VISUALIZING UNCERTAINTY WITH CHROMATIC ABERRATION

by

Md Rashidul Islam

Submitted in partial fulfilment of the requirements
for the degree of Master of Computer Science

at

Dalhousie University
Halifax, Nova Scotia
May 2022

Dalhousie University is located in Mi'kma'ki, the
ancestral and unceded territory of the Mi'kmaq.
We are all Treaty people.

Table of Contents

List of Tables

List of Figures

List of Symbols

Abstract:

In recent years an increasing array of research are being conducted by researchers in the field of uncertainty visualization that attempt to determine the impact of representations on users' perception and evaluate its effectiveness in decision making. Uncertainties are often an integral part of data and by nature model predictions also contain significant amounts of uncertain information. A prominent example of uncertainty, COVID-19 is a respiratory infectious disease caused by novel coronavirus. Due to its unprecedented challenges over time and frequent changes of strains, scientists and researchers are investigating the available data to discover the patterns in different demographic areas and examine the effect of vaccinations against different variants. In this study, we explore a novel idea for a visualization to present predictive model uncertainties using Chromatic Aberration (CA). We first utilized existing machine learning models to obtain predictive results using Covid-19 pandemic data and calculated the corresponding model uncertainties for the most impacted countries with respect to number of new-cases, new-deaths, and new-vaccination for different countries. We then visualized the data itself and its associated uncertainties with an artificially spatially separated channels of red, green, and blue color components. This chromatic aberration representation has been evaluated in a comparative user study. From quantitative analysis it is observed that user is able to identify targets in CA method more accurately than the state-of-the-art Value-Suppressing Uncertainty Palettes (VSUP) approach. In addition, their speed of target identification was significantly faster in CA as compared to the VSUP method. But their preference between the two does not vary significantly.

List of Abbreviations

AI	-	Artificial Intelligence
API	-	Application Programming Interface
D3	-	Data Driven Documents
HCI	-	Human Computer Interaction
JSON	-	JavaScript Object Notation
ANN	-	Artificial Neural Network
CNN	-	Convolutional Neural Network
RNN	-	Recurrent neural networks
MLP	-	Multilayer Perceptron
LSTM	-	Long Short-Term Memory
MAE	-	Mean Absolute Error
RMSE	-	Root Mean Square Error
WHO	-	World Health Organization
REB	-	Research Ethics Board
LIC	-	Line Integral Convolution
OWID	-	Our World in Data

Acknowledgements

I would like to express my heartiest gratitude to my supervisor Dr. Stephen Brooks due to his cordial supervision from the very beginning of the journey. His unparallel thoughts and unlimited patience of listening with highly sophisticated approaches of sharing ideas and clarifying problems or providing suitable directions greatly helped me to reach stage of my thesis component. I revere from my heart for his endless support and guidance and wish for his long and healthy life.

Secondly, I would like to thank my most beloved parents, for their eternal, natural and demand free love and care throughout my life. In addition to that, I can't forget to contribution of my siblings, friends, and all teachers from my very childhood.

Finally, I would like to appreciate to the user-study participants who helped me to conduct the survey by dedicating their invaluable time and effort.

Chapter 1

1 Introduction

Uncertainty visualization is an ongoing area of research but a topic that many practitioners avoid due to the additional complexity that it introduces [21, 35]. There are various studies conducted for uncertainty representations, for example: textual representation such as captions or tooltips [51], graphical representations such as glyphs [21, 54], custom color palettes such as VSUP [35], bivariate choropleth maps [43] and texture patterns [29]. But as far we know, no uncertainty representation has made use of Chromatic Aberration (CA). To accomplish the purpose, we have categorized the scope of the research with several core components: firstly, collect relevant data from reputable sources. Secondly, generate uncertainty information from predictions based on the data (accomplished by feeding collected data into machine learning models and calculated from the resultant forecasts [6]). Thirdly, visualize the uncertainty and data using CA, as well as competing existing methods. Fourthly, conduct a controlled human-computer interaction experiment to evaluate the effectiveness of the new visual representation. Fifthly, explain experimental results with numerical analysis and draw conclusions.

1.1 Background and Motivation

The outbreak of coronavirus COVID-19 first emerged in China in December 2019 and the expansion has propagated all over the world, being declared as an international public health crisis by World Health Organisation (WHO). Since then, the world has been affected in almost all respects. Various preventive health measures were and are imposed, and different short-term restrictions are applied to the habitants in different countries at different times. But the mortality rate was not mitigated significantly until immunizations were introduced. Tragically, over 318 million people have been infected and 5.5 million have died the world over. The infection and death rate have oscillated in different countries due to a variety of reasons. Moreover, the strain of the virus has changed frequently in different geographical locations with more power and variations and a few of the variants like the British variant, the Delta variant, the Indian variant and more recently the Omicron variant became the prime concern for the world community. Though a great deal of research is being conducted and wide range of immunization processes have impacted the severity of the pandemic, at the time of writing this thesis, it is unknown if and when the world will be rid of this severe pandemic.

Recently, many studies have been conducted to forecast the trend of the spread of the COVID-19 pandemic using various statistical models as well as machine learning models. The autoregressive integrated moving average (ARIMA) model has been widely used in previous studies to analyze and predict the spread of diseases such influenza [1], Cholera [5], along with many other popular machine learning algorithms [2, 3, 9]. The pandemic started very abruptly and so during the first year, it was difficult to develop efficient systems to forecast trends due to the lack of required data. But after more than one year, we have data to explore, analyze and forecast with the help of modern machine learning algorithms. The ability to identify the expansion rate at which the disease is spreading is very important to confront it and help governments' regarding contingent policymaking to properly address the consequences of the pandemic and encourage people to be cautious and follow the rules and health guidelines to achieve the maximum benefit by saving valued lives. That is why one of the objective's behind the current research is develop new tools for uncertainty visualization. We use property driven predicted results of COVID-19 as a test case for exploring chromatic aberration as a visual representation of uncertainty. If we can develop more effective representations of uncertainty, then it might help community administrators with planning or at least improve the means of communication with the general public. And more generally, the development of better uncertainty visualizations could be of use in many other areas as well.

1.2 Background Concepts & Technologies

We will now introduce related terms used in the dissertation so that the reader can better understand the work.

1.2.1 Machine Learning (Predictive models)

Machine learning is an approach of artificial intelligence (AI) to provide automatic learning through the uses of data. What separates this from other solutions is it does not need explicit programming to perform the task since the algorithms are designed to themselves learn from data. From machine learning, we mention major three types of algorithms here [34]:

i. Supervised Learning - In this type, the machine learning algorithm is trained on labeled data. Even though the data needs to be labeled accurately for this method to work, supervised learning is extremely powerful when used in the right circumstances.

ii. Unsupervised Learning - This is a type of algorithm that learns patterns from untagged data. This type of learning does not have labels to work from, resulting in the creation of hidden structures. Relationships between data points are perceived by the algorithm in an abstract manner, with no input required from users.

iii. Reinforcement Learning - This learning directly takes inspiration from how human beings learn from data in their lives. It features an algorithm that improves upon itself and learns from new situations using a trial-and-error method.

