Freie Universität Berlin

Bachelor thesis, Institute of Computer Science, Freie Universität Berlin

Human-Centered Computing (HCC)

Animated Transitions to Support Visualization of Missing Data

Lea Tschiersch

Primary Reviewer: Prof. Dr. Claudia Müller-Birn

Secondary Reviewer: Prof. Dr.-Ing. Volker Roth

Berlin, 22.07.2019

Eidesstattliche Erklärung

Ich versichere hiermit an Eides Statt, dass diese Arbeit von niemand anderem als meiner Person verfasst worden ist. Alle verwendeten Hilfsmittel wie Berichte, Bücher, Internetseiten oder ähnliches sind im Literaturverzeichnis angegeben, Zitate aus fremden Arbeiten sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungskommission vorgelegt und auch nicht veröffentlicht.

Berlin, den 22.07.2019

Lea Tschiersch

Abstract

Large datasets often contain missing data, which needs to be dealt with when visualizing the data, either by deletion or by depicting missing data in a certain way. These visualizations are often interactive. Animated transitions can be implemented to smoothen the switch between views.

In the context of the IKON project, which goal it is to facilitate knowledge transfer, animated transitions and missing data visualizations for a line chart will get conceptualized, implemented and evaluated. The general goal of the thesis is to enhance the way how users work with visualizations consisting missing data. Additionally, more information is tried be to collected regarding the user's interaction with the chosen missing data visualizations.

This thesis explores the theory behind missing data, why data can be missing and what can be done to handle missing, including imputation methods and visualization techniques. It also examines already existing visualization techniques and studies, presenting their findings regarding preference of visualization types and different criteria such as accuracy and confidence in data.

Next, this thesis delves into the theory of animated transitions for data graphics. This includes the different types of transitions in data graphics and rules which should be followed to create successful animated transitions. Two different studies about animated transitions are be presented and their results are summarized, showing the possible positive effect animations can have for transitions.

Building on the theory of missing data and animated transitions in addition to the presented existing studies and the goal of the thesis, design requirements are set up. In the implementation "dashing" is chosen as missing data visualization and slow and simple-looking transitions are added to the line chart of the IKON project.

For evaluation a pilot study is conducted in which three participants interact with the implementation. It finds that all participants prefer having an animated transition over a static one. They were able to discern missing data from existing data. The pilot study, however, did also suggest that confidence in the visualization with missing data was high since participants were not hesitant to trust the missing data visualization.

With the pilot study conducted and an implementation in place this thesis presents a first glimpse on how animated transitions for missing data visualizations can be implemented, serving as a possible foundation for further studies.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Topic and Context

Data visualizations have been used successfully for presentation and analysis of data of all kinds for centuries. The rise of today's known visualizations started in the nineteenth century where statistical depictions such as bar charts became an easy way to share information visually [9]. Since then the amount of accessible data has increased rapidly with larger and larger datasets becoming available, aided through the rise of digital technology [7].
However, with the size of datasets the complexity rises as well, making it difficult for users to overlook every piece of information and draw conclusions.
Consequently, these vast quantities of data often call for the creation of several visualizations to aid end users in exploring and understanding the data and its correlations. In addition, visualizations are often interactive, giving the end user the chance to change for example the year in a bar chart or to add the sections which are shown in a pie chart.

Often large datasets are composed of data from different sources, which can all vary in quality and completeness. Sources might, for example, be measurements taken in the field, data collected via crowd-sourcing or logged data. As a consequence, datasets can often contain missing data, meaning that no data is stored where there should be. Handling these occurrences is one of the challenges of visualizing large amounts of data. However, it is a challenge that cannot be avoided unless all data in connection to the missing data gets omitted. Babad and Hoffer[3] mention as early as 1984 that missing data has to be represented, suggesting standard methods for dealing with it. Yet even nowadays there are no evaluated standard methods set.

When visualizing data in several views these depictions are usually separate from each other, meaning that when one switches to the next visualization the page will reload, resulting in an abrupt interruption for the user. The same disruption commonly happens while changing data within one visualization. This can impede the user in connecting the visualizations mentally as their train of thought might be lost[12].
To combat this disruption, animated transitions can be utilized to guide the user in their interaction with visualizations by linking the different views together in a coherent way, as it has been put to use in several different implementations [12], [15].

The role of missing data for animated transitions has yet to be determined. It is still unclear how the use of animated transitions can benefit the user

in working with datasets which include missing data. This thesis will try to tackle this challenge by conceptualizing, implementing and testing animated transitions for missing data in the context of the IKON project, which goal it is to support researchers at the National History Museum Berlin in knowledge transfer activities, including communication of knowledge to the public. An important means to come closer to this goal is the development of a visual prototype that helps analyze and identify existing and potential ways of knowledge transfer for different target groups[13].

1.2 Goal of the Thesis

This thesis pursues the objective of designing, implementing and evaluating both the visualization of missing data and the creation of transitions for them to obtain a functioning prototype which has been evaluated by a brief pilot study. The implementation will be built upon the already existing line chart of the IKON project, which depicts the amount of research projects conducted per year. Per research area there is one line, resulting in one line chart with several colored lines. The IKON project has not yet implemented any solution for the visualization of missing data.

The implementation and evaluation of a missing data visualization try to answer the following questions:

- Will the user realize that data is missing?
- Will the visualization assist in noticing trends in the data?
- Will the missing data visualization create bias?
- Will the user have higher confidence while interpreting the data than if missing data was just not visualized?

The last two questions are a direct continuation of what Song and Szafir evaluated in "Where's My Data? Evaluating Visualizations with Missing Data"[18], where their results showed that highlighting and downplaying missing data respectively raised and lowered quality perception. Finding a visualization which produces only little bias is one of the challenges of this work, as it is desirable that the perceived quality is close to the actual data quality.
The implementation of a fitting transition will be the second goal of this thesis. The following questions will be tried to answer:

- Does creating an animated transition help the user to keep focus on the data during changes?
- Does the creation of an animated transition assist in working with missing data?

The first question is following the study of Heer and Robertson, in which they found that animated transitions can facilitate "graphical perception of changes"[12, p.1247].

1.3 Approach

For a successful implementation it is necessary to evaluate already existing literature about both missing data and animated transitions in the context of data visualizations. By examining existing visualization and animation types and extracting their features, benefits and disadvantages a fitting visualization type and animated transition type can be found. These will need to work well with the already existing visualization of the IKON project.

To ensure the success of the chosen visualization and animation, which will be implemented into the IKON project, an evaluation is planned. The scope of this thesis lends itself for a qualitative rather than quantitative evaluation. A small pilot study will be set up and conducted. In it, the interviewees will interact with the implemented version of the missing data visualization and animated transition. This will be an attempt to answer the questions of the last chapter and represent foundation for future studies.

1.4 Structure of the Thesis

This thesis starts by giving an overview over the theory behind missing data, which will facilitate the understanding when presenting existing missing data visualization types. The same will be done for animated transitions. First the theory of animated transitions will get discussed before animated transitions for data graphics will get evaluated.

In the next chapter the existing chart of the IKON project will be presented to lead the way for the discussion of the implementation, including the design requirements and rationales. Lastly, the evaluation via pilot study will be explained, followed by the discussion of the result of the implementation and a conclusion.

