Otto-von-Guericke University Magdeburg

Faculty of Computer Science
Department of Simulation and Graphics

Master Thesis

Communicating Uncertainty through Line Charts

Author:
Apoorva Karagappa

September 21, 2023

Examiners:

Prof. Dr.-Ing. habil. Bernhard Preim
Faculty of Computer Science
Otto-von-Guericke University
Universitätsplatz 2
39106 Magdeburg, Germany

Prof. Dr. rer. nat. Andreas Gerndt
Institute for Software Technology
German Aerospace Center
Lilienthalplatz 7
38108 Braunschweig, Germany

Supervisor:

Dr.-Ing Pawandeep Kaur Betz
Institute for Software Technology
German Aerospace Center
Lilienthalplatz 7
38108 Braunschweig, Germany

Karagappa, Apoorva:
Communicating Uncertainty through Line Charts
Master Thesis, Otto-von-Guericke University
Magdeburg, 2023.

Contents

Abstract

In recent years, the world has increasingly relied on mathematical models to predict the spread of COVID-19. This information is crucial for healthcare professionals in preparing to ensure that necessary care is available to patients when needed, for policymakers in making vital decisions and implementing policies to mitigate the spread of the disease, and for individuals in making important choices regarding their personal and professional lives.

These prediction models can be highly intricate, considering numerous variables and scenarios. It is essential to acknowledge that the results of these models are inherently uncertain and never completely deterministic. Effectively communicating this uncertainty to users of the predictions is vital to ensure that informed decisions are made based on this information.

Interpreting complex visualizations is known to be influenced by the visualization techniques as well as the individual differences of the users. Thus, this thesis conducts a comprehensive user study of how to effectively communicate uncertainty in time series prediction, such as COVID-19 prediction that are visualized as line charts. In addition to accurately interpreting uncertainty, this study also aims to assess whether it fulfils the informational needs of the users and whether the provided information is sufficient to motivate users to make informed decisions. If these needs are not adequately addressed, the thesis endeavours to understand why and explore ways in which they can be better met.

This thesis is a part of ESID, which is an abbreviation for 'Epidemiological Scenarios for Infectious Diseases´, a visual analytics application for epidemiological analysis, developed by the Institute for Software Technology at the German Aerospace Center (DLR).

Acknowledgements

I want to extend my gratitude towards my supervisor, Dr. Ing Pawandeep Kaur Betz for her support and guidance throughout the completion of my thesis, to M.Sc. Jonas Gilg and M.Sc. Moritz Zeumer for their time and ideas, and to M.Sc. Anna Klein, M.Sc. Lena Rothenhäusler and M.Sc. Christoph Knoll for helping me in reaching the target audience for my user study.

I am also thankful to Prof. Dr.-Ing. habil. Bernhard Preim for his invaluable feedback, and to Prof. Dr. rer. nat. Andreas Gerndt for giving me the opportunity to write a thesis on this interesting research topic in his group.

I would also like to acknowledge and thank the participants of my user studies who generously dedicated their time and efforts.

1 Introduction

Line charts are a common means of representing time series prediction, allowing for the observation of trends, patterns, and irregularities within continuous data, including the representation of uncertainty. The uncertainty can be difficult to interpret and estimate from its graphical representation owing to its complexity. This thesis provides a comprehensive examination of effectively communicating this omnipresent uncertainty inherent in time series predictions. This chapter establishes the motivation for this thesis, lays out its objectives and research questions, presents its contributions, and introduces the structure that will guide this thesis.

1.1 Motivation

On 31 December 2019, the World Health Organization (WHO) China Country Office was informed of detected pneumonia cases of unknown causes[1]. On 22 January 2020, WHO issued a statement saying that there was evidence of human-to-human transmission of this disease[2], since named COVID-19, and as quickly as 11 March 2020, it was declared as a pandemic[3]. In this time period, research was already under way to ascertain the disease dynamics and predict its impact with the help of mathematical modelling, machine learning and information visualization.

The COVID-19 Projections dashboard[4] by the Institute for Health Metrics and Evaluation (IHME), an independent research centre within the University of Washington School of Medicine which was launched in March

[1] WHO Timeline - COVID-19
[2] Mission summary: WHO Field Visit to Wuhan, China 20-21 January 2020
[3] WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020
[4] COVID-19 Projections Dashboard

2020 is likely the first dashboard to be available that was aimed at providing forecasts and projections related to the COVID-19 pandemic. The importance of these predictions can be seen clearly as on 7 February 2022, WHO/Europe and the IHME signed an agreement in continuing their ongoing collaboration in the areas of health estimates, projections and global health data forecasting[5]. WHO/Europe stated that since the start of the COVID-19 pandemic, it has increasingly referenced IHME forecasts in providing guidance and recommendations to countries in the WHO European region with respect to containing the pandemic. Simultaneously, there are concerns over the validity of the predictions and their usefulness to policymakers. One particular concern is whether the graphical representation of uncertainty in line charts is effectively conducive to understanding uncertainty in peak daily death or hospital admission dates, with there being instances where the uncertainty bounds, and consequently the projections, are being interpreted incorrectly in both formal and social media [JEWELL et al. (2020)]. This concern becomes more pronounced knowing that other national-level COVID-19 dashboard development teams have reported the task of visualising data in a clear and understandable way to be challenging, and have left the duty of interpreting the data to the media and data enthusiasts among the public [BARBAZZA et al. (2022)]. While these challenges and concerns persist, the dashboards have been an important tool in making data available to wider and more diverse audiences, and have consolidated their place in guiding public health action [DASGUPTA and KAPADIA (2022), THORPE and GOUREVITCH (2022)]. Thus, the aim of this thesis is to fill in some of the blanks of why these challenges exist and offer potential solutions, in the context of visualizing uncertainty in time series prediction.

1.2 Objectives

With having knowledge that COVID-19 dashboard development teams have reported visualising their data to be challenging [BARBAZZA et al. (2022)] and that visualization authors generally tend to omit uncertainty from their visualizations due to lack of canonical forms of uncertainty visualizations and expected difficulty for users in interpreting them [HULL-

[5] WHO/Europe and IHME sign agreement cementing collaboration on forecasting of health data

MAN (2020)], the primary objective of this thesis is to assess whether current visualization techniques effectively fulfil users' information needs for extracting uncertainty from time series prediction and utilizing it in their decision-making processes, and to provide insights on how authors of visualization can utilize this information in enhancing the effectiveness of their uncertainty visualizations.

Additionally, these visualizations are intended for a broad range of audiences, which underscores the challenges as users are characterized by individual differences which are aspects such as experiences, background, personality and cognitive abilities that differentiate an individual from everyone else. The more complicated a task, the more pronounced the effects of individual differences appear [ZIEMKIEWICZ et al. (2012)]. While it might be impossible to study the effects of the current visualization techniques on every combination of the individual differences, it can be valuable to study the effects of current visualization techniques on individual differences that characterize different expected user groups.

Lastly, it is not only important to recognize the relationships between visualizations and individual differences, but also the effects of the characteristics of the visualizations that affect user perception and problem-solving ability, such as the effect of visual clutter on cognitive load, of which little has been done [SACHA et al. (2016)].

1.3 Research questions

With regard to the objectives mentioned in the previous section, the research questions for this thesis are as follows:

R1: What is the impact of different visualization techniques on the participants' uncertainty estimation in time series predictions visualized in line charts?

R2: Is there a discernible correlation between users' preferences for specific visualization techniques in meeting their information needs and the resulting task performance accuracy when employing these varied visualization techniques?

R3: Do individual differences shared among target user groups, such as area of study, frequency of visualization use, the highest degree achieved or numeracy have an effect on their task performance accuracy?

R4: How do the effects of varying visualization techniques, like clutter and aesthetics, influence users' evaluations of task difficulty and their perceived level of success in task performance? Consequently, is there a correlation between users' assessments of task difficulty, their perceived level of success in task performance, and their actual task performance accuracy?

1.4 Contribution

The main contributions made by this thesis are as follows:

- **Comparison of visualization techniques on uncertainty estimation of COVID-19 predictions in line charts.**

- **Identification of the need for consistent conceptualizations and depictions of uncertainty representation in time series predictions.**

- **Recognition and classification of users informational needs in estimating and utilizing uncertainty from epidemiological predictions.**

1.5 Structure

The overview of the thesis is described in this section.

- **Chapter 2** briefly explains why information visualizations are helpful in communicating complex information, why information visualization is challenging to effectively execute, what time series data and forecasting are, how uncertainty can arise in the forecasting process and how it is expressed.

- **Chapter 3** describes ESID, an Epidemiological Visual Analytics Application, of which this thesis is a part. It also briefly describes the forecast model that is behind ESID.

- **Chapter 4** covers the related work. It follows a top-down approach, where review and survey works are described first, in order to understand what approaches exist when studying uncertainty in the context of visualization, as well as commonly used taxonomies. This is followed by works that study the different factors that affect uncertainty visualizations and, finally, papers that focus on uncertainties specifically in the context of in time series prediction.

- **Chapter 5** describes the objectives, design and results of the preliminary user study. The results of the preliminary study contribute towards the design of the concluding study.

- **Chapter 6** describes objectives and design of the concluding user study. It presents and discusses the results in context of the research questions.

- **Chapter 7** closes the thesis by summarizing the findings and providing potential directions for future work.

2 Background

Information Visualization is defined as *"the use of computer-supported, interactive, visual representations of abstract data in order to amplify cognition"* by CARD et al. (1999). GERSHON et al. (1998) explain that there is a lack of natural physical representations for many classes of information that may arise in fast streams or in large volumes, and that is what differentiates Information Visualization from Scientific Visualization.

CARD et al. (1999) suggest that Information Visualization aids cognition by:

- **Increased Resources**
  - Human gaze combines high spatial resolution and wide aperture for sensing visual environments.
  - Visualizations enable parallel processing of certain attributes, unlike text.
  - Symbolic cognitive inferences can be simplified into perceptual operations.
  - Visualizations expand working memory.
  - They store large amounts of information in an easily accessible form.
- **Reduced Search**
  - Visualizations group related information, reducing the need for extensive searches.
  - They can represent a significant amount of data in a compact space.

  - Grouping data in visualizations can eliminate the need for symbolic labels.

- **Enhanced Recognition of Patterns**
  - Recognizing information within visualizations is easier than recalling it from memory.
  - Visualizations simplify and organize information, providing higher cognitive centres with abstracted forms through selective omission.
  - Organizing data structurally, such as by time, enhances pattern recognition.

- **Perceptual Inference**
  - Visualizations support numerous perceptual inferences that are effortless for humans.
  - They enable complex, specialized graphical computations.
  - Visualizations allow for the monitoring of multiple potential events if displayed in a way that distinguishes them by appearance or motion.

- **Manipulation Medium**
  - Unlike static diagrams, visualizations permit exploration of a parameter space and enhance user interactions.

Challenges

Information Visualization is a very important tool in conveying large and complex information in a manner that is intuitive and deliberate. However, given its nature and extensive applications, it is not without its challenges. GERSHON et al. (1998) explain the key research problems are identifying visual metaphors which accurately represent this information and understanding what analytical tasks they support. The authors elaborate that unlike Scientific Visualizations that are intended for trained experts, Information Visualization is often intended for the utilization of a broad range of users, whose specific needs and specific problems may differ. Visualizations are not a stand-alone entity, and therefore, it is fundamentally necessary to understand what may be a complex system that the visualization is a part of. When specific needs of users are fully understood, knowledge of how users interact with information, how visualizations are perceived and how this perceived information is utilized in solving problems can still pose challenges. Uncertainty associated with the information can be a significant part of this, as it introduces an additional layer of complexity that requires careful consideration. Uncertainty can lead to an increase in the cognitive load for users. A visualization author may also expect uncertainty to provoke psychological anxiety in users [HULLMAN (2020)].

Communicating uncertainty through visualization uses the same basic principles as most other information visualizations, by encoding data using visual cues, but in order to do so effectively, it is necessary to understand the nature of the data and its source. Accordingly, the remainder of this chapter will outline the fundamental aspects required to achieve this understanding.

2.1 Time series data

Time series data refers to data that is obtained across a time period. Generally, the time period will consist of sequential timestamps that are evenly spaced by any chosen unit of time. One may think of the timestamps as indices for the dataset. The data that is indexed by timestamps may be continuous or discrete. An example of this is the COVID-19 dataset from the Robert Koch Institute (RKI) in Germany, shown in Table 2.1.

The dataset shows the following variables:

- Population *(Bevoelkerung)* - Number of residents in the reference group, data from the population statistics of the Federal Statistical Office *(Bevölkerungsstatistik des Statistischen Bundesamtes)* as of December 31, 2021

- Cases_total *(Faelle_gesamt)* - COVID-19 cases since data collection began

- Cases_new *(Faelle_neu)* - Number of cases that are published for the first time in the reporting of the RKI

Table 2.1: Example of COVID-19 dataset. [ROBERT KOCH-INSTITUT (2023)]

Meldedatum	Bundesland _id	Altersgruppe	Bevoelkerung	Faelle _gesamt	Faelle _neu	Faelle _7-Tage	Inzidenz _7-Tage
2020-01-03	01	00-04	130181	0	0	0	0
2020-01-03	01	00+	2922005	0	0	0	0
2020-01-03	01	05-14	264345	0	0	0	0
2020-01-03	01	15-34	628290	0	0	0	0
2020-01-03	01	35-59	1001480	0	0	0	0
2020-01-03	01	60-79	666851	0	0	0	0
2020-01-03	01	80+	230858	0	0	0	0
2020-01-03	02	00-04	98074	0	0	0	0
2020-01-03	02	00+	1853935	0	0	0	0
2020-01-03	02	05-14	170331	0	0	0	0
2020-01-03	02	15-34	491372	0	0	0	0

- Cases_7 *(Faelle_7.Tage)* - Number of COVID-19 cases reported within the last seven days

- Incidents_7_days *(Inzidenz_7.Tage)* - COVID-19 cases reported within the last seven days per 100,000 inhabitants

is measured for various age groups *(Altersgruppe)* across different timestamps *(Meldedatum)*. The dataset is available to the public on Github [ROBERT KOCH-INSTITUT (2023)]. It should be noted, that as with most datasets, time series datasets too may sometimes be incomplete.

2.2 Sources of uncertainty

Uncertainty can arise at many stages, be it data collection, pre-processing, modelling and visualizing. The uncertainties aggregate and are carried forward to the end. This is well described by D. Sacha et al. in Figure 2.1. This figure has been edited to show only the desired data pertaining to sources of uncertainty. This rest of this section describes how the uncertainty can arise.

Figure 2.1: Sources of uncertainty. [SACHA et al. (2016), edited]

Data source

Depending on how the data is collected, uncertainties can arise in many ways. Data can be authoritative or non-authoritative. Non-authoritative data is collected from voluntary sources which do not adhere to strict standards or gate-keeping, they may consist of inaccuracies and high uncertainties. Alternatively, authoritative data can also incur uncertainties due to incorrect measurements, human errors or incomplete data [SACHA et al. (2016)]. In the real world, it can be a combination of both. Consider the instance of COVID-19 cases: those experiencing symptoms or having been in proximity to infected individuals are obligated to get tested at local test centres, or test at home and report it to local health authorities. However, individuals might not comply with this obligation due to unawareness regarding their potential exposure to other individuals carrying infections, or unawareness of their own infection due to the absence of known symptoms of the disease. Additionally, personal reservations or other social factors might also stop individuals from being tested. Moreover, in the event of universal testing compliance, the potential for errors can stem from factors such as testing kit inaccuracies and human errors, among others.

Data processing

Once collected, data undergoes preprocessing before being moulded into a desired output. This step addresses missing data, handles incompleteness, and ensures data quality, among other considerations. Filtering out data that have missing values, or filling in missing values by interpolating or other estimation techniques, can also result in uncertainties [BONNEAU et al. (2014)]. Sampling data might be necessary for resource management, unit conversions may be necessary for homogenizing data from different sources, bucketizing continuous data might be necessary for modelling the data, and dimension reduction may be necessary for making the data more workable, all of which introduce uncertainty into the data.

Model

As all models are essentially approximations, the predictions made by a model on data that it has not been exposed to during its training always has a lower accuracy than predictions on data that it has been trained on [CHATFIELD (2014)]. C. Chatfield states two sources of uncertainty resulting from prediction models alongside uncertainty arising from the source and data itself:

- Model Uncertainty - A true model is very unlikely to exist and even if it were to exist, its structure and mathematical working is unknown a priori, which makes it just as unlikely that the model that fits the data will be selected. Often, analysts will pick a model that provides suitable enough approximations for their desired task at hand, resulting in uncertainty.

- Uncertainty from parameter estimates - Estimating parameters can be difficult because selecting values that fit the training data can result in loss of efficiency and increased variability. So when estimating the parameters of a model, a trade-off has to be made between bias and variance, along with accounting for noise in the data and model complexity, introducing more uncertainty different to that of the model uncertainty.

Visualization

BONNEAU et al. (2014). highlight the importance of understanding how the visualization process itself may impact the propagation of uncertainty owing to several factors including the perceptual and cognitive influences in understanding uncertainty, or the effects of differences in individual traits of different users. While a substantial body of research has been dedicated to the visualization of propagated uncertainties, SACHA et al. (2016) have highlighted a notable research gap in the assessment of the effects of visualizations such as visual clutter on user perceptions of uncertainty and further how this consequentially affects the users' problem-solving efficacy.

2.3 Expressions of uncertainty: Confidence and credible intervals

The uncertainty described in the previous section may be collectively expressed as Confidence or Credible intervals. While both are important statistical tools in understanding and communicating uncertainty, they differ in their interpretation and applications. However, the different intervals carrying different meaning are depicted similarly in visualizations. Therefore, it is important to understand what the terms mean.

The interpretation of Confidence Intervals seems to be a subject of ongoing debate among researchers. HOEKSTRA et al. (2013) asserts that researchers misinterpret Confidence Intervals as often as students, even when students might not have received education on statistical inference. They do so, based on the responses to a study. The study presents a hypothetical scenario in which a professor conducts an experiment and reports a 95% CI for the mean that ranges from 0.1 to 0.4. The questions in the survey were are follows, where the respondents could answer either True or False to each question:

1. The probability that the true mean is greater than 0 is at least 95%.
2. The probability that the true mean equals 0 is smaller than 5%.
3. The "null hypothesis" that the true mean equals 0 is likely to be incorrect.

4. There is a 95% probability that the true mean lies between 0.1 and 0.4.

5. We can be 95% confident that the true mean lies between 0.1 and 0.4.

6. If we were to repeat the experiment over and over, then 95% of the time the true mean falls between 0.1 and 0.4.

The authors assert that the correct answer to all the questions is false, as statements 1, 2, 3, and 4 assign probabilities to parameters which are not allowed within the frequentist statistics and statements 5 and 6 mention the boundaries of the CI whereas a CI can only be used to evaluate the procedure and not a specific interval. Bayesian statistics deals with posterior distribution based on prior distribution, whereas frequentist statistics deals with sample data. The concept of Confidence Intervals can be understood better considering the explanation by TAN and TAN (2010) with reference to Figure 2.2. The figure shows the results for a study conducted twenty times using 20 different samples from the population. Nineteen out of the twenty 95% Confidence Intervals contain the true population mean. Nineteen is 95% of the total number of samples, hence the term 95% Confidence Interval. It can be seen that the intervals change over every sample, thus exact statements cannot be made about any one interval. However, MILLER and ULRICH (2016) regard the statement 5 as true and offer a refutation but acknowledged that the statement is ambiguous due to the nature of the term "confidence" that is associated with a 95% Confidence Interval.

