanted to provide the user with the ability to choose between them. However, as the standard annotation I generally decided to add an annotation that highlights observations in the current visualization. Using observational annotations as a first choice is unusual, as previous works often recommended to use additive annotations instead [17, 24, 27]. However, in my case additive annotations as first choice just confused the viewer on how to interpret the visualizations. Using observational annotations instead provided the viewer with a clear starting point on how to read and interpret the visualization. This phenomenon might arise from the fact that this work employs a combination of multiple visualizations, necessitating their utmost simplicity. In contrast, prior studies mostly utilized a single visualization, affording viewers more time to familiarize themselves with its interpretation. The way I use observational annotations is more inspired by the work of Wang et al. [60] using textual descriptions next to visualizations to describe and explain them. Future work could investigate if my observations are correct and if observational annotations are generally more useful in scenarios with multiple visualizations. I also leave to future work the use of annotations for narrative techniques, as this would expand the scope of this work too much. For now, I provide the user with the ability to add custom annotations for this purpose.

As I generally prefer observational annotations and have a fixed list of annotations, I kept the scoring of annotations to determine the best annotation simple. I manually selected scores for each annotation type and choose from all included annotations the annotation with the highest score.

For simplicity, annotations are generally indifferent to most attributes of the visualization. The only exception is the unit used in the visualization as this is important for consistency between visualization and annotation.

Annotations are generated for significance graphs, impact graphs, similarity graphs, and context graphs. I will now go into more detail on each type separately.

Impact Graph Annotations

For impact graphs, I found two observational annotations that might be of interest to the user. The first annotation highlights the bin with the most persons in it. This hightlights the bin that is most common and therefore most relevant to the highest amount of persons. It also gives an example how the visualization can be read, which is crucial for graph understanding. Therefore, I rated it highest from all annotations. The second annotation highlights all bins who have less than 100 persons and who therefore might not be of a significant enough size for further analysis. In terms of additive annotations, I added an annotation stating the amount of persons in the current risk group to be used in alignment with the significance graph.

For the example of a diabetes data set with risk factor BMI, the following annotations will be generated.

1. "Most people have a BMI of <30"
2. "These groups each have fewer than 30 people"
3. "253 people have a BMI of >=40"

Significance Graph Annotations

For significance graphs, the observational annotation simply describes the bin with the highest risk, stating their name and risk likelihood. This highlights the most important information and gives an example of how the visualization can be interpreted. However, I actually rated this annotation the lowest for the following reason.

Significance graphs are based on calculations on top of the information available on each individual group. This makes additive annotations more useful as they help to interpret the reliability of the visible information. The most important information here is therefore the annotation informing the viewer when the visible differences between bins are not statistically relevant. If they are statistically relevant, the highest-scoring annotation selects all bins belonging to the risk group and states their names and risk. This not only imparts information to the viewer regarding the groups most vulnerable to risk, but also again helps to interpret the visualization by giving an example how it can be read. Because of this function as an interpretation example, I decided against displaying the mean likelihood of these groups but chose the smallest likelihood of the risk groups instead, as this number will also be visible in the graph. For example, when the risk group contains likelihoods of 20% and 30%, I will state that the likelihood is "20% or higher". The last additive annotation displayed to the user marks all bins whose calculation refers to bins with under 100 persons. This helps the user to interpret the reliability of the visible information.

For the example of a diabetes data set with risk factor BMI, the following annotations will be generated.

1. "Not statistically relevant!" or "people with a BMI of >=30 have a 20% or higher risk of diabetes"
2. "Based on only 2 people."
3. "29% of people with a BMI of >=40 have diabetes"

Similarity Graph Annotations

For similarity graphs, I only found one relevant annotation. This observational annotation simply states the names of the factors with whom the current factor correlates strongly with, again aiding in readability.

For the example of a diabetes data set with risk factor BMI, the following annotations will be generated.

1. "BMI strongly correlates with Weight Category"

Context Graph Annotations

For the context graphs, I decided to use one observational annotation each giving examples as to how the graph can be read. I decided on this simple approach as each context visualization is only used once and therefore the user has the least amount of time to learn how to read them. Additionally, user tests revealed them to be rather difficult to understand at first glance. With each context graph showing different data I created one annotation each.

For the example of a diabetes data set with risk factors BMI and Fruits per day, in case BMI is rated the highest in all context visualizations, the following annotations will be generated.

1. "people with BMI: >=30 have a 2.5 higher likelihood of diabetes than the rest"
2. "people with BMI: >=30 have a 22% likelihood of diabetes"
3. "BMI has the strongest influence on the model"

Chapter 4

Implementation

In this chapter, I will describe the implementation of my tool RACCOON. The name RACCOON is chosen as an easily rememberable acronym for Risk fACtor COmmunicatiON. The website is hosted on github pages at https://akleinau.github.io/raccoon/. The source code is available on github at https://github.com/akleinau/raccoon. I will first present the data set that is used in the tool and throughout this thesis as an example. I will then describe the architecture of the tool and how it is implemented. Finally, I will describe how a user can interact with the tool.

4.1 Data Set

Whilst the tool is usable with all epidemiological data sets that are in a CSV format, it is specifically designed for data sets of cross-sectional studies. To provide a user with an example data set to explore the tool, as well as to have a data set that the tool can be evaluated with, I searched for publicly available data sets of cross-sectional studies. I finally decided to use data set of the BRFSS study [16].

The Behavioral Risk Factor Surveillance System (BRFSS) [16] is a yearly telephone survey conducted by the Centers for Disease Control and Prevention (CDC) in the United States. The survey collects a multitude of data on different risk factors, related behaviors, diseases and preventative actions. The data is publicly available with multiple preprocessed data sets containing a subset of the information provided on kaggle.com[1]. I am using the data from 2015 as it was most easily available in csv format.

The data set provided as an example on the tool starting page is a subset of the BRFSS data set containing diabetes risk factors. I created the diabetes data set by selecting columns from

[1] https://www.kaggle.com/datasets

the BRFSS data set that vary from being very relevant to less relevant for diabetes, in order to show the user the capabilities of the tool to determine the relevant risk factors. I preprocessed the selected columns to have descriptive names and values as in this case the user will probably not be familiar with the data set. If the user is interested in learning more about the data set or downloading the csv file, links are presented on the starting page of the website.

I selected the risk of diabetes as it has high prevalence rates and is a chronic disease, making it an important risk to educate a broad audience about. The data set contains the following columns:

- the categorical variables *diabetes*, *weight category*, *general health*, *sex*, *exercise*, *smoker*, *high blood pressure*, *high blood cholesterol* and *physical activity*. These variables have two to five bins.

- the continuous variables *BMI*, *age*, *fruit per day*, *vegetable per day*, and *alc per day*

The categorical columns *exercise* and *physical activity* correlate strongly with each other. Depending on the bins of the continuous column *BMI*, it may also correlate with the categorical column *weight category*. To ensure a good performance on all devices, a subset of 5000 participants was randomly chosen.

4.2 Architecture

The architecture implemented for the tool is shown in a summarized format displaying the most important components in Figure 4.1. It consists of multiple components that are responsible for different aspects of the tool. The implemented architecture is inspired by the similar architecture from the visualization generation system *Voyager* [62] that separates the interface from the recommendation and visualization components.

My architecture is structured as follows. The user views and interacts with the frontend of the tool. This interface interoperates with the backend of the tool by providing user input and receiving recommendations and visualizations. The components of the backend can be roughly connected to the different steps of the risk factor pipeline from risk factor calculation over fact selection and visualization generation to annotation generation.

The first part of the interface is the dashboard. User-selected risk factors and visualizations are collected here. They can be further customized and exported.

The settings component of the interface is responsible for allowing the user to customize global aspects of the pipeline like how risk factors are calculated, which intention is used for the visualizations and how the overall design should look like.

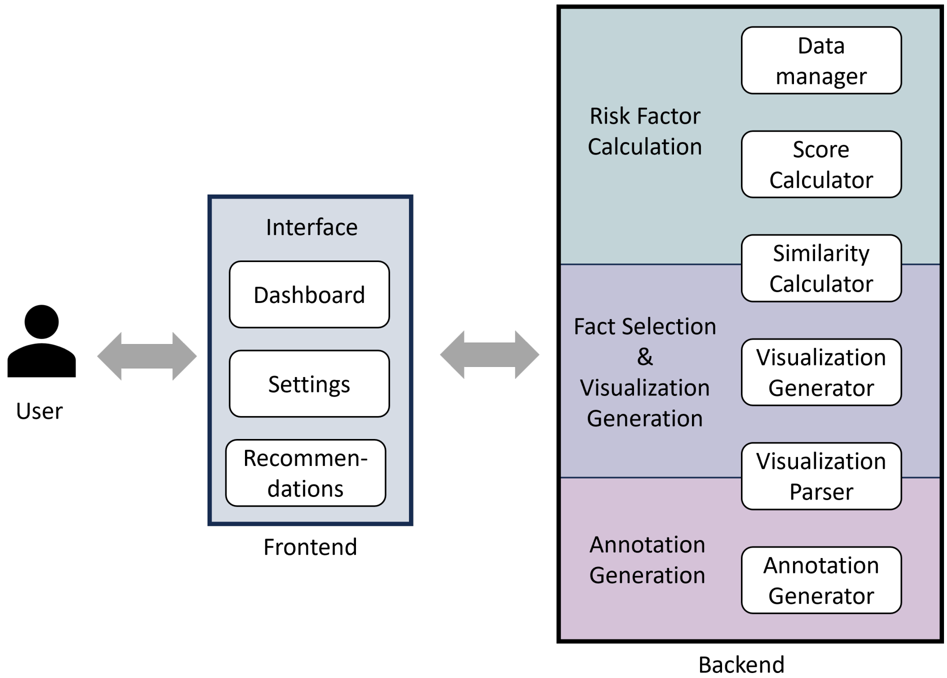


Figure 4.1 Architecture of Raccoon. The user interacts with the user interface, which then interacts with the other components of the tool. The components can be roughly separated into the individual phases of the risk factor pipeline.

The final part of the interface is the browser. The user receives recommendations of visualizations and risk factors here. Risk factors are recommended according to their score, additionally general and context visualizations are provided.

The first component of the backend is the data manager. It is part of the first phase of risk factor calculation. The data manager prepocesses the data set that is provided as input. The component manages the information of each column of the data set. A summary is created for each column to calculate all necessary information to determine the potential of the column as a possible risk factor for the outcome. The data is summarized so that subsequent visualizations of individual risk factors require minimal further computation. This entails the creation of data structures for significance and impact information, and binning of continuous variables.

The score calculator component compares the information of each column provided by the data manager to sort the potential risk factors according to a scoring metric. Metrics provided are regression, correlation and the maximal difference in significance of the groups of a column. It is part of the first phase of risk factor calculation. A separate part of the score calculator component designated to the regression metric is the regression calculator. It uses logistic regression to calculate risk factor scores considering already selected factors as confounding factors.

The similarity calculator component calculates correlations between variables to determine their similarity. Similarity between variables and the outcome is used in risk factor calculation for a pre-selection of promising risk factors. It is also used in the fact selection and visualization generation phase to recommend similar factors to the user and provide the according visualizations.

Based on the data of each column, visualization descriptions are created by the visualization generator component. The descriptions are saved as formal language descriptions. Here, each fact and their according standard visualization are manually defined, combining the phases of fact selection and visualization generation.

The visualization parser is responsible for parsing the formal language descriptions of the visualizations to create visual mappings, and combining them with the provided annotations to create the final visualizations. It is therefore part of the visualization generation and annotation generation phases. The final visualizations are created as SVGs.

Finally, for each visualization, appropriate annotations are created and scored by the annotation generator component.

4.2.1 Implementation of the Architecture

My tool is implemented as a static website using html, css and javascript. It was developed using the javascript framework Vue[2]. With all data processing being performed locally on client-side, data protection is ensured. No data is sent to a server. For UI elements, the library Vuetify[3] is used. For visualizations, the javascript library d3[4] was used. PDF exports are possible through pdfmake[5]. The tool is usable on all devices, but only optimized for the screen sizes of laptops and desktops. Due to some export functions not working properly on Mozilla Firefox, the browsers Google Chrome and Microsoft Edge are recommended.

4.2.2 Implementation of the Visualization Generation

The implementation of the visualization generation required careful consideration of how global design settings can be implemented despite each visualization having the option for local changes. The goal of this architecture was to enable full customizability of each visualization whilst keeping consistency between the visualizations a priority. This is implemented by only saving for each visualization the attributes that are specific to that individual visualization. All other information is saved in the global formal language description of that visualization type.

[2]https://vuejs.org/
[3]https://vuetifyjs.com/en/
[4]https://d3js.org/
[5]http://pdfmake.org/

The visualization parser then fills in any missing information in the individual visualization description using the global description. This way, any update to the global description will update all previously generated visualizations seemlessly. The user is able to choose with each customization of a visualization if this customization should stay specific to the current visualization or if it should be applied to the global description and therefore applied to all visualizations of the same type.

The descriptions in the formal language can be characterized as a collection of key-value pairs that elucidate the visualization. They are saved in JSON-format. For example, a json object describing a bar chart might look like this:

```
{
    "graph": "bar",
    "size": 1,
    "range": [0, 100],
    "axis": [{"text": "amount of people", "color": "black"}],
    "title": [{"text": "People per BMI", "color": "black"}],
    "unit": "percent",
    "context": true,
    "color": 0
},
```

4.3 User Experience

In this section, I want to describe how the user can interact with my tool RACCOON. I will start with how a typical workflow may look like and the accompanying views, followed by a description of how the user is guided through the tool and the creation of their presentation or data story.

4.3.1 Typical Workflow

The first view the user sees when opening the RACCOON website is the start screen. Here, the user can first upload their data set or select the example data set. Then, they specify the outcome variable and select the intention they want to start with. After clicking the calculate button, the tool will start the initial computation and the user will be redirected to the dashboard.