2 Background

2.1 Missing Data

2.1.1 Missing Data Theory

When working with datasets it often happens that the available data is incomplete. This can have numerous reasons. If the data is taken from a survey it might be because participants did not answer some questions. If the data is measured by a weather station missing data might occur during a power outage. If several datasets are combined into one there might be missing values due to different data quality of the datasets. Also during the storage phase data might be lost due to a memory error[18]. It is necessary to take a closer look at the reasons why missing data is occurring to handle it properly and find efficient ways of visualizing it. This chapter will first get into the mechanisms of missing data and its consequences before focusing on existing examples of visualizations, which have been implemented before and are helpful for the goal of this thesis.

Mechanisms of Missing Data

When handling missing data, it is important to observe reasons and mechanisms behind missing data, as different mechanisms might require different ways of dealing with them. In addition, it can be helpful to minimize the occurrence of missing data in a project when the reason is known[1].

Commonly it is possible to discern between three different categories, as done by several authors[1], [11], [17], [20]. These categories will be explained in the following and will include examples for further illustration. All examples will share the context of a dataset which consists of data about research projects, their conduction dates and further information, similar to the IKON project.

- **Missing Completely at Random (MCAR)**

  Missing data falls under this category when there is no underlying pattern of missingness and it does not depend on other variables or data.

  As an example, a person would be filling in project information and would forget at random points to add data to some projects.

- **Missing at Random (MAR)**

  Much like MCAR the data of MAR is missing randomly, however, there is a reason outside of the actual dataset. It is a reason for missing data that can be ignored while interpreting the data.

This would correspond to several projects missing data because the person filling in the information quit their job.

- **Missing Not at Random (MNAR)**

  Different than both MCAR and MAR this mechanism of missing data is not ignorable when dealing with a dataset due to a relation between the missingness and the actual variables.

  If some projects deliberately had values which were not added to the dataset it would be MNAR. As an example, this could be related projects which were not linked to in the project's data entry because they were not deemed necessary to fill in.

Handling Missing Data

Since missing data can often not be prevented, it has to be dealt with in one way or another. Just as the reasons for the occurrence of missing data can be multitudinous so can the ways for dealing with it be. The easiest way would be ignoring the missing data and omitting it. However, this can lead to inaccurate interpretations as several authors note [11], [16]: Data which is missing (completely) at random may have the possibility of having the same relation between variables because the likelihood of each variable missing is the same. However, with data missing not at random, the bias between parameters can rise and other methods of handling with missing data can become necessary.

Imputation

When visualizing missing data in line charts one must decide whether to not show any missing data points, thus delete them, or to replace the missing data points with estimated values[20]. This process is called imputation.

There are numerous imputation methods with different qualities which can be chosen from. Depending on the specific dataset some might be better fit than others. This is also subject of a study conducted by Song and Szafir[18] where three imputation methods get compared to each other in the context of time series data and which findings will be discussed further along in this chapter. While the following list of imputation methods is by no means exhaustive it gives a small overview over some of them[18], [19], [20].

- **Zero-Filling**

  Missing values are substituted with zero.

- **Marginal Mean**

  The mean of all values is calculated and imputed every time there is a data point missing. Similar to zero-filling it is always the same value missing data points are substituted with.

- **Linear Interpolation**

  The missing data value is decided by linear interpolation between the nearest existing values of it in the line chart. In time series data this equals usually to the data points left and right next to the missing value(s).

- **Last Observation Carried Forward (LOCF) and Next Observation Carried Backward (NOCB)**

  By repeating the closest available data point next to the missing value, the missing data point gets imputed. This method only works for time series data. For LOCF this is the time-wise "older" value which gets used for imputation, for NOCB it is the time-wise "newer" value.

Visualization Techniques

In addition to the imputation technique a visualization technique for the missing data visualization can be chosen. The easiest option is using the same technique for existing and missing data, thus letting the whole line chart have the same visualization type. However, missing data which is not labelled sufficiently will make it difficult for the user to realize which of the data is missing. In the worst case users might not even realize they are working with non-existent data points at all, as was found in [8]. Looking at already existing visualizations I sorted these techniques into two different categories:

- In cases where missing data was chosen to be omitted a label can be added which makes note of data being missing[2], [8].

- In cases where missing data points were chosen to be interpolated a different formatting of the missing data points can be chosen to notify the user of non-existent values[2], [18], [22]. This can include for example the shape of the line in the line chart, the color or opacity.

A more detailed view over the different visualization techniques can be found in the list of existing visualizations in the following section.

Choosing different visualization techniques can have an effect on how the user perceives missing data, as found in several studies[2], [18]. Depending on visualization the user might be heavily biased regarding the perceived data quality and their overall confidence in the dataset. This can go in both directions, as data quality might be perceived lower than it actually is, and thus discourages the user from working with the data. Song and Szafir[18] note that finding a middle ground between these two and avoiding bias is critical to effective visualization.

2.1.2 Existing Missing Data Visualizations

Song and Szafir's "Where's My Data? Evaluating Visualizations with Missing Data"

In "Where's My Data? Evaluating Visualizations with Missing Data" Song and Szafir[18] conduct crowd-sourced studies to compare several visualization techniques and imputation methods for both bar- and line charts in their effect on perceived data quality and confidence in the data set. The chosen imputation methods are zero-filling, marginal mean and linear interpolation. It was found that perceived data quality and confidence were highest for linear interpolation and worst for zero-filling. The visualization techniques were separated into four categories: highlighting, in which a different and saturated color was chosen for visualizing missing data points; downplaying, in which a less opaque visualization was chosen; annotation, in which the visualization was extended with error bars which gave additional information, and information removal, in which missing values were not depicted at all. Both the visualization techniques and the imputation methods can be seen in figure 2.1. The result of the survey showed that datasets with highlighted missing values were perceived to have higher quality, while downplaying and information removal lowered it. This led to the highest perceived quality when missing values were linearly interpolated and both highlighted and annotated.

The findings of this study are useful for the implementation since it observes the same case of a line chart and missing data and gives a good overview over several different types of visualizations. However, it has to be noted that the study does not determine whether any visualization or imputation is more useful than another as it only showed a link between visualization type and perceived quality.

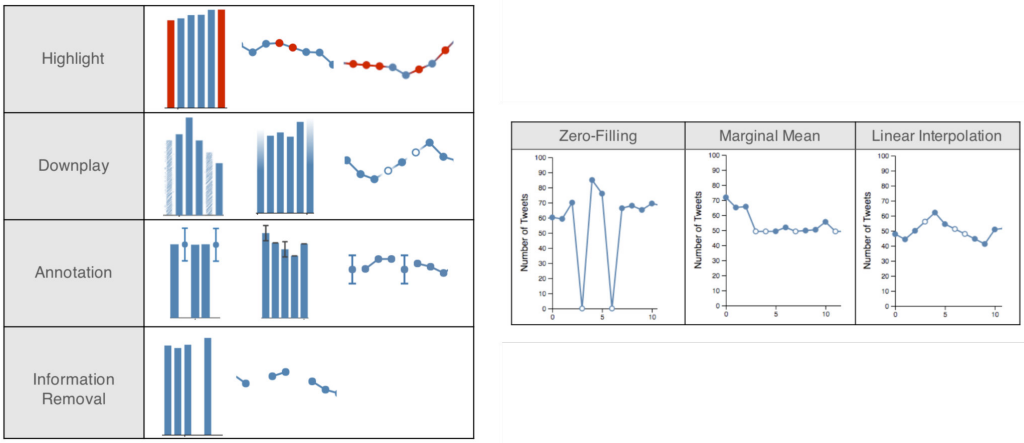

Figure 2.1: Song and Szafir's visualization techniques used in their study (left), as well as the tested imputation methods (right). Source: [18]

Eaton et al.'s "Visualizing Missing Data: Graph Interpretation User Study"

In Eaton et al.'s "Visualizing Missing Data: Graph Interpretation User Study"[8] a study is conducted in which three different visualizations are compared regarding the participant's ability to answer questions concerning the data. The three visualizations chosen were zero-fill with no other formatting to notify a value was missing, omission of missing data points and omission of missing data points with annotation that data was missing, as seen in figure 2.2. Users were asked questions about trend detection as well as simple value comparisons. Both the correctness of the answers and the user's perceived confidence in them were measured and compared, as well as the user's preference for a certain depiction. The study found that zero-fill was the worst visualization regarding correctness of answers and preference. While both absent and annotated visualization had similar values for correct responses users preferred being informed about missing data via annotation. Despite these differences between the visualizations, confidence in the answers by the users was high throughout all three of them.