However, it is not debated that the Confidence Interval provides a range of values whereas the Credible Interval provides a probability distribution. Therefore, it is important that we do not represent the two intervals in similar ways, leading to further confusion and misinterpretation of what the intervals mean. Knowing this difference is key in designing visualization techniques for visualizing uncertainty, as well as assessing a user's interpretation of the graphical representations of uncertainty.

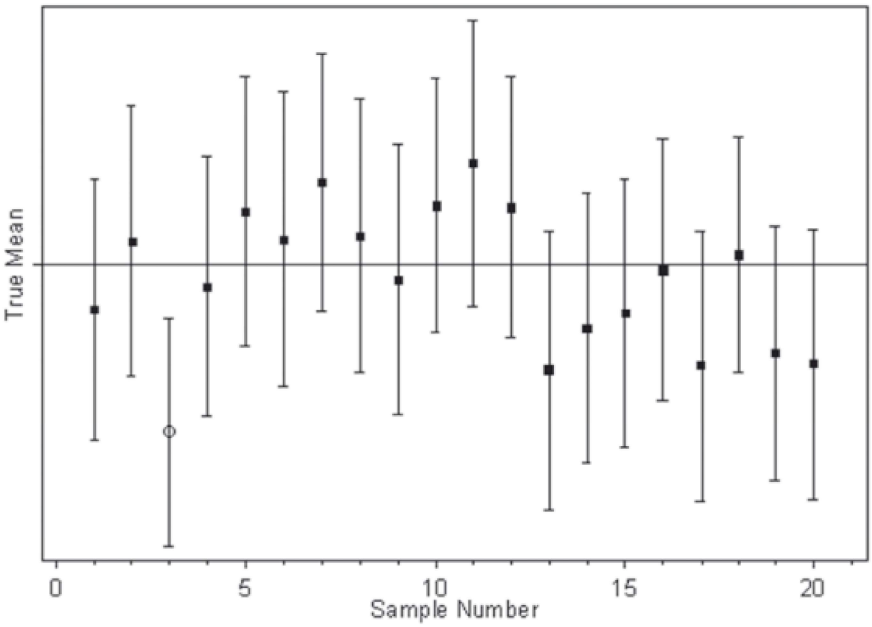


Figure 2.2: 95% CI for the population mean for 20 independent samples drawn from the population. [TAN and TAN (2010)]

3 ESID: An Epidemiological Visual Analytics Application

Epidemiological Scenarios for Infectious Diseases, abbreviated as ESID [GILG et al. (2023)], is a visual analytics application for epidemiological analysis, developed by the Institute for Software Technology at the German Aerospace Center (DLR). While some of the early COVID-19 Public Health Dashboards might not have undergone a full user-centric design process, owing to time constraints and the fast changing nature of the disease, ESID results from a carefully curated user-centric design process described by BETZ et al. (2023). The application is current deployment to show COVID-19 simulations resulting from the hybrid graph-equation-based model briefly described in Section 3.2, however, it is designed to visualise and analyse the spread of any infectious diseases, resulting from any model as long as it adheres to predefined data format. The code for this application is available as open source and can be accessed on GitHub[1]. At this juncture, it should be highlighted that ESID is part of a pilot project, 'Integrated Early Warning System for Local Recognition, Prevention, and Control for Epidemic Outbreaks', abbreviated as LOKI by the Helmholtz Association to assist the German health authorities with a local control system for monitoring and investigating epidemic outbreaks.

3.1 A close look at ESID

ESID is designed with the intention of facilitating analysis and decision-making by policymakers, healthcare professionals, and the public, based on real-time and simulated data [BETZ et al. (2023)]. Based on require-

[1] ESID GitHub Repository

ments gathered through interviews and analysis of previously existing, verified, published speeches, and surveys from members of the target groups, ESID's user interface has been structured with three key components; A Choropleth Map, Scenario Cards and a Timeline. The prototype consisting of these three components can be seen in Figure 3.1.

Figure 3.1: ESID visualization tool. [BETZ et al. (2023)]

The three components are fine-tuned to incorporate specific features that facilitate user requirements.

- The **choropleth map** assists the user identifying the Geo-spatial distribution of the pandemic across counties/districts. This is important in implementing local containment measures. Counties can be selected by click on the choropleth map or via the search bar, in order to filter data in the other components.
- The **scenario cards** facilitate the user in understanding the dynamics of the pandemic with respect to different scenarios. This encompasses estimating the impact of specific policies, non-pharmaceutical interventions, testing strategies, and related measures.
- The **timeline** component consists of a line chart, which contains simulated future infectious disease dynamics, based on the filtered counties and scenario card(s), resulting in support of decision-making based on foresight of the disease dynamics as well as visual comparison of different scenarios. The uncertainty that accompanies the

simulated results are shown with coloured half-transparent areas on the line charts and through tool tips on hover. In their paper, BETZ et al. (2023), highlight the importance of visualizing uncertainty to mitigate misinterpretation of data and the potential distrust that could arise from it among citizens and decision makers. As such, it is important to use visualization techniques that most effectively communicate the arisen uncertainty. This thesis is positioned to aid in the advancement of this goal.

3.2 Simulation model

Compartmental models compartmentalize a population into separate groups based on their disease status. The compartments may be of classes M, S, E, I or R. With regard to a certain disease, class M corresponds to newborn infants that are born with temporary immunity passed on to them from their mothers, class S refers to individuals who are susceptible to being infected by the disease, class E is the group exposed to the disease but have yet to become infectious, class I is the group that are infectious and class R are the set of individuals who have recovered and are immune to the disease. The type of disease characterizes the model compartments and the flow pattern are often used to describe the models, such as MSEIRS, SEIR, SEIRS, SIR, etc [HETHCOTE (2000)]. KÜHN et al. (2021) define the compartments SECIHURD, which in addition to the previously mentioned compartments include class C for carrier who carry the virus and are infectious to others but do not yet show symptoms, class H for the hospitalized, class U for those in intensive care unit and class D for the dead. A simplified image of the flow pattern from KÜHN et al. (2021) is shown in Figure 3.2. The figure has been simplified by omitting information on model equations describing the compartments, as they are not essential for comprehending the contents of this thesis. Covid-19 affects people of different age groups differently and to account for this, the model is extended such that each compartment is further characterized by an age group. The age groups are described as [0–4, 5–14, 15–34, 35–59, 60–79, 80+], with all numerical values denoting years.

Additionally, SIR-type models homogenize the disease dynamics over the entire population, and in order to ensure that important characteristics of

Figure 3.2: Flow pattern between compartments [KÜHN et al. (2021), edited]

the spread of the disease such as infection clusters are not lost, the age-resolved model is assigned to each county. Further, to ensure that cross county transmissions are not lost, the model utilizes travel information from the German Federal Employment Agency and Geo-referenced Twitter data, with some assumptions made on the frequency of travel. Each county is first compartmentalised individually, followed by a commute simulation between counties resulting in a midday compartmentalization, further followed by a back-home commute simulation, resulting in the final county compartmentalization. A simplified version of the depiction of the spatial heterogeneity implementation taken from KÜHN et al. (2021) is shown in Figure 3.3. The figure is edited to simplify the image by removing an alternative scenario and information on the numerical solver to model equations, as it is not vital to the understanding of this thesis.

It is mentioned in the paper that uncertainty is accounted for with Monte-Carlo Ensemble runs with a thousand runs for each scenario. The resulting uncertainty information is in the form of median values and percentiles obtained from the simulation runs. The figures showing simulation values representing uncertainty as *Simulation Percentiles* visualized in the form of overlapping coloured bands bounded by dotted lines, shown in Figure 3.4.

Figure 3.3: Spatial heterogeneity implementation. [KÜHN et al. (2021), edited]

Figure 3.4: Prediction of infected patients; X-axis: Date, Y-axis: Number of infected patients. [KÜHN et al. (2021)]

This simulated values along with uncertainty is presented in the line chart of the **timeline** component of ESID. This thesis attempts to ensure that the uncertainty is presented to the users, allowing them to best estimate this uncertainty and take it into account in their decision-making.

4

Related Work

Visualizing uncertainty can be difficult because of its nature; it falls in the intersection of communicative visualization and cognitive decision-making. This chapter delves into the related literature that guides this thesis.

4.1 Taxonomies in uncertainty visualization

Taxonomies are important to research as they help in categorizing and understanding complex problems. It helps in identifying previous studies that focus on similar goals and questions, to aid one's pursuit in finding answers to their own questions and meeting their own goals. For this reason, this sections details out taxonomies proposed by uncertainty visualization authors.

POTTER et al. (2012) state that expressing uncertainty through visualizations can be particularly challenging owing to the limited number of visual channels. To bridge this gap, it is important to understand and describe how uncertainty data is translated in relation to the certain data, so that it may be expressed visually in a manner that is effective.

POTTER et al. (2012) categorize uncertainty visualizations by two categories: data dimension and data uncertainty dimension. The data dimension defines the space occupied by the data of interest, which may or may not be the same as the data uncertainty dimension, which defines the space occupied by the uncertainty. Data dimensions can be one of the following:

- **1D**: This describes a variable that can take on a single value

- **2D**: This describes data defined by a joint probability distribution
- **3D**: This describes data whose position can be expressed in spatial volume
- **ND**: This describes data that is non-spatial, multivariate, and time-varying. Here, the relationship between the input parameters, which are multivariate and the output is of interest, but the output can be modelled to exist in any of the previously mentioned spaces. This is where our dataset of interest, time series prediction, lies.

Data uncertainty dimensions, on the other hand, can be either scalar, vector or tensor. They are defined as follows:

- **Scalar**: This includes uncertainty that can be expressed as a 1D probability density function. This accurately describes time series prediction, where at each timestamp, uncertainty can be expressed as a density function.
- **Vector**: This includes uncertainty that is described by a value as well as a direction
- **Tensor**: This includes uncertainty that is high dimensional

In summary, time series prediction is an ND dataset with scalar uncertainty. Farther in the paper, literature on the parameter-space analysis of ND data with visualization is described which unfortunately does not fit this thesis, as we are interested in the output space, i.e. the prediction.

As difficult as it may be to describe and express uncertainty visually, it can be just as difficult to evaluate the efficacy of the visualization. HULLMAN et al. (2019) present a taxonomy to identify differences between evaluation methods. This is done at six levels; Behavioural Targets, Expected Effects, Evaluation Goals, Measures, Elicitation and Analysis. Behavioural Targets and Expected Effects are grouped together as *Research Values and Aims*, and the rest are grouped into *Research Design*. Each of these levels are explained below.

- **Behavioural Targets**: Describes which behaviour outcome from the user is expected when presented with an uncertainty visualization. Example: Better performance

- **Expected Effects**: Identifies user traits that elicit the desired behavioural target. Example: Increased accuracy

- **Evaluation Goals**: Approach taken in the study to evaluate research goals. Example: Visualization comparison

- **measures**: Tool to estimate the degree of expected effects. Example. Probability estimation

- **Elicitation**: Tool that user interacts with to provide a response. Example: Slider

- **Analysis**: Method of interpreting results. Example: Null hypothesis significance testing

In addition to expressing and evaluating visualizations, it is important to take into account the individual traits of a user in addition to the nature of the data. A comprehensive review provided by LIU et al. (2020) shows that an individuals' trait affects their visualization use. They provide a taxonomy based on four dimensions; individuals' traits, visualization types, tasks and measures. The dimension of individual traits is categorized as follows:

- **Extroversion:** Engagement with the world

- **Neuroticism:** Tendency to experience negative emotions

- **Openness to Experience:** Propensity to understand and use information

- **Agreeableness:** The tendency to consider the harmony among a group of individuals

- **Conscientiousness** Propensity to exercise discipline

- **Locus of Control:** Belief in having control of outcome to events around them

- **Need for Cognition** Tendency to engage in thinking

- **Spatial Ability:** Understanding of spatial relations among objects

- **Perceptual Speed:** Rate of making accurate visual comparisons between objects
- **Visual/Spatial Memory:** Capacity to remember the visual traits of an object
- **Working Memory:** Capacity to store information for immediate use
- **Associative Memory:** Ability to remember relationships between unrelated items

The author conclude from the literature survey that, with the exceptions of conscientiousness and agreeableness, all individual traits are shown to impact visualization use. Therefore, it is important to account for the variability among these traits when designing and evaluating visualizations.

Lastly, MCCUAIG et al. (2005) provide a typology to categorize types of uncertainty, particularly in Geospatial Information that can be extended outside to other areas, to aid identification of visual metaphors that are effective for each type of uncertainty, as well as for a composite uncertainty. The authors are hopeful in finding visualizations that map each type of uncertainty into a different visual dimension that will allow analysts to identify the implications of each type of uncertainty individually as well as grouped uncertainty.

Table 4.1: Analytic Uncertainty Typology [MCCUAIG et al. (2005)]

Category	**Subcategories**
Accuracy/error	Accuracy of data collection
	Data processing errors
	Wrong data resulting from malicious intent
Precision	Precision of the data collection tools or methods
Completeness	Incomplete data that result from missing chunks
	Incomplete data where certain attribute values maybe missing
	Incomplete sequence
Consistency	Multiple sources may conflict
	Model values and observation values, maybe inconsistency
Lineage	Translation
	Data transformation of collected data
	Interpretation of collected data
Currency/timing	Temporal gaps
	Versioning
Credibility	Reliability
	Proximity
	Appropriateness
	Motivation (of the source)
Subjectivity	Differing analytic judgment with interpretation of uncertainty
Interrelatedness	Source dependence on derived information

4.2 Visualizing uncertainty

Related work on uncertainty visualization which is important to this thesis but does not directly deal with time series data is described here.

A comprehensive classification of visualization techniques that account for various types of uncertainty information and techniques is presented by PANG et al. (1997), with five characteristics:

- The datum value which may be a scalar, vector, tensor or multivariate and its associated uncertainty value. For a multivariate variable, each component may be a scalar, vector or tensor with its own uncertainty.
- The datum location and associated positional uncertainty, which may be 0D, 1D 2D, 3D or time.
- Extent of the datum location, that may be either discrete or continuous.
- Extent of the visualization, that may also be either discrete or continuous. This is dependent on the visual primitives used, for example, points and glyphs are grouped into discrete whereas curves, surfaces, volumes are grouped into continuous.
- Axes mapping, which describes the mapping of variables or groups of variables onto different axes. This may be experiential or abstract. Experiential mapping replicates the viewer's experience, whereas abstract mapping may result in additional insight and understanding. Further, they use axes mapping to differentiate information visualization (abstract mapping) and scientific visualization (experiential mapping).

Predicted COVID-19 cases are scalar, time dependent datum, which when visualized as line charts, have the extent of visualization as continuous and abstract axis mapping. Pang et al. developed a variety of new uncertainty visualization methods and present them according to their classification, of which none accurately describe predicted COVID-19 cases visualized as line charts. However, for scalar values they advocate for the following types of continuous visualization extent techniques that are of interest to this thesis; Multivariate Glyphs (ellipsoids), ribbons, blurring, texture and bump mapping, shown in Figure 4.1. As a way forward, JOHNSON and SANDERSON (2003) call for a formal, theoretical error and uncertainty visualization framework and to investigate and explore new visual representations for characterizing error and uncertainty. The authors assert the necessity for a formal evaluation of these visual techniques with incorporation of statistical, numerical, and/or measurement errors. They also note that better visual representation of and interaction with statistical data should constitute the new techniques.

Figure 4.1: (Left) Animation with motion blurring to indicate uncertainty. (Right) Surface illumination differences mapped to 2D circular textures.[PANG et al. (1997)]

In a tool designed to interactively visualize ensemble outputs of a numerical weather model and its associated uncertainty, Noodles, SANYAL et al. (2010) study the implications of ribbon and glyph-based uncertainty visualization (Figure 4.2) in conjunction with spaghetti plots. The glyph based visualization designed with the intention of avoiding obscuring of underlying data and the ribbon visualization designed for showing the variation of uncertainty along a contour were found to represent the overview of uncertainties well, with the ribbon being more effective than the glyphs and the graduated glyphs being difficult to interpret. In the graduated visualizations, the concentric distances were calculated as the absolute difference between each member from the ensemble mean, which the authors theorize as possible too much information to present in a glyph resulting in comprehension difficulty.

Figure 4.2: (Top-left) Glyphs showing the width of the 95% CI of the ensemble mean on the entire grid
(Top-right) Graduated glyphs along the ensemble mean
(Bottom-left) Uncertainty ribbon showing the bootstrap Inter Quartile Range
(Bottom-right) Graduated ribbon to illustrate uncertainty [SANYAL et al. (2010)]

The acceptance of algorithmic prediction in the presence and absence of uncertainty, and whether it is dependent on the method of visualization, is evaluated by LEFFRANG and MULLER (2021). Users, separated into three groups, were asked to make their own prediction for a use case before and after being presented with an algorithmic prediction. The algorithmic prediction was presented differently to each of the three groups, either as a point estimate without uncertainty representation, with uncertainty depicted by a confidence interval or as an ensemble visualization. The authors define two measures for evaluating the acceptance, which are:

- Adjustment - The difference between the users' initial prediction, before viewing the algorithmic prediction and the second prediction, after viewing the algorithmic prediction.
- Disagreement - The difference between the algorithmic prediction and the users' second prediction.

The results of the study concluded that users adjusted more and disagreed less with the algorithmic prediction when presented as either point estimates or with uncertainty visualized with a Confidence Interval band, when compared to the ensemble visualization. However, no significant differences were found between the point estimates and Confidence Interval band.

MCGRATH et al. (2020) study the acceptance of the algorithmic prediction under varying uncertainty distributions. The design of the study is similar to that of the previously described user study in that the participants are assigned to one of five groups, and are asked to make a prediction for a use case before and after being shown the algorithmic prediction. However, each group is shown either no uncertainty information or uncertainty with one of the following distributions: Normal distribution with low variance, normal distribution with high variance, bimodal distribution and skewed distribution. The type of uncertainty was shown to have a marginally significant effect ($p = 0.0715$) on how close the people's second estimate was to that of the algorithm. The participants were most influenced by uncertainty that is normal distributed with low variance, and were least influenced when presented with no uncertainty. As the parameters of uncertainty distribution vary along the different timestamps of a time series model prediction and the users' acceptance is dependent on it. I want to include different variances of a normal distribution to understand if this difference in acceptance is a result of being able to estimate probabilities under uncertainty and if some visualization techniques will yield a uniformly accurate estimation.

Although not evaluating uncertainty directly, ELHAMDADI et al. (2022) study how modifications to the clarity of visualizations such as blur, opacity, outlines to markers, jitter, gridlines, and manipulation of the scale affect the user's trust. This study is of value to this thesis for two reasons; first, these modifications are often used techniques to communicate uncertainty and second, trust is an important factor for whether users will make decisions based off the information conveyed to them through an uncertainty representation. The findings show that the speed and accuracy with which users performed tasks, correlate with their trust in a visualization.

4.3 Visualizing uncertainty in line charts

This section discusses related work that expresses uncertainty statistically in line charts or as ensemble visualizations over time. The evaluations are typically done as user studies, sometimes also accounting for individuals traits.