We have chosen three supervised learning algorithms (MLP, CNN and LSTM) and one (ARIMA) statistical algorithm. We discuss further detail about these algorithms in Chapter 3.

1.2.2 D3.js

D3 (d3js.org) is a JavaScript library for manipulating web documents based on data. It creates visualizations by binding the data and graphical elements to the Document Object Model and eventually produce dynamic and interactive data visualizations in web browsers with the help of standard web technologies like HTML, CSS, SVG. The visualizations developed in this thesis were all created using the D3 visualization library.

1.2.3 Uncertainty

Uncertainty is an essential part of life and is defined by lack of sureness or certainty in data. The lack of certainty is a state of limited knowledge where it is impossible to exactly describe the existing state or a future outcome. In practice, uncertainty is a complex concept and there are many kinds of uncertainty that decision makers must face. It covers a broad range of concepts like inconsistency, doubtfulness, reliability, inaccuracy, or error (unknown or not quantified error). Hence, it is difficult to give a precise and generally accepted definition of uncertainty [45].

Uncertainty describes a comparison that can often be understood visually, such as the difference between surfaces generated using different techniques, or a range of values that a surface might fall in. A simple approach to the visualization of this type of information is a side-by-side comparison of data sets [48].

Different types of uncertainty also result in differing interpretations and misinterpretations and so different people perceive and explain it differently. For example, participants in a survey [53] used phrases like 'imperfect knowledge,' 'inadequate information' and 'lack of absolute knowledge' to describe uncertainty. Some participants saw uncertainty as a condition when the probability of something is not 1.0, or when more than one event could happen, this was also considered uncertainty. One participant articulated this as a 'partial belief' in something. In general, uncertainty is understood as a composition of different concepts [18] such as:

- error – outlier or deviation from a true value,
- imprecision – resolution of a value compared to the needed resolution (e.g. values are highly accurately given for countries but are needed for states),
- accuracy – size of the interval a value lies in,
- lineage – source of the data (e.g. at first hand or at second hand)
- subjectivity – degree of subjective influence in the data,
- non-specificity – lack of distinctions for objects (e.g. an attribute value is known to be one of several alternatives but not which one)
- noise – undesired background influence.

Data uncertainty is the degree to which it is inaccurate, imprecise, or unreliable. It can come from source (e.g.: data provider), data lineage (e.g.: from calculation), noise (e.g.: inaccurate post in social media), abnormalities (e.g.: two sources give different values) to name a few. In this work we are considering only the uncertainties calculated from machine learning model predictions.

1.2.4 Streamgraphs

Stream graphs are an approach to visualization which are suitable for displaying high-volume datasets, to discover shapes, trends, and patterns over time across a wide range of numerical groups side by side. For example, seasonal peaks in the stream shape can suggest a periodic pattern. They often work better when there is an interactive component involved that enables the following of each separate "flow" or allows filtering the view in some way.

Figure-1 shows a streamgraph prototype of number of movies for the period of time. We see a side by side comparison of the number of movies among seven countries for the duration of

1900 to 2000. It's representing individual values through time by providing a continuous 'flow' from one temporal point to the next.

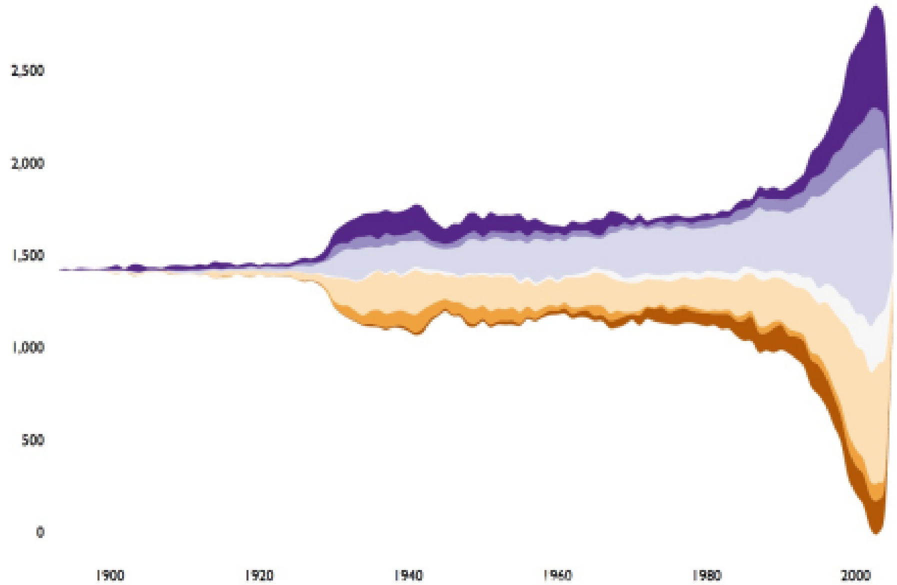


Figure-1.1: Streamgraph prototype [58]

1.2.5 Texture

Texture is the perceived surface quality. It can be used in the analysis of images or charts in several ways: in the segmentation of scenes into distinct objects and regions, in the classification or recognition of surface materials, and in the computation of surface shape [25]. It has been studied extensively in the field of computer vision, computer graphics, and modeling the low-level human visual system in cognitive psychology. Researchers have used different methods to study the perceptual features inherent in a texture pattern [22, 56].

In the field of visualization, people have studied methods for using texture patterns to display information. Although different research groups have concentrated on different tasks, it is advantageous to consider the interdisciplinary integration of these research efforts and apply it in new areas, e.g., data visualization [57]. Textures can be generated in different ways but since our research work is implemented in web (to facilitate remote evaluation), we have used the JavaScript and CSS driven textures called SVG patterns. The SVG <pattern> element allows us to define patterns inside of our SVG markup and use those patterns as a fill. Each pattern has specific shape and we have mostly used circle and rectangle pattern to represent our texture. We will further discuss the generation procedure and algorithm in chapter 3.

1.2.6 Chromatic Aberration

Chromatic aberration is a color distortion or alteration that is sometimes seen on high contrast edges of objects in photographs. Since different colors of light refract to different angles upon traveling through materials with refractive indices [9] (see Figure 1.2), the resulting images may appear to be distorted [10]. It happens when the light of certain wavelengths becomes relatively bent. It usually appears in the form of purple, red, blue, cyan, green fringes. It can be seen alongside deep contrast edges and traditionally it means finding colors where they should not be or found in an unexpected form of color.