This study is helpful as it also uses line charts with missing data as subject. However, while the visualization types are not as vast as in Song and Szafir's study, it does make clear that missing data has to be made noticeable to be recognized as such.

Figure 2.2: Eaton et al.'s visualization techniques used in their study. From left to right: zero-filling, omission of data, omission of data including annotation. Source: [8]

Andreasson and Riveiro's "Effects of visualizing missing data: an empirical evaluation"

Andreasson and Riveiro present a study in "Effects of visualizing missing data: an empirical evaluation"[2] where three different visualization techniques are compared to each other. Different from the previously mentioned studies this evaluation lays focus on user preference and willingness to answer questions about the missing portions of the data set, describing these as "risky decisions".

Chosen visualization techniques for this were omission of missing values, omission of missing values with annotation and "fuzziness" which can be seen in figure 2.3. The study found that omitted values including annotations were the most preferred and fuzziness the least preferred visualization types. Missing values with annotations resulted in the highest percentage of "risky decisions" while completely missing values resulted in the lowest one.

This study is interesting as it is the first to try using "fuzziness" to convey missingness even though the preference for it was low. It also replicates Eaton et al.'s findings about preference for notifications about missing data.

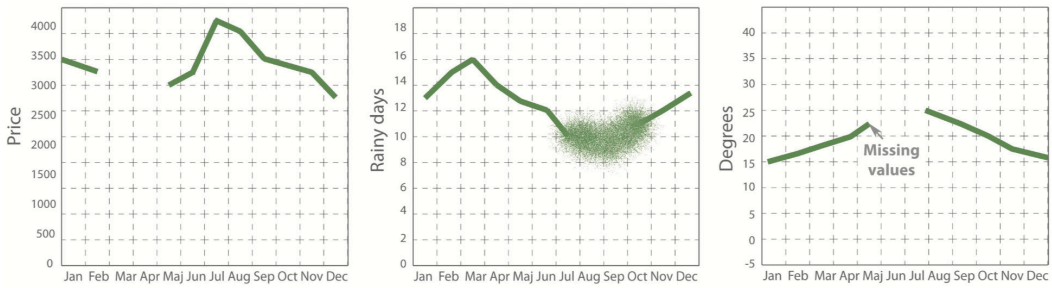

Figure 2.3: Andreasson and Riveiro's Visualization techniques used in their study. From left to right: omission of missing data, fuzziness and missing values with annotation. Source: [2]

Boukhelifa et al.'s "Evaluating Sketchiness as a Visual Variable for the Depiction of Qualitative Uncertainty"

While not specifically about missing data this paper is focusing on the depiction of uncertainty in data. According to Boukhelifa et al.[6] uncertainty is tightly linked to the user's confidence in data. In their paper they work with four different techniques to depict uncertain data, namely dashing, blur, gray scale and sketchiness, as seen in figure 2.4. Sketchiness in this case was a way to mimic hand-drawn lines. In a study they compared these types with each other to find out which technique was the preferred one and the one which could be read with most accuracy. For this they used a bar chart, a family tree, a network, a Venn diagram and two maps as scenarios. The result was that dashing was the most preferred type and sketchiness the least preferred one, with participants noting that it seemed "unprofessional". Sketchiness was also the technique which was the least accurate to read.

Missing data and uncertain data are closely linked. Andreasson and Riveiro[2] and Song and Szafir[18] speak of both uncertain data and missing data when discussing their visualization techniques. This makes these visualizations just as viable for the implementation. While no line charts are used in Boukhelifa et al.'s paper the visualization techniques can easily be used for them since they are all in line-form.

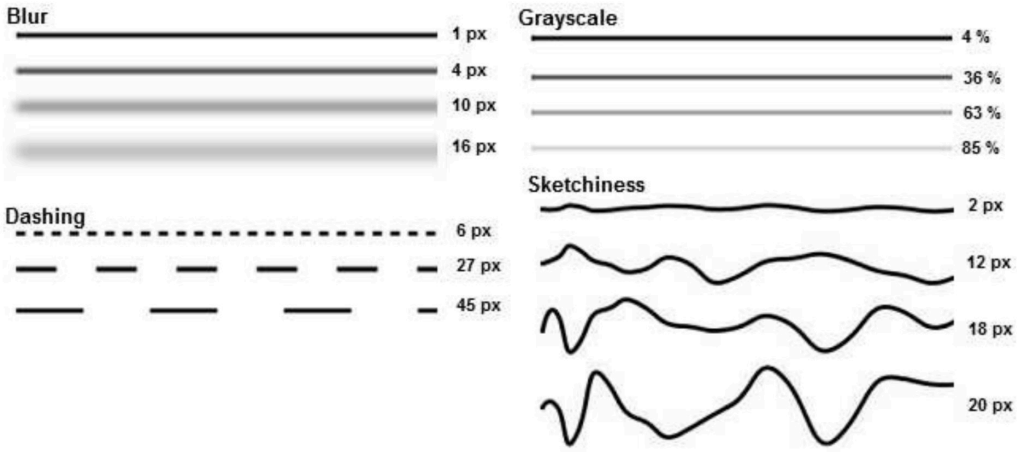

Figure 2.4: Boukhelifa et al.'s visualization techniques used in their study. Depicts several severities of the individual techniques. Source: [6]

2.2 Animated Transitions

2.2.1 Transition Theory

When switching between different visualizations or when changing input data, for example by changing the shown years in line chart, the view on the data changes by transitioning. These transitions can get animated, which studies found makes it easier for users to follow[12], [14]. There are several properties of animated transitions which will be discussed in the next sections before specific transition examples will be examined.

Transition Types

Transitions in data visualizations can be used in different ways. Heer and Robertson[12] created a taxonomy of transition types to describe differences between them. In addition to listing these it will be determined if these types of transitions are needed in the implementation.

- **View Transformation**

  When neither the actual data nor the visualization type change but instead only the field of view is adjusted, for example by zooming in or out or by scrolling along, it is called view transformation. As the implementation does not have any scrolling or zooming features this transformation will not be used.

- **Substrate Transformation**

  Substrate transformations happen by changing the spatial view, for example by scaling the axes, logarithmic transformations or distortions of

the field of view. These transformations are needed in the implementation, mostly for re-scaling purposes after other transformations were applied.

- **Filtering**

  When filtering is applied some of the data in a visualization might be turned visible or invisible depending on certain conditions. Heer and Robertson note that a filtering transition might be accompanied by a substrate transformation to make sure the axes are properly re-scaled. In this thesis's implementation of transitions for a line chart this might happen when toggling projects of a specific faculty on or off, resulting in a path in the line chart either vanishing or appearing.

- **Ordering**

  During an ordering transition data will be rearranged and ordered depending on different conditions. These transitions are not needed in this thesis's implementation as there are no ways to re-order the data.