The dashboard will contain previews of all currently selected facts and visualizations. As a starting point, the prevalence information on the outcome is already added. The dashboard is shown in Figure 4.2. Directly below the dashboard are settings to change global aspects of the

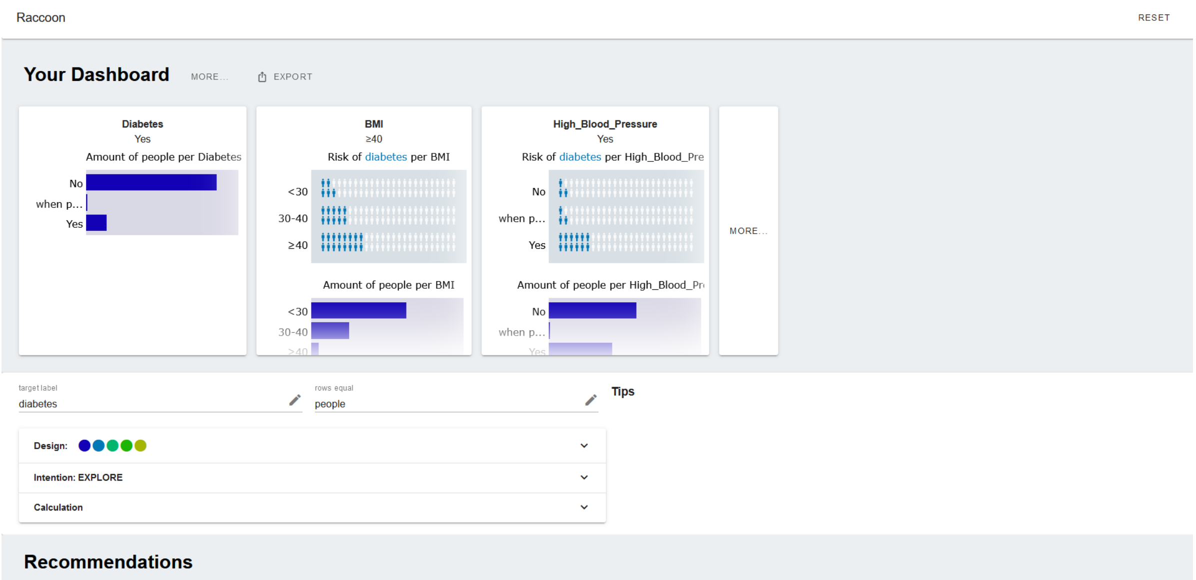


Figure 4.2 The dashboard with three selected fact groups. For each group, a preview of their visualizations is shown. Below the dashboard, the settings are visible, as well as tips on how to improve the visualizations.

tool or visualizations. In a *design tab* the user can change the overall color scheme and font of the visualizations. In an *intentions tab* the user can change the current intention. Lastly, using the *calculations tab* the user can change the data processing and regression parameters. The user is also able to change the phrasing used in the visualizations for the outcome and rows. Next to the settings are relevant tips listed how the visualizations might be improved. They will be explained in further detail in Section 4.3.2.

When scrolling down, the user will be presented with fact group recommendations. Again, to save space, all visualizations are shown as simplified previews. Fact groups are used to group visualizations by topic, for example with one fact group per risk factor containing a significance and an impact visualization. The recommendations will include the five most relevant risk factors according to the selected score. As a standard, the regression score is selected. There is also a button to open an overlay displaying all potential risk factors and providing a functionality to search factors by name. Fact groups can be added to the dashboard. The recommendation view is shown in Figure 4.3.

When clicking on a fact group, a detailed fact view is opened showing the complete visualizations and multiple customization options. For each risk factor, for example, the user can modify the groups or bins of the factor and add or remove visualizations. Additionally, statistical information on that risk factor is provided and similar risk factors are recommended. When selecting an individual visualization, additional customization options specific to that graph are available. The fact view is shown in Figure 4.4.

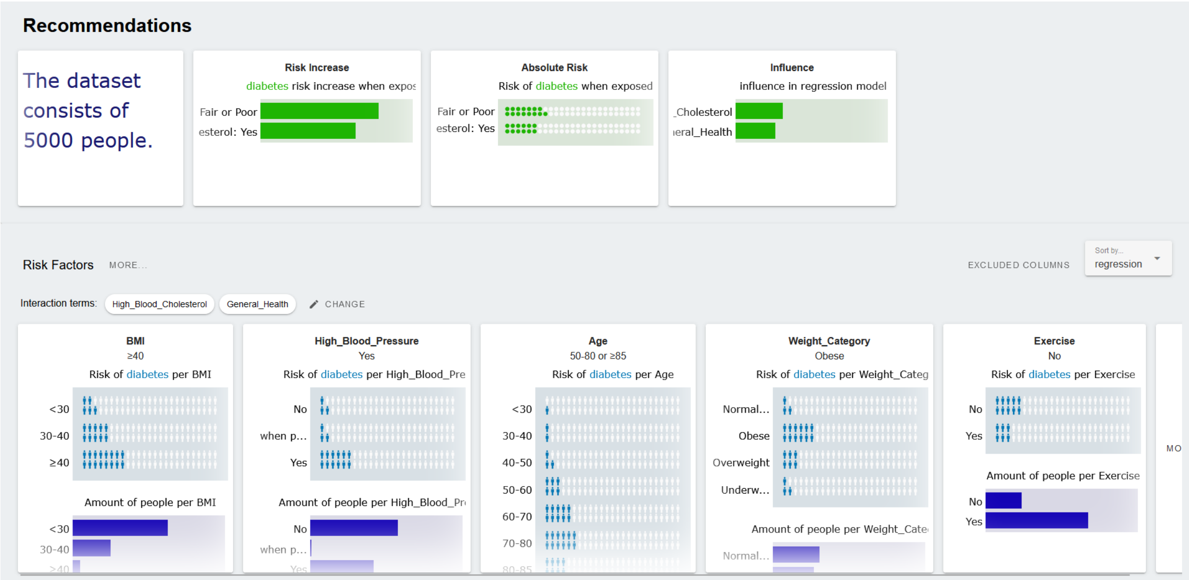


Figure 4.3 The recommendation view. The user is provided with previews of each fact group and can add them to the dashboard. Fact groups include general and context facts on top, as well as risk factors below, sorted by their score.

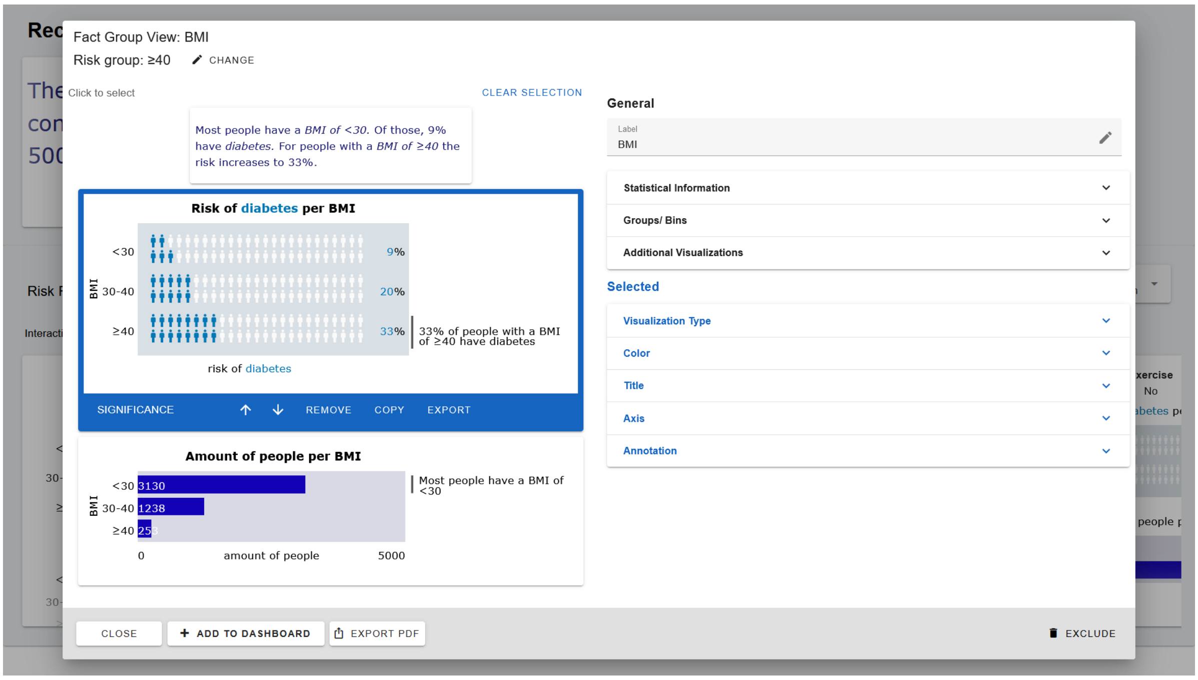


Figure 4.4 The fact group view displayed when a fact group is selected. The complete visualizations are displayed on the left side, with customization options for the fact group and the individual visualizations on the right side.

The fact view allows the user a broad range of customization options for their graphs, including the graph type, color and phrasings of text and axis of the visualization. Additionally,

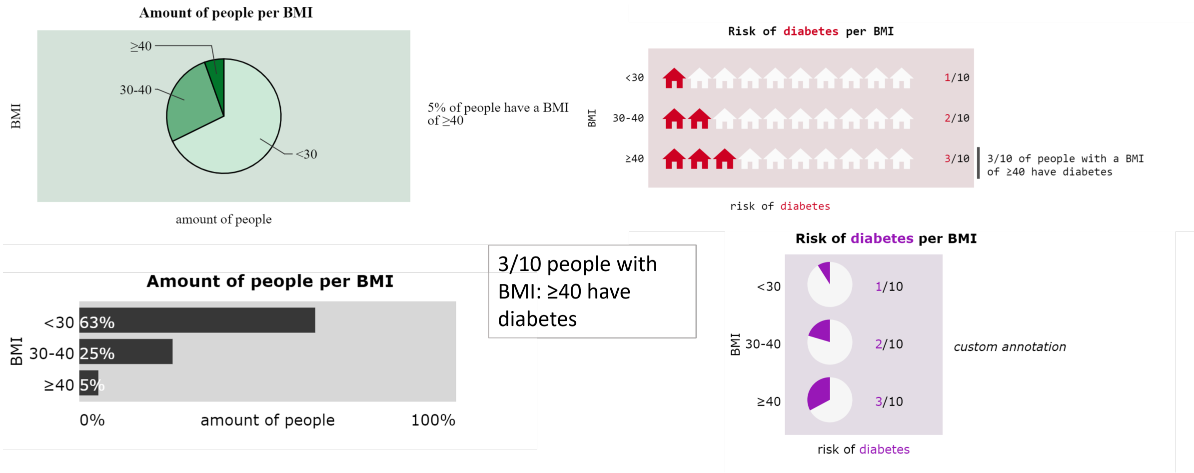


Figure 4.5 The tool provides a broad range of customization options for the visualizations. Some example visualizations created with the tool are shown here.

different icons for the pictographs can be selected, the unit can be changed and the annotation can be modified or removed. Figure 4.5 shows some examples of visualizations created with the tool.

When the user is satisfied with their visualizations, they are provided with multiple export options. Each individual visualization can be exported as a png image or text file, depending on the type of visualization. They can also be directly copied to the clipboard for easy import into other applications like the widely used presentation software PowerPoint. The user can also export each fact group as a pdf file. This pdf file will contain all visualizations of the fact group, as well as the fact group title and overall annotation. Lastly, the user can export the whole dashboard. When clicking on the export button, first a preview of all fact groups and their visualizations is shown. The user can use this view for a final check of their visualizations. Then, they can either export the dashboard visualizations as a pdf file or as individual images. An example of how the pdf export looks like is shown in Figure 4.6.

4.3.2 User Guidance

The tool will also provide the user with guidance on how to use the tool, what to pay attention to in risk factor calculation, and support in visualization creation.

First of all, the user is supported in using the tool itself. Help boxes are used in the start screen to explain the input options. Additionally, explanations are provided for features like the interaction terms and scoring methods.

When determining and interpreting risk factors, the user is supported by hints on top of the fact view when, for example, calculated significance information is based on a small data

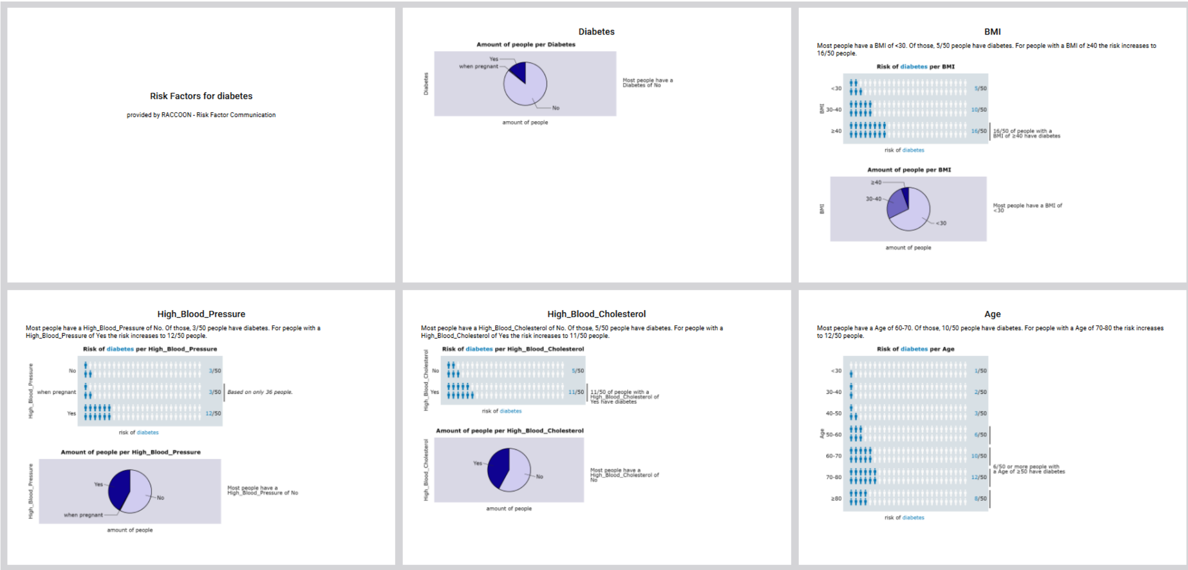


Figure 4.6 PDF Export. Each fact group is exported as one or multiple pages, depending on the size of the visualization. For each group, the title, if available the overall annotation, and all visualizations are included.

set size or a factor does not improve the current regression model. Hints on small data set sizes or when data is not statistically significant is also provided through the annotations of the visualizations themselves.

For the creation of the final visualization, the tool mostly uses direct feedback to enable the user to just try out a feature and see the result. For example, when changing the color of a visualization, the user will see the result immediately. Additionally there is a section below the dashboard that provides tips on how the visualizations might be improved. This includes hints when there is inconsistency between visualizations of similar facts or when a graph type like pie charts is used that may be not the best choice in common scenarios. These tips are generated by checking the selected visualizations based on a fixed set of rules for inconsistencies and possible improvements.

Chapter 5

Evaluation

I have evaluated the results of my thesis through multiple user studies and a performance study. First, I will describe the methodology of the studies. Then, I will present the results obtained from these studies. Finally, I will discuss the implications of the results.

5.1 Evaluation Studies

To evaluate my tool I first performed a pre-user study to gather first feedback and implement the most important changes. Then, I conducted two formal user studies, one with experts and one with the general public. Conducting two studies allows me to evaluate the usability of the tool from the perspective of possible users, as well as the perception of the resulting visualizations by a general audience. Lastly, I present some information about the performance of the tool.

5.1.1 Pre-User Study

Before conducting the actual user studies I started with two informal interviews involving a free exploration and discussion about the tool, as well as open-ended questions. These interviews were conducted with a visual analytics expert and an epidemiologist. Both interviews used the diabetes data set provided as an example in the tool. The purpose of these interviews was to gather initial feedback on the intuitiveness and user experience of the tool, as well as identify any missing features. Based on this feedback, I was able to implement the most pressing changes and address any bugs before conducting formal evaluations.

5.1.2 Expert User Study

The objective of the first formal evaluation is to evaluate the user interface and visualizations from the perspective of possible users.

I interviewed experts from different fields including visual analytics, visualization, medicine and risk communication. This allowed me to get diverse perspectives on the tool. The participants were invited via email, which was either sent to partners at different universities or forwarded by them to others.