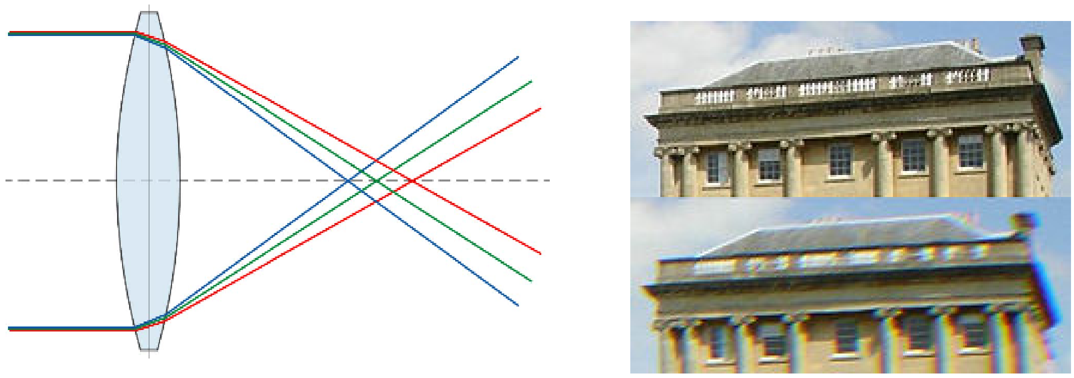


Figure 1.2: Left - different colors of light refract to different angles [10], Right – example of chromatic aberration due to poor quality lens (from wikipedia.com).

CA is a phenomenon that can cause image distortions when viewed through lenses. Since light of various colors refract at various angles on traveling through materials with refractive indices (Figure 1.2-left), the resulting images may appear to be distorted (right). Since more and more people undergo impaired vision due myopia or astigmatism, the usage of corrective lenses increases, making more people vulnerable to this type of visual distortion.

CA is an image quality problem so most of the research surrounding CA are conducted to fix the problem and improve image quality thereby. On the other hand, uncertainty is the problem of data quality and relevant research are conducted mostly regarding reducing it to improve data certainty. But existing research conducted to visualize uncertainty is done with traditional approaches such as glyphs. Since our goal is neither to improve image quality nor data quality, we borrowed the term CA for our research to represent uncertainty as a novel approach in the field of visualization.

Typically displays use three colors (RGB) of light, because it provides a convenient conversion process between human color vision and the color space and hence it creates a very special phenomenon where the misperception comes from aberration of three distinct lights [10]. Conforming to the aberration formation concept, we have chosen three color (RGB) channels to form blended shapes (circle, rectangle, etc.) where each of the 3 component colors are internally laterally shifted from each other by the amount of uncertainty.

1.3 Problem statement

The primary objective of this research is to present and evaluate a novel concept of employing CA to represent uncertainties. For our test case we use uncertainty values generated from predictive machine learning algorithms by amassing and feeding the COVID-19 data into the models. We hypothesized that our proposed system would potentially offer a more effective means of visualizing this type of information.

To implement the system, we needed to consider the following aspects:

i. How to generate realistic uncertainty data?
ii. Which platform or framework was to be chosen to implement the visualization?
iii. What is the design process of representing uncertainty with CA?
iv. How to evaluate CA representation?
v. What is the applicability of this representation?

Considering the above aspects, we have chosen to use recent WHO authorized COVID data to feed into four machine learning predictive models and one statistical model to obtain forecasted results for a certain period [3, 6]. Then calculated uncertainties from the predicted results and those are depicted as CA in D3 based visualizations as well as existing alternative options such blur, noise, and palette-based uncertainty visualizations [35]. We conduct a comparative user study and conduct numerical analysis to assess the effectiveness of our novel design of uncertainty representation with CA. The survey is conducted online given potential issues with in-person contact during the pandemic.

1.4 Approach

At the first step we sought a suitable dataset in terms of completeness and accuracy. By analyzing numerous data repositories, we determined that the WHO approved OWID dataset is the most comprehensive one among all others.

Secondly, we had to study an extensive set of existing work about forecasting from temporal data using machine learning models and chose four popular modeling algorithms for our research. Since finding and comparing the effectives of algorithms' is out of our scope of work, we simply chose a reasonable set of the models because we needed to generate the uncertainty data for the countries by using the predictive models.

Thirdly, having the data generation component in python, we needed to write APIs to connect and pull the data when drawing the charts. Since the model training and data generation for all countries are long running processes, we precompiled the models to generate the data and stored the data into json files so that they can be input readily and sent back to the client on demand.

Fourthly, we have chosen D3.js as our front-end library for drawing the charts because it is an efficient platform for visualization prototyping and widely used. Since developing the basic drawing algorithms is not our goal, we relied on the existing library features but the aggregate data collection, preparation, manipulation, correction and drawing algorithms were developed specifically for this thesis.

Fifthly, we conducted an experiment to evaluate the approach approved by the Research Ethics Board (REB) of Dalhousie University and with the participation of the members of the community.

Finally, in we conducted a numerical analysis and offer a discussion on the survey responses and compare alternative perspectives of reference studies to consolidate and explore the research outcomes.

1.5 Thesis Outline

The remainder of this thesis is organized as follows. In **chapter 2**, we review the relevant literature on Predictive Machine Learning Models, Texture, Uncertainty, and CA. The literature review is subdivided into several sub-sections based on the contents. **Chapter 3** presents data collection, processing, introducing predictive machine learning algorithms and the necessary arrangements to setup models, a brief description of time series forecasting, and snapshots of uncertainty data. **Chapter 4** focuses on visualization component calculations, methods, background architecture, examples of CA with different shapes, techniques and algorithms of pattern and texture generation. **Chapter 5** describes experimental designs with Chromatic Aberrations and Texture Patterns. **Chapter 6** explains the user study design and administering procedure. It introducing the technology used, study components, the counter balancing mechanism, the recruitment criteria and hiring procedure, the color blindness test, questionnaire formation process, and finally, the data collection and storing mechanism. **Chapter 7** shows results obtained from the user study and its numerical analysis for evaluation. Quantitative analyses are conducted with statistical methodologies for questionnaire results, SUS results and NASA-TLX results. Finally, in **Chapter 8**, we pointed out the thesis outcome as a conclusion, and suggested potential directions of future work and associated improvements.

Chapter 2

Literature Review

2.1 Introduction

This study involves three major components i. Generate time series forecasted data from COVID-19 data using four machine learning predictive models, ii. Calculate corresponding uncertainties for different countries and visualize uncertainties in terms of Chromatic Aberration (CA) in a graphical presentation, and iii. Conduct user studies to evaluate user perceptions and applicability with commonly used visualizations. In this section, we are going to discuss related studies of each component separately conforming to the aspects of the research.

2.2 Overview

We stated three major components in the previous section where every component is a significant research area on its own domain. Since our research work is novel, there is no fully connected prior work but there are several component-based connections. For example: there are many research which were conducted only for time series forecasting from machine learning models. But of the research conducted regarding Chromatic Aberration they are not related to time series forecasting and many uncertainty visualizations techniques are not related with the other related areas. Since our research encompasses all those aspects, we consider them closely related to our research. Hence, to facilitate better understanding of literature reviews, we categorize the entire scope of prior work into the following four technical areas:

i. Prediction in Machine Learning Models
ii. Chromatic Aberration
iii. Texture
iv. Uncertainty

We then organize the reviews in the same order for better readability and consistency.