- **Timestep**

  Timestep transitions occur when the data is temporally changed. A substrate transformation may follow for re-scaling purposes. In the implementation that is the case when different years are selected to be shown.

- **Visualization Change**

  This transformation type occurs when the data visualization changes, for example when a donut chart is changed into a bar chart. These transitions are not part of this implementation, however, as the focus lies on transitions within a visualization and not between different ones.

Viability of Animated Transitions

Several sources have studied the usefulness of animated transitions and found it aided in graphical perception both regarding object tracking and estimation of change, leading to less errors when asked about the data[12] and aided in "building mental maps of spatial information"[4].

However, Tversky et al.[21] are critical towards animations being used instead of static views as they can easily become too complex and fast-paced to be of use. To counteract these problems they suggest two rules for the design of animated transitions, which are also being considered and refined in in Kim et al.'s[14] and Heer and Robertson's[12] studies containing animated transitions.

The first rule, that of "Congruence", determines that during animations visualizations should be kept valid at all times and visualizations used should be

as consistent and non-ambiguous as possible. Following this rule aids in making the transition understandable in every frame with no room for speculation what specific intermediate frames depict.

The second rule, that of "Apprehension", determines that transitions should be kept simple. This entails making sure animated objects are not occluding each other and making transitions as simple and predictable as possible.

Another tool which can be used to enhance apprehension is staging, as used by Heer and Robertson[12] in their study as well. In case an animated transition is too complex, for example when too many objects are being moved or scaled, or when several transition types happen at once, it can be viable to break this large and complex transition in several smaller transitions. These can then be chained together to make it easier for the user to follow. However, Heer and Robertson also note that while simple staging animations proved beneficial in their case there have not yet been detailed studies on the timing of staging, including pauses in between the stages.

Also important for apprehension is the transition duration, which defines how long the switch between the views takes. Heer and Robertson[12] note that both too short and long duration can be a hindrance and set their own animation duration to 1.25 seconds, while Kim et al.[14] set theirs to 2 seconds.

2.2.2 Existing Uses of Animated Transitions

Heer and Robertson's "Animated Transitions in Statistical Data Graphics"

Heer and Robertson's "Animated Transitions in Statistical Data Graphics"[12] introduces *DynaVis*, a framework for creating visualizations including animated transitions. In this, transitions are created between different views, including scatter plots, bar charts and donut charts. Heer and Robertson claim that using animated transitions has a positive effect on the user's perception. They support their theory by conducting two experiments, which findings were that users were indeed more perceptive when animation was employed. In addition, animated transitions were also preferred by users over static ones. However, different types of animated transitions were found to have different levels of effectiveness with e.g. staged animations being more powerful than animations where everything moved at once.

While Heer and Robertson do not work with line charts in their study and thus do not have exact transitions viable for this implementation their findings about animated transitions and their usefulness in general are still helpful.

Kim et al.'s "Designing Animated Transitions to Convey Aggregate Operations"

In "Designing Animated Transitions to Convey Aggregate Operations" Kim et al.[14] design animated transitions for several aggregate operations to convey

the actual aggregation type more clearly. Observed aggregation types were among others count, sum, maximum, minimum and standard deviation. They follow Heer and Robertson's[12] and Tversky et al.'s[21] design principles and in a corresponding study find that animated transitions are consistently preferred over static transitions. The accuracy of correctly guessing the animated aggregation type, however, varied depending on the type, but never was worse than the static version.

3 Development of Animated Transitions for Missing Data

3.1 The Existing Application and Visualization

The IKON project aims to help researchers at the National History Museum Berlin with knowledge transfer. The creation of a visual prototype is an important part of this[13]. Next to other visualization forms a line chart has been created, which consist of several lines, all corresponding to a specific research area. The x-axis corresponds to the time, split into years, while the y-axis corresponds to the number of research projects. A depiction of this can be seen in figure 3.1. Interactivity is added by allowing the user to set a specific timescale or by toggling research areas on and off, thus hiding the corresponding lines in the line chart. The research areas are additionally separated into subcategories which can be filtered as well. When toggled off, the corresponding line will only contain projects which are not in that specific subcategory.

Figure 3.1: The existing line chart of the IKON project with no missing data visualization. It allows lines to be toggled on and off and filtered by choosing to include or exclude several sub-categories. Furthermore, the timescale can be changed via slider. Source: [10]

Missing data can occur in several places in this dataset as the it also contains information about the individual research projects. These possible missing

values are not necessarily visible in the line chart itself because the line chart serves as an overview, without taking the actual project contents into account. However, if the actual number of projects is missing in a research group, the missing data is becoming visible (or invisible, depending on the visualization type) in the line chart.

The visual prototype is created with JavaScript, using React for the user interface. In addition, D3 is used for the creation of the data visualizations themselves. These technologies will also be used in this thesis' implementation to work well with the existing line chart.

3.2 Design Requirements

While, as already discussed in the last chapter, there are several studies about the visualization of missing data and about the use of animated transitions in data visualizations, there has been no study yet which combines these two. However, as these two topics are linked, the requirements and considerations for this thesis's implementation are supported by the studies of both. First, there will be a focus on the requirements of the missing data visualization, then on the transitions, before the requirements of the combination will be discussed.

Missing Data Visualization

The chosen visualization technique for missing data needs to work with the already existing visualization of the IKON project. The requirements are closely linked to the goals of the thesis, as they were discussed in section 1.2. The implementation has the following requirements:

- The user needs to be able to determine where data is missing.
- Users should not feel disrupted when transitions appear. They need to be able to keep working with the data.
- As little bias as possible should be produced by the visualization.

Transitions

The first step for the design of transitions is to properly describe what transitions are. The line chart itself consists of several data points which can be connected by lines depending on the state of each data point. States consist of the value of the data point, which describes the position, and the type. Values of missing data get calculated depending on the chosen imputation type. The types possible in this specific implementation are non-missing (t_n), missing (t_m) and hidden (t_h). Non-missing data points are all points whose values are known and which are not filtered out. Filtering out might happen when the research area is toggled off or if the specific year is not part of the current view.

If a data point gets filtered out its type becomes hidden. Missing data points which are not filtered out have the type missing.

A transition is a switch from one state s_1 to another s_2. If the value of s_1 is different from s_2's value a translation will happen during the transition to move the data point to the new position. If the type of the data point changes as well there will be another aspect to the transition. Since each type can turn into every other available type this results in 3^2 different transitions. The transition types can be inspected in table 3.1.

t_n	$\longrightarrow$	t_n	t_m	$\longrightarrow$	t_h
t_m	$\longrightarrow$	t_m	t_n	$\longrightarrow$	t_h
t_h	$\longrightarrow$	t_h	t_h	$\longrightarrow$	t_n
t_n	$\longrightarrow$	t_m	t_h	$\longrightarrow$	t_m
t_m	$\longrightarrow$	t_n			

Table 3.1: All possible transitions between types. The types are: t_n (non-missing), t_m (missing), t_h (hidden).

All types of transitions need to function properly in the prototype.

Following Heer and Robertson[12] and Tversky et al.[21], these animated transitions should follow the rules of congruence and apprehension, as described in section 2.2.1.2. This is to make it easy for the user to understand what is happening and to follow the transition. As stated in the goals of the thesis, the animation should help in letting the user keep their focus on the data during changes. The animated transition needs to work with missing data, in the best case even make it more clear that data is missing. This will have to be evaluated by the pilot study.