The issue of representing different types of uncertainty; statistical and bounded by the same visual indicators is recognized by OLSTON and MACKINLAY (2002), who first propose what we now know as the Confidence Interval Band. The authors offer suggest using graphical fuzz, giving an effect that resembles ink smearing to define a region that ambiguous, seen on the right side of Figure 4.3. They offer this as a visualization technique for bounded uncertainty, arguing that the then commonly used error bar are typically used in conjunction with an estimated exact value and that users have been trained to interpret them as a statistical uncertainty. The terminology used to differentiate between the two types of uncertainty, however, might continue to cause confusion as a Confidence Interval is a statistical interval but is not a probabilistic distribution, as the authors define statistical uncertainty.

Figure 4.3: Error Bar showing statistical uncertainty in a line chart (left). Ambiguation showing bounded uncertainty in a line chart (right). [OLSTON and MACKINLAY (2002)]

Further, VAN DER LAAN et al. (2015) compare the two ways of expressing uncertainty in line charts, confidence band and error bars, both with and without the point estimate and additionally, a line chart without uncertainty, shown in Figure 4.4. The task to evaluate the visualizations consisted of identifying trends with 5 different degrees of certainty. Two datasets consisted of a combination of high and low uncertainty with either no trend or an increasing trend, and one other dataset where the trend was decreasing, followed by an increase. A user study conducted, by assigning participants into 5 groups, each shown 5 unique visualization and dataset combinations showed that users best identified trends when shown the visualization of a confidence band with point estimate and worst when shown the visualization of a line chart with no uncertainty.

SANYAL et al. (2009) compare less common, novel techniques for comparing uncertainty, by encoding uncertainty as glyph size, glyph colour and line colour, Figure 4.5. Given a region of interest, users are asked to identify the data point with the highest and lowest uncertainty. Results show that glyph size most effectively assisted users in identifying points of the least uncertainty whereas glyph and surface colour worked better for points of high uncertainty, to which the authors theorize that human perception of these visual expressions may not be uniform.

TAK et al. (2014) study the interpretation of uncertainty under different visualization methods, whereby each user is shown time series prediction with the mean, upper and lower bounds in the form of either solid border, dashed border, confidence interval band, thinning lines, random lines,

Figure 4.4: Techniques for visualising uncertainty. (Top-left) No Uncertainty (Top-right) Confidence band with mean line (Middle) Error Bars with mean line (Bottom-left) Confidence band without mean line (Bottom-right) Error Bars without mean line [VAN DER LAAN et al. (2015)]

gradient or error bars. The user is shown, at a time, one of nine points equidistant from the mean line and made to estimate the likelihood that the actual value might occur at this point. It is important to note here that the visualizations are not labelled to understand the users' inherent interpretation. The results show that users have one of three interpretations, given uncertainty:

- Equal probabilities to the points that are equidistant from the centre line.
  - Peaked Interpretation: User interpret a normal distribution, the closer a point is to the mean line, the more likely it is.

Figure 4.5: Visualizing Uncertainty by a) Glyphs size b) Glyph colours c) Surface colour, and d) Error bars. [SANYAL et al. (2009)]

  - Flat interpretation: Users interpret that all points lying inside the bounds are equally likely.

- Linear mapping: Users interpret that the higher a point is, the higher the likelihood and the lower a point is, the lesser the likelihood, or vice versa.

The results show dashed border, random lines, and gradient visualizations to have the best fit with a normal curve. As the fit of the given visualizations do not improve with quadratic or cubic functions, the authors infer that the participants' internal model of uncertainty is not linear and remains unaffected by the form of representation.

The examination of individual variances and user metrics in prior research influences the design of visualizations, as well as the decisions regarding the structuring of user studies and the selection of parameters and assessment criteria. This thesis will use some of these techniques in the context of COVID-19 simulations and in addition to determining whether users are inclined to estimate and accept these visualizations, this thesis will investigate the specific attributes of visualizations that facilitate this process. It also seeks to assess whether the standard presentation of information via line charts, as typically employed in these studies, is sufficient for users examining COVID-19 predictions.

5
Preliminary User Study

Previous studies described in Chapter 4 have primarily focused on assessing users' trust in algorithmic predictions when visualizations incorporate varying levels of uncertainty. However, it's important to note that visualization authors do not have much control over the calculated uncertainty ranges. Additionally, there is limited understanding of the specific characteristics of these visualizations that influence user trust. Furthermore, it cannot be assumed that users, when trusting a visualization, necessarily consider or factor in the projected uncertainty in their decision-making process.

Some studies in this field also investigate the impact of participants' numeracy skills on their task performance. This metric is valuable in assessing whether all users benefit equally from visualizations. Moreover, it is worthwhile to explore whether individuals with different levels of numeracy are aware of how this affects their task performance and whether they are confident in using this information.

To address the existing knowledge gap, we have devised a two-step research approach. The initial step involves conducting a preliminary user study, followed by a concluding user study. The preliminary study targets a broad and diverse participant pool with the aim of refining both the research questions and the methodology for the concluding user study, which will focus on a more specific and homogeneous audience. This chapter will focus on the preliminary user study.

5.1 Objectives

The preliminary study provides an opportunity to gain familiarity with task design within the research context, tackle potential challenges, and evaluate the methodology of the user study. Its primary objective is to identify any gaps in the study, allowing us to address them before proceeding to the concluding study.

- **Identify effective visualization techniques for estimating uncertainty:** Evaluate a range of visualization techniques and shortlist those that are most effective in helping users estimate uncertainty in the context of COVID-19 predictions.
- **Refine user study design:** Refine the research questions and the data collection methods used in the user study, ensuring that they are aligned with the chosen visualization techniques and research objectives.
- **Assess feasibility of user study platform:** Test the feasibility of the user study platform by piloting it with the initial group of participants to identify and address any technical or logistical challenges that may arise during the concluding user study.
- **Evaluate feasibility of evaluation metrics:** Assess the feasibility of the chosen evaluation metrics and methodologies to measure the effectiveness of the selected visualization techniques in conveying uncertainty to users.

5.2 Visualizations

The initial study utilized synthetic data, while the visualizations were created using Python's Numpy and Pandas libraries to generate randomized data, and Plotly was employed to produce the graphical representations, both using the Python programming language. Figures 5.1 - 5.5 show the data visualized with different uncertainty visualization techniques.

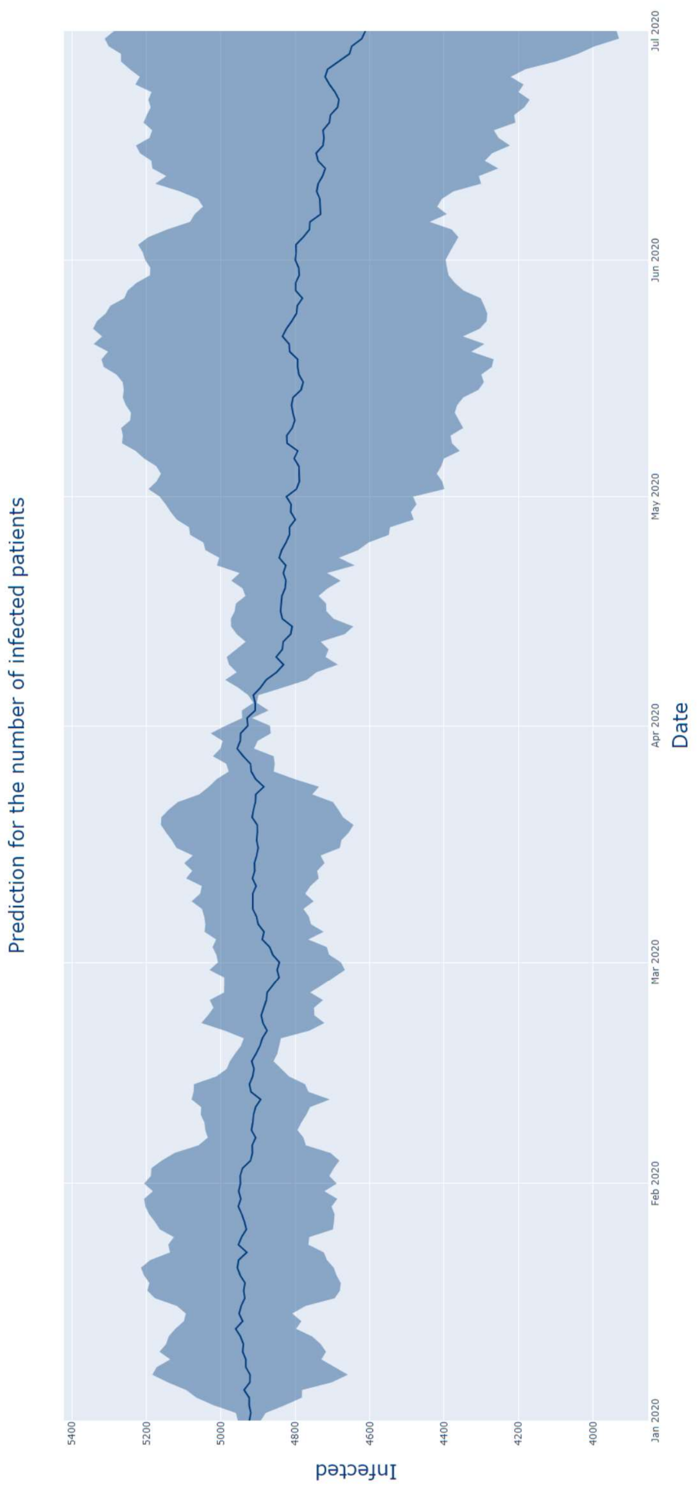


Figure 5.1: Image showing the data used in preliminary study tasks, with uncertainty visualized as Confidence Band.

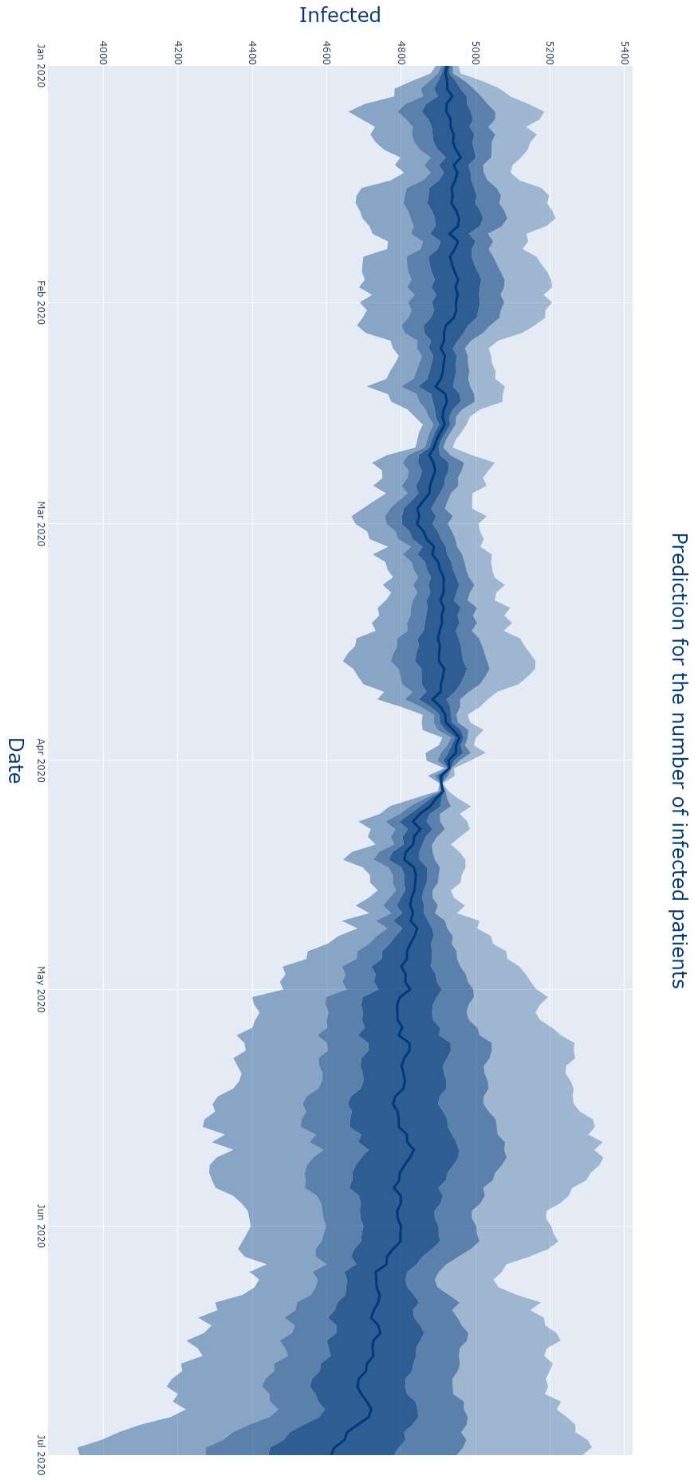


Figure 5.2: Image showing the data used in preliminary study tasks, with uncertainty visualized as Overlapping Bands.

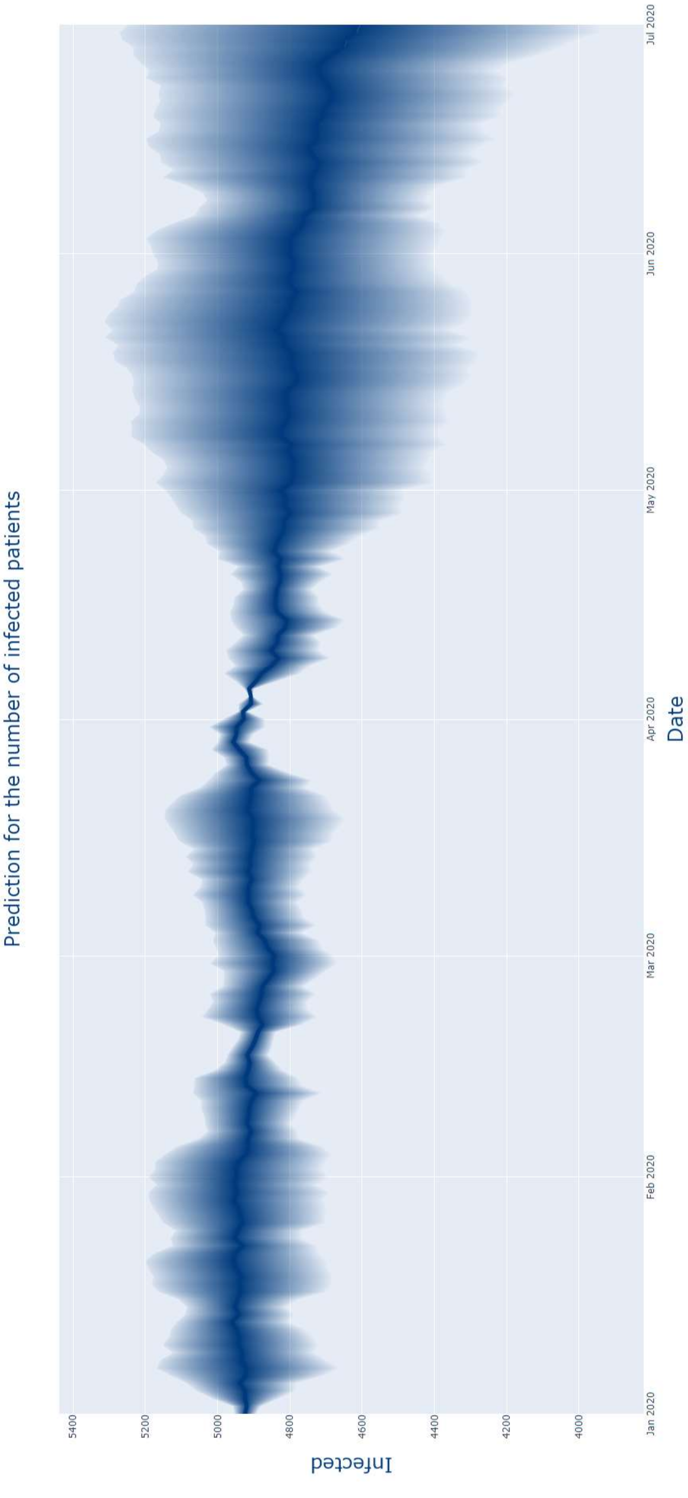


Figure 5.3: Image showing the data used in preliminary study tasks, with uncertainty visualized as Blur.

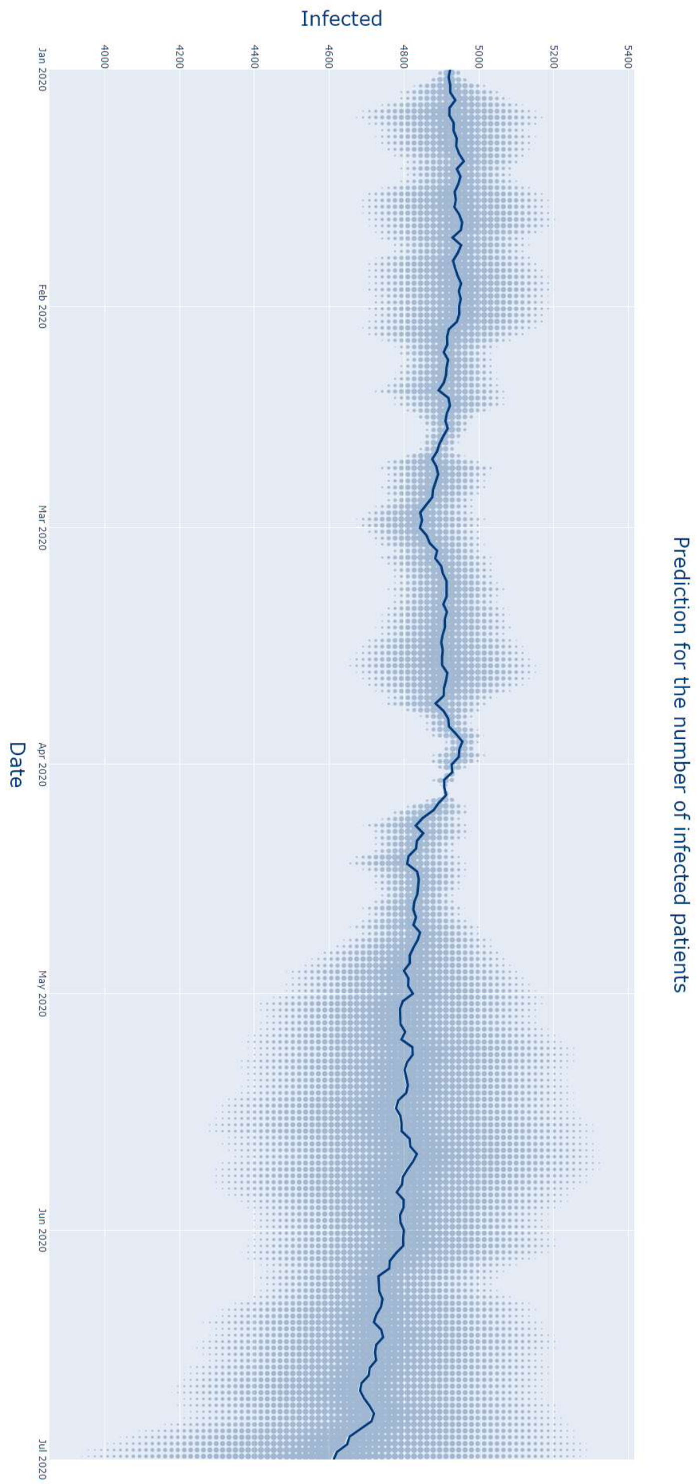


Figure 5.4: Image showing the data used in preliminary study tasks, with uncertainty visualized as Texture.