We have gained inspiration and technical ideas from each of these areas of research but to investigate the efficacy of our work, we also needed to find research against which we would compare Chromatic Aberration for visualizing uncertainty. The recent state-of-the-art technique for uncertainty visualization is Value-Suppressing Uncertainty Palettes (VSUP) [35] which we will discuss in some detail at the end of the above-mentioned four categories.

2.2.1 Prediction in Machine Learning Models

Related to model predictions, Song et. al. [1] compiled monthly data of influenza incidences from all provinces in mainland China from January 2004 to December 2011, comprehensively evaluated and classified these data, and then randomly selected 4 provinces with higher, median and lower incidences, using time series analysis to construct an ARIMA model. The same model but different analysis and forecasting approaches was conducted on the coronavirus disease by other researchers [2] but they conducted their research only with a statistical ARIMA model where they suspect it may perform poorly in case of nonlinear trends. Recent studies of [3, 4] use Facebook's Prophet Forecasting Model and ARIMA Forecasting Model to compare their performance and accuracy on the dataset containing the confirmed cases, deaths, and recovered numbers, obtained from the Kaggle website. The forecast models are then compared to the last 2 weeks of the actual data to measure their performance against each other. The result shows that Prophet generally outperforms ARIMA. Researchers in [1, 4, 6] used different versions of ARIMA such as ARMA, SARIMA, PROPHET models to conduct time series analysis but have not used any machine learning or deep learning algorithms to compare with. In [5] researchers have formulated a model of the XGBoost machine learning algorithm for cholera epidemics predictions linked with weather variable, but we noticed that they have not worked with real world data from health-care systems.

Several neural network predictive models were also used to evaluate their performance against more common machine learning models in a Dengue forecasting project [7] and suggested for further work to obtain more competitive results with more sophisticated neural network architectures. They concluded that neural network models (MLP, LSTM, GRU) significantly outperforms traditional machine learning models but they have not given analysis or background reasoning and no indication of if they tried with optimal hyperparameter settings, since they play a key role in such modeling. A decision-supporting tool [8] for medical centers and health-care services has been proposed for influenza prediction with limited data for Belgium which could be tested with more sophisticated and diverse dataset and the similar issue noticed in [9] where they conducted their study on performance evaluation of prediction of machine learning models of five different classifiers (Naïve Bayes, logistic regression, support vector machines, Random Forest, K Nearest Neighbour) with liver disease by taking some sample data. According to their findings, the Logistic Regression classifier demonstrated its ability in foreseeing with the best outcomes, regarding precision and least execution time. But since recent neural network models outperforms traditional machine learning models [7], there is wide scope to test the performance of Logistic Regression against neural network models.

2.2.2 Chromatic Aberration

From a vision perspective, chromatic aberration leads to various forms of color imperfections in the image. When tampering with an image, these aberrations are often disturbed and fail to be consistent across the image. Koh et. al. [10] presented a user study to observe the effect on users' judgment with Lateral Chromatic Aberration (LCA) for Chart Reading in Information Visualization on Display Devices and suggested guidelines for information visualization designers to avoid such issues. LCA occurs when the lens does not focus all lights with different wavelengths to the same convergent point. Although the effect can be observed from natural scenes, they focus on LCA on modern display devices, and they present a series of controlled user experiments to show how people can misjudge information due to LCA. Although humans can compensate for the error especially with monochromatic aberration, the ability to correct errors caused by polychromatic aberration is still limited. There is an open task to investigate different degrees of aberration. A quantitative prediction on the amount of aberration depending on the wavelength and the power of eyeglasses will let us estimate the threshold on which viewers start to misinterpret the chart. We noticed that the limitation of this study is that the solution works only with certain eyeglasses and some common objects and does not extend to a real and generalised environment.

Colour is widely used in information visualisation to deliver different types of information such as extreme values, patterns and attribute values. Colour coding is known to be a particularly effective way to represent extreme values for human viewers due to the nature of pre-attentive vision. Therefore, Hyun Seung Yoo et. al. [11] aim to identify appropriate interventions and propose design guidelines for information visualisation, especially in applications where size judgement is critical. The colour size illusion was replicated on an LCD monitor, revealing that yellow images appeared the smallest among a series of red, yellow, green and blue images on a white background. They address the use of image warping to reduce the illusion effect but without calibrating the model for different zoom/focus level, displacement, and deformation. Their proposition investigated with only limited number of domain experts. As per our understanding, only lateral effect with image warping is considered in a proposed system [12, 13] to resolve such problems but not considered for longitudinal, geometric, or other forms of optical distortions.

Lens flare is an effect caused by light passing through a photographic lens in any other way than the one intended by design. In the paper [14] Matthias Hullin et al. present a novel method to interactively compute physically plausible flare renderings for photographic lenses where underlying model covers many components that are important for realism, such as imperfections, chromatic and geometric lens aberrations, and anti-reflective lens coatings. They mention a common problem arises when triangles become smaller than one pixel and their rendering mechanism is limited to single light sources and undergoes rasterization aliasing effects.

Real cameras have an aperture through which light falls on an image plane containing receptors to register an image. For a sharp image, a small aperture is preferable, but then less light would hit these sensors and diffraction becomes an issue. Sungkil Lee et al. [15] present a novel rendering system for defocus blur and lens effects. The efficient solution achieved by approximating the image-capturing process by considering not only aperture but also aspects of the lens interaction itself. They approximate optical aberrations, which is a unique feature for real-time approaches, and sometimes considered as crucial for realism. More precisely, the major contributions of the paper are: i an efficient algorithm for DOF and lens blur effects, ii. an interactive and intuitive focus control system and, iii. a generalized method for expressive DOF rendering. They argue that combining their approach with single-pass depth peeling can be an interesting avenue for future work and mentioned single-pass decomposition of their depth peeling is slower, but their cache-efficient ray tracing mechanism helps to achieve better quality with a strong speedup.

One of the interesting research projects conducted by Micah K. Johnson et al. [13] shows that inconsistencies in lateral chromatic aberration can be used to detect tampering in visually plausible forgeries. They describe a computational technique for automatically estimating lateral chromatic aberration and show the efficacy of the approach for detecting digital tampering in synthetic and real images. They considered only lateral chromatic aberration for their study where the lateral aberration can be modeled as an expansion/contraction of the color channels with respect to one another. When tampering with an image, these aberrations are often disturbed and fail to be consistent throughout the image. They describe a computational technique for automatically estimating lateral chromatic aberration and show its efficacy in detecting digital tampering. Then mention, the current approach only considers lateral chromatic aberrations that are well approximated with a low-parameter model based on maximizing the mutual information between color channels. They expect future models to incorporate longitudinal chromatic aberrations and other forms of optical distortions.

2.2.3 Texture

Some of our early experiments in visualization designs involved textures. So, we also discuss aspects of textures in this section.