3.3 Implementation of Animated Transitions to Support Missing Data

When creating a visualization containing missing data it needs to be made apparent to the user where data points are missing so that the user can properly work with the visualization. This was noted by Eaton et al.[8], who suggested specific visual attributes to indicate the location of missing data. If missing data is not made visible enough for the user to notice they might start an unnecessary search for it. This was something considered when finding a missing data visualization.

Since Song and Szafir[18] did find that information removal lowered the confidence in the data and it was not the preferred visualization technique for Andreasson and Riveiro's[2] and Eaton et al.'s[8] studies, mere deletion of missing values was not the chosen visualization. While annotation received high marks in preference, especially for Eaton et al.'s and Song and Szafir's studies, and was considered for the implementation it was ultimately not chosen as visualization either. The reason for this was the amount of lines available in the line

chart, which was feared to lead to a confusing and crowded display. Highlighting data with brighter colors, as it was subject of Song and Szafir's[18] studies, was not chosen as the visualization type, either, because of similar reasons. With a lot of differently colored lines already displayed adding more colors could make it more difficult for the user to read the chart quickly[5]. This led to a test-wise implementation of Boukhelifa et al.'s visualization techniques[6], namely sketchiness, blur and dashing, into the existing line chart, which can be seen in figure 3.2. Sketchiness was implemented using the same model as Boukhelifa et al. did to ensure a similar result as theirs.

The imputation methods considered were zero-filling, marginal mean and linear interpolation as used by Song and Szafir in their study[18]. For all three tested implementations linear interpolation was chosen. Because research projects often continue over several years there usually are no large variations between adjacent years. Because blur is barely visible and had the lowest

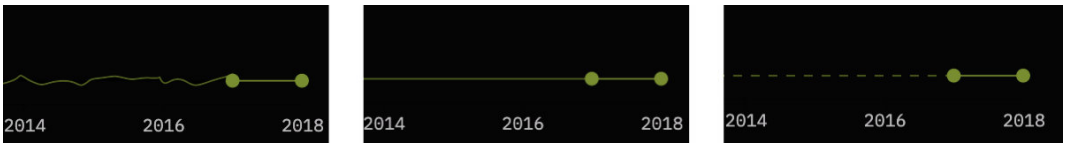

Figure 3.2: The implemented sketchiness, blur and dashing for the line chart.

preference in Boukhelifa et al.'s study it was not chosen. Sketchiness is a promising visualization. However, it is a very uncertain and not yet very well researched area and Boukhelifa et al. explicitly state that it can be difficult "for line graphs (...) where geometry is also perceived as related to values change"[6, p. 2778]. Furthermore, missing data is not the sole focus of the thesis and instead needs to interact with the animated transition. This is why dashing was chosen as missing data visualization.

As discussed in the design requirements there are 9 different transition types which need to be considered, in addition to the translation. Tversky et al.'s[21] rule of apprehension states that animated transitions should be kept simple. To follow this there was no additional effect added to transitions where the type stayed the same, meaning transitions from hidden to hidden, missing to missing and non-missing to non-missing were decided to just get translated if the values changed. This lowered the number of needed transition types from 9 to 6. In addition, one part of the rule of congruence for animation states that similar changes should have similar animations to keep it consistent and as clear to the user as possible[21]. This led to the decision to group the transitions further together, as can be seen in figure 3.3. To keep it simple and coherent these groups act similarly.

visible to hidden	hidden to visible	visible to visible
$t_m \longrightarrow t_h$	$t_h \longrightarrow t_m$	$t_n \longrightarrow t_m$
$t_n \longrightarrow t_h$	$t_h \longrightarrow t_n$	$t_m \longrightarrow t_n$

Table 3.2: The different types of transitions grouped into three different groups.

The duration for all transitions was chosen to be the same and was set to 1.25 seconds, equal to Heer and Robertson[12] as they did not note any problems with this duration. However, the animation duration will also be part of the pilot study for the evaluation and will be discussed further in the next chapter.

To keep animations simple and thus conform with the rule of apprehension they were set to be a simple linear transition between the two states, both for the translation and the type change. This means that the transition group of "visible to hidden" fades slowly out and the "hidden to visible" group fades slowly in. The "visible to visible" group changes by letting the dashes become bigger until it becomes a solid line or vice versa.

Staging was not added because the transitions were considered simple enough to follow, even with the rule of apprehension in mind. When changing the timescale there is only translation happening and all data points are moving into the same direction. When toggling a research area on or off only one line is changing. With further filtering options in the future filtering might become viable but it was not deemed necessary so far. The timing of staging and the order in which the points transition can certainly be an interesting topic for further study.

The finished implementation can be seen in 3.3 and 3.4.

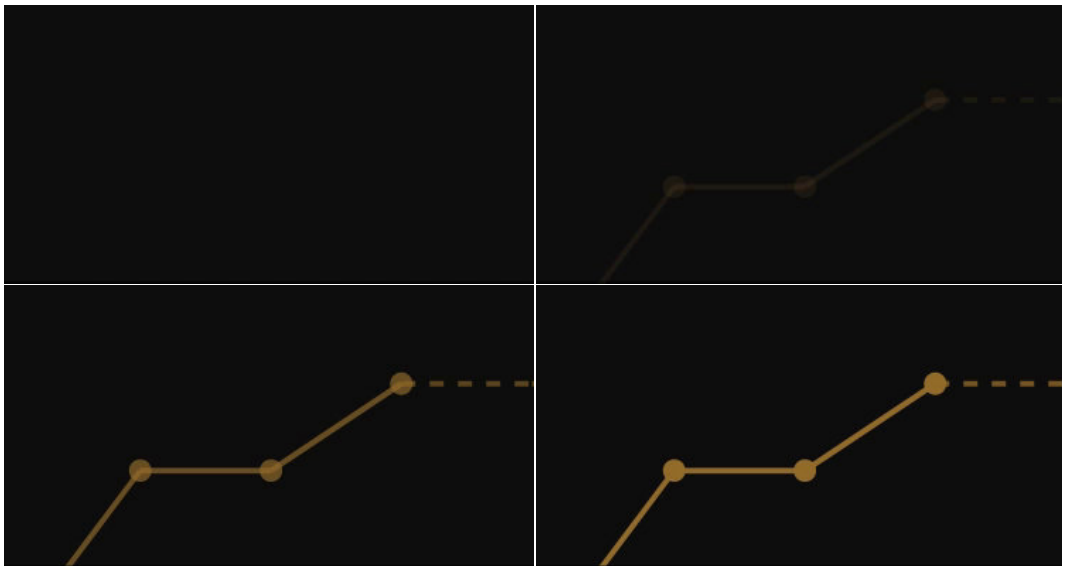

Figure 3.3: The animated transition between hidden and visible data points (left to right, top to bottom).