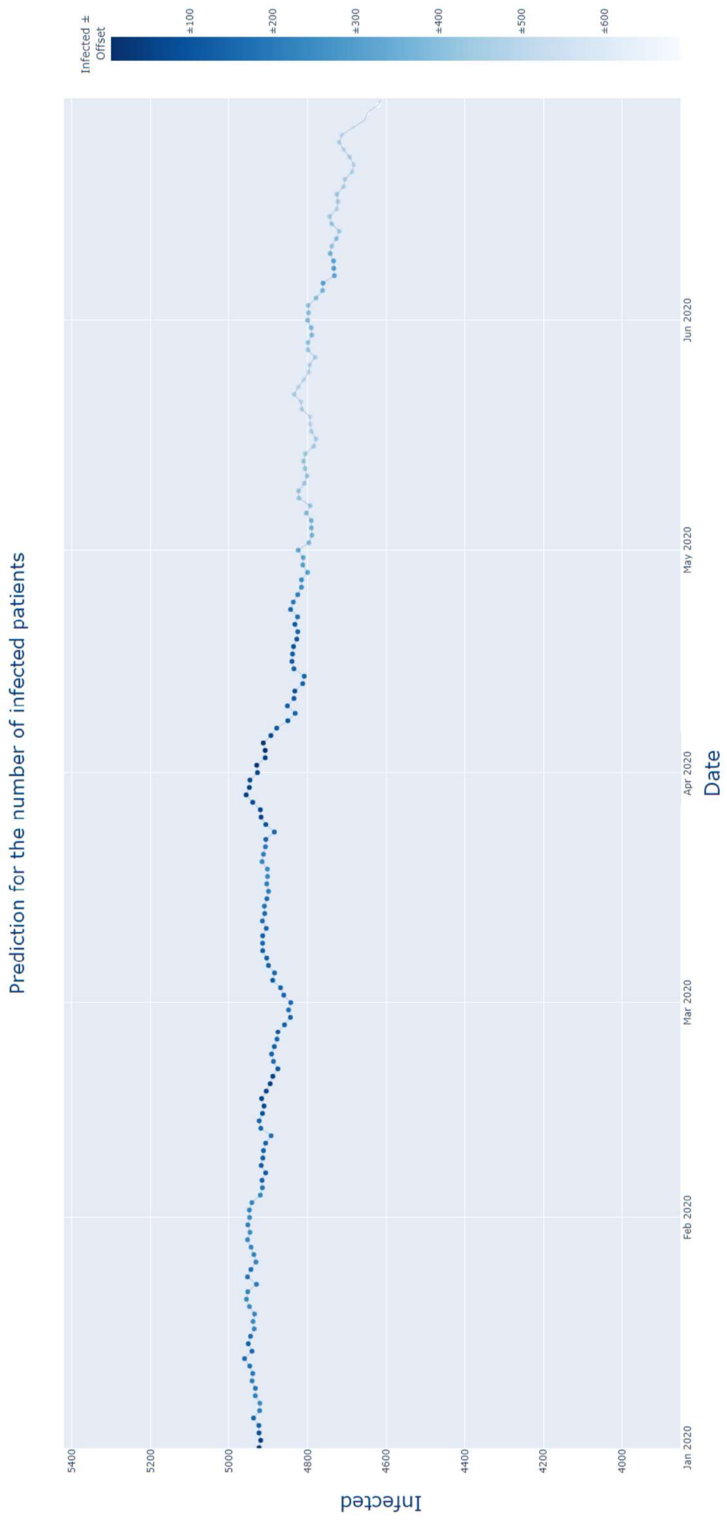


Figure 5.5: Image showing the data used in preliminary study tasks, with uncertainty visualized as Coloured Markers.

5.3 Design

The user study begins with a brief explanation of the goal of the user, as to understand how different variants of line charts affect the quality of the information conveyed to the user. Technical phrases like 'uncertainty information' or 'probability distribution' were intentionally avoided, so the tone of the study is in line with how participants interact with information in non-technical environments.

The concluding user study consists of 7 parts:

- **Participant's data** - Collection of information that will allow us to identify the effects of reported individual differences.
- **Numeracy task** - Collection of information that will allow us to identify the effects of numeracy, a measured individual differences.
- **Visualization task 1** - Collection of participants' uncertainty estimation and user reported metrics under *visualization 1*.
- **Visualization task 2** - Collection of participants' uncertainty estimation and user reported metrics under *visualization 2*.
- **Visualization task 3** - Collection of participants' uncertainty estimation and user reported metrics under *visualization 3*.
- **Visualization task 4** - Collection of participants' uncertainty estimation and user reported metrics under *visualization 4*.
- **Visualization task 5** - Collection of participants' uncertainty estimation and user reported metrics under *visualization 5*.

Visualization 1, *visualization 2*, *visualization 3*, *visualization 4* and *visualization 5* employ one of five techniques — Confidence Band, Overlapping Bands, Blur, Texture, or Coloured Markers — to convey uncertainty information. To prevent participant bias, the assignment of these visualization techniques to the visualization tasks is done randomly. This randomized approach ensures impartiality and allows for an unbiased evaluation of how each visualization technique influences participant understanding and interpretation of uncertainty. All participant information is collected anonymously.

Participant's data

In order to consider the influence of individual factors on uncertainty assessment, participants' highest degree, field of study and the frequency with which they interact with visualizations were inquired about, comprising three out of the four individual factors examined.

Numeracy task

Numeracy is described as a persons' ability to understand, manipulate and use numerical information. A person's numeracy is known to play an important role in their ability to process information with uncertainty and, consequently, weigh the risks and benefits in choices following this information [PETERS et al. (2007)]. The disparity in processing uncertainty information based on numeracy is also true for graphical representations [TOET et al. (2019), TAK et al. (2014)] There are two widely used performance based numeracy scales; the Lipkus numeracy scale and the cognitive reflection test (CRT). Given that the median scores on the Lipkus measure approach the maximum range of scores and the reverse is true for the CRT scale, WELLER et al. (2013) develop a shorter numeracy measuring 8 items of varying difficulty that maintains a normal distribution across different educational groups. To keep the preliminary study short, three questions from the J. Weller et al. numeracy measure were selected whose difficulties span over easy, intermediate and difficult.

The selected questions are:

> **Easy**: If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?
>
> **Intermediate**: In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?
>
> **Difficult**: A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

Visualization task

The visualization task is divided into two distinct segments to comprehensively assess user estimation and perceptions.

The first segment features three hypothetical questions, each with three components. In this segment, participants are tasked with making estimations and asserting whether their informational needs have been met. This segment is designed to probe users' ability to estimate uncertainty and their awareness of the gap between the information provided and the information needed. The hypothetical questions are as follows:

Q1.1 You have two important appointments, one on the 1st Feb and the other on the 1st Apr. You can only make each of those appointments if the infected number on the previous day is below 5000.

a) What is the likelihood that you are able to make the appointment of Feb 1st?

b) What is the likelihood that you are able to make the appointment of Apr 1st?

Q1.2 I feel informed enough to make this decision.
◯ Yes
◯ No

Q2.1 As of Jan 1st 2020, a ban on travelling exists. However, it is announced that if the number of patients remains the same or decreases at the end of May, the ban will then be lifted effective on June 1st. With the given prediction, how likely do you think this scenario is, so that you are able to travel in the June?

Q2.2 I feel informed enough to make this decision.
◯ Yes
◯ No

The second segment of the task focuses on gathering user-reported metrics. Participants are invited to provide feedback on several aspects, including the perceived difficulty of the task, their perceived success in completing it, the aesthetic quality of the visualizations presented, and the level of clutter or complexity they perceive within the visualizations. This segment aims to capture user perspectives on various subjective elements related to the task. The questions in this segment are as follows:

Q3. Please answer the following questions in reference to the current task.

	1	2	3	4	5
How hard did you have to work to accomplish your level of performance? (1: Not hard at all - 5: Very hard)	○	○	○	○	○
How successful do you think you were in accomplishing the goals of the task? (1: Poor – 5: Good)	○	○	○	○	○
Visual representation is not cluttered (1: Strongly disagree – 5: Strongly Agree)	○	○	○	○	○
Visualization is aesthetically designed (1: Strongly disagree – 5: Strongly Agree)	○	○	○	○	○

Recent authors recommend eliciting decisions from the users as hypothetical actions, in the context of uncertainty to evaluate whether, and how users take into account uncertainty when making decisions [HULLMAN et al. (2019); CASTRO et al. (2022)]. For the purpose of the preliminary study, no decision is elicited, as the goal is to identify a set of visualizations that allow users to extract probabilistic information best. However, queries, Q1.1 a, Q1.1 b and Q2.1, are presented in the guise of a ’plausible’ scenario involving a hypothetical prediction resembling a pandemic, which might remind users about decisions that might have thought about in the recent past. The answers are elicited in the form of a range slider with the min and max marked at 0 and 100, respectively.

Some authors look towards self-reported metrics of trust, confidence or visual appeal [ELHAMDADI et al. (2022); CASTRO et al. (2022)]. Similarly, we include two parameters, Performance and Effort, of the self reported NASA-TLX workload scale, along with user satisfaction with the visualization in terms of clutter and aesthetics in Q3.

5.4 Implementation

The user study was designed to be web based as it included interactive visual elements and was intended at reaching as many people as possible. PT-Survey which is DLR’s in-house extension of LimeSurvey was chosen as the tool for implementation largely because it met the requirements of the Data Privacy policy that was necessary, along with providing an easy means to the study development and execution. The visualizations were created in Python with the Plotly package. The visualizations were the uploaded as JPEGs to PT-Survey.

5.5 Evaluation

The evaluation of the results was predominantly conducted using the R programming language and associated statistical tools and libraries.

The performance measure chosen for evaluating user estimations was squared error. This selection was made because squared error provides a measure of the magnitude of deviations between the estimated values and the ground truth while capturing both the direction and magnitude of errors, making it suitable for identifying both overestimations and underestimations. Additionally, squared error emphasizes larger errors, giving them more weight in the evaluation process. This is particularly important when dealing with estimations where larger errors can have a more significant impact or when the goal is to penalize substantial inaccuracies.

The answers to the questions in the numeracy section were graded not equally, but based on their difficulty to avoid potentially flattening the responses, to account for the varying levels of complexity or the fact that some questions may be more challenging than others. This was important, as respondents who answered easier questions correctly and respondents who answered harder questions correctly may differ in their level of knowledge or skill.

In finding correlation between the attributes of the user study results, non-parametric tests are implemented as the responses do not follow a normal distribution. Kendall Tau-b is used to find correlation between ordinal-ordinal values, and the Pearson's coefficient is used in finding the correlation between continuous-ordinal values.

5.6 Analysis

The study had a total of 94 Participants. Most participants held a higher degree with either Masters, PhD or higher, and also used visualizations more than once a week, either in work or in their everyday activities.

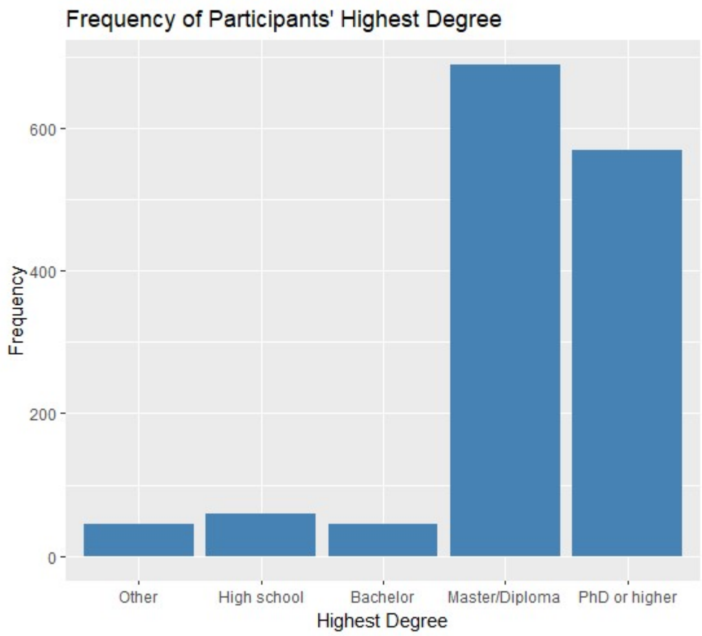


Figure 5.6: Count of users by highest degree.

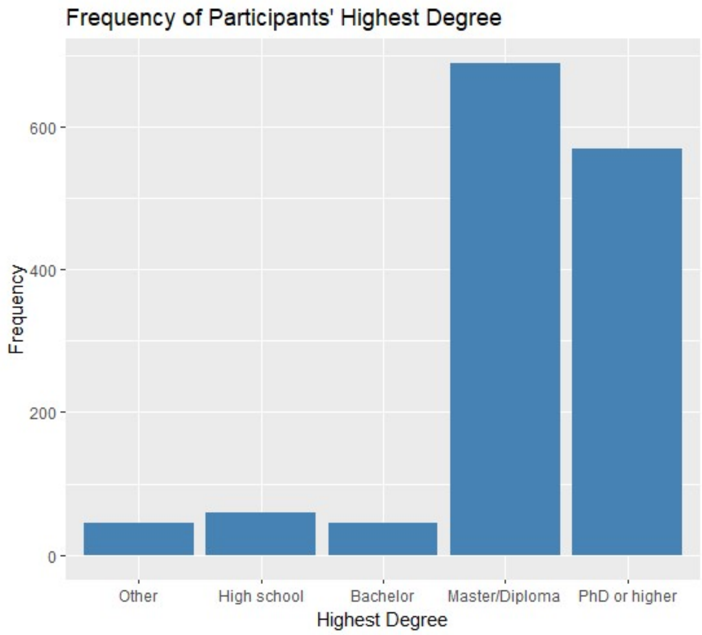


Figure 5.7: Count of users by frequency of visualization use.

Global Statistics

Repeated Measures ANOVA is often used as a statistical tool to find insight into tests when the same conditions are measured under the different treatments, but it assumes sphericity and that the dependent variables are close to a normal distribution. The results of the study violate these assumptions, and to overcome this, the R package npmv BURCHETT et al. (2017), which computes global statistics with ANOVA with permutation and randomization p values is used.

The function **nonpartest()** performs global statistics. In code snippet 5.1, the squared error resulting from participants' responses to Q1.1 a, Q1.1 b and Q2.1, binary variables from responses to Q1.2 and Q2.2, and ordinal responses to Q3 are treated as dependent variables under the *treatment* of different visualizations (explanatory variable). The global test statistics resulting from code snippet 5.1 is shown in Table 5.2 and the relative effects is shown in Table 5.3.

Listing 5.1: nonpartest() function from npmv package in R

```
nonpartest(Error|Percieved_Difficulty|Percieved_Difficulty|
Percieved_Clutter|Percieved_Aesthetic|Information_needs_met
~ Visualization, data = survey_results_as_dataframe,
permreps = 1000)
```

Test	Test Statistic	df1	df2	P-value	Permutation Test p-value
ANOVA type test p-value	11.90	16.9	5946	< 0.5	< 0.5

Table 5.2: Results from the global statistics test in code snippet 5.1

The p-values in Table 5.2 tell us that given the collective influence of the response variable between the visualizations is significant. The values in Table 5.3 which are the relative effects show us that none of the individual response variables can uniquely explain the explanatory variable. The relative effects quantify the tendencies observed in the data in terms of probabilities.

Visualization	Difficulty	Success	Clutter	Aesthetic	Error	Information needs met
Blur	0.50632	0.49958	0.38974	0.38524	0.49680	0.48652
Coloured Marker	0.56571	0.45121	0.44191	0.44481	0.53914	0.48475
Confidence Band	0.45559	0.49344	0.59001	0.51810	0.50519	0.48298
Overlapping Bands	0.49724	0.52941	0.57472	0.56744	0.48416	0.51312
Texture	0.47515	0.52636	0.50363	0.58442	0.47470	0.53262

Table 5.3: Influence of each response variables on the visualization resulting from code snippet 5.1

Posthoc test

The **ssnonpartest()**, from the same package is used for post-hoc analysis, in identifying response variable(s) that are statistically significant with $\alpha \leq 0.05$ (user definable) in their contribution to the differences between different treatments. The posthoc test is configured in code snippet 5.2 to use ANOVA to find significant dependent subsets.

Listing 5.2: ssnonpartest() function from npmv package in R

```
ssnonpartest(Error|Percieved_Difficulty|Percieved_Difficulty
|Percieved_Clutter|Percieved_Aesthetic|Information_needs_met
~ Visualization , data = survey_results_as_dataframe,
test = c(1, 0, 0, 0), alpha = 0.05, factors.and.variables =
TRUE)
```

The results of npmv text (code snippet 5.2), show that the response variables 'Perceived Clutter' and 'Perceived Aesthetic' show significant difference between the visualization, and both variables show significant pairwise comparison with Error. Plotting these two variables against squared error (Figure 5.8 and 5.9) show that visualizations with the user reported least clutter and best aesthetic, have the smallest error and least variability.

Both perceived clutter and aesthetic do not seem to be influenced by the user traits recorded; numeracy, degree of education and frequency of visualization use. However, there is moderate to good correlation between the user reported metrics for the visualizations. It follows that the lower the visual clutter, the better is the aesthetic, following which the users feel that the task was easier to accomplish and that they were successful in their estimation. This is also confirmed by the comparison of perceived aesthetic and clutter against error.