Based on the time required in the pre-user study each interview was scheduled for an hour. The participants were motivated to share their honest feedback in their role as experts of their field. To ensure important aspects were not missed during exploration of the web-based tool, I prepared a set of tasks to guide them through the tool. Participants were invited to share their thoughts using the think-aloud protocol. I included the following tasks:

1. Please load the example data set to create a data story about risk factors for diabetes.
2. Select the intention “Explore” to first explore the data.
3. Add the risk factor “BMI” to the dashboard.
4. Add the risk factor “High Blood Pressure” to the dashboard.
5. Correct any wording errors in your selected visualizations.
6. Adjust the bins of BMI to more useful bins.
7. Add one of the visualizations that compare multiple risk factors to the dashboard
8. Change the colors of the design to your liking
9. Compare the intentions “Educate” and “Convince”
10. Export your dashboard as a PDF file

After completing all tasks, they were asked to fill out a questionnaire collecting quantitative feedback and giving them the ability to give feedback anonymously. This includes their feedback on the intuitivity, accuracy, likeability and wording of the visualizations. They were asked to give an overall rating, and separate ratings for the significance pictographs, impact bar graphs, impact pie charts and summary texts.

Considering the UI, they were asked to rate the following aspects on a likert scale:

- intuitivity

- success in accomplishing goals
- helpfulness
- motivation to use again
- recommendation to others

They were also asked to state how satisfied they were with the risk factor recommendations, visualization customization options, and export options.

5.1.3 Public User Study

The objective of the second formal evaluation is to evaluate the perception of the visualizations by the general public. I particularly evaluated the two intentions I specifically designed for the general public, "Convince" and "Educate". I excluded the "Explore" intention because it is mainly targeted at the domain experts, not the general public.

As the general public is a very diverse user group I decided to perform a quantitative user study, allowing me further outreach. I collected participants through passing the study to friends, colleagues, family and multiple local student or online groups with the request to further distribute it to others.

To enable easy access, the study is designed as an online survey. The tool is designed to support an expert in creating data visualizations for their data story. As my tool will only generate the visualizations instead of the whole data story, I had to decide between the following options. First, I could evaluate the images independently, without a data story. The feedback obtained from this study would directly apply to the visualizations. However, the visualizations will miss context making them harder and probably less engaging to interpret. Additionally, the evaluation would focus on a scenario different than the use case that the tool was developed for. Secondly, the most realistic study would require experts to create a data story using the tool and evaluate these. However, this would also be the study type with the most outer influences as preferences of the expert, their customizations and their presentation skills will impact the study. The last option which I decided to implement, is evaluating the images as part of a presentation, but creating the presentation by myself and as close to the standard templates used by the tool as possible. This will still introduce other data story elements as outer influences, but the visualizations will be evaluated closely aligned with the preconfigurations provided by the tool. The only adaptations done to the visualizations will be adjusting bin sizes to domain conventions and fixing grammatical errors in the phrasing of the tool, as these are based on automatic templates.

I decided to implement the study in German instead of English, because the study will be conducted in Germany. Creating the study in English would not only introduce a language barrier but also create a bias based on the English competency of the participants. To minimize bias, I translated all textual elements of the visualizations as close as possible to the English templates.

Considering the topic of the data stories, I decided to use the diabetes example data set already included in my tool, see Section 4.1. This dataset is publicly available and based on a telephone survey, thereby reaching a broad range of individuals. Diabetes affects over 10% of the global population and is associated with multiple behavioral risk factors. It is a good example of an area where behavior change could be fruitful in reducing the individual's risk. To improve performance, I used a sample dataset of 100k randomly selected participants instead of the original dataset, which contains over 440k participants.

The final study is based on two data stories about diabetes that just differ in the used visualizations. One data story will use the visualizations created by using the "Convince" intention, the other data story those of the "Educate" intention. I kept the story intentionally short and to a minimum as to provide context for the visualizations but not introduce too much outer influence. With short stories I also wanted to keep the user engaged, which is important when working with a general public. The story starts with an eye-catcher headline "Diabetes - Liegts am Zucker?", in English "Diabetes - is it the sugar?". It then provides some general information about diabetes and the first visualization showing the diabetes prevalence. Then, a short paragraph introduces the risk factors *obesity* and *no exercise*, which are then both presented individually. *Obesity* is visualized through the factor *BMI* to present visualizations of a continuous risk factor, and *no exercise* through the factor *Exercise* with a categorical classification of yes and no. With BMI being binned to five bins and exercise consisting of two groups, I also show different group sizes. For both risk factors, the significance and impact visualizations are shown. Lastly, the risk factors are compared by showing the context visualization displaying the relative risk increase. At the end of the page, the data and information sources are listed. All visualizations contain the standard annotations provided by the tool.

The online questionnaire will begin by asking about personal information and previous experience of the user. I consider the following aspects:

- age
- gender
- previous knowledge about diabetes
- familiarity with graphs

ID	Operating system	CPU	RAM	GPU
1	Microsoft Windows 10	Intel Core i5-2500	8GB	NVIDIA GeForce GT 730
2	Microsoft Windows 11	12th Gen Intel Core i7	16GB	Intel Iris Xe Graphics

Table 5.1 Specifications of the compared PCs

- attention to a healthy lifestyle

Then, one of the two data stories will be randomly assigned and presented. I decided for an in-between-subject-design because both data stories are very similar which would significantly decrease the engagement with the second story. Lastly, the user is asked to rate the data story using likert scales and provide additional comments if they wish so. I consider the following aspects of the data story:

- detail
- need for more explanations
- likeability
- motivation for behavior change
- trust
- understandability

The complete questionnaire with both data stories is available in the appendix.

5.1.4 Performance Study

As the last part of the evaluation I want to present some information about the performance of the tool. This is motivated by the design decision to not use a backend, which makes sure data security is considered but might decrease performance. Therefore, I want to present some information on how long the calculations take depending on hardware and data set. I will present timing information on the longest calculation of the website, which is the initial calculation time happening at the end of the start screen. This includes the initial binning and scoring of risk factors, the logistic regression, fact generation and visualization.

I will compare the two PCs listed in Table 5.1 using the diabetes data set as it is publicly available. I will compare their performance on different amounts of rows ranging from 1k to 400k rows.

5.2 Results

I will now present the results of the user studies. I will start by describing the pre-user study, followed by the expert user study and the public user study. I will then present the results of the performance study.

5.2.1 Pre-User Study

The main finding of the pre-user study was the complexity of the tool. The complexity is mostly horizontal, meaning that all in all the tool is easy to use but it simply has a lot of features, customization options and visualizations. This mainly resulted in both interviews taking around an hour just to explore most of the features. Apart from that the interviews revealed many aspects of the tool that needed further explanations or more user guidance to be understood. In consequence, I added help boxes, help texts, and redundant buttons to make features more visible, as well as a tutorial for the expert evaluation.

5.2.2 Expert User Study

In total, I interviewed five experts (3 female, 2 male). The experts had a mean age of 37 ranging from 28 to 42. All experts opted for an online meeting. One expert was a clinician. The remaining four experts are visualization experts, with different focuses. Two of them have a background as computer scientists with expertise relevant to this work in either medical visualization or narrative visualization. The other two have a background in design, with one of them having expertise in risk communication and one specializing in interaction and interface design.

Previous experience was quite varied, see Figure 5.1. All experts were *familiar* or *very familiar* with data visualizations. The experience with risk communication ranged from *unfamiliar* to *very familiar*. Considering experience with risk factor calculation, three experts were *unfamiliar* or *very unfamiliar*, and two experts *familiar*. All experts had at least some familiarity with visualization generation tools, with one expert *somewhat familiar*, three *familiar* and one *very familiar*.

Interview results

I collected valuable feedback during the interviews using the think-aloud protocol. The experts generally found the tool helpful because it allowed them to generate and customize visualizations easily. In particular, the clinician expressed a strong liking for the tool as it enabled them to quickly create visualizations on their own, without programming knowledge

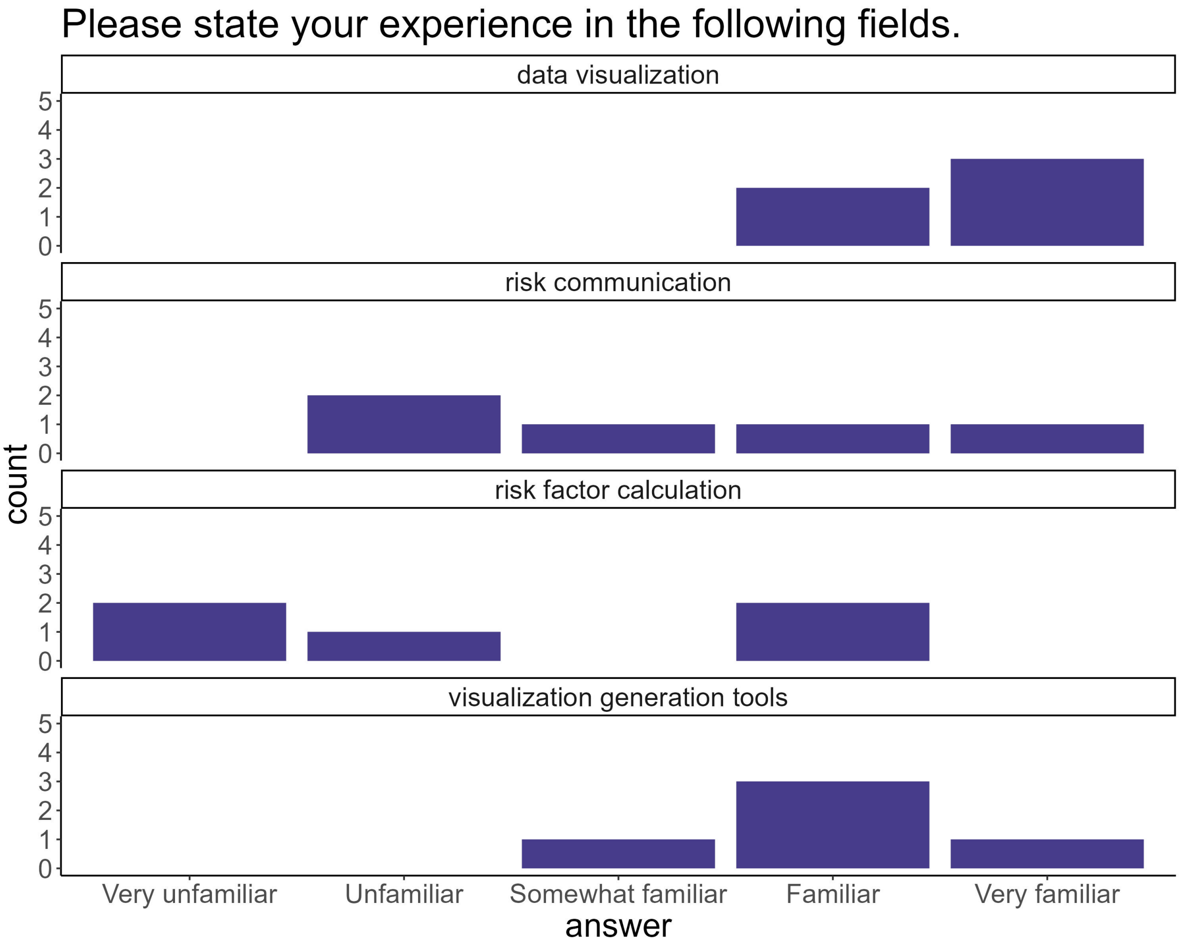


Figure 5.1 Bar charts showing the distribution of answers per question about previous experience.

required. They explained that visualizations can greatly assist in risk communication. During the interviews some minor bugs were found, for example, that annotations were not always displayed correctly when bins were changed. Considering user guidance, despite the tutorial, for most experts the usage of the tool was not directly clear and I had to provide tips on how to use it. This applies in particular to the meaning of the dashboard, the ability to click on fact groups for more detailed information, and the ability to then click on individual visualizations to change them. Furthermore, the concept of risk group was not clear to most experts. However, after initial guidance, the experts appreciated how easy they could use the tool to customize the visualizations, with one expert stating that they "really like how easy I can change it". Especially the changing of icons for the pictograph, the different provided graph types and the easy changing of texts were praised. One expert commented on how the tool can enhance collaboration by enabling swift implementation of changes. Additionally, the experts also had some suggestions for improvement. Most experts requested a more direct manipulation of the aspects of the visualization, instead of using a menu. One expert also expressed a desire to be able to change the colors of the bins individually, while another suggested assigning a unique color to each risk factor.

Regarding the visualization design, most experts liked the usage of pictographs because, as one stated, they are “reflective of current research”. However, multiple experts wondered why 50 persons were used as the denominator, finding 100 or 10 to be more intuitive. The experts also appreciated the automatic calculation of bins and the connectivity that ensures bins are changed automatically in all visualizations. Two experts also liked the option to show or hide missing values. However, multiple experts requested a feature to delete individual bins. The chosen two facts for each risk factor were considered important by most experts. However, one expert found it challenging to relate the two visualizations of each risk factor together. They suggested an integrated visualization. Experts liked how the context visualizations adapted to the risk factors selected in the dashboard. However, all experts required some time to understand the visualizations. Improvements could be made by simplifying the graphs, such as considering color saliency in the graph design and removing redundant axis descriptions. Additionally, it was unclear to one expert whether the impact graph referred to all people or only those affected by the outcome. One expert also suggested visualizing different risk factors as icons, like a heart for heart disease. All experts liked the annotations, especially as they give examples for how the visualizations can be interpreted. Two experts specifically mentioned the value of being notified when there were insufficient data points. The experts liked how easily annotations could be changed, with different templates to choose from. However, one expert wished for more salient annotations to draw the attention of the viewer. The intentions were explored with curiosity and different preferences as to which intention and changes they liked or disliked. Three experts expressed concerns about misleading the audience with the *convince* intention. They provided various suggestions for improvement, such as using alarm colors for the convince intention and exploring the use of color saliency and icons.

Further comments were made regarding different aspects of the tool. One expert recommended to improve the visual hierarchy of the interface, visualizations and pdf export. Multiple suggestions were provided on how the pdf export could be improved or expanded, for example by utilizing a 16:9 slideshow format. However, in general, the given export options were well-liked by all experts. One expert suggested that the branding of the tool could be improved, for example, by incoroporating a logo. Additionally, the given tips could be improved by not only identifying problems, but also offering solutions, and providing more specific tips.