3.4 Evaluation

To evaluate the implemented visualization a small pilot study was created and conducted. While a large study is not in the scope of this thesis, the pilot

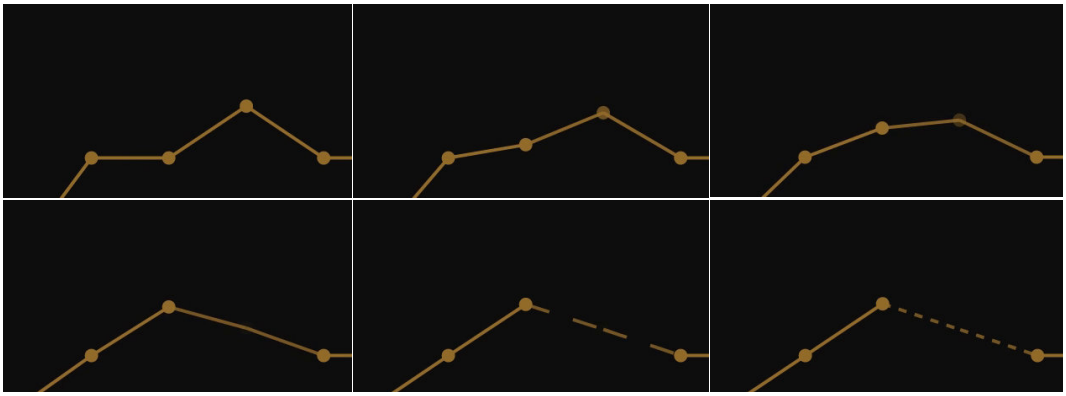

Figure 3.4: The animated transition between existing data and missing data, including a translation along the y-axis (left to right, top to bottom).

study can give valuable insight whether the implementation had the intended effect and can pave the way for further studies.

3.4.1 Setup of the Pilot Study

For the pilot study a standalone visualization was created, mimicking the implementation in the IKON project. A depiction of this is visible in figure 3.5. In addition to the line chart some options were added. In these the missing data visualization could be switched on and off, resulting in either blank spaces where data was missing or in the chosen "dashed" missing data visualization. One further option was to add circles to the missing data visualization, resulting in unfilled circles along the interpolated line, mimicking the full circles of the existing data. A slider allowed manipulating the duration of transition in ms, with the two extremes being an instant non-animated transition at 0ms and a two second long animated transition at 2000ms. Steps of the slider were 10ms.

With a button the user could filter the data, resulting in a transition being triggered, changing the values in the line chart. The last option was a time scale slider, giving the option to look at only parts of the line chart at once.

For every interview the setup and layout were first explained to the interviewee to enable them to work with it and understand the process. They were free to use the options in any way they pleased throughout the whole study. The exact explanations and the setup for the study can be found in the appendix, however, as the study was conducted in German the explanation is in German as well. After the introduction and the sign of a consent form, interviewees were shown a total of 6 different datasets. For each dataset they were asked three questions:

- If you look at only the year 2012, how would you describe what happens when turning the filter on? (Question about single value change)

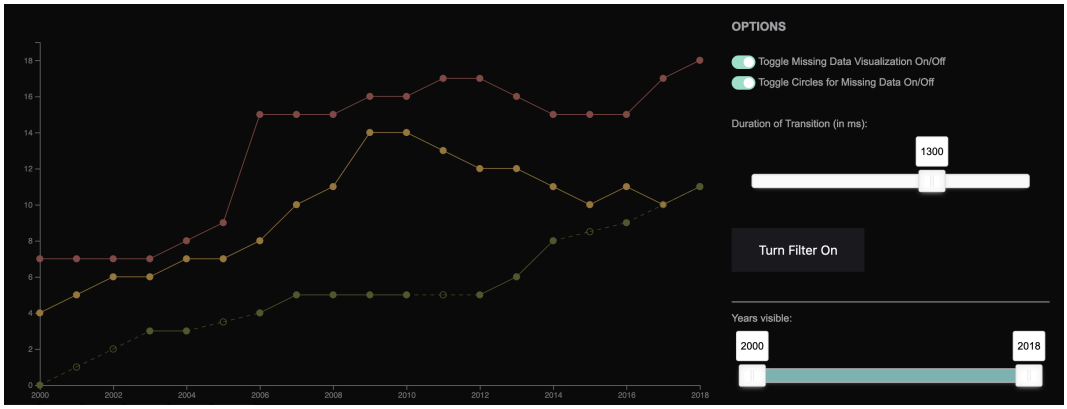

Figure 3.5: The setup of the pilot study. Depicts a line chart with several lines and missing data points. Options on the right allow the toggling of a missing data visualization, the addition of circles to the missing data visualization. Further options are the setting of the transition duration, being able to turn a filter on and off and changing the depicted years.

- Were there in general more research projects before or after the transition? (Question about averaging)
- Can you detect a trend before the transition? If yes, does it change after the transition? (Question about trend detection)

The questions about averaging and trend detection were based on Song and Szafir's study[18], which asked for the same for missing data visualization. They were changed for this study to instead include the transition as well. Interviewees were asked to comment on their thoughts and actions while answering the questions. This was done to collect more information about their solution path and their thoughts about the visualization.

There were two datasets with only one line in the line chart, two different datasets with two lines and two datasets with three lines. The first of each group of datasets consisted of lines without missing data while the second of each group contained big portions of missing data. This was done this way to have be able to compare the interviewee's responses to existing and missing data and to see whether the amount of different lines made a difference.

3.4.2 Goal of the Pilot Study

While statistics about confidence and ability to detect trends would be interesting it is not possible to achieve good statistics with only a few interviews. The goal of the pilot study is to get a first overview over how users interact with transitions and the visualization of missing data. This includes preference for one type. In the following section the findings will be presented.

3.4.3 Results of the Pilot Study

Interviewees

Before the study started a test-run was conducted to collect feedback. After the test-run one point of data in one dataset was changed (so that the first data set with two lines had no missing data at all) and the order of the options was changed slightly as well. In the following the results of the interviewees, which were interviewed after the test-run, will be presented. At times the test-run will get alluded to. This will only happen explicitly.

Three different people were interviewed. They all had academic backgrounds and were used to working with computers. They rated their knowledge in statistics, data analysis and data visualizations on a scale from 1 to 5 and throughout rated higher than 1 with an average of 3 for statistics, 2.33 for data analysis and 3.66 for data visualizations.

Usage of Given Options

Throughout the questions interviewees were able to switch in between the given options as much as they wanted to. During the first question interviewees checked out all options. One of the three interviewees was not using animated transitions while there was only one line available. During the test-run animated transitions were not used for the line chart with one line, as well. As soon as two lines were visible at the same time, however, all interviewees switched to animated transitions. The chosen transition duration was between 1000 and 1730 and was not changed after they were set to a duration which worked for the interviewee.

The first time missing data was present in the line chart interviewees switched between the visualization options as well. Two interviewees kept the missing data visualization that way until the end while one person switched the visualizations on and off while answering questions.

Interaction with Missing Data

For graphs without missing data interviewees were confident in averaging and trend detection. They were able to describe the change in single point values as well. However, as soon as missing data was part of the visualization the interaction changed. Especially in the beginning users were fast to point out instances of missing data during the questions while in the second half of the interview they were not mentioning it a lot during their answers. During averaging and trend detection all interviewees chose the state depending on the overall view, not caring about whether or not the state had more missing data in it. One user mentioned that “even if the data point was missing the existing data points outside of it were suggesting where the missing data point would be if it was existing”. There was no interviewee who was expressing worry that the missing data visualization and interpolation method was producing

bias. However, one interviewee was mentioning that "exactly because of the visualization" they were reminded that "data was missing. Without transition I would just interpolate in my head. With the visualization I am realizing that that might not be the right way". While no user mentioned feeling biased that is no proof that they were not indeed biased, especially since there was no question regarding bias.

Preference

Interviewees all preferred the animated transition over the static one. All of them stated that they were able to follow the change far better like that. They made no negative comments regarding the animation. However, during the test-run it was stated that the transition was taking longer than the static transition.

All interviewees preferred the missing data visualization, as well. While one was indifferent to whether or not the circles were added to the missing data visualization two interviewees expressed their preference. One mentioned that it made it easier to see "exactly how many data points were missing" and both mentioned that it enabled them to read the line chart better.