Particle Tracing and Line Integral Convolution (LIC) in Netzel et al. [22] are parallelly and independently used on every pixel of the texture to reduce the computational cost. On top of that a Gaussian low-pass filter with sparse input noise is used for phase shifting along the streamlines. But there is no indication of how high pass filter and/or variable input noise impacts on the result and performance in terms computation and rendering. Streamline computations were replaced by texture advection that works well for both steady and unsteady flow and provides extremely quick results. But according to the authors, the disadvantage of this setup is coupling exponential filter that cannot handle trends properly.

Existing techniques are not capable of accurately aligning and tracking dynamic time-varying data because of the segmentation problem, key feature identification or absence of overlap in consecutive timestep. So, Caban et al. [23] introduces a texture-based feature tracking technique capable of tracking multiple features over time by analyzing local textural properties and finding correspondent properties from synthetic and real-world time varying volumetric data. The main limitation specified in the paper is the cumulative error issue that is caused from the “drifting problem” which exists when small errors are introduced to the texture-based multi-dimensional feature vector over time.

The authors Bachthaler et al. [24] have introduced a new technique of utilising the overlay of two different LIC (line integral convolution) textures to combine the visualization of the tangential and orthogonal vector fields. They have applied a weaving of high-frequency spatial textures of different colors and avoided avoid a direct color blending for compositing. Different filter kernels and filter methods are compared and discussed in terms of visualization quality and speed to obtain a consistent and temporally coherent animation. A perception study was carried out to measure the discrimination and perceived speed of moving patterns under realistic settings. Also, they noticed, it doesn't support higher dimensions and a more refined investigation is required to quantify the effectiveness of conveying flow structures.

To avoid color blurring and inconsistencies in the popular Line Integral Convolution (LIC) scheme and mitigate the expensive computation or memory cost Huang et al. [25] have introduced a novel image-space surface flow visualization approach that preserves the coherence during user interactions. They have employed a precomputed sequence of triangle textures on coordinates of each vertex to ensure noise textures under different viewpoints remain coherent. Although the approach works fine for most models to mitigate expensive computation, memory cost but suffers with popping artifacts (too far/close viewpoint), the authors say.

Kratz et al. [26] have presented a method for the generation of anisotropic sample distributions in the planar and the two-manifold domains. They also presented interactive rendering of anisotropic Voronoi cells. They have used a special sampling approach to generate sample distributions that cover the underlying domain densely while significant holes and cluttered areas are avoided. They use quadratic textures as GPU data structures, which results in some redundant storage that consumes higher memory than it should be required. They state that the most time-consuming step during initial sampling and relaxation in the two-manifold domain is the back-projection and the influence of imposing noise to the cell boundaries are not well tested in their experiment but plan to address this in the future.

To improve the use of color in combination with motion where the author Weiskopf [27] has distinguished between the detection of patterns in motion (seeing the existence) and the actual perception of motion (recognizing speed and direction). It discussed on how calibration is needed to represent data by the perceived speeds of colored patterns and demonstrated how the guidelines of design of animated graphics and the calibration approach can be used. Finally, they have mentioned several of possible future works, firstly - user studies could be conducted to test the proposed guidelines for various application scenarios, secondly - evaluate the calibration process in more detail by statistically significant user tests, and thirdly - address specific combinations of chromatic motion and further perceptual features like texture. They have kept open for future research conducting a study on the proposed guidelines for applications in visualization and computer graphics and to evaluate the calibration process in more detail by statistically significant user tests. Another line of future research could address specific combinations of chromatic motion and further perceptual features like texture.

Healey et al. [28] presents a new method for using texture to visualize multidimensional data elements arranged on an underlying three-dimensional height field. Perceptual texture elements are built by controlling three separate texture dimensions: height, density, and regularity. They present a method for combining three texture dimensions (height, regularity, and density) to form perceptual texture elements (or pexels) but did not investigate the effectiveness of orientation for encoding information, and the interactions that occur when multiple texture and color dimensions are displayed simultaneously. Experimental and real-world results showed that the pexels can be used to rapidly, accurately, and effortlessly analyze large, multi-element displays. They mention that sufficient care is needed to ensure that the data to texture mapping builds upon the fundamental workings of the low-level human visual system. Otherwise, an ad-hoc mapping will often introduce visual artifacts that actively interfere with a user's ability to perform their visual analysis tasks.

2.2.4 Uncertainty

Botchen et al. [29] focuses on uncertainty that occurs during data acquisition and demonstrates the usefulness of the methods for the example of real-world fluid flow data measured with the particle image velocimetry (PIV) technique. They present two novel texture-based techniques to visualize uncertainty in time-dependent 2D flow fields where in the first method, texture advection is employed to show flow direction by streaklines and convey uncertainty by blurring these streaklines and in a second method isotropic diffusion implemented by Gaussian filtering to continuous change of the density of flow representation. Flexibility and generality are two main advantages of this system and hence they can be applied to any density of texture representation ranging from dense noise-based up to sparse dye-based methods. Moreover, these approaches could be directly mapped to GPUs in order to achieve real-time visualization to facilitate interactive user exploration of the flow field. Finally, they think a further extension of the system to 3D flow will be a challenging task.

Error in data is inherent so it cannot be ignored in visualization. Improper or eliminated presentations in visualizations can mislead decision making for data analysts. The goal of uncertainty visualization is to minimize the errors in judgment and represent the information as accurately as possible. This survey Kamal et al. [30] discusses state-of-the-art approaches such as the quantification approach to uncertainty visualization, along with the concept of uncertainty and its sources. They grouped the approaches into two categories: quantification and visualization - to make it easier for the readers to distinguish between different expressing uncertainty methodologies. One of the goals of this survey was to provide insight into the topic of uncertainty visualization by conveying recent accomplishments in the visualisation area and additionally offers a creative forecast of what is to follow in the coming years.

Bonneau et al. [16] explores uncertainty in the visualization domain by comparing different results, such as a weather forecast generated with different parameters and to detect similarities or differences in the results a comparative visualization technique is employed. To compare certain regions in more detail, e.g., borders, they suggested to consider larger comparison areas than individual pixels and it is crucial that data sets which should be compared are visualized next to each other to get a direct comparison for a certain area. They mention two open problems. The first is Perceptual and Cognitive Implications - Since visualization often relies heavily on the use of colors to convey information, it can be quite challenging for individuals with color vision deficiency. And the second is Comparative Visualizations – Uncertainty of measurement, simulation, or process provides an additional data stream which generates further visualization challenges. For a dense representation of uncertainty, comparative visualization seems to be a promising emerging area.

Objective uncertainty of a visual system is evaluated by Barthelme et al. [17] where they discuss the natural perceptual systems involvement with systematic uncertainty because sensory information is imperfect and insufficient to uniquely designate the environment. In their experiment, observers were presented with pairs of images of oriented objects embedded in high levels of noise and had to report the orientation of the image of their choice. In their experiment, they compare objective uncertainty (computed using the Bayesian framework) with subjective uncertainty (the confidence observers report about their visual perception). To this end, they used a visual task with well-defined statistical properties, discrimination under noise. They report a surprising degree of agreement between objective and subjective uncertainty and discuss possible computational models that could explain this ability of the visual system. Even though the two images contained the same extent of noise, one particular noise structure made an image orientation more obvious than the other. Eventually, observers reliably chose the more obvious of the two images, thereby providing evidence of a capacity to accurately evaluate objective uncertainty. They conducted the study with a custom noise model and kept open to test with more generalised noise models.