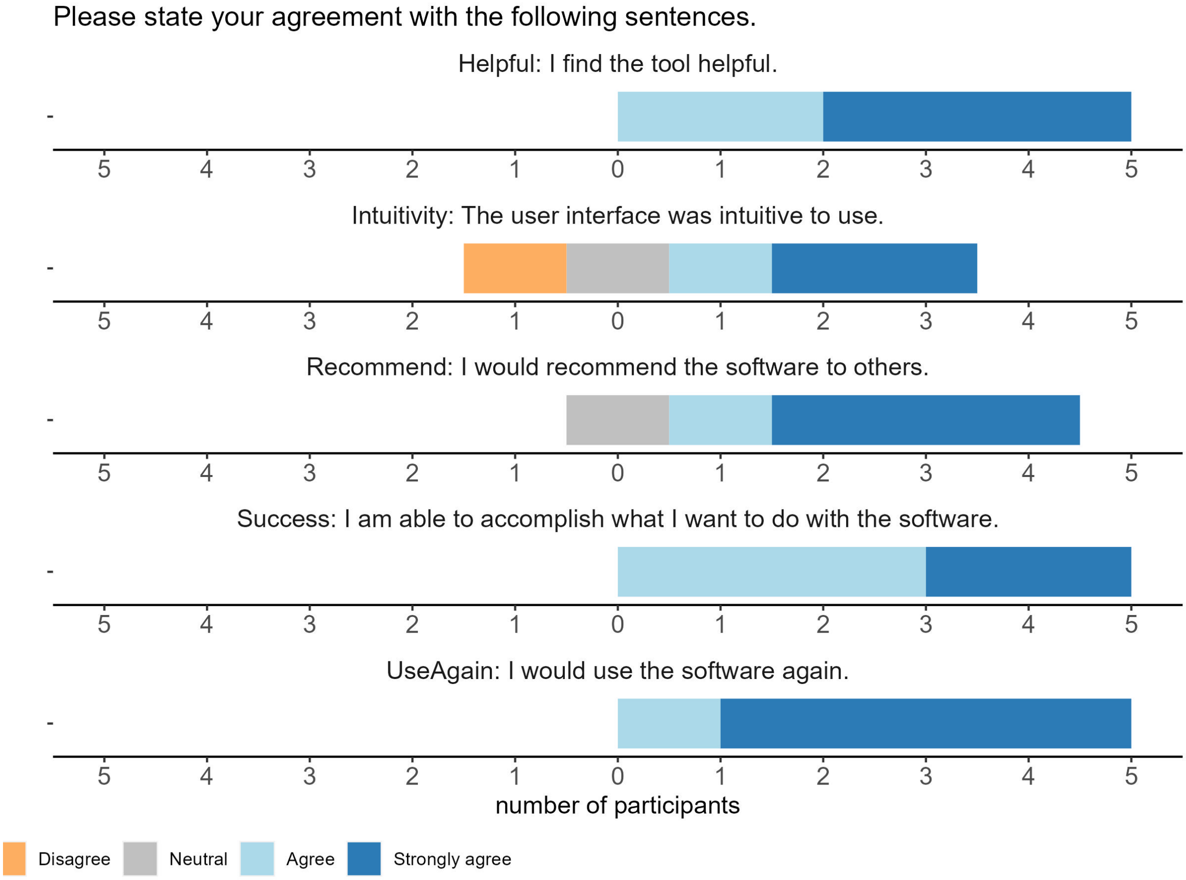


Figure 5.2 Results of the likert scale questions about the tool in general. For each question, the number of people per answer is shown, centered around the neutral answer with generally disagreeing answers (disagree or strongly disagree) in red colors on the left side and generally agreeing answers (agree or strongly agree) in blue colors on the right.

Questionnaire results

I will now describe the results of the feedback questionnaire filled out by the experts after the interview. Regarding the tool, the experts were first asked to rate the tool in general, see Figure 5.2. The feedback was generally positive. Of the five experts, two *agree* and three *strongly agree* that the tool is helpful. Opinions on intuitivity were more varied, with one expert each *disagreeing*, *neutral*, and *agreeing*, and two experts *strongly agreeing*. Most experts would recommend the software to others, with one expert being *neutral*, one *agreeing* and three *strongly agreeing*. Most experts felt like they were able to accomplish what they wanted to do with the software, with three experts *agreeing* and two experts *strongly agreeing*. Most experts would use the software again, with one expert *agreeing* and the rest *strongly agreeing*.

Afterwards, the experts were asked to rate specific aspects of the tool. The results are shown in Figure 5.3. Most experts were satisfied with the visualization customization options, with

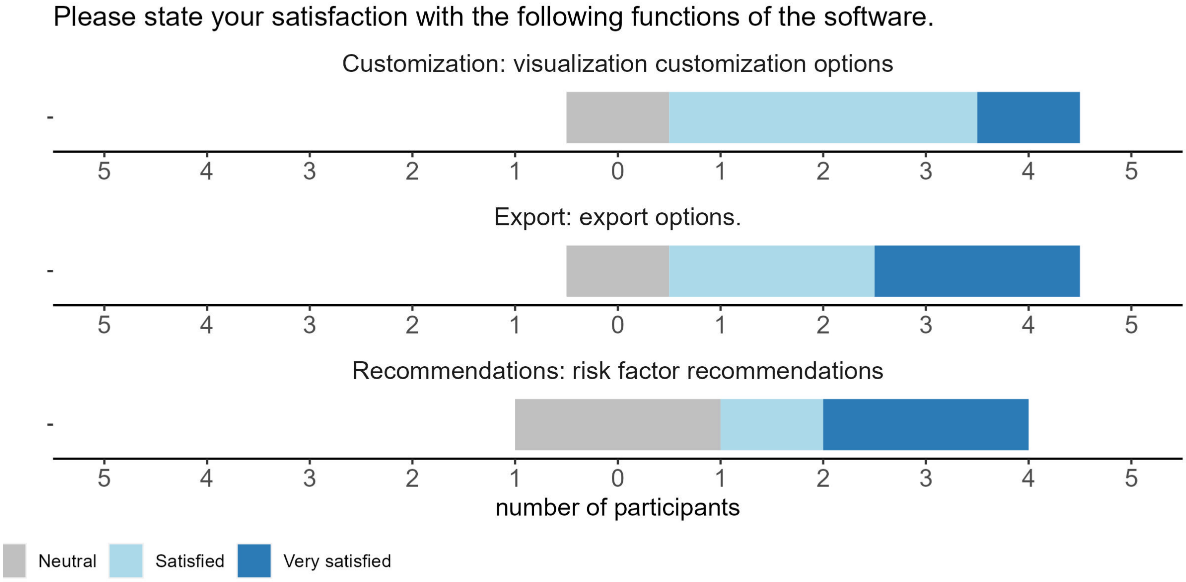


Figure 5.3 Results of the likert scale questions about specific aspects of the tool. For each question, the number of people per answer is shown, centered around the neutral answer with generally disagreeing answers (disagree or strongly disagree) in red colors on the left side and generally agreeing answers (agree or strongly agree) in blue colors on the right.

one expert *neutral*, one expert *very satisfied* and the rest *satisfied*. The export options were also well-liked, with one expert *neutral*, and two experts each *satisfied* and *very satisfied*. The risk factor recommendations were rated by two experts as *neutral*, by one as *satisfied* and by two as *very satisfied*.

Regarding the visualization design, the experts were asked to rate the generated visualizations based on their accuracy, intuitivity, likeability and wording. Overall, the feedback was positive, see Figure 5.4. Two experts stated the overall accuracy of the visualizations to be *good* and three as *very good*. The intuitivity was desribed by three experts as *good* and two as *very good*. The likeability was also stated by three experts as *good* and two as *very good*. The wording was rated slighly lower, with two expers *neutral*, two *good* and one *very good*. When examining the experts' opinions on the individual visualization types, slight differences can be seen. For accuracy, opinions on the impact visualizations varied the most, and the textual summaries were rated the lowest with only one expert rating them as *very good*. For intuitivity, the significance pictographs and impact bar charts were rated slightly more positive than the other visualizations. For likeability, the impact visualizations and textual summaries were rated lowest and the significance pictographs highest. For wording, the individual visualizations were all rated slightly better than the overall wording, with almost no difference between them.

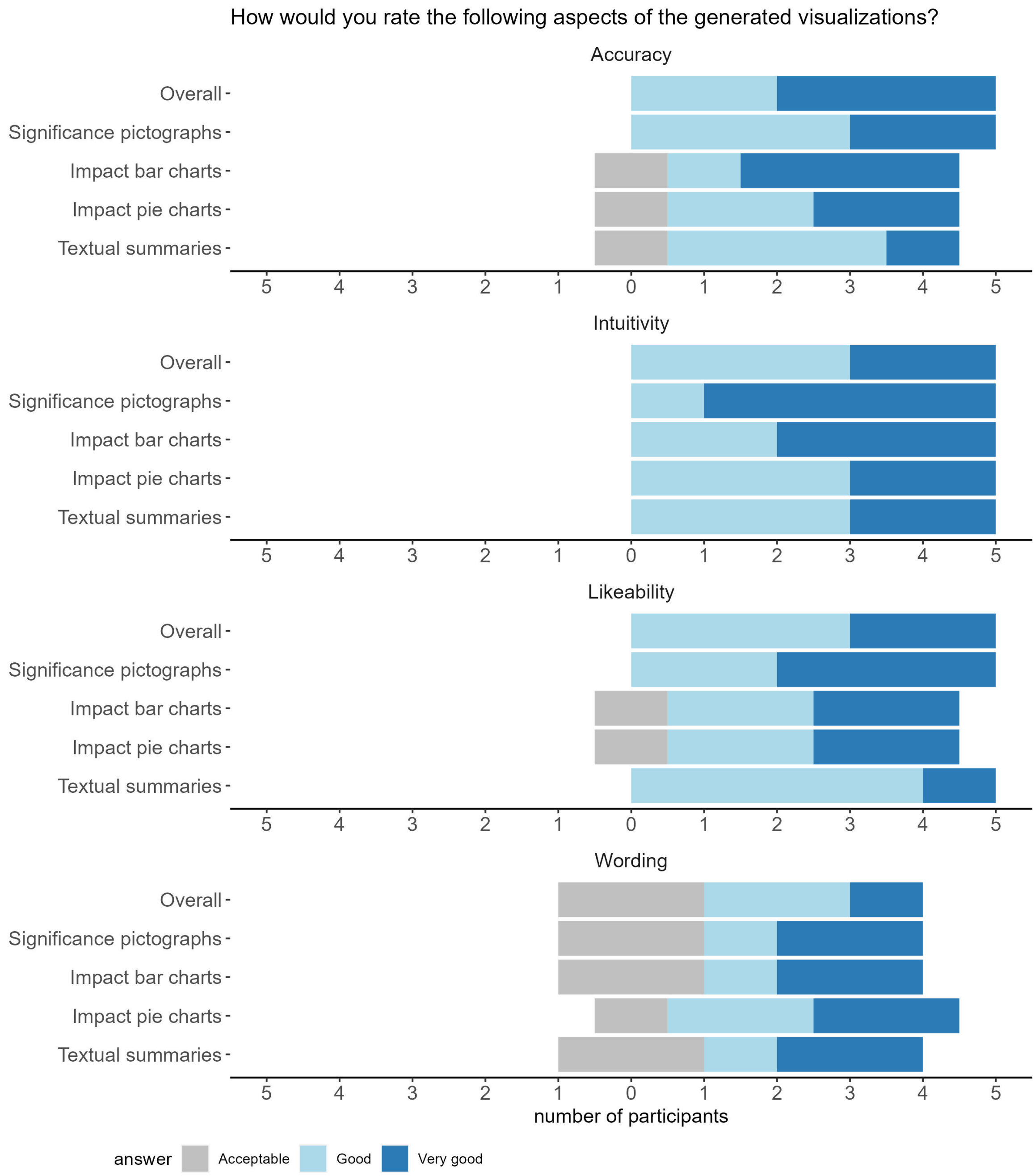


Figure 5.4 Results of the likert scale questions about the visualization templates generated by the tool. For each question, the number of people per answer is shown, centered around the neutral answer with generally disagreeing answers (disagree or strongly disagree) in red colors on the left side and generally agreeing answers (agree or strongly agree) in blue colors on the right.

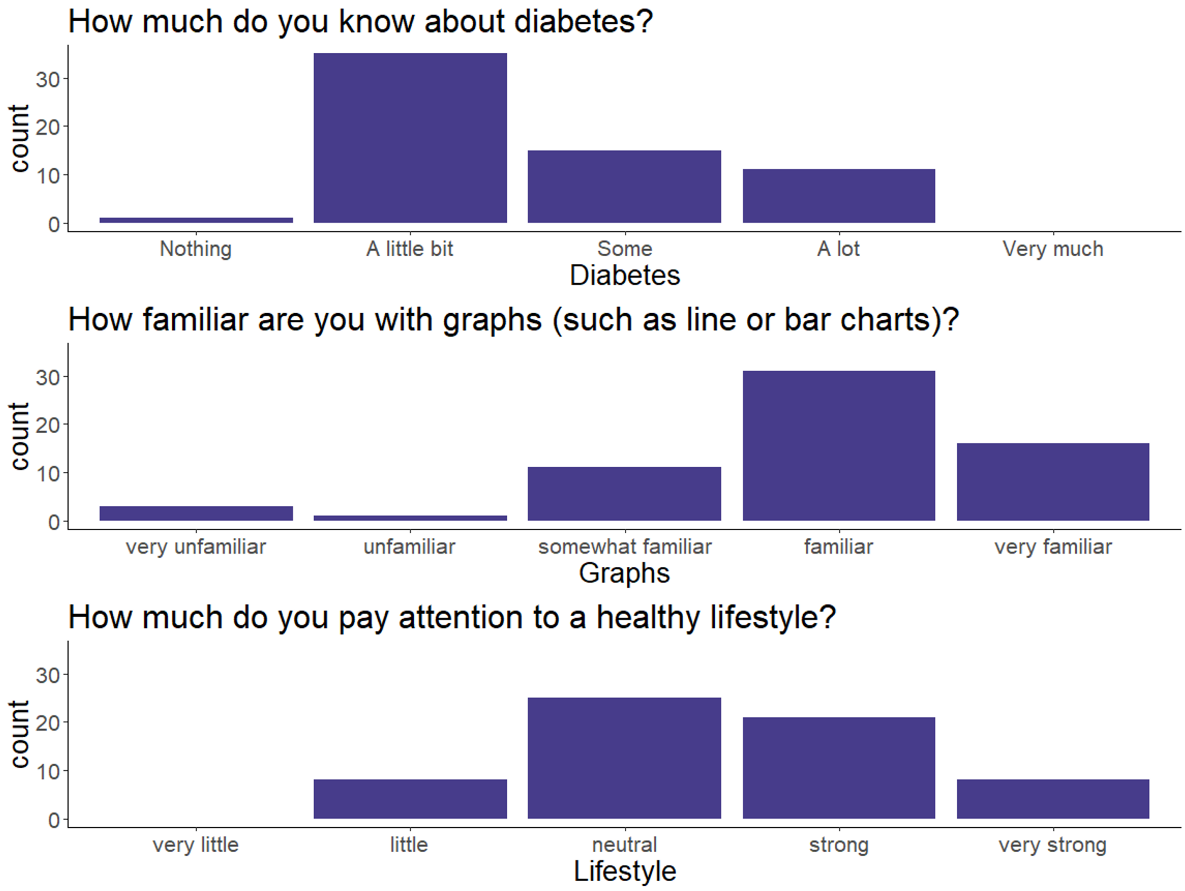


Figure 5.5 Bar charts showing the distribution of answers per question about previous experience.