3.5 Discussion

A visualization for missing data was decided on and implemented into a line chart. In addition to that, an animated transition was implemented as well to work together with the missing data visualization. In the introduction of this thesis a few questions were asked. Now that the implementation and evaluation were explained it is time to revisit these questions:

- **Will the user realize that data is missing?**

  The missing data visualization which was chosen was different from the existing data visualization in an attempt to make it noticeable. According to the reactions of the participants in the pilot study they did realize that data was missing. Especially in the end of the study there were not many mentions of missing data during their answering process. This might have to do with the repetitiveness of the questions or with the fact that users did get used to the visualization. But it could very well be a sign that the chosen visualization was too similar to the existing data visualization, resulting in users overlooking data the less they were paying explicit attention.

- **Will the visualization assist in noticing trends in the data?**

  The study did find that the majority of users were willing to trust the linear interpolation and guess on overall trends when missing data was available, leading to the assumption that more trends were noticeable

with the missing data visualization. Depending on the dataset this might be desired or not. However, one person did express that they were more hesitant with the missing data visualization, as mentioned in the last section. Since the number of participants is not big enough to make meaningful statistics this is a factor which could be considered in a further study.

- **Will the missing data visualization create bias? / Will the user have higher confidence while interpreting the data than if missing data was just not visualized?**

  While there was no mention of confidence during the pilot study it was apparent that even with data missing they were still working with the parts of the dataset which had the bigger portions of missing data in them. They did not express much worry in their choice. Song and Szafir[18], who did include a question to trend detection in their study, did find that the perceived quality was higher with linear interpolation as imputation method. This could be a reason why participants were not hesitant to work with the data and it does suggest that users were indeed biased in their decision. However, without further study this will not be able to be said for sure.

- **Does creating an animated transition help the user to keep focus on the data during changes?**

  All participants did express that they did prefer the transition because it helped them visualize the change more clearly. It enabled them to let their eye follow data points, meaning that according to the pilot study the animated transition indeed helps. This is in accord with Heer and Robertson's[12] study, which found that their implemented animated transitions did help users to perceive changes better.

- **Does the creation of an animated transition assist in working with missing data?**

  Users were able to follow single data points more easily due to the animation. While it is unclear whether it explicitly helped, they were able to easily and quickly identify the switch from existing to missing data in the questions about single point values in the pilot study.

Looking at the studies of Eaton et al.[8], Song and Szafir[18] and Andreasson and Riveiro[2], there are no findings in their studies which are opposed to the findings of this thesis. In [2] they worked with "riskiness" as measure, meaning that users were making decisions based on actual missing data. A similar thing was observable in this thesis's pilot study with participants choosing certain views despite data being missing. Similar to [8]'s findings this pilot study's participants preferred having a visualization which showed that data was missing.

However, as there was no option for creating statistics in this data a lot of things like confidence measure, accuracy of answers and perceived data quality, as they were measured in [18], could not be compared. This will have to be part of a bigger quantitative study.

4 Conclusion and Outlook

This thesis explored the backgrounds of missing data and animated transitions. Existing versions of visualizations were presented and examined. Based on these a missing data visualization and animated transitioned were implemented into an existing line chart. The general goal of the thesis was to answer several questions about how users work with the chosen missing data visualization and the animated transitions which were chosen. This was evaluated via a small pilot study.

Missing Data visualization has been a topic for several decades but there is still not one single standard method for dealing with it. While this thesis's implementation certainly seems to be a viable option it is not argued that it is the only one. It is difficult to choose one visualization type as many studies still introduce new ways of visualizing missing data, like Boukhelifa's sketchiness[6] or Andreasson and Riveiro's fuzziness[2]. It also seems like in missing data visualization one always has to be aware of creating bias. Song and Szafir[18] found that highlighting missing data leads to higher perceived quality in data. However, depending on the data set the perceived quality might be too high. In the pilot study this was also a concern as participants generally seemed confident as they worked with missing data. When is high confidence too high? This was one question that could not be answered in the scope of this thesis. Perceived data quality is objective and might vary from person to person. It also might vary depending on the data set which level of perceived quality should be reached. This is a subject for further studies.

There have not been bigger studies on the effect animated transitions have on datasets with missing data yet. This thesis's pilot study gave first ideas on how to approach it. It also showed the participants' clear preference for animated transitions and missing data visualizations. However, a larger study would open the way for comparing different types of animations with each other. Especially highlighting missing data through animation is an interesting topic which was also suggested by Eaton et al.[8]. A larger study would also enable the collection of statistics like accuracy of answers when transitions are in place, or if perceived data quality rises with animations in place compared to visualizations without animated transitions.

Bibliography

[1] Wale Akinfaderin. *Missing Data Conundrum: Exploration and Imputation Techniques*. 2017. URL: https://medium.com/ibm-data-science-experience/missing-data-conundrum-exploration-and-imputation-techniques-9f40abe0fd87 (visited on 07/05/2019).

[2] Rebecca Andreasson and Maria Riveiro. "Effects of visualizing missing data: an empirical evaluation". In: *2014 18th International Conference on Information Visualisation*. IEEE. 2014, pp. 132–138.

[3] Yair M Babad and Jeffrey A Hoffer. "Even no data has a value". In: *Communications of the ACM* 27.8 (1984), pp. 748–756.

[4] Benjamin B Bederson and Angela Boltman. "Does animation help users build mental maps of spatial information?" In: *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis' 99)*. IEEE. 1999, pp. 28–35.

[5] Meg Blanchette. *What to Consider When Choosing Colors for Data Visualization*. 2018. URL: https://www.dataquest.io/blog/what-to-consider-when-choosing-colors-for-data-visualization/ (visited on 07/18/2019).

[6] Nadia Boukhelifa et al. "Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty". In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2769–2778.

[7] Kenneth Cukier and Viktor Mayer-Schoenberger. "The rise of big data: How it's changing the way we think about the world". In: *Foreign Aff.* 92 (2013), p. 28.

[8] Cyntrica Eaton, Catherine Plaisant, and Terence Drizd. "Visualizing missing data: Graph interpretation user study". In: *IFIP Conference on Human-Computer Interaction*. Springer. 2005, pp. 861–872.

[9] Michael Friendly. "A brief history of data visualization". In: *Handbook of data visualization*. Springer, 2008, pp. 15–56.

[10] FUB-HCC. *IKON-projektor on Github*. 2019. URL: https://github.com/FUB-HCC/IKON-projektor (visited on 07/18/2019).

[11] John W Graham. "Missing data analysis: Making it work in the real world". In: *Annual review of psychology* 60 (2009), pp. 549–576.

[12] Jeffrey Heer and George Robertson. "Animated transitions in statistical data graphics". In: *IEEE transactions on visualization and computer graphics* 13.6 (2007), pp. 1240–1247.

[13] *IKON - Visualizing the potential for knowledge transfer in research museums.* www.mi.fu-berlin.de/en/inf/groups/hcc/research/projects/ikon/index.html. Accessed: 2019-05-12.

[14] Younghoon Kim, Michael Correll, and Jeffrey Heer. "Designing Animated Transitions to Convey Aggregate Operations". In: *Computer Graphics Forum (Proc. EuroVis)* (2019). URL: http://idl.cs.washington.edu/papers/animated-aggregate-operations.

[15] George Robertson et al. "Effectiveness of animation in trend visualization". In: *IEEE transactions on visualization and computer graphics* 14.6 (2008), pp. 1325–1332.

[16] Donald B Rubin. "Inference and missing data". In: *Biometrika* 63.3 (1976), pp. 581–592.