Additionally, it is interesting to note, that whether the users feel that their information needs have been met are only weakly correlated with the users' degree and frequency of visualization. The user traits themselves also show very weak correlation when compared with each other. Both these statements could possibly be attributed to the fact the participants

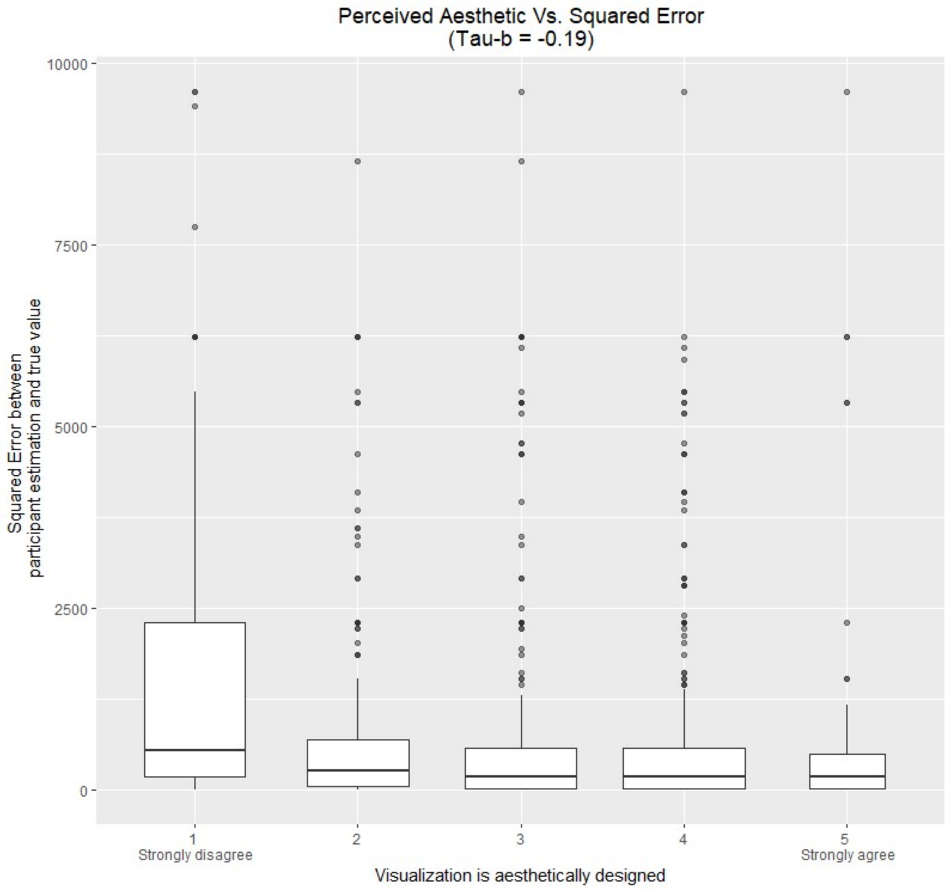


Figure 5.8: Scatter plot showing the squared error between true value and user estimation against user reported value for aesthetic. (5 -> Most pleasing aesthetically)

are homogeneous in that the majority have a higher degree of education, showed to have higher numeracy and frequently used visualizations. Another reason for numeracy not correlating with error, or other traits could be due to the selection of a limited number of numeracy questions from a scale that was perhaps not meant to be used this way.

Plotting the squared error, with (Figure 5.10) and without (Figure 5.11) log scale against visualization, shows us that Overlapping bands, Texture and Blur have the median values closest to zero. From the user study, the three mentioned visualizations will be taken forward for the final study. It should be noted, however, that the pairwise comparison between visualization and error is not statistically significant.

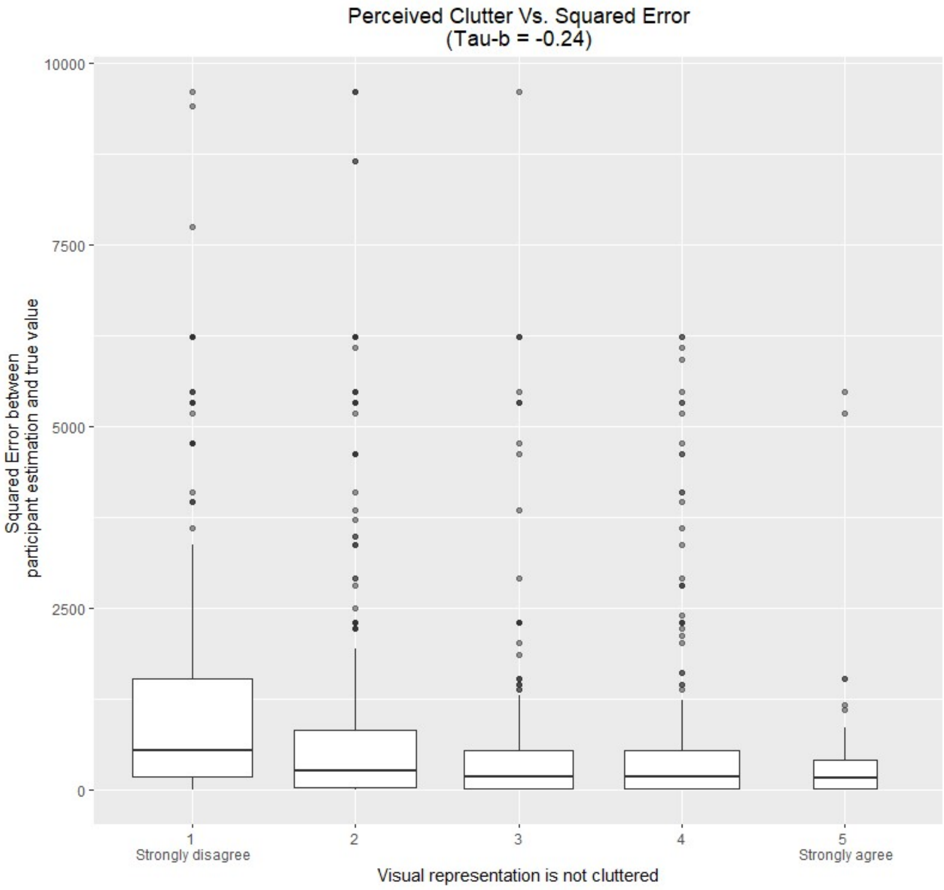


Figure 5.9: Scatter plot showing the squared error between true value and user estimation against user reported value for clutter. (5 -> Least clutter)

Based on the analysis, some minor changes were made to the design of the concluding user study that will be discussed in the next chapter. The implementation and dissemination of the preliminary study posed no significant challenges, and this implementation will be carried forward as such. The corresponding evaluation metrics also posed no significant challenges in the preliminary study.

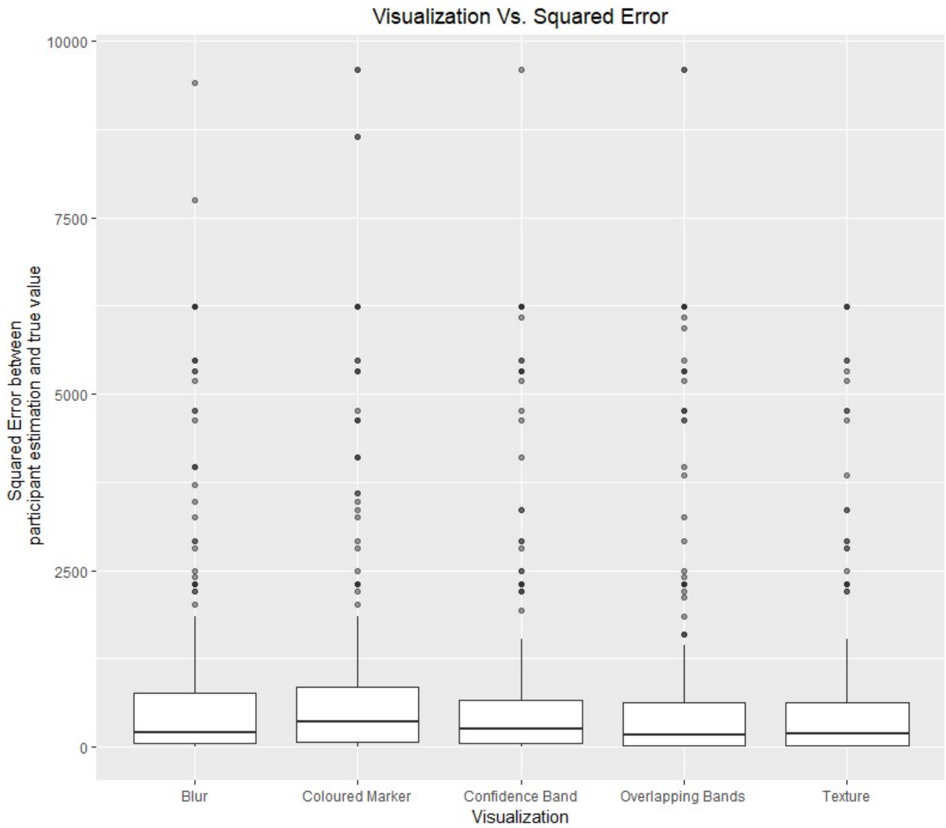


Figure 5.10: Box plot showing distribution of squared error for each visualization

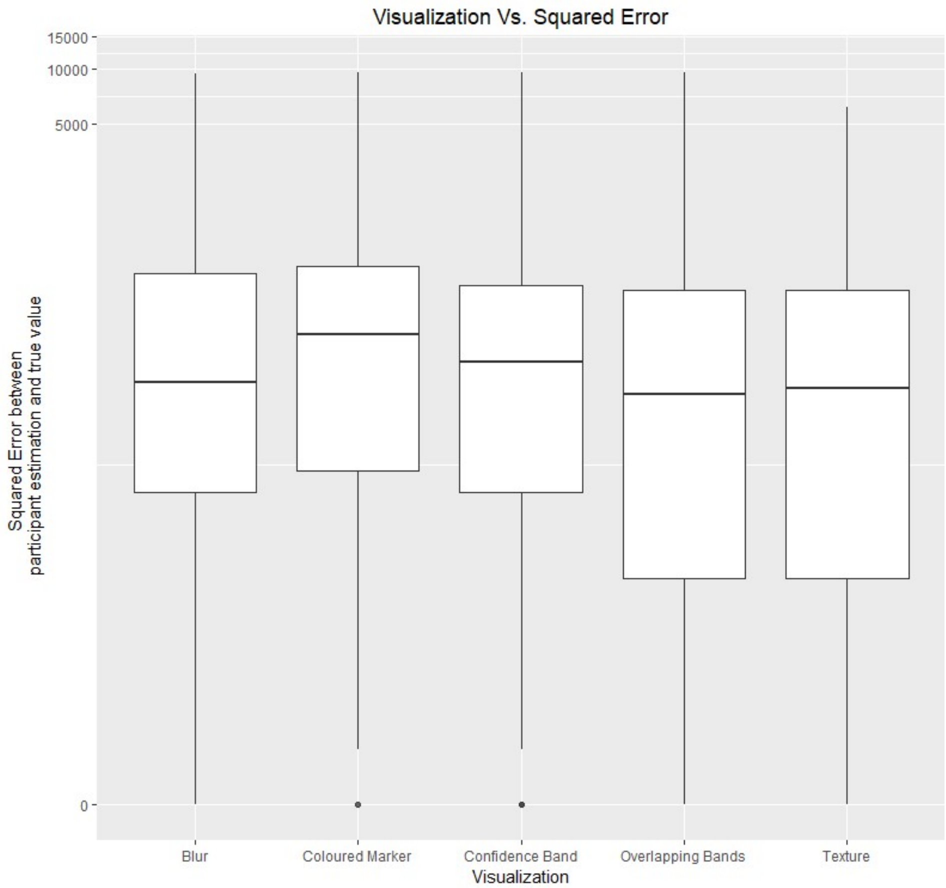


Figure 5.11: Box plot showing distribution of log of squared error for each visualization

6 Concluding User Study

Following the analysis of the preliminary study, the concluding study was carried out to connect with the intended target group, the local health authorities.

Comparison of multiple uncertainty visualizations is a frequently used approach for estimating how effectively a user can extract information, make inferences, or make decisions with a visualization [HULLMAN et al. (2019)]. Thus, the concluding user study follows this approach with the three visualizations selected from the preliminary study: Blur, Overlapping Bands and Texture. This chapter will detail the changes made to design and implementation of the concluding study from the preliminary study, followed by the evaluation and results. The last section discusses the results in the context of the research questions, and what the results mean towards effectively communicating uncertainty in time series visualization.

6.1 Objectives

The primary objective of the concluding user study is to identify the visualization technique that best allows a user to estimate uncertainty in time series prediction data, if at all, the selected visual techniques differ widely in their permissibility of uncertainty estimation. It is possible that there is a gap between the informational needs of users in confidently estimating uncertainty and the information communicated by existing uncertainty visualizations of time series prediction, however widely they differ. To this end, the secondary objective of this study is to first, identify this gap and second, to offer possible solutions that might contribute toward closing this gap.

KINKELDEY et al. (2014) report that in cases of uncertainty visualisation user studies where user reported confidence is included in assessing the usability of visualizations, user performance and confidence were in agreement. As an extension of this, the user study also aims to identify if user reported factors that affect usability such as clutter and aesthetic correlate with user reported success and difficulty, and consequently, user performance.

6.2 Design

Much of the design and procedure of the preliminary user study is carried forward to the concluding user study, with a few changes. The study can be found in its entirety in appendix A. However, this section provides of a concise overview to ensure a better understanding of the results. The concluding user study consists of 3 visualization tasks, instead of 5 as two visualizations from the preliminary study. So, three visualization tasks employ one of three techniques — 'Blur,' 'Overlapping Bands,' or 'Texture' — to convey uncertainty information. The randomization of the assignment of these visualization techniques to the visualization tasks persists to account for participants bias.

Participant's data

In order to consider the influence of individual factors on uncertainty assessment, the recorded information in the concluding study is the same as the preliminary study, except for the frequency with which they interact with visualizations is now changed to frequency with which they interact with line charts.

Numeracy task

Given the influence of numeracy of uncertainty estimation described in section 5.3, it is again measured as an individual factor. The same three numeracy questions of varying difficulty as in the preliminary study are asked, along with two additional questions intended to evaluate probabilistic understanding from its graphical representation. Given an image of a probability density function of a normal distribution (Figure 6.1), participants were asked two questions to assess their comprehension regarding the concept of what a normal distribution looks like, the understanding that, given a median, 50% of data points have a value smaller or equal to the median, and 50% of data points have a value higher or equal to the median, as well as the observation that approximately 95% of values fall within 2 standard deviations from the mean. Additionally, the use of Mini-VLAT [PANDEY and OTTLEY (2023)] was considered to evaluate visual literacy, but eventually dismissed due to time constraints.

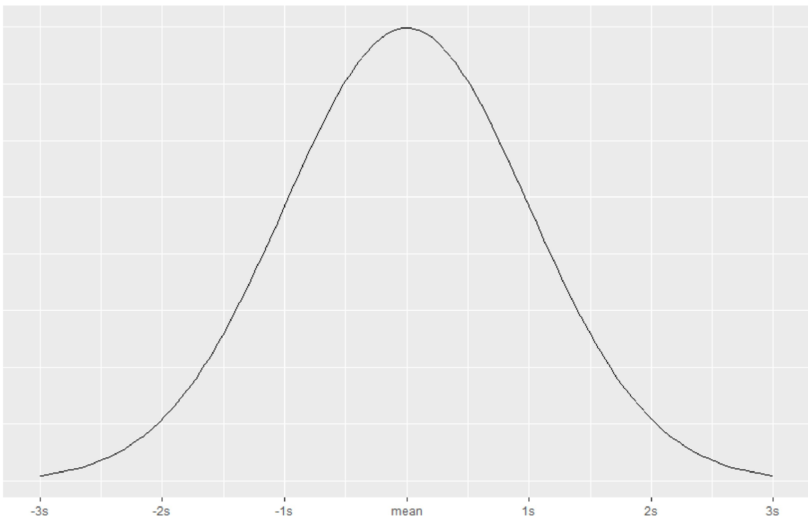


Figure 6.1: Probability density function of a normal distribution

Visualization task

The visualization task is divided into the same two distinct segments as before. In the first segment, the hypothetical questions are extended to have an additional component. In addition to making estimations and gauging whether their informational needs have been met, participants are now asked to state the information missing that they believe is necessary for making accurate assessments. The second segment remains the same.

The visualizations themselves show the true number of weekly cases reported by the European Centre for Disease Control and Prevention (ECDC) for Germany between 28.01.2023 and 04.03.2023 which can be found on GitHub [SHERRATT et al. (2023)] in a line chart, along with point estimates of predictions from three models for the weeks between 11.03.2023 and 01.04.2023 along with their 95% Credible interval, which are also present in the same GitHub dataset.

6.3 Implementation

The concluding user study is executed in PT-survey as before. The visualizations were created in Python, with the Plotly package in a Dash application. For the concluding study, the elements of interaction were intended to be kept in the study. While Plotly visualizations can be saved and uploaded as an HTML object, PT-Survey does not allow for this. Thus, the Dash application was hosted on Google Cloud and inserted into PT-Survey as an inline frame.

6.4 Evaluation

The evaluation of the results of the concluding study are also conducted in R. The performance measure chosen for evaluating user estimations continues to be the squared error, and the answers to the questions in the numeracy section continue to be graded on their difficulty. The new information obtained on the information gap entered by users as free text is evaluated manually, as the sample size is too small for any kind of machine based text analysis. Additionally, some of the responses were in German, and others in English, adding further complexity. The remaining evaluation remains the same; Kendall Tau-b is used to find correlation between ordinal-ordinal values, and the Pearson's coefficient is used in finding the correlation between continuous-ordinal values.

6.5 Results

The study had a total of 31 Participants. Most participants held a higher degree with either Masters, PhD or higher, and the frequency with they interacted with line charts is shown in Figure 6.3. The response from one participant was removed from the study as they explicitly stated having misinterpreted the questions in their response.

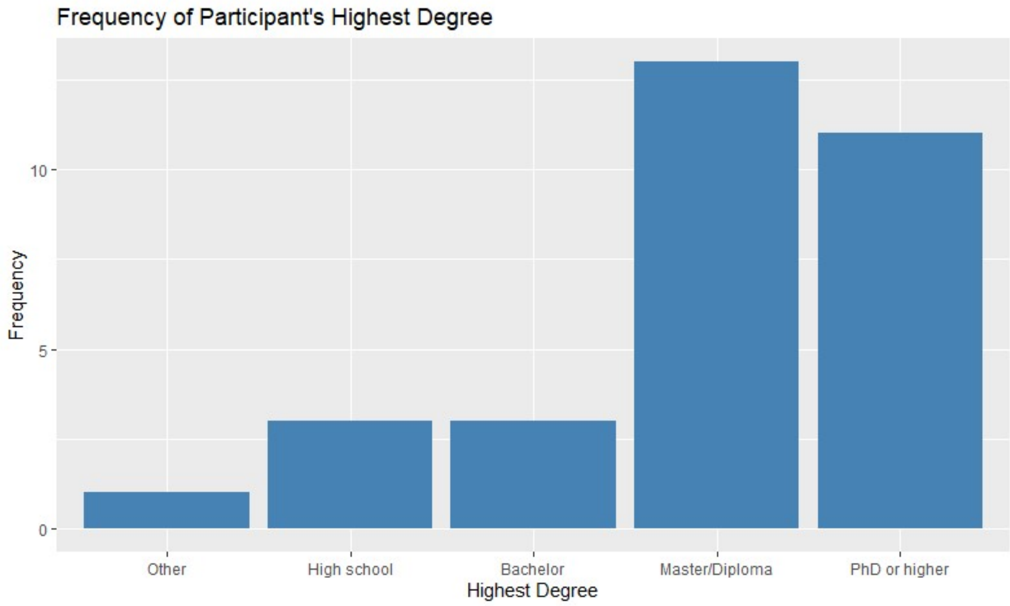


Figure 6.2: Count of users by highest degree.