A statement on the position of uncertainty visualization today is explained in Griethe et al. [18] that defines the basic concept of uncertainty and discusses sources and necessary measures. Visualization is an indispensable approach to the exploration and communication of large data sets of different domains where data sets may contain an unavoidable amount of uncertainty that needs to be included in the visualization process to enable the correct cognition of hidden facts and figures. In addition, it explains how existing approaches could be systematically presented to the acquisition and display of uncertainty can be transferred to new fields, e.g., the visualization of uncertainty in structures. Despite impressive advances in the field, they think there is still a great need for further research on how to convey content from different application areas to the user as accurately as possible.

State-of-the-art visualization techniques have been successfully engaged in diagnostic medical imaging and Direct Volume Rendering (DVR) sectors and attained maturity in regular clinical works. However, still a major problem is the lack of information on the uncertainty of the tissue classification, which is addressed in the paper Lundstrom et al. [20] by proposing animation methods to convey uncertainty in the rendering. The rendering is animated by sampling the probability domain over time that allows direct user interaction with the classification and it outperforms traditional rendering in terms of assessment accuracy but still that needs to be studied in real clinical environments, the authors added.

Most of the visualization research has ignored the presentation of uncertainty from data because of the inherent difficulty in defining, characterizing and controlling the uncertainty in the visualization process stated by Pang et al. [21]. They introduced a wide variety of new uncertainty visualization methods like adding glyphs, adding geometry, modifying attributes, modifying geometry, animation and applied to many applications such as environmental visualization, surface interpolation, global illumination with radiosity, flow visualization, and figure animation. The results of the research show that there are a wide variety of possible means to map uncertainty into a scene. The methods presented in the paper represent significant steps toward achieving the goals of uncertainty visualization. They believe that these uncertainty visualization methods will prove valuable to people who need to make informed decisions based on imperfect data.

Being a complex topic, most of the authors try to eliminate the existence of uncertainty from their visualization outcome, so the researcher Hullman conducted a survey and interviewed over 103 visualization authors in [36]. They identified that perceptions, practices, challenges, and attitudes are associated with uncertainty visualization and the majority of them agreed that or at least were sympathetic about the importance of uncertainty communication.

Data analysts also face unique challenges in interpreting the results on applying machine learning and statistical methods to timestamped event sequences to tackle various problems. Through a controlled study, the researchers Guo et. al [37] found that users experience more confidence in making decisions when alternative predictions are displayed alongside uncertainty information, and they consider the alternatives more when deciding between two options with similar top predictions. There are several limitations of this research, for example: they have used darkness to address uncertainty but that is not suitable to determine exact uncertainty values and make accurate decisions, so new research could be conducted by replacing darkness with color or glyphs. Also, it requires the participants to be domain experts and data requires alternatives.

Since uncertainty is a multi-faceted concept, there are various kinds of uncertainties, and the visualization of such uncertainties are applied in many contexts with different objectives, so there may not be optimal uncertainty visualization technique. The study of Korporaal et al. [38] investigates how data uncertainty visualized in maps might influence the process and outcomes of spatial decision-making, especially when made under time pressure in risky situations. According to researchers, the limitation of the research is that they have not considered the effect of stress along with time constraints, they didn't test with experts such as helicopter pilots, it was limited to a cartographic display and neglected the diversity of uncertainty. In addition, they have used only one type of texture (dotted) in their visualization experiment. So, the result cannot be generalized with non-texture, non-color based or gradients.

Earthquake models can produce aftershock forecasts but research on uncertainty visualization is often missing from earthquake science. So, Schneider et al [39] conducted research where three different uncertainty visualizations were produced: (1) forecast and uncertainty maps adjacent to one another; (2) the forecast map depicted in a color scheme, with the uncertainty shown by the transparency of the color; and (3) two maps that showed the lower and upper bounds of the forecast distribution at each location. They mention the limitations of the paper includes: they needed to fix either the forecasted aftershock rate or its uncertainty and in the comparative judgment task, geographical features, such as roads and landmarks were omitted from the maps to avoid potential confounding effects on judgments which lowers the ecological validity of the study.

The authors Brodlie et al. [40] have reviewed the state of the art in uncertainty visualization, looking at both the visualization of uncertainty (which considers how to depict uncertainty specified with the data) and the uncertainty of visualization (which considers how much inaccuracy occurs in data processing through the pipeline of Haber and McNabb uncertainty reference model). They note that the visualization research community has enthusiastically taken up the challenge of uncertainty and most of the popular visualization techniques have been extended in some way to handle uncertain data. In future work, they hope to extend the interactive tool to support more varied types of visualizations and are interested in understanding how the difference between systematic and random error and larger inconsistencies than tested in this experiment influence humans' inference.

When making an inference or comparison with uncertainty, noise, or incomplete data, measurement error and confidence intervals can be as important for judgment as the actual mean values of different groups. One study [41] investigates drawbacks with the standard encoding and considers a set of alternatives and conducted a series of crowd-sourced experiments that confirms the encoding of mean and error significantly changes and by which viewers make decisions about uncertainty. They use gradient plots with transparency to encode uncertainty and violin plots with width as better alternatives. They think, one area is not well-covered by their experimental tasks was decision making and did not collect a great deal of qualitative data such as viewer preferences for different chart types which could be an important consideration for how data are perceived and used, especially for issues of trust and uncertainty.

In daily life, people regularly make decisions based on uncertain data. The authors Greis et al. [42] published a web-based game on Facebook and compared four representations that communicate different amounts of uncertainty information to the user and compared. The results show that an abundance of uncertainty information leads to taking unnecessary risks. However, an absence of uncertainty information reduces the risk taking and leads to more won turns, but with the lowest money gain. Representations with aggregated detailed uncertainty provide a good trade-off between being understandable by the players and encouraging medium risks with high gains. We noticed that the research does not visualize the uncertainties but uses aggregated detailed uncertainty to offer a good compromise between understandability, encouraging educated risks and achieving credible winning criteria with high gains.

In statistics, people usually quantify uncertainty to help determine the accuracy of estimates, yet this crucial piece of information is rarely included on maps visualizing real data estimates. Lucchesi et al. [43] develop and present three approaches to include uncertainty on maps: (1) the bivariate choropleth map repurposed to visualize uncertainty; (2) the pixelation of counties to include values within an estimate's margin of error; and (3) the rotation of a glyph, located at a county's centroid, to represent an estimate's uncertainty. They have not conducted user studies to determine whether these three methods effectively communicate uncertainty. In the approaches of bivariate choropleth map, map pixelation and glyph rotation, they not only visualize uncertainty but also demonstrate why it is so important to include uncertainty on maps displaying real data estimates. They say, although users can see which counties have high uncertainties, they cannot determine the exact quantities of the margins of error by looking at the pixelated map and have not conducted study to determine whether the methods effectively communicate uncertainty.