[17] Joseph L Schafer and John W Graham. "Missing data: our view of the state of the art." In: *Psychological methods* 7.2 (2002), p. 147.

[18] Hayeong Song and Danielle Albers Szafir. "Where's My Data? Evaluating Visualizations with Missing Data". In: *IEEE transactions on visualization and computer graphics* 25.1 (2019), pp. 914–924.

[19] Alvira Swalin. *E-Learning Project SOGA Statistics and Geospatial Data Analysis. Department of Earth Sciences, Freie Universitaet Berlin.* 2018. URL: https://www.geo.fu-berlin.de/en/v/soga/Geodata-analysis/time-series-analysis/Dealing-with-missing-values/Imputing-missing-values/index.html (visited on 07/06/2019).

[20] Alvira Swalin. *How to Handle Missing Data.* 2018. URL: https://towardsdatascience.com/how-to-handle-missing-data-8646b18db0d4 (visited on 07/05/2019).

[21] Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. "Animation: can it facilitate?" In: *International journal of human-computer studies* 57.4 (2002), pp. 247–262.

[22] Ray Twiddy, John Cavallo, and Shahram M Shiri. "Restorer: A visualization technique for handling missing data". In: *Proceedings of the conference on Visualization'94.* IEEE Computer Society Press. 1994, pp. 212–216.

Appendix

4.1 Setup of Pilot Study

4.1.1 Anmerkungen

Fachbegriffe und bereits geprägte Begriffe wie “Line Chart” und “Missing Data” werden für dieses Dokument weiterhin Englisch belassen. Ansonsten wird Deutsch verwendet, da die Studie auf Deutsch durchgeführt wurde.

Zwischenfragen

Dem/der Befragten werden einige Fragen gestellt. Er/Sie soll seine/ihre Gedankengänge kommentieren. Wenn jedoch nicht viele Kommentare gegeben wurden währenddessen, kann nach der Beantwortung der Frage noch einmal nachgehakt werden, wieso er/sie etwas an den Optionen verstellt hat. Zum Beispiel:

- Du hast, während du überlegt hast, die Missing Data Visualisierung angestellt. Wieso?
- Wieso hast du die Transition Duration erhöht bevor du geantwortet hast?

4.1.2 Einführung

Zunächst ist ein einleitendes Line Chart mit mehreren Linien und mehreren Missing Data Punkten zu sehen, an dem alles erklärt wird.

Zunächst würde ich dich bitten die Einwilligungserklärung zu unterschreiben. Die Daten werden anonymisiert ausgewertet.

Dies ist ein Line Chart, welches die Anzahl an diversen Forschungsprojekten pro Jahr beschreibt. Die verschiedenen Linien sind jeweils unterschiedliche Forschungsbereiche. Also: Eine Linie zeigt nur Forschungsbereiche aus z.B. den Naturwissenschaften und eine andere nur aus den Geisteswissenschaften. Du bist in diesem Szenario ein Forscher, der mit diesem Line Chart arbeitet, um mehr über die Forschungsaktivitäten der letzten Jahre zu erfahren. Ich erkläre zunächst das Interface.

Im Datensatz befinden sich einige Punkte, bei denen Daten fehlen. Diese sind in dieser Ansicht nun ausgeblendet, aber es kann eine Visualisierung der Linie eingeblendet werden (Toggle Missing Data Visualization On/Off).
Zusätzlich dazu kann man die einzelnen fehlenden Punkte einblenden (Toggle Circles for Missing Data On/Off).

Weiter unten findest du einen Slider, welcher anzeigt wie lange die Transition zwischen den Ansichten dauern. Im Moment ist er auf 0 eingestellt, das heißt dass die Transition direkt passiert. Wenn die Zeit höher eingestellt wird, wird der Übergang animiert werden. Die Einheit ist in Millisekunden und du kannst bis zu 2000 Millisekunden Länge einstellen, was 2 Sekunden entspricht.

Mit dem "Turn Filter On/Off" Button kannst du zwischen zwei verschiedenen Ansichten wechseln. In diesem Szenario ist das z.B., dass du erst alle Forschungsprojekte betrachtest und danach durch Filtern bestimmte Forschungsprojekte miteinander vergleichst.

Dort drunter ist noch ein Slider für die angezeigten Jahre.

Ich werde dir gleich Fragen zu dem Line Chart stellen und du kannst frei heraus antworten. Bitte kommentiere auch deine Gedankengänge währenddessen. Die Daten ändern sich bei jeder Frage etwas, aber werden immer noch den gleichen Hintergrund haben. Du kannst jederzeit die gegebenen Optionen nach deinem Belieben verändern, sobald ich mit meinen Fragen starte. Aber davor möchte ich dir noch kurz die Möglichkeit geben selbst Fragen zu stellen falls du mehr Informationen haben willst oder etwas unklar ist. Du kannst auch während der Befragung jederzeit Fragen stellen, natürlich.
(Warten auf Fragen)

4.1.3 Fragen über das Line Chart

Anm.: Dem Befragten wird ein Line Chart gezeigt. Dies passiert für die im Folgenden aufgezählten Situationen mit der genannten Anzahl an Linien. "↔" stellt die Transition zwischen den beiden States dar. Es kann zwischen beiden States jederzeit hin- und hergewechselt werden.

1. ***One Line**: Existing Data ↔ Existing Data*
2. ***One Line**: Existing Data ↔ Half of the Data is Missing*
3. ***Two Lines**: Existing Data ↔ Existing Data*
4. ***Two Lines**: Existing Data / Less Than Half of the Data Missing ↔ Half of the Data Missing / Same Amount of Data Missing as Before*
5. ***Three Lines**: Existing Data ↔ Existing Data*
6. ***Three Lines**: Two Lines with Existing Data / ↔ $\frac{1}{3}$ of the complete Data is Missing*

Die folgenden Fragen werden für jede dieser Situationen wiederholt.

1. Wenn du dir nur das Jahr 2012 anschaust, wie würdest du beschreiben was passiert?

2. Gab es insgesamt über alle Jahre mehr Forschungsprojekte im Zustand vor der Transition oder danach?

3. Siehst Du im Anfangszustand einen Trend in der Anzahl der Forschungsprojekte über die Jahre? Wenn ja, ändert sich dieser nach der Transition?

4.1.4 Abschluss

Jetzt habe ich noch einige allgemeine Fragen:

1. Wenn du die nicht animierte und die animierte Transition vergleichst, was fällt dir auf? Ziehst du eine Variante vor?

2. Stell dir vor diese Visualisierungsart wird in einer Software eingebaut. Was für Vor- und Nachteile könnte das haben?

Fragen zur Person

1. Wie schätzt du dein Wissen über Statistik ein?
   (1- sehr gering, 5- sehr umfangreich)

2. Wie schätzt du dein Wissen über Datenanalyse ein?
   (1- sehr gering, 5- sehr umfangreich)

3. Wie schätzt du dein Wissen über Datenvisualisierungen ein?
   (1- sehr gering, 5- sehr umfangreich)

4. Alter?

5. Geschlecht?

6. Fachlicher Hintergrund?

4.2 Result of Pilot Study

Das Transkript der Studie kann auf ofs.io gefunden werden. Der Link zum Projekt ist:
`https://osf.io/zr3sf/?view_only=8dcfd15d5f274a6aaf13dded6d45f7d2`
Teilweise wurden Bemerkungen gekürzt, welche nicht zu der Studie gehörten und keinen Einfluss auf das Ergebnis hatten. Einige Sätze wurden zusammengefasst, wenn die Sätze nicht vollständig waren.