5.2.3 Public User Study

In total, 62 persons (30 female, 19 male, 8 divers, 5 not stated) finished the user study. The mean age was 35 years, ranging from 19 to 91. Of those, 31 were assigned to the convince version and 31 to the educate version. Previous experience is shown in Figure 5.5. Most persons had *a little bit* of experience with diabetes (35 out of 62) and were *familiar* with graphs (31 out of 62). Only 3 persons were *unfamiliar* or *very unfamiliar* with graphs. How much persons payed attention to a healthy lifestyle was wider distributed, with 25 persons *neutral*, 21 persons *strong* and 8 persons each *little* and *very strong*.

The results of the likert scale questions are displayed in Figure 5.6. For each question, the number of people per answer is shown, centered around the neutral answer with generally disagreeing answers (disagree or strongly disagree) in red colors on the left side and generally agreeing answers (agree or strongly agree) in blue colors on the right. The answers are separated by the intention of the data story, with the convince version displayed below the educate version. The results show that the data stories were generally perceived as detailed, trustworthy and understandable. Most persons also liked the info page (35 out of 62), however some persons did not like it (11 out of 62) and some were neutral (16 out of 62). Participants differed on there opinion if they would have liked more explanations, with most not wanting more (32 out of 62) but also 22 out of 62 agreeing that they would have liked more explanations. Persons also differed in if they were motivated to strive for a healthier lifestyle. Here, the strongest differences between the two versions can be seen. The educate version had few neutral answers (3 out of 31), many disagreeing answers (16 out of 31) and also many agreeing ones (12 out of 31). In contrast, the convince version had many neutral answers (10 out of 31), many disagreeing ones (16 out of 31) but only 5 agreeing ones. The differences between groups for the other answers were less pronounced. However, generally the convince version had a wider spread of answers with often more strongly disagreeing and agreeing answers than the educate version.

I analyzed correlations between previous experience, age or watch time with the results of the likert scale questions and watch time. For this analysis, I used a scatterplot matrix and a matrix of the pearson correlations. Both feature previous experience, age and watch time as the columns and the results of the likert scale questions and watch time as the columns. For the pearson correlation, gender was excluded as it has no inherent order. The scatterplot matrix in Figure 5.7 shows only low correlations with previous experience. This is also visible when calculating the pearson correlations as seen in Figure 5.8. The plot also shows that the time does not correlate with any of the previous experiences or results. The strongest correlations are seen for the age, with the maximum correlation being age and detail ($r=0.38$) with older persons rating the visualizations as more detailed.

In summary, persons looked at the visualizations with a mean of 128 seconds. As presented in Figure 5.9 the distribution of watch time in convince has a greater variance and higher mean than that of convince. When performing a welch's t test the difference in mean is statistically significant ($p=0.037$) with the mean of the convince version being 148 seconds and the mean of the educate version being 108 seconds.

Further textual comments were left by 17 participants. Some of the comments referred to the info page itself. Two persons wanted more information on the used data set, and two others on the chosen risk factors. One person wanted a distinction between the different diabetes types. One person criticized that the references were not directly linked to the text. One person did not see the aim of the info page. One person stated, that they already pay attention to a healthy lifestyle and therefore the page will not change their motivation. three persons stated that they found the website overloaded with graphs leading to confusion.

Some of the comments referred to the graphs in general. Two persons did not fully understood the calculations and one person wished for more intuitive graphs. Two persons criticized

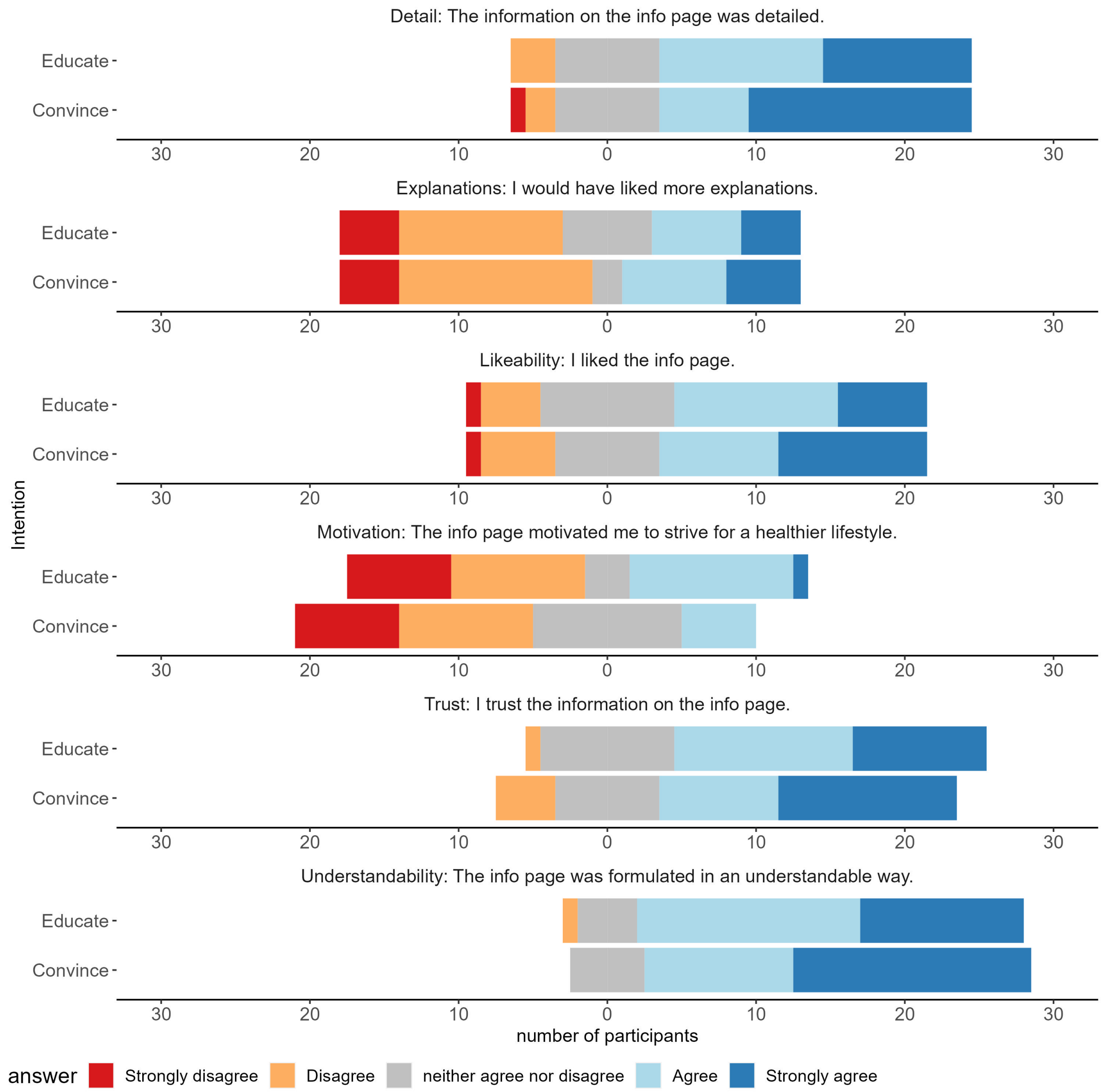


Figure 5.6 Results of the likert scale questions. For each question, the number of people per answer is shown, centered around the neutral answer with generally disagreeing answers (disagree or strongly disagree) in red colors on the left side and generally agreeing answers (agree or strongly agree) in blue colors on the right. The answers are separated by the intention of the data story, with the convince version displayed below the educate version.

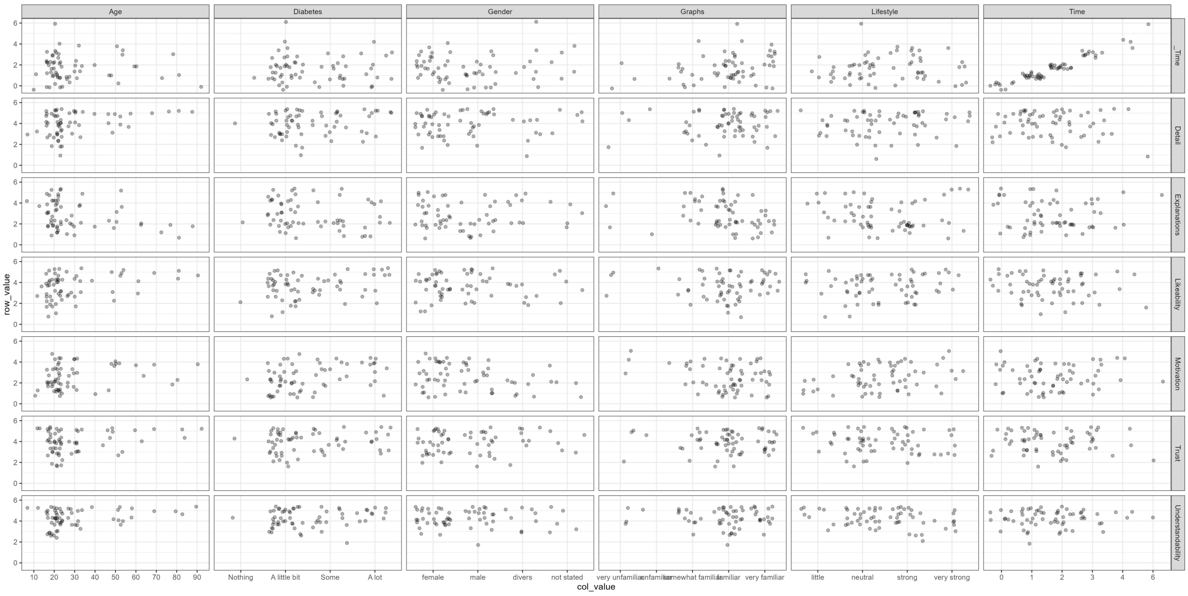


Figure 5.7 Scatterplot matrix of correlations between previous experience, age, or watch time with the results of the likert scale questions and watch time. Only low correlations are visible.

Figure 5.8 Matrix of the pearson correlations between previous experience, age or watch time with the results of the likert scale questions and watch time. Only low correlations are visible. Gender was excluded as it has no inherent order.

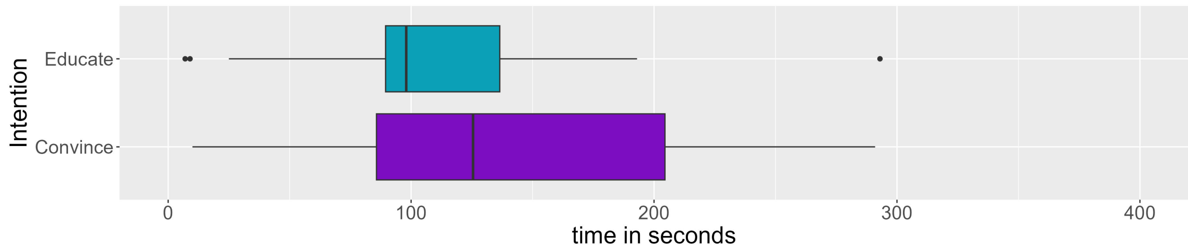


Figure 5.9 Watch time difference between versions. For each version, the distribution of watch time is shown as a boxplot of the time in seconds. The convince version had a higher variance and longer watch time than the educate version.

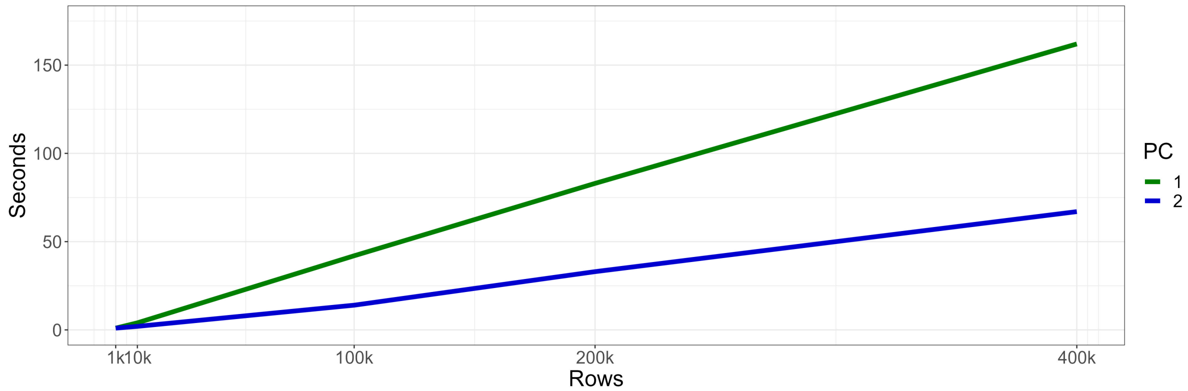


Figure 5.10 Performance of the tool on two different PCs. The y-axis shows the time in seconds to complete the initial calculation. The x-axis shows the number of rows in the dataset. On both PCs, the time increases linearly with the number of rows. PC 2 performs better with under half the calculation time from 10.000 rows upwards than PC 1.

the chosen denominator of 50. One person found the coloring of the graphs inconsistent. The annotations were remarked by one person as weird and by one person as helpful.

One comment specific to the convince version found it not intuitive that only 5 of the 50 persons were shown in the graph. Two comments were specific to the educate version, wishing for more distinct colors for the pie charts and numbers visible in the graph, not just in the text.

5.2.4 Performance Study

The performance on the two PCs is shown in Figure 5.10. The time to complete the inititial calculation increases linearly with the number of rows. PC 2 performed better with under half the calculation time from 10.000 rows upwards than PC 1. For 1k rows both PCs had a waiting time of around a second. With 100k rows the waiting times were 14s for PC 2 and 42s for PC 1. With 400k rows the waiting times were over 1 minute for PC 2 and over 3 minutes for PC 1.

5.3 Discussion

To start this section, I will first discuss the results of the expert evaluation regarding the tool itself. Next, I will discuss the results of both studies regarding the generated visualizations. Finally, I will address the limitations and summarize the main learnings from the evaluation.

5.3.1 Evaluation of the Tool

The evaluation of the tool by expert users generally indicated good usability and perceived helpfulness. Four experts strongly agreed that they would use the software again, suggesting that the tool is well-received by experts from different fields and fills a gap. Especially the clinician valued the tool for its ability to create visualizations without requiring visualization skills. The varied responses on the intuitivity of the tool and the initial confusion when starting to use it imply a need for better user guidance and an improvement in the tool's design. Despite these initial difficulties, the experts highly valued the customizability provided by the tool to adapt the visualizations to their specific needs.

The performance study demonstrated that the tool is able to handle large data sets of up to 10,000 rows without any significant delays for the user, despite running exclusively on the local machine of the user. However, this also results in the performance depending greatly on the device used. The calculation time increases linearly with the number of rows. These results support the design decision to uphold a user's privacy by not using a server backend.

5.3.2 Evaluation of the Visualizations

Both evaluation studies revealed that the visualizations were generally well-received. The experts particularly liked the pictograph visualization for communicating risks. However, the visualizations contained a lot of text elements, which led to confusion and overload for multiple participants in the public user study. Multiple experts proposed reducing the amount of text by deleting specific elements of the visualization. Also, both experts and participants of the public user study questioned the choice of 50 as the denominator for the pictograph visualization. The denominator was chosen as a compromise between accuracy and not overloading the graph with too many icons, but it was repeatedly mentioned in the feedback that it was confusing. A denominator of 10 or 100 would have been more intuitive. The annotations were generally well received in both studies, especially to give an example for interpretation of the visualization. However, they also contributed to the perceived complexity and overload of the visualizations.