Uncertainty is a fact of information; many types of information contain uncertainty, usually of heterogeneous categories. While there have been many calls for research about uncertainty visualization, the understanding of when and why one uncertainty visualization strategy should be used over others remains incomplete. To address the gap MacEachren el al. [44] presents two linked conceptual perspectives focused on uncertainty visualization. First, a typology of uncertainty is used to delineate kinds of uncertainty matched with space, time, and attribute components of data. Second, concepts from visual semiotics are applied to representing different categories of uncertainty. They address representation intuitiveness and relative performance, considering visual variables and iconic representations of uncertainty. The study does not cover both data and uncertainty at the same symbol and didn't test the impact of symbol size.

Many information fusion applications process and present huge quantities of data to enable an operator to make effective decisions. Reveiro [45] provides a general overview on uncertainty representations techniques and explains why the recognition of uncertainty plays an important role in decision making. In addition, it suggests the techniques developed in information visualization can be applied in information fusion and outlines how information fusion research might proceed further. The major contributions of this paper are (1) to highlight the importance of uncertainty visualization in decision-making, (2) to briefly review relevant modern uncertainty visualization techniques, (3) to propose general theories and results of user experiments for their theoretical analysis, (4) to suggest that techniques developed in information visualization can be applied in information fusion and (5) to outline how information fusion research might proceed further. The evaluation of uncertainty visualization techniques should include theoretical cognitive/perceptual analysis and usability tests with actual users. In this paper, a small set of cognitive and perceptual theories as well as empirical studies documented in the literature have been presented. They can be used in theoretical evaluations of existing and newly developed techniques for displaying uncertain information, and providing insights into their weakness and strengths. We think future empirical research is needed regarding the validity of the given theoretical guidelines, to cross check the validity of the presented preliminary study.

Visual representations of information are challenged to incorporate uncertainty because the factors that influence the uncertainty of information vary with the type of information. Visualization researchers have no abstract model or framework for describing and constructing visualizations of uncertainty as it relates to intelligence analysis. The paper [46] of Judi Thomson presents a typology describing the aspects of uncertainty related to intelligence analysis, drawing on existing frameworks for uncertainty representation. They do not conduct any uncertainty visualization work themselves but organize the uncertainties into a logical framework or typology and then explores frameworks for uncertainty that have been developed for representation within the geosciences and the scientific visualization community.

Instead of professional data scientists, the authors Boukhelifa et al. [47] engage domain experts with varying skill levels to find pertinent patterns and build a new uncertainty-aware sensemaking model. They describe their various coping strategies to understand, minimise, exploit, or even ignore the uncertainty influenced by accepted domain practices, but it appears to depend on the types and sources of uncertainty. They noted one limitation as participants of the study have different technical skill levels which may have had an impact on their behaviour and coping strategies and their recruitment scheme may have introduced potential bias due to snowball and social network effects.

Evaluating the impact of an uncertainty visualization is complex due to the challenge of defining correct behavior with uncertainty information and difficulties of interpreting uncertainty by people. Hullman et al. [48] present a taxonomy of methods for evaluating uncertainty visualizations and describe the results of a qualitative analysis applying their own framework to 86 publications which represent the state of uncertainty visualization evaluation. The taxonomy differentiates six levels of decisions that comprise an uncertainty visualization evaluation: the behavioral targets of the study, expected effects from an uncertainty visualization, evaluation goals, measures, elicitation techniques, and analysis approaches. They characterize overall trends in evaluation paths of uncertainty visualization which indicate distinctions between methods for measuring accuracy and decision, as well as different methods for eliciting and assessing subjective confidence. They recommend specific steps that researchers should take when designing uncertainty visualization evaluations to strive for valid and transparent findings.

Understanding how effectively to display uncertain information has become increasingly important because uncertain information can be shown in many formats ranging from simple text to graphical representations. The paper [49] describes two studies in which degraded or blended icons were used to convey uncertainty regarding the identity of a radar contact as hostile or friendly. A classification study first showed that participants could sort, order and rank icons from five sets intended to represent different levels of uncertainty. Contacts and probabilistic estimates of their identities were depicted on a simulated radar screen in one of three ways: with degraded icons and probabilities, with non-degraded icons and probabilities and with degraded icons only. Results showed that participants using displays with only degraded icons performed better, which means the presence of numeric probabilities did not provide a statistically significant advantage in this task. Future research can be conducted to determine the suitability of the display techniques across different and more realistic task situations such as defence applications. The authors specified limitation of the paper is the uses of icons in combination with numerical probabilities causes decision-makers to hesitate and they expect more assistive information.

Since many visual depictions of probability distributions, such as error bars are difficult for users to accurately interpret, the authors Hullman et al. present a study [50] of an alternative representation, Hypothetical Outcome Plots (HOPs). In contrast to the many static representations of distributions, HOPs require relatively little background knowledge to interpret. Results showed that with HOPs, users made more accurate judgments than error bars and violin plots. Authors suspect that viewers of HOPs could make even more accurate probability hypothesis if provided with interactive graphical annotations. They noted limitations of the paper as: i. they have not tested all abstract, static and special purpose representations of concrete outcomes, ii. they did not ask subjects to explain their conclusions about data and uncertainty and they know relatively little about the subject pool.

Authors Kay et al. [51] present a novel mobile interface design and visualization of uncertainty for transit predictions on mobile phones based on discrete outcomes. To develop it, they identified domain specific design requirements for visualizing uncertainty in transit prediction through 1) a literature review, 2) a survey of users of a popular real-time transit application, and 3) an iterative design process. In a controlled experiment they found that quantile dotplots reduce the variance of probabilistic estimates by ~1.15 times compared to density plots and facilitate more confident estimation by end-users in the context of real-time transit prediction scenarios. Fernandes et al. [52] noticed that when using uncertainty displays, decision quality may ameliorate over time. In the real world, bus riders decide to leave for a bus using a real-time transit prediction application and everyone's utility function remains personal and changes according to each situation dynamically. But participants of their studies use the same utility functions for all which may make people feel complicit in bad decisions leading to missing buses. Respondents gave mixed opinions about the usefulness of the uncertainty information provided by the app and so future work is necessary to see how widespread such reactions may be in real-world deployments. They both suggested that the presented designs should be evaluated in longitudinal field studies to assess actual acceptability and use.

By developing ways to include uncertainty in traditional information visualizations, we can provide more accurate depictions of critical data sets so that people can make more informed and accurate decisions. Skeels [53] reviewed existing work from several domains on uncertainty and created a classification of uncertainty based on the literature. They empirically evaluated and improved upon their classification by conducting interviews with participants from several domains. Their classification better describes the broad range of uncertainty across domains and provides a structure for more readily understandable uncertainty visualization. One of the most promising aspects of their classification is the concept of 'layers' of uncertainty that add complexity to data and is not simple to conceptualize or convey with current techniques and due to complexity it is kept as open task to visualize. They suggest a lab study to evaluate how well a sense of certainty is conveyed, but more relevant and rich data would probably be discovered by a study of real-world situations when people use data that matters to them. In addition, one of the most promising aspects of their classification for uncertainty is the concept of 'layers' of uncertainty where layers add complexity to data that is not simple to convey with current techniques, so it opens up opportunities for visualization.