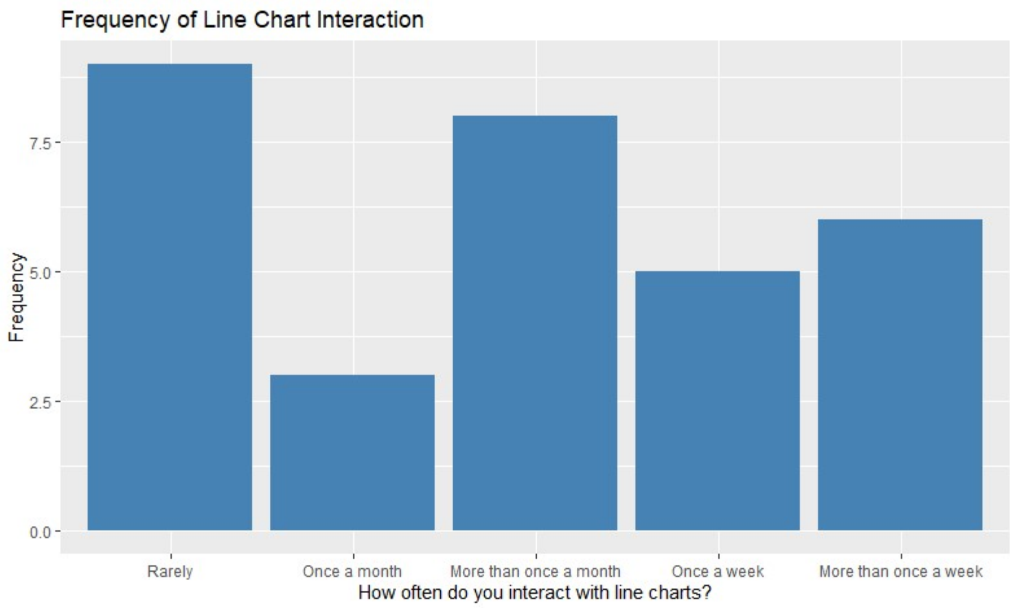


Figure 6.3: Count of users by frequency of line chart use.

This section discusses the results of the concluding user study. The results are grouped so that they may be addressed directly under the appropriate research questions.

R1: What is the impact of different visualization techniques on the participants' uncertainty estimation in time series predictions visualized in line charts?

When visualizing uncertainty with the method introduced in this thesis, *Texture*, the squared error between the participants' estimation and the true value averages lower than in comparison to Blur and Overlapping Bands. It also shows lesser variability in the measure of squared error, see Figure 6.4.

R2: Is there a discernible correlation between users' preferences for specific visualization techniques in meeting their information needs and the resulting task performance accuracy when employing these varied visualization techniques?

It appears from Figure 6.5, squared error between participant estimation of likelihood and true value show more variability, but also a lower median value for instances when participants felt that they had enough information to make an estimation as opposed to when they did not. The proportion of Yes to No is not different under the different visualization.

There is a weak negative correlation between participants' numeracy and their information needs. i.e, the higher their numeracy, the more likely they felt they were able to make an estimation with what information they were given. The results also show a correlation between their field of work, with participants in *Public Health, Neuroscience, Political and Administrative Sciences* requiring additional information. However, the distribution of the number of participants in each field is very skewed to be able to make such a blanket statement. The participants' highest degree and the frequency with which they interact with line charts do not play a role in how they answered this question.

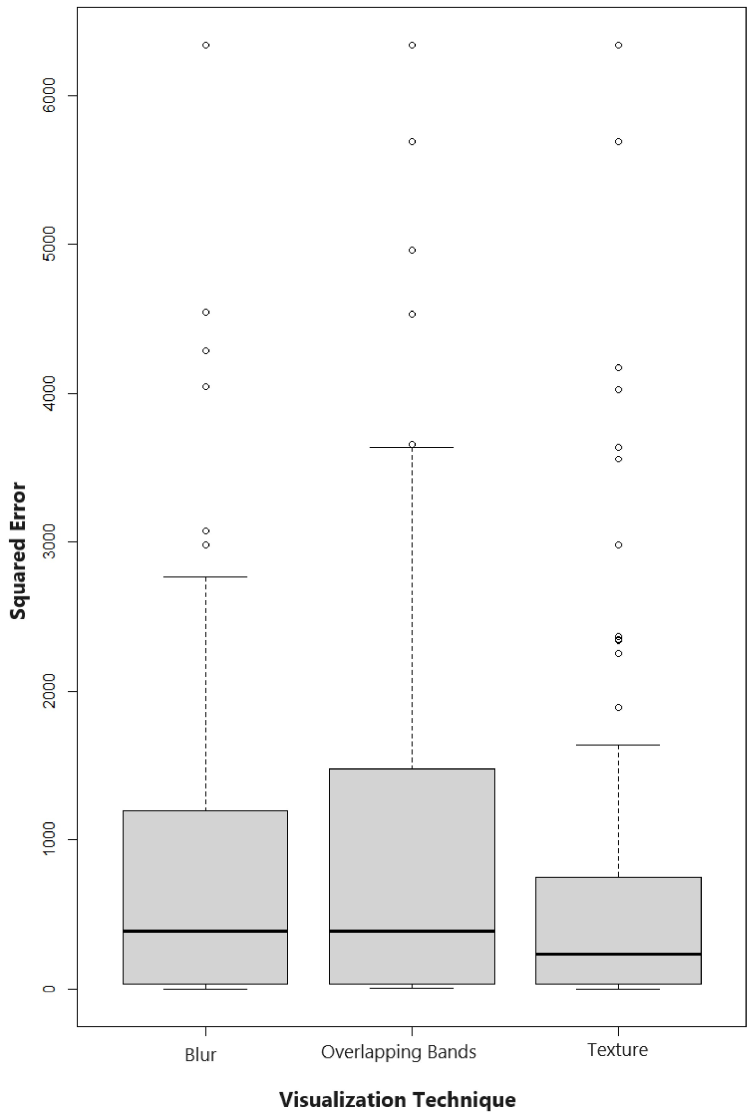


Figure 6.4: Box plots comparing the distribution of squared error between participants' estimation and ground truth for different visualization techniques.

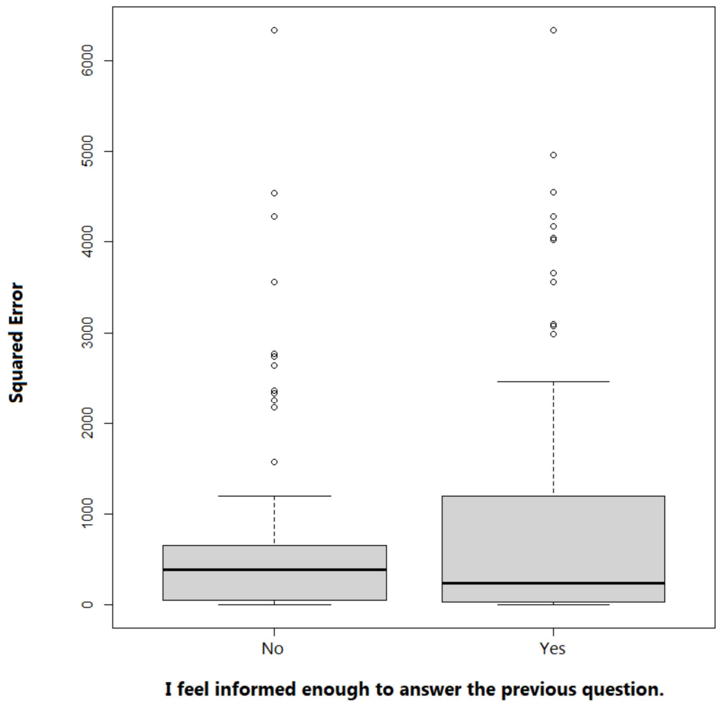


Figure 6.5: Box plots comparing the distribution of squared error between participants' estimation and ground truth when informational needs met and not.

Out of 279 instances (31 participants * 3 visualizations * 3 likelihood estimations), in approximately a third of the instance (96 instances), participants reported that they did not have the necessary information to make the required estimation. These instances were spread over responses from 19 participants. The information needed by participants can be divided largely into two groups:

- Statistical Information

  In 44 instances, 6 participants either explicitly stated or hinted towards needing more statistical information that would help them link the terms likelihood and uncertainty range. *likelihood to the uncertainty interval.* Two participants suggested the inclusion of additional interactive tools to adequately make an estimation, such as a *calculation tool* or *a table with upper and lower limits estimated by varying the size of the uncertainty interval.* Moreover, two participants stated that they made an assumption that 50% of the range fell above the point estimate and 50% below.

  In 30 instances spread over the responses of 5 participants, the uncertainty interval was referred to as the either *Confidence Interval* or *CI*, despite it being stated that the shown uncertainty was a Credible Interval, with one participant even asking if a Credible Interval was the same as a Confidence Interval.

- Model Information

  In 32 instances, spread over the responses of 5 participants, it was stated that the information provided about the parameters of the model and its **historic accuracy** was necessary to adequately make an estimation. Some participants also mentioned that the width of the credible intervals in some cases were too large to make an estimate.

R3: Do individual differences shared among target user groups, such as frequency of visualization use, the highest degree achieved or numeracy, have an effect on their task performance accuracy?

There is a weak negative correlation between numeracy and error. The higher the numeracy of a user, the lower the squared error between their estimation and the true value 6.6.

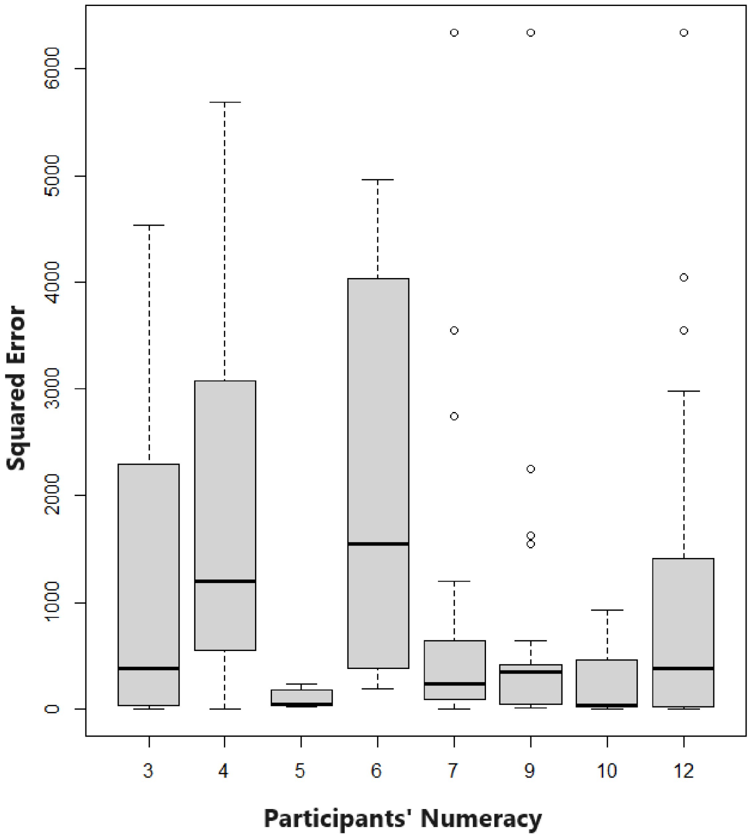


Figure 6.6: Box plots comparing the distribution of squared error between participants' estimation and ground truth for participants with different numeracy scores.

There is a weak negative correlation between users highest degree and error. The higher the numeracy of a user, the lower the squared error between their estimation and the true value 6.7.

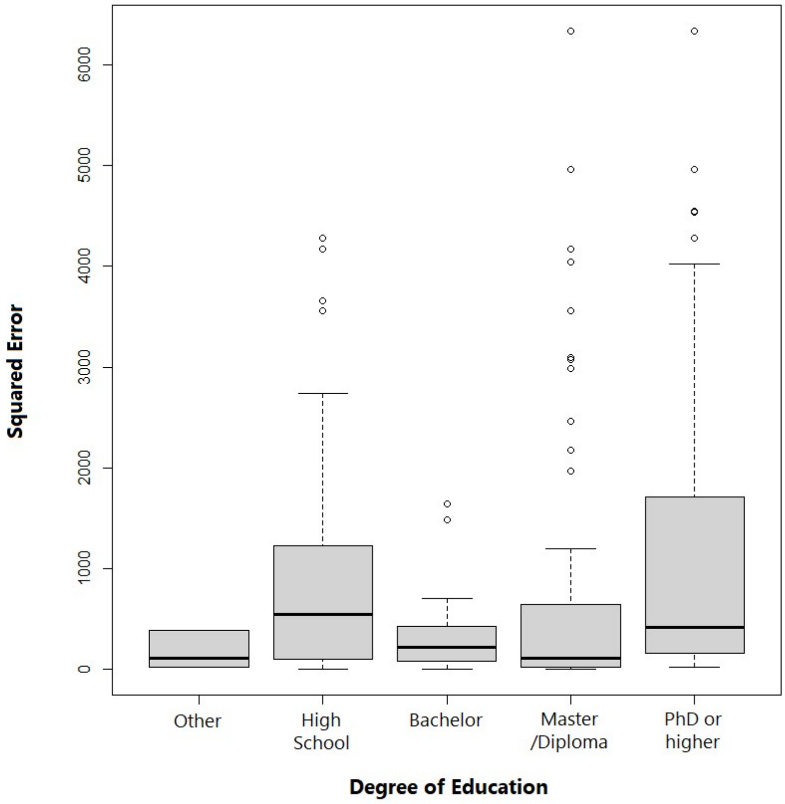


Figure 6.7: Box plots comparing the distribution of squared error between participants' estimation and ground truth for participants with different degrees.

There is no significant correlation between frequency with which the participants interact with line charts and the squared error between their estimation and the true value.

Furthermore, there appears to be a weak positive correlation between the participants' Numeracy and the frequency with which they interact with line charts 6.8 as well as a weak positive correlation between the participants' Numeracy and the highest degree they hold 6.9.

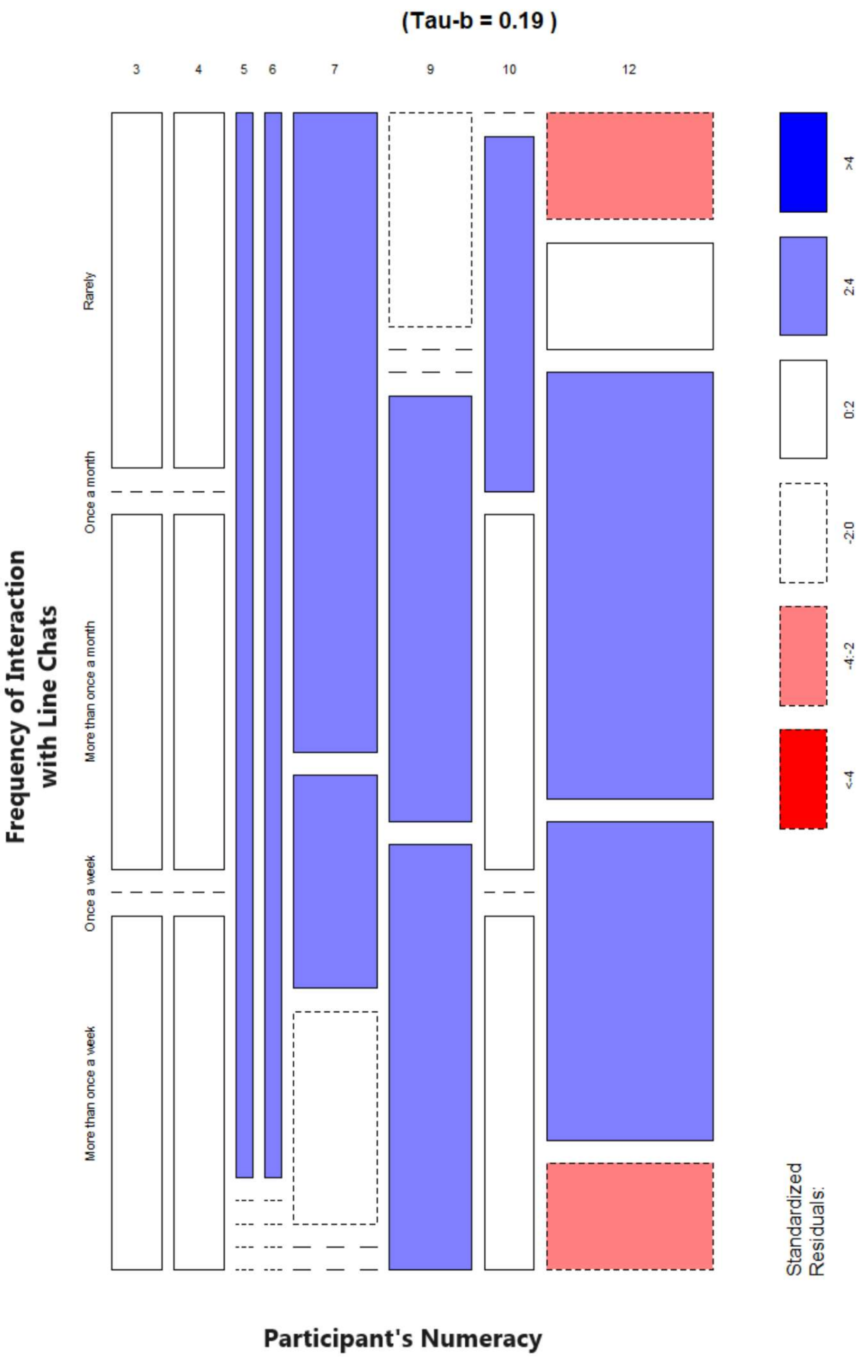


Figure 6.8: Mosaic plots showing the relationship between participants' Numeracy and the Frequency of line chart interaction.

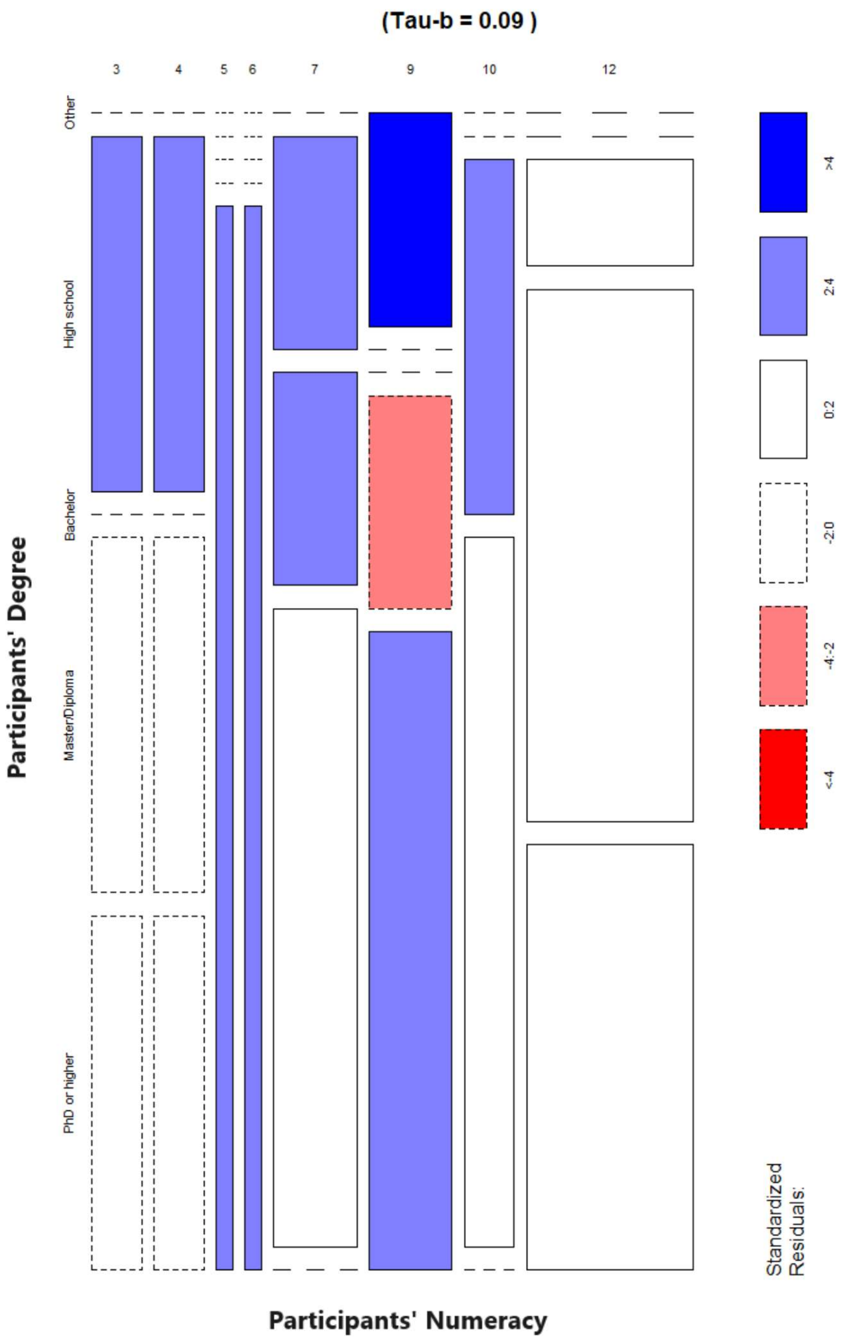


Figure 6.9: Mosaic plots showing the relationship between participants' Numeracy and the highest degree they hold.