While experts valued the choice of visualizations for risk factors containing one significance and one impact visualization, feedback indicated that is was hard to relate the two visualizations to each other. Visualizing both facts separately simplifies the visualizations, but it also makes it harder to understand the connection between the two. Two experts therefore suggested combining both visualizations into one.

One purpose of the visualizations was to affect motivation for behavior change, a goal especially shared by the clinician. However, the data stories evaluated in the public user study did not have a strong effect on motivation to strive for a healthier lifestyle. This suggests that the visualizations can still be improved to better support this intention, but also that visualizations alone will not be sufficient to change the behavior of a person.

A further result of the public user study was that there were no significant correlations between previous experience and questionnaire results. This may be due to the small sample size and limited number of participants with no previous experience with visualizations.

A surprising result was the difference in time spent reading the data story between the two evaluated intentions. The longer time spent on the *convince* version might be attributed to increased user engagement or a higher complexity of the visualizations. However. as both versions had similar answer distributions for questions about detail and understandability, a higher complexity seems unlikely. Instead, the more polarized answers to the *convince* version might indicate that it generated more engagement.

Unfortunately, the differences between the two versions of the data story in the public user study do not support the initial hypothesis about which visualizations would be more effective for each intention. The *educate* version actually convinced slightly more participants to strive for a healthier lifestyle than the *convince* version specifically designed for this purpose. This cannot be explained by differences in the other evaluated aspects, as the *educate* version did not perform better in those areas. The fact that the answers to the convince version were generally more polarized might indicate that the version did achieve a stronger effect through the enhancement of perceived differences between bins. However, this effect did not translate into a stronger motivation to strive for a healthier lifestyle. Concerns about data manipulation through the *convince* version did not result in significantly lower trust values. However, they should still be taken into account when designing visualizations for this intention. Improving the intentions could be achieved by implementing the valuable feedback provided by the experts during the expert evaluation. They recommended the use of more subtle aspects of the visualization, such as colors or icons.

5.3.3 Limitations

Several limitations apply to the results of the public user study.

- **Small sample size.** The sample size of 62 participants with 31 persons per version is small when trying to generalize the results to a larger population. However, the results provide first tendencies and give valuable indications for further research.
- **Biased sample.** The sample was biased towards younger individuals, which may have influenced the results. Younger persons probably have a different perception of the visualizations compared to older individuals, because they are growing up in a digital age. Additionally, the recruitment of participants was biased towards college students and acquaintances,who generally have a higher education level and therefore may have an easier understanding of visualizations. Moreover, as many know me personally, they might be inclined to give more positive feedback.
- **Only one data story.** The results are based solely on one data story about diabetes, which might have influenced the outcomes. The results may differ for other data stories.
- **German translation.** The data story was translated from English to German. This might have influenced the results as despite best efforts the translation might have changed the meaning of certain words.

Limitations applying to the expert evaluation are as follows.

- **Biased evaluation.** Since I acted as both the interviewer and the developer of the tool, the evaluation might be biased towards a positive evaluation of the tool.
- **Small sample size.** The sample size of the expert evaluation was small with only 5 participants. Further interviews may have revealed additional insights or different results.
- **Only one data set.** The results based solely on one data set related to diabetes, which could have affected the outcomes. The results might be different for other data sets or if the participants had to use their own data.
- **Artifical environment.** The tool was utilized as part of a user study, where participants followed specific tasks, rather than creating visualizations for an actual data story. This could have impacted the results, as participants may have used the tool differently in a real-world scenario.

5.3.4 Main Learnings

The main learnings from the evaluation are as follows:

- **Choose a denominator that is a power of 10.** The denominator of 50 was practical for the display of icons, but it was repeatedly mentioned in the feedback that it was confusing. A denominator of 10 or 100 would have been more intuitive.

- **Do not overload the graphs.** Creating intuitive visualizations requires a balance between providing enough information to understand the graph and not overloading the graph with too much visual clutter.

- **Choose terminology carefully.** Both the intuitivity of the tool and the generated visualizations greatly depend on the choice of terminology. It should therefore receive careful consideration.

- **More research into intentions.** Both the results of the public user study and the varied feedback from experts indicate the need for more research on how to best support different intentions with adaptations of the visualizations. Subtle changes to the visualizations, such as the use of percentages or natural frequencies, or the choice of color, can have a large impact on the perception of the visualizations.

- **Motivating persons to change their behavior is hard.** While visualizations may have a positive contribution, they are unlikely to change a person's behavior on their own.

- **Interpreting multiple visualizations together requires special considerations.** When using multiple visualizations together, for example, when trying to understand the connection between two visualizations of the same risk factor, or when comparing multiple risk factors, it is not sufficient to just consider each visualization on its own. Instead, their interplay has to be carefully considered.

- **New methods required to evaluate tools for the general public.** Evaluating the effectiveness of visualizations for a diverse user group like the general public requires new evaluation methodologies. Merely gathering individuals of the user group is not enough to ensure a diverse sample reflective of the entire group. The methods used in epidemiological studies to accurately represent the general public are too costly for smaller projects.

Chapter 6

Conclusion

6.1 Summary

In my thesis, I aim to support experts in communicating risks to a general public by creating effective visualizations for their presentations or data stories.

Based on current research in risk communication and narrative visualizations, I created visualizations for the risks and risk factors of diseases. These visualizations are specifically designed for a general audience.

As part of this thesis, I also investigated the use of annotations to simplify the visualizations or provide additional information.

I integrated this research into my risk communication tool Raccoon. It enables an expert to create data-driven and highly customizable risk visualizations from a given data set without requiring any expertise in risk visualization or design. The tool also takes into consideration the user's intention to explore the dataset, convince an audience, or educate the public, and adapts the visualizations accordingly.

The usability of the tool was evaluated in an expert evaluation, which confirmed its usability and helpfulness. The visualizations were evaluated with an online questionnaire, which showed that the visualizations were generally well-liked and considered to be detailed. However, more research is needed on how the visualizations should be adapted to the intention of the user.

6.2 Future Work

My work could be extended in a variety of ways to improve the calculation, visualization or annotation of risk factors and their combination for the final presentations. Additionally improvements could be made to the usability of the tool and user guidance through the provided intentions and support in risk communication.

- **Filtering**. A highly requested feature during evaluations was to add filter functions, for example, by age or gender. This would allow the user to create visualizations for specific subgroups. This would be especially useful for risk factors that are only relevant for a specific subgroup, for example, diseases exclusive to female individuals.
- **Non-linear correlations.** The calculation of risk factors could be improved by supporting more types of correlations, like non-linear correlations.
- **Natural language processing.** It would also be interesting to generate visualizations directly from text descriptions, for example by using natural language processing techniques. This would enable the user to add their own risk factors to the tool.
- **Layouting & sequencing.** The usability of my tool could be further improved by better support of the layouting and sequencing of visualizations. This would also enhance the ability of the tool to create a narrative through the visualizations. A narrative could also be incorporated by adapting the annotations of each risk factor to create a story.
- **Further data types & visualizations.** The tool could also be extended to support futher data types like temporal or geospatial data. This would enable the tool to be used for a wider range of data sets. Further data input abilities could also be added to support more visualizations like risk ladders or scales.
- **Individualization.** Another interesting addition would be to support individualized visualizations to visualize risks of example persons or a narrative character.
- **Improved intentions.** Lastly, the intentions provided by the tool could be improved by using more aspects of the visualization to support them. For example, for the intentions different icons could be used, as well as different colors, visual cues and annotations.

Bibliography

[1] Shiva Alemzadeh, Uli Niemann, Till Ittermann, Henry Völzke, Daniel Schneider, Myra Spiliopoulou, Katja Bühler, and Bernhard Preim. Visual analysis of missing values in longitudinal cohort study data. *Comput. Graph. Forum*, 39(1):63–75, 2020. doi: 10.1111/cgf.13662. URL https://doi.org/10.1111/cgf.13662.

[2] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani. Authoring data-driven videos with dataclips. *IEEE transactions on visualization and computer graphics*, 23(1):501–510, 2016.

[3] Jessica S Ancker, Yalini Senathirajah, Rita Kukafka, and Justin B Starren. Design features of graphs in health risk communication: a systematic review. *Journal of the American Medical Informatics Association*, 13(6):608–618, 2006.

[4] Dominic Balog-Way, Katherine McComas, and John Besley. The evolving field of risk communication. *Risk Analysis*, 40(S1):2240–2262, 2020.

[5] Arriel Benis and Moshe Hoshen. Disepi: compact visualization as a tool for applied epidemiological research. *The Practice of Patient Centered Care: Empowering and Engaging Patients in the Digital Era*, 244:38, 2017.

[6] Scott Bissett, Sharon Wood, Richard Cox, Donia Scott, and Jackie Cassell. Calculating alcohol risk in a visualization tool for promoting healthy behavior. *Patient education and counseling*, 92(2):167–173, 2013.

[7] Chris Bryan, Kwan-Liu Ma, and Jonathan Woodring. Temporal summary images: An approach to narrative visualization via interactive annotation generation and placement. *IEEE transactions on visualization and computer graphics*, 23(1):511–520, 2016.

[8] Stuart K Card, Jock Mackinlay, and Ben Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[9] Jawad Ahmed Chishtie, Jean-Sebastien Marchand, Luke A Turcotte, Iwona Anna Bielska, Jessica Babineau, Monica Cepoiu-Martin, Michael Irvine, Sarah Munce, Sally Abudiab, Marko Bjelica, et al. Visual analytic tools and techniques in population health and health services research: scoping review. *Journal of Medical Internet Research*, 22(12):e17892, 2020.

[10] William S Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554, 1984.

[11] S Crovato, A Pinto, Paolo Giardullo, G Mascarello, Federico Neresini, and L Ravarotto. Food safety and young consumers: Testing a serious game as a risk communication tool. *Food control*, 62:134–141, 2016.

[12] Zhe Cui, Sriram Karthik Badam, M Adil Yalçin, and Niklas Elmqvist. Datasite: Proactive visual data exploration with computation of insight-based recommendations. *Information Visualization*, 18(2):251–267, 2019.

[13] Goodarz Danaei, Eric L Ding, Dariush Mozaffarian, Ben Taylor, Jürgen Rehm, Christopher JL Murray, and Majid Ezzati. The preventable causes of death in the united states: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS medicine*, 6(4):e1000058, 2009.

[14] Rui Ding, Shi Han, Yong Xu, Haidong Zhang, and Dongmei Zhang. Quickinsights: Quick and automatic discovery of insights from multi-dimensional data. In *Proceedings of the 2019 International Conference on Management of Data*, pages 317–332, 2019.

[15] Michael Alin Efroymson. Multiple regression analysis. *Mathematical methods for digital computers*, pages 191–203, 1960.

[16] Centers for Disease Control and Prevention (CDC). Behavioral risk factor surveillance system survey data. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2015.

[17] Tong Gao, Jessica R Hullman, Eytan Adar, Brent Hecht, and Nicholas Diakopoulos. Newsviews: an automated pipeline for creating custom geovisualizations for news. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3005–3014, 2014.

[18] Irene Garcia-Martí, Raul Zurita-Milla, Arno Swart, Kees C van den Wijngaard, Arnold JH van Vliet, Sita Bennema, and Margriet Harms. Identifying environmental and human factors associated with tick bites using volunteered reports and frequent pattern mining. *Transactions in GIS*, 21(2):277–299, 2017.

[19] Anzu Hakone, Lane Harrison, Alvitta Ottley, Nathan Winters, Caitlin Gutheil, Paul KJ Han, and Remco Chang. Proact: iterative design of a patient-centered visualization for effective prostate cancer health risk communication. *IEEE transactions on visualization and computer graphics*, 23(1):601–610, 2016.

[20] Sarah T Hawley, Brian Zikmund-Fisher, Peter Ubel, Aleksandra Jancovic, Todd Lucas, and Angela Fagerlin. The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient education and counseling*, 73(3):448–455, 2008.

[21] Ulrich Hoffrage, Samuel Lindsey, Ralph Hertwig, and Gerd Gigerenzer. Communicating statistical information, 2000.

[22] Gareth J Hollands, Juliet A Usher-Smith, Rana Hasan, Florence Alexander, Natasha Clarke, and Simon J Griffin. Visualising health risks with medical imaging for changing recipients' health behaviours and risk factors: Systematic review with meta-analysis. *PLoS Medicine*, 19(3):e1003920, 2022.

[23] Jessica Hullman and Nick Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE transactions on visualization and computer graphics*, 17 (12):2231–2240, 2011.

[24] Jessica Hullman, Nicholas Diakopoulos, and Eytan Adar. Contextifier: automatic generation of annotated stock visualizations. In *Proceedings of the SIGCHI Conference on human factors in computing systems*, pages 2707–2716, 2013.

[25] Jessica Hullman, Steven Drucker, Nathalie Henry Riche, Bongshin Lee, Danyel Fisher, and Eytan Adar. A deeper understanding of sequence in narrative visualization. *IEEE Transactions on visualization and computer graphics*, 19(12):2406–2415, 2013.

[26] Ulrich John, Elke Hensel, Jan Lüdemann, Marion Piek, Sybille Sauer, Christiane Adam, Gabriele Born, Dietrich Alte, Eberhard Greiser, Ursula Haertel, et al. Study of health in pomerania (ship): a health examination survey in an east german region: objectives and design. *Sozial-und Präventivmedizin*, 46:186–194, 2001.

[27] Eser Kandogan. Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 73–82. IEEE, 2012.

[28] Paul Klemm, Kai Lawonn, Sylvia Glaßer, Uli Niemann, Katrin Hegenscheid, Henry Völzke, and Bernhard Preim. 3d regression heat map analysis of population study data. *IEEE transactions on visualization and computer graphics*, 22(1):81–90, 2015.

[29] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. Internal and external visual cue preferences for visualizations in presentations. In *Computer Graphics Forum*, volume 36, pages 515–525. Wiley Online Library, 2017.

[30] Nicholas Kong and Maneesh Agrawala. Graphical overlays: Using layered elements to aid chart reading. *IEEE transactions on visualization and computer graphics*, 18(12): 2631–2638, 2012.

[31] William Leiss. Effective risk communication practice. *Toxicology letters*, 149(1-3): 399–404, 2004.

[32] Li Li, Zhigui Liu, Xiao Wang, Xiaoqian Zhang, and Lili Wang. Pregnancyline: A visual analysis system for pregnancy care and risk communication. *IEEE Access*, 10: 111493–111506, 2022.

[33] Jock Mackinlay. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2):110–141, 1986.

[34] Jock Mackinlay, Pat Hanrahan, and Chris Stolte. Show me: Automatic presentation for visual analysis. *IEEE transactions on visualization and computer graphics*, 13(6): 1137–1144, 2007.

[35] Monique Meuschke, Laura A Garrison, Noeska N Smit, Benjamin Bach, Sarah Mittenentzwei, Veronika Weiß, Stefan Bruckner, Kai Lawonn, and Bernhard Preim. Narrative medical visualization to communicate disease data. *Computers & Graphics*, 107:144–157, 2022.

[36] Habshah Midi, Saroje Kumar Sarkar, and Sohel Rana. Collinearity diagnostics of binary logistic regression model. *Journal of interdisciplinary mathematics*, 13(3):253–267, 2010.

[37] Sarah Mittenentzwei, Laura A Garrison, Eric Mörth, Kai Lawonn, Stefan Bruckner, Bernhard Preim, and Monique Meuschke. Investigating user behavior in slideshows and scrollytelling as narrative genres in medical visualization. *Computers & Graphics*, 2023.

[38] Sarah Mittenentzwei, Veronika Weiß, Stefanie Schreiber, Laura A Garrison, Stefan Bruckner, Malte Pfister, Bernhard Preim, and Monique Meuschke. Do disease stories need a hero? effects of human protagonists on a narrative visualization about cerebral small vessel disease. *Computer Graphics Forum*, 42(3), 2023.

[39] Christopher JL Murray, Aleksandr Y Aravkin, Peng Zheng, Cristiana Abbafati, Kaja M Abbas, Mohsen Abbasi-Kangevari, Foad Abd-Allah, Ahmed Abdelalim, Mohammad Abdollahi, Ibrahim Abdollahpour, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The lancet*, 396(10258):1223–1249, 2020.

[40] Steffen Oeltze-Jafra and Bernhard Preim. Survey of labeling techniques in medical visualizations. In *Proc. of Eurographics Workshop on Visual Computing for Biology and Medicine (EG VCBM)*, pages 199–208, Wien, 2014.

[41] Daniel Perry, Bill Howe, Alicia MF Key, and Cecilia Aragon. Vizdeck: Streamlining exploratory visual analytics of scientific data. In *iConference*, 2013.

[42] Timothy B Plante and Mary Cushman. Choosing color palettes for scientific figures. *Research and Practice in Thrombosis and Haemostasis*, 4(2):176, 2020.

[43] Matthias Plaue. Supervised machine learning. In *Data Science: An Introduction to Statistics and Machine Learning*, page 232. Springer, 2023.

[44] Douglas Alan Powell and William Leiss. *Mad cows and mother's milk: the perils of poor risk communication*, volume 10. McGill-Queen's Press-MQUP, 1997.

[45] Donghao Ren, Matthew Brehmer, Bongshin Lee, Tobias Höllerer, and Eun Kyoung Choe. Chartaccent: Annotation for data-driven storytelling. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pages 230–239. Ieee, 2017.

[46] Meike Ressing, Maria Blettner, and Stefanie J Klug. Data analysis of epidemiological studies: part 11 of a series on evaluation of scientific publications. *Deutsches Arzteblatt International*, 107(11):187, 2010.

[47] Purvi Saraiya, Chris North, and Karen Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE transactions on visualization and computer graphics*, 11(4):443–456, 2005.

[48] Arvind Satyanarayan and Jeffrey Heer. Authoring narrative visualizations with ellipsis. *Computer Graphics Forum*, 33(3):361–370, 2014.

[49] Doris Schaeffer, Eva-Maria Berens, Dominique Vogt, Svea Gille, Lennert Griese, Julia Klinger, and Klaus Hurrelmann. Health literacy in germany—findings of a representative follow-up survey. *Dtsch Arztebl Int*, 118(43):723–9, 2021.

[50] Ralf Schmälzle, Britta Renner, and Harald T Schupp. Health risk perception and risk communication. *Policy Insights from the Behavioral and Brain Sciences*, 4(2):163–169, 2017.

[51] C Mary Schooling and Heidi E Jones. Clarifying questions about "risk factors": predictors versus explanation. *Emerging themes in epidemiology*, 15(1):1–6, 2018.

[52] Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics*, 16(6):1139–1148, 2010.

[53] Arash Shaban-Nejad, Maxime Lavigne, Anya Okhmatovskaia, and David L Buckeridge. Pophr: a knowledge-based platform to support integration, analysis, and visualization of population health data. *Annals of the New York Academy of Sciences*, 1387(1):44–53, 2017.

[54] Danqing Shi, Xinyue Xu, Fuling Sun, Yang Shi, and Nan Cao. Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):453–463, 2020.

[55] Arjun Srinivasan, Steven M Drucker, Alex Endert, and John Stasko. Augmenting visualizations with interactive data facts to facilitate interpretation and communication. *IEEE transactions on visualization and computer graphics*, 25(1):672–681, 2018.

[56] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15(1):1–28, 2015.

[57] Lyndal J Trevena, Brian J Zikmund-Fisher, Adrian Edwards, Wolfgang Gaissmaier, Mirta Galesic, Paul KJ Han, John King, Margaret L Lawson, Suzanne K Linder, Isaac Lipkus, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. *BMC medical informatics and decision making*, 13(2):1–15, 2013.

[58] ER Tufte. Small multiples. *Envisioning Information*, pages 67–80, 1990.

[59] Xiaofeng Wang and Zhenshun Cheng. Cross-sectional studies: strengths, weaknesses, and recommendations. *Chest*, 158(1):S65–S71, 2020.

[60] Yun Wang, Zhida Sun, Haidong Zhang, Weiwei Cui, Ke Xu, Xiaojuan Ma, and Dongmei Zhang. Datashot: Automatic generation of fact sheets from tabular data. *IEEE transactions on visualization and computer graphics*, 26(1):895–905, 2019.

[61] Steven Woloshin and Lisa M Schwartz. Communicating data about the benefits and harms of treatment: a randomized trial. *Annals of internal medicine*, 155(2):87–96, 2011.

[62] Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE transactions on visualization and computer graphics*, 22(1): 649–658, 2015.

[63] Kanit Wongsuphasawat, Zening Qu, Dominik Moritz, Riley Chang, Felix Ouk, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. Voyager 2: Augmenting visual analysis with partial view specifications. In *Proceedings of the 2017 chi conference on human factors in computing systems*, pages 2648–2659, 2017.

[64] Ning Yu, Mingwen Zheng, Ximena Andrade, and Ron Patane. Visual analysis for exploring the relation between air pollution, environmental factors and respiratory diseases. In *Workshop on Environmental Health and Air Pollution*, pages 232–238, 2018.

[65] Sujia Zhu, Guodao Sun, Qi Jiang, Meng Zha, and Ronghua Liang. A survey on automatic infographics and visualization recommendations. *Visual Informatics*, 4(3):24–40, 2020.

[66] Brian J Zikmund-Fisher, Mark Dickson, and Holly O Witteman. Cool but counterproductive: interactive, web-based risk communications can backfire. *Journal of medical Internet research*, 13(3):e1665, 2011.

[67] Carlo Zocchetti, Dario Consonni, and Pier A Bertazzi. Relationship between prevalence rate ratios and odds ratios in cross-sectional studies. *International journal of epidemiology*, 26(1):220–223, 1997.

Appendix - Public Questionnaire

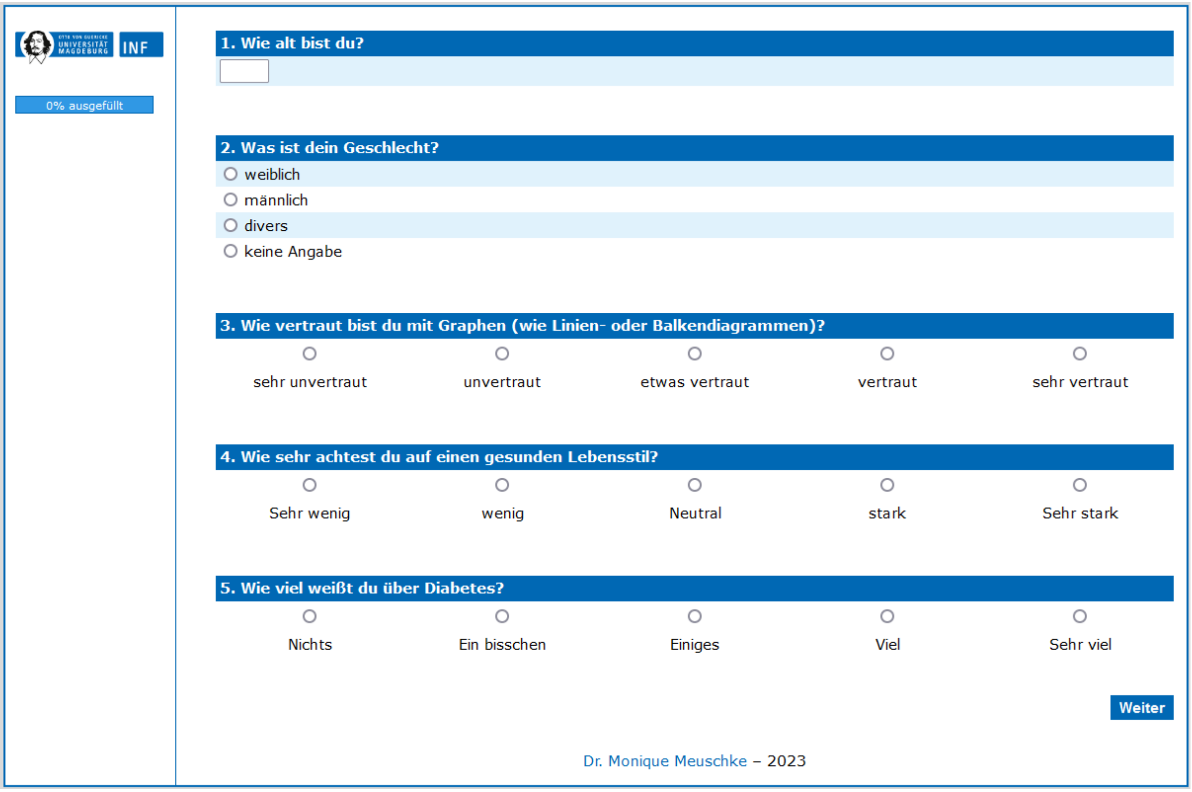
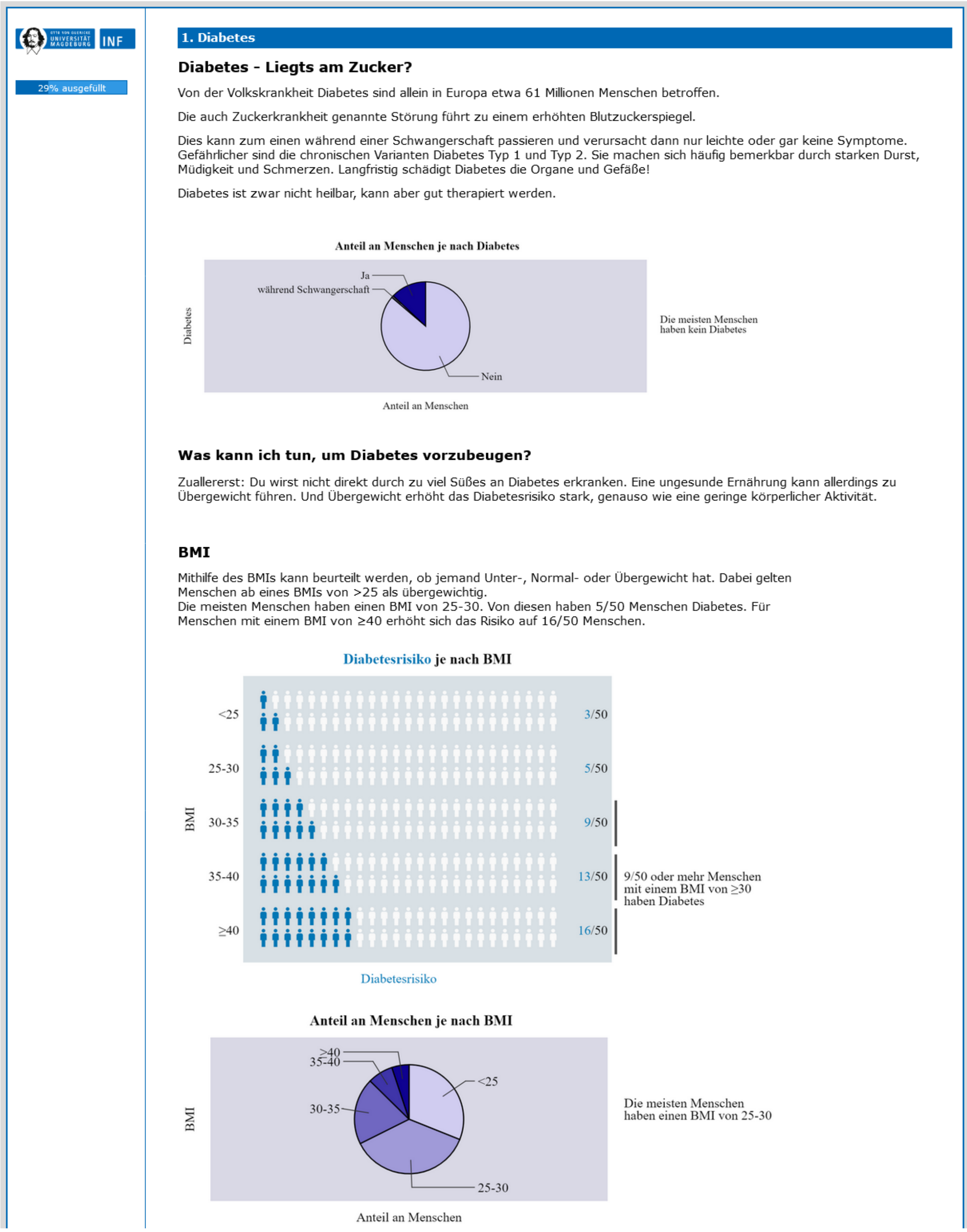


Figure A.1 Experience questions of the questionnaire



29% ausgefüllt

1. Diabetes

Diabetes - Liegts am Zucker?

Von der Volkskrankheit Diabetes sind allein in Europa etwa 61 Millionen Menschen betroffen.

Die auch Zuckerkrankheit genannte Störung führt zu einem erhöhten Blutzuckerspiegel.

Dies kann zum einen während einer Schwangerschaft passieren und verursacht dann nur leichte oder gar keine Symptome. Gefährlicher sind die chronischen Varianten Diabetes Typ 1 und Typ 2. Sie machen sich häufig bemerkbar durch starken Durst, Müdigkeit und Schmerzen. Langfristig schädigt Diabetes die Organe und Gefäße!

Diabetes ist zwar nicht heilbar, kann aber gut therapiert werden.

Was kann ich tun, um Diabetes vorzubeugen?

Zuallererst: Du wirst nicht direkt durch zu viel Süßes an Diabetes erkranken. Eine ungesunde Ernährung kann allerdings zu Übergewicht führen. Und Übergewicht erhöht das Diabetesrisiko stark, genauso wie eine geringe körperlicher Aktivität.

BMI

Mithilfe des BMIs kann beurteilt werden, ob jemand Unter-, Normal- oder Übergewicht hat. Dabei gelten Menschen ab eines BMIs von >25 als übergewichtig.
Die meisten Menschen haben einen BMI von 25-30. Von diesen haben 5/50 Menschen Diabetes. Für Menschen mit einem BMI von ≥40 erhöht sich das Risiko auf 16/50 Menschen.