R4: How do the effects of varying visualization techniques, like clutter and aesthetics, influence users' evaluations of task difficulty and their perceived level of success in task performance? Consequently, is there a correlation between users' assessments of task difficulty, their perceived level of success in task performance, and their actual task performance accuracy?

The results show a strong negative correlation between perceived clutter and perceived aesthetic 6.10 and a strong positive correlation between perceived success and perceived difficulty 6.11. Both perceived success and perceived difficulty have a weak negative correlation with clutter and weak positive correlation with aesthetic. However, no relationship was found between the reported metrics and user performance.

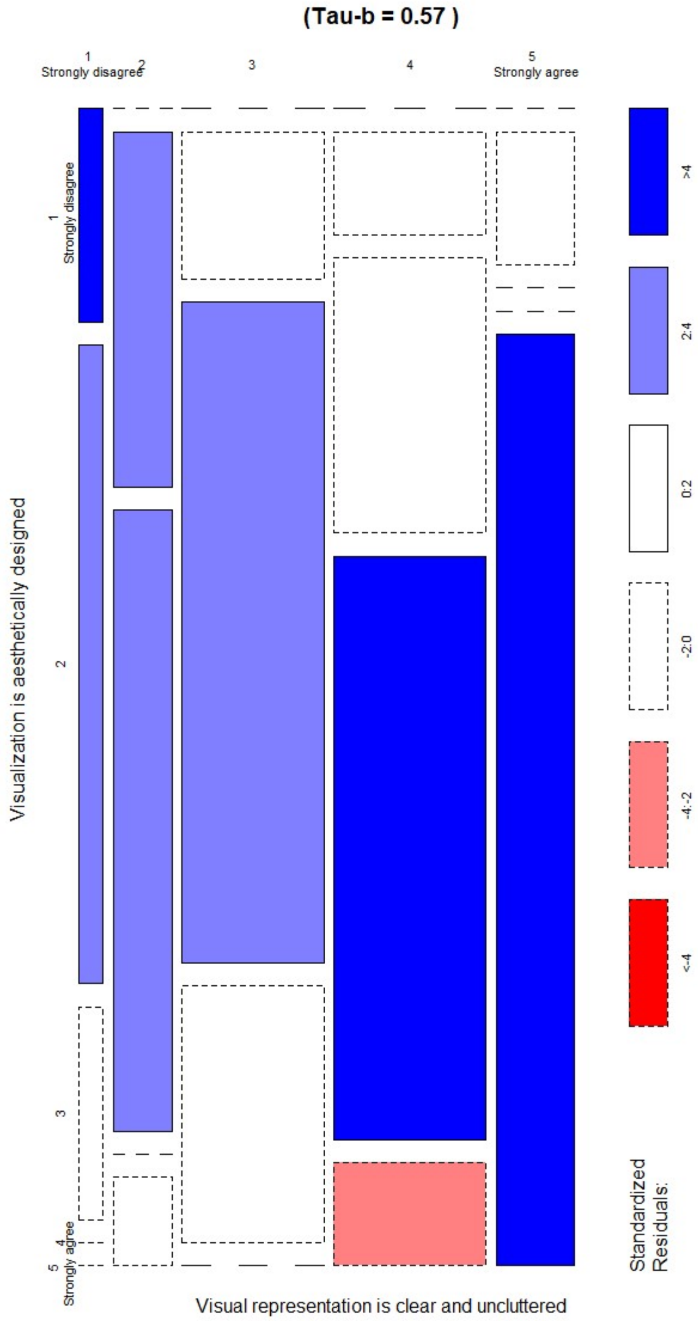


Figure 6.10: Mosaic plots showing the relationship between perceived aesthetic and perceived clutter.

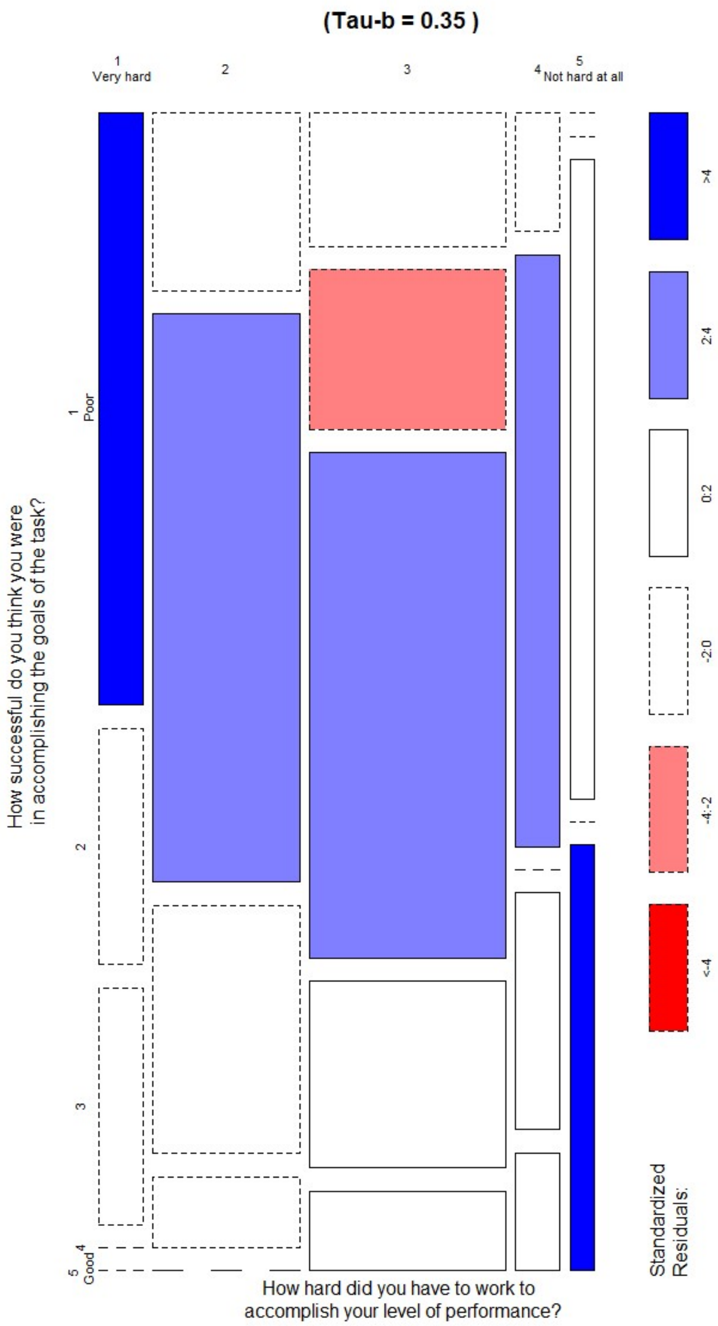


Figure 6.11: Mosaic plots showing the relationship between perceived success and perceived difficulty.

6.6 Discussion

R1: What is the impact of different visualization techniques on the participants' uncertainty estimation in time series predictions visualized in line charts?

Texture performs slightly better than Blur and Overlapping bands in the concluding user study, but not significantly. It is challenging to provide an explanation for this because there was no observed correlation between the visualization technique and the factors that influence usability, possibly owing to the small sample size. According to empirical findings presented in the work of CLEVELAND and MCGILL (1984), it is demonstrated that, in quantitative perceptual tasks, the use of area outperforms colour hue. This observation may offer insights into why texture is more effective than blur. Furthermore, Mackinlay's study lends support to this notion by suggesting that minor inaccuracies in estimating the size of an area result in only minor misperceptions of the associated quantitative value being encoded MACKINLAY (1986). Under the Overlapping Bands technique, one participant made an assumption that *the credibility of 50 is evenly distributed over the range below and above the point estimate.* While this assumption holds true, it highlights the fact, that when using terms like point estimate or statistical jargon, it may not always be evident to those who may be unfamiliar with modelling techniques or statistical terminology. It may also be less apparent in the context of Overlapping Bands, which utilize shaded areas. This differs from Blur or Texture techniques, where visual elements from both sides converge toward the point estimate, thereby providing an additional perceptual cue in understanding that where they converge may be the 'centre', or in statistical terms, the median.

R2: Is there a discernible correlation between users' preferences for specific visualization techniques in meeting their information needs and the resulting task performance accuracy when employing these varied visualization techniques?

It appears that the visualization techniques studied do not influence whether the participants' felt that they had enough information to make an estimate. That suggests that this is then influenced by individual differences. Participants with lower numeracy skills or those specializing in public health, political sciences, and administrative sciences tend to require more information when making estimations. For those in the mentioned fields, this tendency is likely due to the higher stakes and professional obligations associated with their fields of study in context to epidemiological predictions.

Now, let's examine the specific type of information required.

Statistical information

When participants are unfamiliar or have difficulty with interpreting statistical intervals, indeed, providing an interactive basis of calculating uncertainty may be useful. For instance, a visualization may incorporate two movable horizontal lines that allow users to render a calculated probability of the predicted value of each displayed model falling within the selected range defined by these lines.

Additionally, the participants who referred to the credible interval as confidence interval, bring attention to the lack of uniformity in which uncertainty is represented in time series prediction. In the realm of COVID-19 predictions, there exist various terms to describe uncertainty ranges in graphical representations. These terms include Confidence Interval, Credible Interval, Prediction Interval, Simulation Percentile, Uncertainty Interval and +/- 2 Standard Deviation among others. LEFFRANG and MULLER (2021) state that *"Since the predictive model used in this study was a Bayesian model, the uncertainty intervals were actually credibility intervals. Yet, our participants were more familiar with confidence intervals than credibility intervals due to their previous statistics courses. Furthermore, the visualization of credible intervals is similar to the one for confidence intervals. Therefore, we labelled credible intervals as confidence intervals in our experiment."*. The choice of terminology used for these uncertainty ranges often depends on the specific modelling approach and how uncertainty is quantified. However, when creating visual representations of these uncertainty ranges, it is advisable to strike a balance between clarity and comprehensiveness. The uncertainty range maybe a strict interval or a probabilistic distribution, the graphical representation and the terminology used in the visualization should convey this. Furthermore, it is necessary to establish visual representations that cannot be

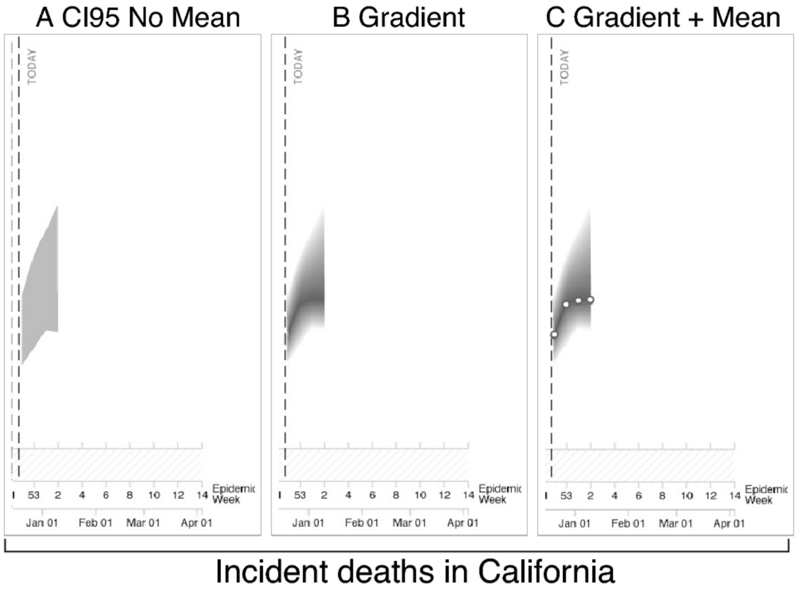


Figure 6.12: Confidence interval visualized as A. Confidence band without mean, B. Gradient without mean and C. Gradient with mean. [PADILLA et al. (2022)]

used interchangeably. PADILLA et al. (2022) acknowledge that *"widespread misconceptions about the meaning and interpretations of standard statistical concepts related to uncertainty (e.g., frequentist confidence intervals, Bayesian creditable [sic] intervals, and variability) form the foundation of misunderstandings of uncertainty visualizations."* but study the visualization of confidence intervals with gradient, see Figure 6.12. A gradient indicates that the data closer to the mean is somehow different from the data farther away, which is untrue for a Confidence Interval.

Model information

Model information regarding scenarios represented, parameters used, and parameter values are entities separate from of uncertainty information, and will not be discussed as part of this thesis. However, ESID is in the process of incorporating model parameter into its application to provide more context to users. In some applications, the reliability of a model is effectively conveyed through the presentation of its past predictions and the uncertainty. This serves as a means to demonstrate the trustworthiness and performance history of the model.

R3: Do individual differences shared among target user groups, such as frequency of visualization use, the highest degree achieved or numeracy, have an effect on their task performance accuracy?

Both numeracy and degree have a weak positive correlation with participants accuracy. This lends credence to the assertions made by TOET et al. (2019). Although numeracy is an individual difference, research, particularly in the field of medical information, has explored ways to enhance visualizations in order to mitigate the impact of numeracy differences. KELLER and JUNGHANS (2017) conclude from their experiment, individuals with low numeracy can be trained by providing appropriate instructions to improve their graph-processing efficiency. PETERS et al. (2014), suggests 'doing the math' for them. This is improbable to do keeping in mind the interests of every target user group, but a calculation tool (mentioned earlier) might be helpful. Such a tool can help users of the visualization focus on their problem. Rather than policymakers having to examine the given information and make educated guesses about the uncertainties associated with various intervention strategies to identify the most suitable one, they can adopt a reverse approach based on their existing data. They can start by stating, for example, that they currently have "x" hospital beds available and the goal is to limit the number of infections to "y" to accommodate this capacity. Then, they can evaluate the various predictions resulting from varying intervention plans and select the one that best aligns with these predefined requirements. In essence, it simplifies decision-making by working backward from their existing constraints and goals.

R4: How do the effects of varying visualization techniques, like clutter and aesthetics, influence users' evaluations of task difficulty and their perceived level of success in task performance? Consequently, is there a correlation between users' assessments of task difficulty, their perceived level of success in task performance, and their actual task performance accuracy?

Participants who believed they had succeeded in their estimations tended to perceive the tasks as less difficult. However, it's worth noting that in this user study, there was no observed correlation between participants' self-assessed success and their actual performance. Nevertheless, the participants' perceived level of success and the perceived difficulty of the tasks can still play a significant role in determining whether they choose to rely on their estimations when making decisions. When the perceived clutter was higher, the perceived difficulty is higher and lower perceived success. Similarly, OGNJANOVIC et al. (2019) suggest that judgment performance and visual information processing in novices or laypeople are negatively impacted by clutter in financial visualizations, in comparison to experts. The perceived clutter did not correlate with visualization technique, which could once again be the result of the low sample size, but it could also allude towards other elements of the visualization that may have contributed to the clutter. ROSENHOLTZ et al. (2007) describe measures of feature and colour variability as visual clutter in information visualization. The concluding user study also shows that perceived aesthetic has a positive correlation to perceived success, and negative correlation to perceived difficulty. CAWTHON and MOERE (2007) show that aesthetic visualizations display a higher level of user patience, resulting in lower task abandonment and erroneous response in their study. While aesthetic may be subjective and difficult to evaluate, the negative correlation with clutter suggests that reducing clutter, which is measurable, may already improve the aesthetic of a visualization.

7 Conclusion

This thesis studied the estimation of uncertainty in time series predictions presented in line charts in a comprehensive manner. The study design was made with careful considerations, drawing from recommendations made by HULLMAN et al. (2019) based on a survey of recent literature for evaluation of uncertainty visualizations, while incorporating multiple aspects of both the visualization and the user that affect uncertainty estimation. To the best of the authors' knowledge, studies so far have focused on the relationship between two or three of these aspects, as described in Chapter 4. The results of the study were discussed in Section 6.6. In the summary section of this chapter, a synopsis of this discussion is provided. With that, there is room for further exploration on this topic. This will be stated in the future work section.

7.1 Summary

The work presented in this thesis assists authors of uncertainty visualizations in improving their visualizations to meet the needs of their diverse audience in achieving more precise uncertainty estimation. The work evaluates a new visual structure that can be used to represent uncertainty in line charts, 2D circular texture, similar to that described by [PANG et al. (1997)] for showing surface illumination differences. This visual structure performed slightly better in allowing for estimation of uncertainty when compared to blurring and overlapping confidence bands.

A disparity emerged between users for whether they felt they had enough information to make an estimation given the credible interval. This is an important factor to consider because users who feel they do not have all the information could potentially not act on the information that is provided. This disparity did not appear to be a result of the visualization technique, nor the individual differences recorded. Nevertheless, the thesis also gathered data regarding the additional information needed by users that could potentially reduce this disparity. The information needed could largely be categorized into: Statistical Information and Model Information. The statistical information comprised of information needed to interpret statistical intervals and the model information comprised of the parameter consideration, source of the data, and reliability in the modelling of the prediction.

The study confirmed the effects of numeracy on uncertainty estimation, previously studied by TOET et al. (2019). This provides the motivation to investigate the work conducted by PETERS et al. (2014), KELLER and JUNGHANS (2017) and others in the field of communicating medical information and its potential applicability to improving uncertainty estimation.

Finally, the thesis examined certain user-reported metrics; success, difficulty, clutter and aesthetic. The perceived success of the user is important because it could again potentially dictate whether they choose to act on the information that is provided. As one would expect, the perceived success of the user correlates negatively with the perceived difficulty. The perceived difficulty correlates positively with clutter, this is expected as clutter can cause cognitive overload, resulting in potential hampering of the tools needed for uncertainty estimation. The perceived difficulty also correlates negatively with aesthetic, which is supported by the study by CAWTHON and MOERE (2007) reporting that higher user patience and lower erroneous response to aesthetically pleasing visualizations.