Figure A.2 Beginning of the data story for the intention educate.

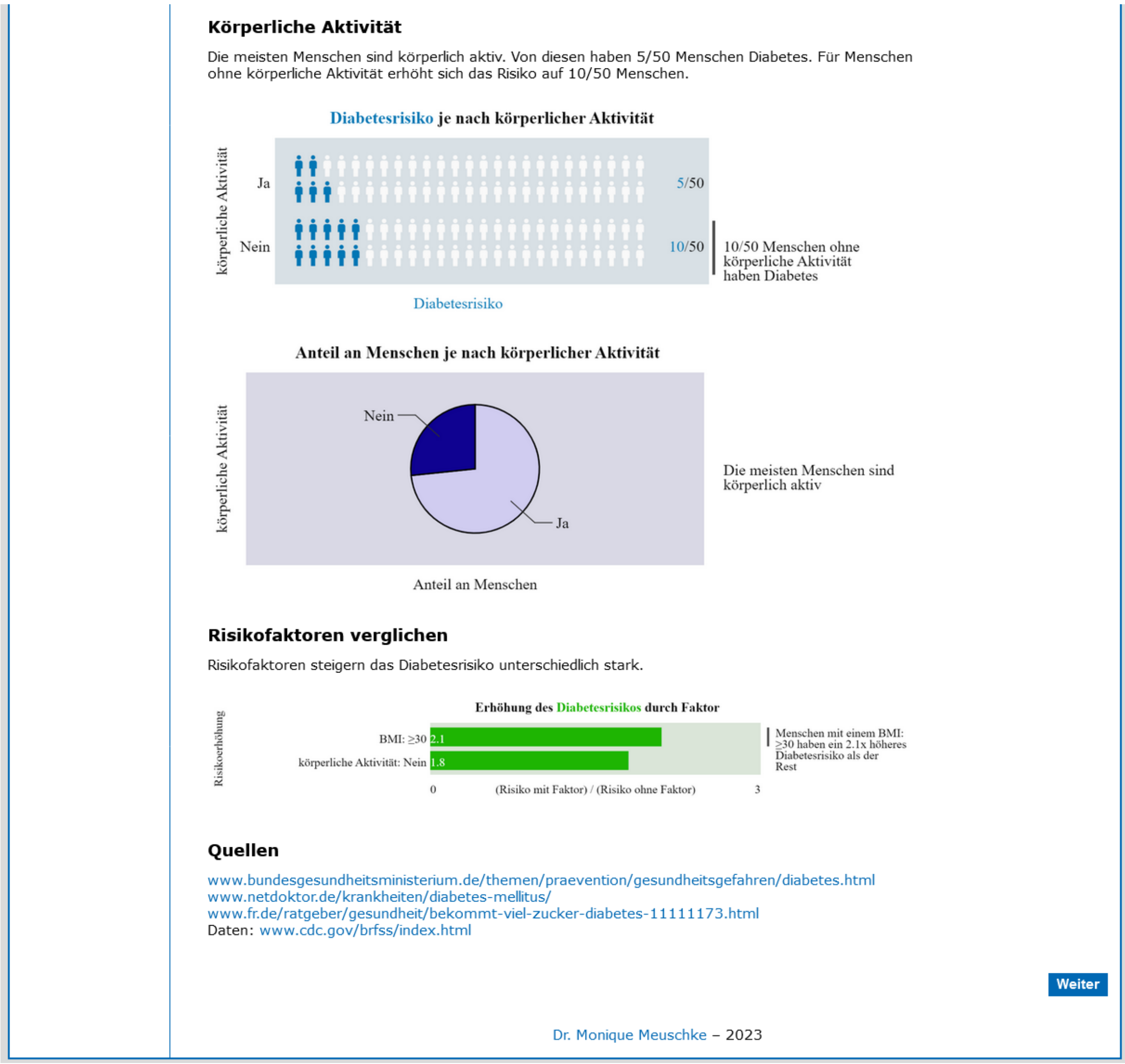


Figure A.3 End of the data story for the intention educate.

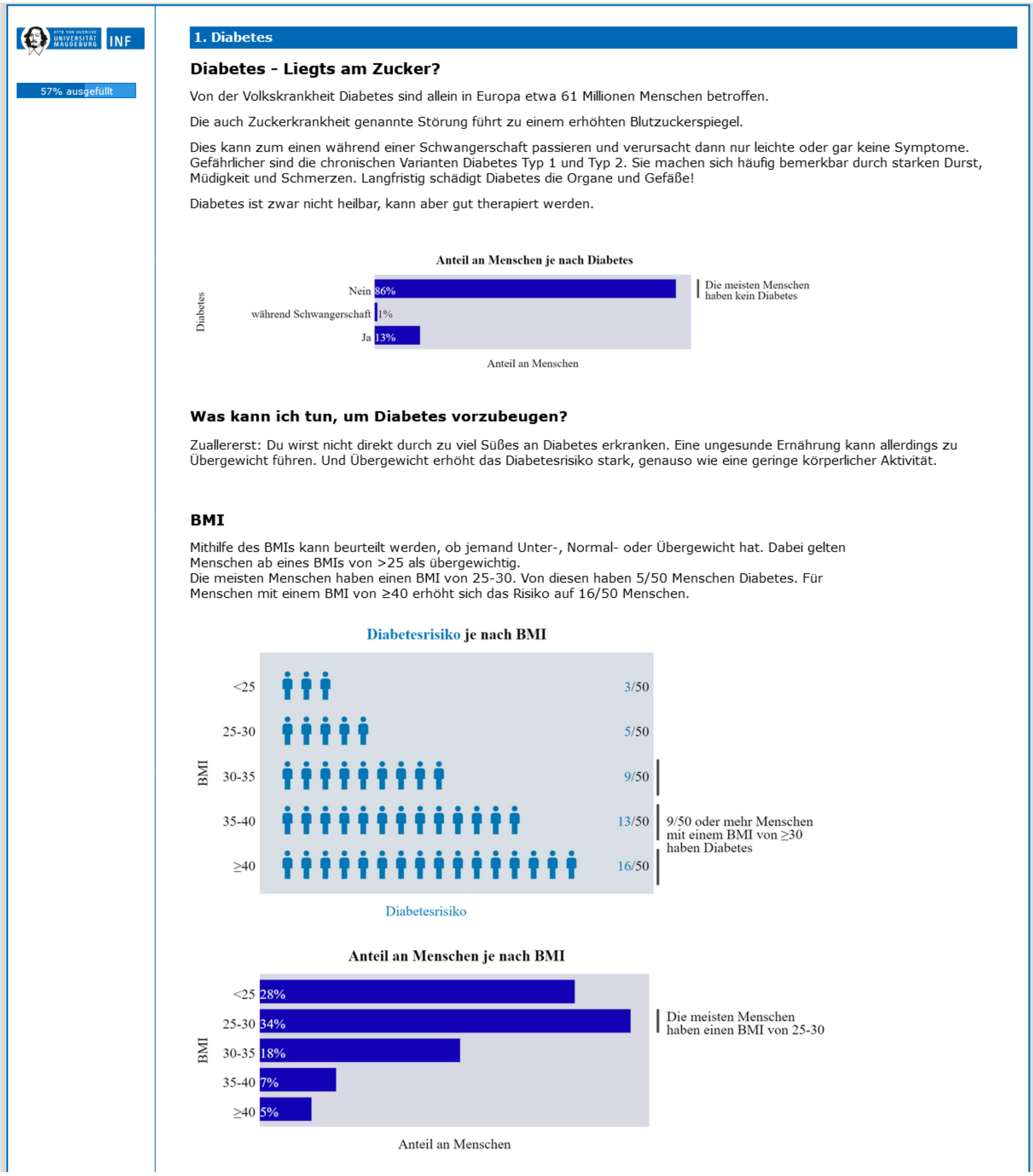


Figure A.4 Beginning of the data story for the intention convince.

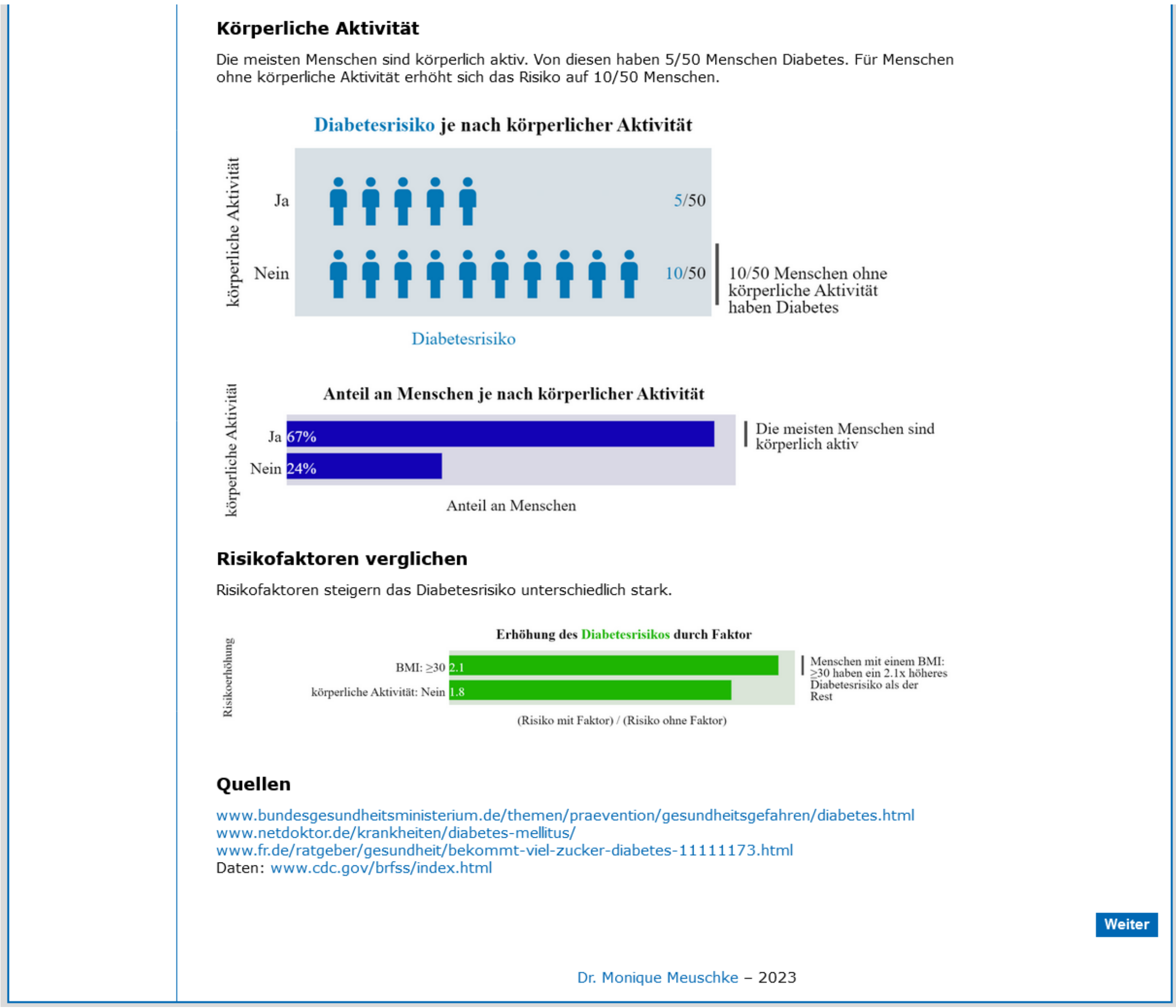
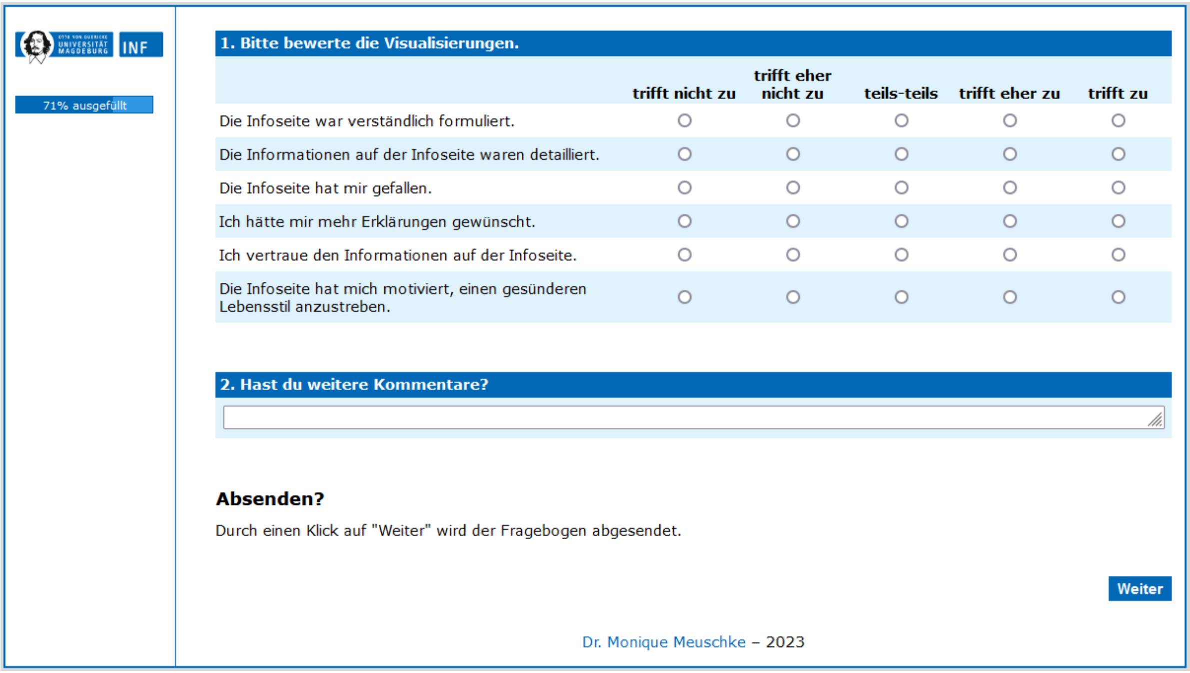


Figure A.5 End of the data story for the intention convince.



INF

71% ausgefüllt

1. Bitte bewerte die Visualisierungen.

	trifft nicht zu	trifft eher nicht zu	teils-teils	trifft eher zu	trifft zu
Die Infoseite war verständlich formuliert.	○	○	○	○	○
Die Informationen auf der Infoseite waren detailliert.	○	○	○	○	○
Die Infoseite hat mir gefallen.	○	○	○	○	○
Ich hätte mir mehr Erklärungen gewünscht.	○	○	○	○	○
Ich vertraue den Informationen auf der Infoseite.	○	○	○	○	○
Die Infoseite hat mich motiviert, einen gesünderen Lebensstil anzustreben.	○	○	○	○	○

2. Hast du weitere Kommentare?

Absenden?

Durch einen Klick auf "Weiter" wird der Fragebogen abgesendet.

Weiter

Dr. Monique Meuschke – 2023

Figure A.6 Feedback questions of the questionnaire