7.2 Future work

The communication of uncertainty is a vast area of research, and there is much potential for further probing in this area. The work that will benefit this area as supported by this thesis are as follows.

- **Standardizing uncertainty terminology and visualization techniques:** There exists much discrepancy in the terminology used to describe uncertainty in the present literature, even when limiting the terminology to those describing uncertainty in time series prediction in line charts. Both researchers and users of visualizations can benefit from standardizing the terminology, that also indicates the difference in bounded uncertainty and probabilistic uncertainty, and does not require any or much knowledge of statistical tools and methodology to interpret. This should be supported by the visualizations structures representing the terminology. Techniques such as confidence bands, dotted boundary lines and blurring are used to define both types of uncertainty resulting in misinterpretation. The visualizations that provide cues to show that the data closer to the point estimate are different from data farther away, could be used exclusively to indicate probabilistic uncertainty, whereas visualizations that do not could be used exclusively for bounded uncertainty. Thus, there is a strong need for well-supported standardization for defining and visualizing the two groups of uncertainty.

- **Equalizing effects of numeracy in uncertainty estimation:** Ongoing research in the field of medical information communication focuses on effectively conveying complex numerical medical information to patients and consumers with varying levels of numeracy. The general idea seems to aim at minimizing the cognitive effort needed and reducing the need for making many inferences. Exploring and adopting this research is worthwhile in the context of uncertainty visualizations. An example of reducing the cognitive load was discussed earlier as a calculation tool in Section 6.6. Such interactive techniques should be explored in the future.

- **Clutter reduction and aesthetic design influence on uncertainty estimation:** The effect of clutter and aesthetic, and their relationship to uncertainty estimation under complex decision-making tasks, can be explored further. Does the addition of uncertainty information result in a uniform increase in perceived clutter, under all uncertainty visualization techniques? Which visual elements cause the most increase in cognitive load? These questions remain open. In a study conducted by PADILLA et al. (2023), greyscale visualizations of ensemble predictions were more trusted than colour-coded visualizations. The authors initially predicted that adding colour may add complexity and reduce clarity, whose effect would increase with the number of forecasts but found it to be consistently true irrespective of the number of forecasts. Nevertheless, this could also suggest evidence of clutter and aesthetic playing a role in perceived trust of a visualization. Such evidence could be important in establishing principles for clutter reduction and aesthetic enhancement as a means for increasing trust.

A
Concluding User Study

Participants' Data

Q.P1* Degree of Education:

◯ High school

◯ Bachelor

◯ Master/Diploma

◯ PhD or higher

◯ Other: ..

Q.P2 Field of Study:

..

Q.P3* How often do you interact with line charts?

◯ Never

◯ Rarely

◯ Once a month

◯ More than once a month

◯ Once a week

◯ More than once a week

Numeracy Task

Q.N1* If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?

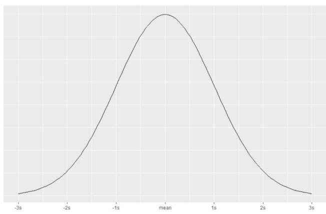
Q.N2* In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?

%

Q.N3* A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

$

Q.N4* The figure below shows the probability density distribution of the variable x. Based on this image, please answer the questions below.

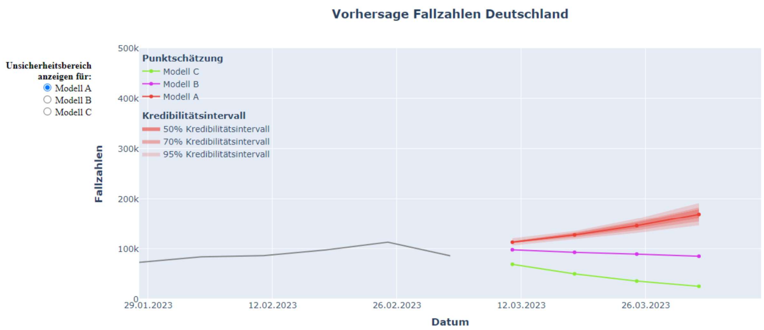


a* If the total area under the curve is 1, what is the area under the curve for x 0 ?

b* If the total area under the curve is 1, what is the area under the curve for -2 x 2 ?

Visualization Task

The visualization below shows the number of COVID cases, in grey-black, in Germany between 19.11.2022 and 04.03.2023. Based on these values, three different prediction models are used to predict data for the 4 weeks following 04.03.2023. The point estimates from the model predictions are show in red, green and magenta. The point estimate is the models' best estimate, but the prediction is not deterministic. You may interact with the chart to see more information with the use of hover, zoom, pan, etc. Additionally, the radio buttons on the left may be used to see the 95% credible interval for each model. The 95% suggests that there is a 95% chance that the true value lies in this range.



Q.T1* You are required to implement a new safety policy for your workplace, for the week following Saturday, 11.03.2023. It is necessary to base this policy on predicted values prior to knowing the actual values. The choice of policies are intended as follows:

Choice of Policy	**Expected Value on 11.03.2023**
Policy I	$0 \leq$ Expected Value $\leq 80,000$
Policy II	$80,000 <$ Expected Value $\leq 90,000$
Policy III	$90,000 <$ Expected Value $\leq 110,000$
Policy IV	Expected Value $> 110,000$

Considering model A, what is the likelihood that the expected value is > 110,000, on 11.03.2023 so you are expected to implement Policy IV.

Q.T2* I feel informed enough to answer the previous question.

- ◯ Yes
- ◯ No

Q.T3* What additional information would make you feel informed enough to answer the question?

Q.T4* Given the same choice of policies as before, and considering model B, what is the likelihood that the expected value on 11.03.2023 is between 80000 and 90000, so you are expected to implement Policy II.

Q.T5* I feel informed enough to answer the previous question.

- ◯ Yes
- ◯ No

Q.T6* What additional information would make you feel informed enough to answer the question?

Q.T7* Given the same choice of policies as before, and considering model C, what is the likelihood that the expected value is < 80000 on 11.03.2023, so you are expected to implement Policy I.

Q.T8* I feel informed enough to answer the previous question.

○ Yes

○ No

Q.T9* What additional information would make you feel informed enough to answer the question?

Q.T10* Please answer the following questions in reference to the current task.

	1	2	3	4	5
a.* How hard did you have to work to accomplish your level of performance? (1: Very hard - 5: Not hard at all)	○	○	○	○	○
b.* How successful do you think you were in accomplishing the goals of the task? (1: Poor – 5: Good)	○	○	○	○	○
c.* Visual representation is not cluttered (1: Strongly disagree – 5: Strongly agree)	○	○	○	○	○
d.* Visualization is aesthetically designed (1: Strongly disagree – 5: Strongly agree)	○	○	○	○	○

B
List of Figures

C
List of Tables

D
Bibliography

[JEWELL et al. 2020] N. Jewell, J. Lewnard and B. Jewell. **Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic**. Annals of Internal Medicine, Vol. 173, 2020.

[BARBAZZA et al. 2022] E. Barbazza, D. Ivanković, K. Davtyan, M. Poldrugovac, Z. Yelgezekova, C. Willmington, B. Meza-Torres, V. L. Bos, Óscar Brito Fernandes, A. Rotar, S. Nuti, M. Vainieri, F. Carinci, N. Azzopardi-Muscat, O. Groene, D. Novillo-Ortiz, N. Klazinga and D. Kringos. **The experiences of 33 national COVID-19 dashboard teams during the first year of the pandemic in the World Health Organization European Region: A qualitative study**. DIGITAL HEALTH, Vol. 8:20552076221121154, 2022.

[DASGUPTA and KAPADIA 2022] N. Dasgupta and F. Kapadia. **The future of the public health data dashboard**. Am. J. Public Health, Vol. 112(6):886–888, 2022.

[THORPE and GOUREVITCH 2022] L. E. Thorpe and M. N. Gourevitch. **Data dashboards for advancing health and equity: Proving their promise?** Am. J. Public Health, Vol. 112(6):889–892, 2022.

[HULLMAN 2020] J. Hullman. **Why Authors Don't Visualize Uncertainty**. IEEE Transactions on Visualization and Computer Graphics, Vol. 26(1):130–139, 2020.

[ZIEMKIEWICZ et al. 2012] C. Ziemkiewicz, A. Ottley, R. J. Crouser, K. Chauncey, S. L. Su and R. Chang. **Understanding Visualization by**

Understanding Individual Users. IEEE Computer Graphics and Applications, Vol. 32(6):88–94, 2012.

[SACHA et al. 2016] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis and D. A. Keim. **The Role of Uncertainty, Awareness, and Trust in Visual Analytics**. IEEE Transactions on Visualization and Computer Graphics, Vol. 22(1):240–249, 2016.

[CARD et al. 1999] S. K. Card, J. D. Mackinlay and B. Shneiderman, Eds. **Readings in Information Visualization: Using Vision to Think**. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[GERSHON et al. 1998] N. Gershon, S. G. Eick and S. Card. **Information Visualization**. Interactions, Vol. 5(2):9–15, 1998.

[ROBERT KOCH-INSTITUT] **7-Tage-Inzidenz der COVID-19-Fälle in Deutschland**. `https://github.com/robert-koch-institut/COVID-19_7-Tage-Inzidenz_in_Deutschland/`.

[BONNEAU et al. 2014] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans and T. Schultz. **Overview and State-of-the-Art of Uncertainty Visualization**. Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization, Springer London, pp. 3–27, 2014.

[CHATFIELD 2014] C. Chatfield. **Model Uncertainty**. Wiley, 2014.

[HOEKSTRA et al. 2013] R. Hoekstra, R. D. Morey, J. N. Rouder and E. Wagenmakers. **Robust misinterpretation of confidence intervals**. Psychonomic Bulletin & Review, Vol. 21:1157–1164, 2013.

[TAN and TAN 2010] S. H. Tan and S. B. Tan. **The Correct Interpretation of Confidence Intervals**. Proceedings of Singapore Healthcare, Vol. 19(3):276–278, 2010.

[MILLER and ULRICH 2016] J. Miller and R. Ulrich. **Interpreting confidence intervals: A comment on Hoekstra, Morey, Rouder, and Wagenmakers (2014)**. Psychonomic Bulletin & Review, Vol. 23(1):124–130, 2016.

[GILG et al. 2023] J. Gilg, P. Kaur Betz, M. Zeumer, L. Spataro, J. Stoll, V. Grappendorf, M. J. Kühn and A. Gerndt. **ESID-0.2.0-alpha**. 2023.

[BETZ et al. 2023] P. K. Betz, J. Stoll, V. Grappendorf, J. Gilg, M. Zeumer, M. Klitz, L. Spataro, A. Klein, L. Rothenhäusler, H. Bohnacker et al. **ESID: A Visual Analytics Tool to Epidemiological Emergencies**. arXiv preprint arXiv:2304.04635, 2023.

[HETHCOTE 2000] H. W. Hethcote. **The Mathematics of Infectious Diseases**. SIAM Review, Vol. 42(4):599–653, 2000.

[KÜHN et al. 2021] M. J. Kühn, D. Abele, T. Mitra, W. Koslow, M. Abedi, K. Rack, M. Siggel, S. Khailaie, M. Klitz, S. Binder, L. Spataro, J. Gilg, J. Kleinert, M. Häberle, L. Plötzke, C. D. Spinner, M. Stecher, X. X. Zhu, A. Basermann and M. Meyer-Hermann. **Assessment of effective mitigation and prediction of the spread of SARS-CoV-2 in Germany using demographic information and spatial resolution**. Mathematical Biosciences, Vol. 339:108648, 2021.

[POTTER et al. 2012] K. Potter, P. Rosen and C. R. Johnson. **From Quantification to Visualization: A Taxonomy of Uncertainty Visualization Approaches**. In: A. M. Dienstfrey and R. F. Boisvert, Eds., Uncertainty Quantification in Scientific Computing, pp. 226–249. 2012, Springer Berlin Heidelberg, Berlin, Heidelberg.

[HULLMAN et al. 2019] J. Hullman, X. Qiao, M. Correll, A. Kale and M. Kay. **In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation**. IEEE Transactions on Visualization and Computer Graphics, Vol. 25(1):903–913, 2019.

[LIU et al. 2020] Z. Liu, R. J. Crouser and A. Ottley. **Survey on Individual Differences in Visualization**. arXiv, 2020.

[MCCUAIG et al. 2005] J. Mccuaig, B. Hetzler, A. MacEachren, M. Gahegan and M. Pavel. **A Typology for Visualizing Uncertainty**. 2005, Vol. 5669.

[PANG et al. 1997] A. T. Pang, C. M. Wittenbrink and S. K. Lodha. **Approaches to uncertainty visualization**. The Visual Computer, Vol. 13(8):370–390, 1997.

[JOHNSON and SANDERSON 2003] C. Johnson and A. Sanderson. **A Next Step: Visualizing Errors and Uncertainty**. IEEE Computer Graphics and Applications, Vol. 23(5):6–10, 2003.

[SANYAL et al. 2010] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn and R. Moorhead. **Noodles: A Tool for Visualization of Numerical Weather Model Ensemble Uncertainty**. IEEE Transactions on Visualization and Computer Graphics, Vol. 16(6):1421–1430, 2010.

[LEFFRANG and MULLER 2021] D. Leffrang and O. Muller. **Should I Follow this Model? The Effect of Uncertainty Visualization on the Acceptance of Time Series Forecasts**. 2021, pp. 20–26.

[MCGRATH et al. 2020] S. McGrath, P. Mehta, A. Zytek, I. Lage and H. Lakkaraju. **When Does Uncertainty Matter?: Understanding the Impact of Predictive Uncertainty in ML Assisted Decision Making**. arXiv, 2020.

[ELHAMDADI et al. 2022] H. Elhamdadi, L. Padilla and C. Xiong. **Using Processing Fluency as a Metric of Trust in Scatterplot Visualizations**. arXiv, 2022.

[OLSTON and MACKINLAY 2002] C. Olston and J. Mackinlay. **Visualizing Data with Bounded Uncertainty**. Proc. of the IEEE Symposium on Information Visualization, Vol. 2002, 2002.

[VAN DER LAAN et al. 2015] D. J. van der Laan, E. de Jonge and J. Solcer. **Effect of displaying uncertainty in Line and Bar charts**. 2015.

[SANYAL et al. 2009] J. Sanyal, S. Zhang, G. Bhattacharya, P. Amburn and R. Moorhead. **A User Study to Compare Four Uncertainty Visualization Methods for 1D and 2D Datasets**. IEEE transactions on visualization and computer graphics, Vol. 15:1209–18, 2009.

[TAK et al. 2014] S. Tak, A. Toet and J. van Erp. **The Perception of Visual UncertaintyRepresentation by Non-Experts**. IEEE Transactions on Visualization and Computer Graphics, Vol. 20(6):935–943, 2014.

[PETERS et al. 2007] E. Peters, J. Hibbard, P. Slovic and N. Dieckmann. **Numeracy Skill And The Communication, Comprehension, And Use Of Risk-Benefit Information**. Health affairs (Project Hope), Vol. 26:741–8, 2007.

[TOET et al. 2019] A. Toet, J. B. van Erp, E. M. Boertjes and S. van Buuren. **Graphical uncertainty representations for ensemble predictions**. Information Visualization, Vol. 18(4):373–383, 2019.

[WELLER et al. 2013] J. A. Weller, N. F. Dieckmann, M. Tusler, C. K. Mertz, W. J. Burns and E. Peters. **Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach**. Journal of Behavioral Decision Making, Vol. 26(2):198–212, 2013.

[CASTRO et al. 2022] S. C. Castro, P. S. Quinan, H. Hosseinpour and L. Padilla. **Examining Effort in 1D Uncertainty Communication Using Individual Differences in Working Memory and NASA-TLX**. IEEE Transactions on Visualization and Computer Graphics, Vol. 28(1):411–421, 2022.

[BURCHETT et al. 2017] W. W. Burchett, A. R. Ellis, S. W. Harrar and A. C. Bathke. **Nonparametric Inference for Multivariate Data: The R Package npmv**. Journal of Statistical Software, Vol. 76(4):1–18, 2017.

[KINKELDEY et al. 2014] C. Kinkeldey, A. M. MacEachren and J. Schiewe. **How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies**. The Cartographic Journal, Vol. 51(4):372–386, 2014.

[PANDEY and OTTLEY 2023] S. Pandey and A. Ottley. **Mini-VLAT: A Short and Effective Measure of Visualization Literacy**. Computer Graphics Forum, Vol. 42(3):1–11, 2023.

[SHERRATT et al. 2023] K. Sherratt, H. Gruson, H. Johnson, R. Niehus, B. Prasse, F. Sandman, J. Deuschel, D. Wolffram, S. Abbott, A. Ullrich, G. Gibson, E. L. Ray, N. G. Reich, D. Sheldon, Y. Wang, N. Wattanachit, L. Wang, J. Trnka, G. Obozinski, ... and S. Funk. **European Covid-19 Forecast Hub**. 2023.

[CLEVELAND and MCGILL 1984] W. S. Cleveland and R. McGill. **Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods**. Journal of the American Statistical Association, Vol. 79(387):531–554, 1984.

[MACKINLAY 1986] J. Mackinlay. **Automating the Design of Graphical Presentations of Relational Information**. ACM Trans. Graph., Vol. 5(2):110–141, 1986.

[PADILLA et al. 2022] L. Padilla, H. Hosseinpour, R. Fygenson, J. Howell, R. Chunara and E. Bertini. **Impact of COVID-19 forecast visualizations on pandemic risk perceptions**. Scientific Reports, Vol. 12(1):2014, 2022.

[KELLER and JUNGHANS 2017] C. Keller and A. Junghans. **Does guiding toward task-relevant information help improve graph processing and graph comprehension of individuals with low or high numeracy? An eye-tracker experiment**. Med. Decis. Making, Vol. 37(8):942–954, 2017.

[PETERS et al. 2014] E. Peters, L. Meilleur and M. K. Tompkins. **Numeracy and the affordable care act: Opportunities and challenges**. National Academies Press, Washington, D.C., DC, 2014.

[OGNJANOVIC et al. 2019] S. Ognjanovic, M. Thüring, R. O. Murphy and C. Hölscher. **Display clutter and its effects on visual attention distribution and financial risk judgment**. Applied Ergonomics, Vol. 80:168–174, 2019.

[ROSENHOLTZ et al. 2007] R. Rosenholtz, Y. Li and L. Nakano. **Measuring visual clutter**. Journal of Vision, Vol. 7(2):17–17, 2007.

[CAWTHON and MOERE 2007] N. Cawthon and A. V. Moere. **The Effect of Aesthetic on the Usability of Data Visualization**. In: 2007 11th International Conference Information Visualization (IV '07), 2007, pp. 637–648.

[PADILLA et al. 2023] L. Padilla, R. Fygenson, S. C. Castro and E. Bertini. **Multiple Forecast Visualizations (MFVs): Trade-offs in Trust and Performance in Multiple COVID-19 Forecast Visualizations**. IEEE Transactions on Visualization and Computer Graphics, Vol. 29(1):12–22, 2023.

Declaration of Academic Integrity

I, Apoorva Karagappa hereby declare that I have written the present work myself and did not use any sources or tools other than the ones indicated.

Datum:	....
21 September, 2023	(Signature)