Inherent uncertainties from environmental data (e.g., Meteorological stations and doppler radars, etc.) is often omitted from visualization. The authors Whittenbrink et al. [54] showed scientific data collected from different sources, derived uncertainty information, and presented some ideas on designing uncertainty vector glyphs. They have developed a new vector glyph to visualize uncertainty in winds and ocean currents. Their approach is to include uncertainty in direction and magnitude, as well as the mean direction and length, in vector glyph plots. They defined visualization overloading and verity visualization, illustrating how their new glyphs represent the latter. They use both quantitative and qualitative methods to compare their glyphs showing they are superior to traditional ones in terms of uses because of their ease of understanding and information presentation. Since no single technique of visualization works best for all data and applications, experiments are still needed to identify effective visualization strategies that work well in a given context.

2.3 VSUP: a Comparator Study

Both uncertainty visualization and understanding uncertainty are complex and critical tasks. One of the most common approaches of uncertainty visualisation is to encode data values and uncertainty values independently, using two visual variables in a bivariate map. These resulting bivariate maps can be difficult to interpret, and the discriminability of marks can be reduced due to the interference between visual channels. To address this issue, Correl et al. [35] introduces Value-Suppressing Uncertainty Palettes (VSUPs) as in Figure 2.1 (right) whereas a traditional bivariate map shown in Figure 2.1 (left). We highlight this prior work as it is the comparator approach in our user study.

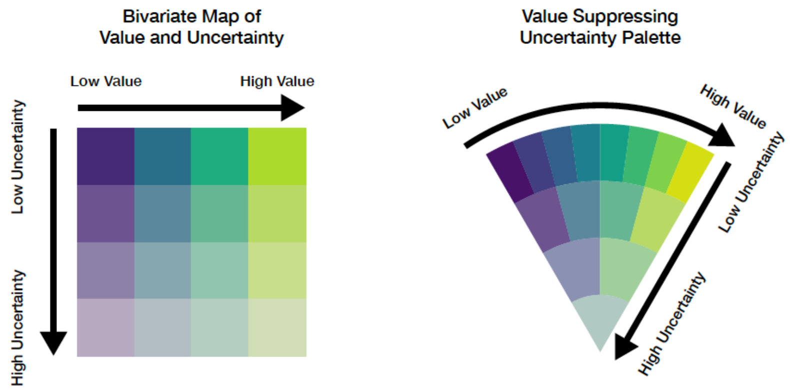


Figure 2.1: A standard bivariate map (left) and a VSUP (right) [35]

We see that VSUPs allocates smaller ranges of the visual channel to data when uncertainty is high and larger ranges when uncertainty is low. This allocation of visual variables promotes patterns of decision-making that make efficient use of uncertainty information, discouraging comparison of values in unreliable regions of the data, and promoting comparison in regions of high certainty. In traditional bivariate maps Figure 2.1(left), outputs for each combination of value and uncertainty might be represented as a 2D square whereas VSUP approaches it as arcs mapping larger number of outputs for smaller and smaller sets of outputs for higher uncertainty.

But one the main requirement of that research is they filter out higher uncertainty values by grouping them altogether which suppresses the values for decision making when uncertainties are high. Due to this higher uncertainty elimination aspect the designers need to carefully consider if this representation is suitable and desirable for certain systems. Another limitation is, since both uncertainty and value are represented by a single color, the perceptual non-separability of color channels are well-known, and which requires the concept of a limited "budget" of distinguishable marks [35]. To achieve the limited budget criteria, it necessitates one to quantize the data. Due to the data quantization, uncertainty visualisation for continuous (or all discrete) values are not possible with limited color budgets.

In our research, we represent uncertainty with Chromatic Aberration (CA) and for the quantitative test we denote uncertainty as quantized values in terms of CA where in VSUP they keep it as uncertainty. Values are used in the same way in both cases to represent the counts (e.g.: number of occurrences). So, in VSUP they represent a specific color by using the combination of value and uncertainty. On the other hand, in our approach the value is represented by a certain color for the target cell's centre fill color and CA represents the outer edge thickness based on the underlying shifting of chromatic circles. So, our research is not to fill any gap of VSUP but to compare the relative usability of two approaches of representing uncertainty.

2.4 Evaluation of Visualization Systems

Visualization evaluation is a complex task since it involves complex data structures or patterns, or it can exhibit various interconnected information. Researchers and practitioners in this field have faced many challenges in different phases when designing, planning, conducting, and executing an evaluation of a visualization systems. It can be a difficult task for an evaluator to design suitable evaluation questions to ask the participants, to pick the right variables from visualization artifacts, decide and develop an efficient way to test data sets and pick the proper methods of evaluation. Existing literature guidelines can help to solve these problems for example, Heidi et al. [64] present seven scenarios of information visualisation from their research that includes evaluating visual data analysis and reasoning, evaluating user performance, evaluating user experience, evaluating environments and work practices, evaluating communication through visualisation, evaluating visualisation algorithms, and evaluating collaborative data analysis. They suggest different approaches to reaching decisions about what could be most effective evaluation of a given visualisation system. Among them some of them are related our study, so we briefly discuss about those in the following section.

User Performance

User performance is mainly measured in terms of objectively measurable metrics such as time, error or accuracy rate, or work quality but the task completion time or task completion accuracy is commonly used. Output of the tasks are generally numerical values analyzed using descriptive statistics such as mean, median or standard deviations. They can also come from the user interactions, perception, and cognition for specific types of visually presented techniques. Most widely used methods are controlled experiments or quantitative evaluation. A controlled experiment requires tasks that can be performed by large number of participants in different study sessions. It is not imperative that the participants be domain experts, hence non-experts can also participate in such experiments. To answer evaluation questions with quantitative and statistically significant results, evaluations in the user performance group require high precision. The commonly used methodologies involve an experimental design with only a small number of variables changed between experiment conditions such that the impact of each variable can be measured J. McGrath [66].

User Experience

Evaluation of user experience is taken in the form of participants' subjective feedback and opinions in written, spoken form or online feedback with a set common questionnaire to all participants. It seeks to understand how participants react to the presented visualisation. A visualisation can be an initial design sketch, a basic prototype, a finished product, or part of a complex system. The goal is to understand to what extent the participants' vision can perceive the intended information conveyed by the system such as perceived effectiveness, perceived efficiency and perceived accuracy. Other measures include satisfaction, trust, features liked/disliked, effort required, and time required. The collected data in such a study helps designers to explore gaps and limitations in the visualised system, as well as allow researchers to take necessary steps to enhance it at a further stage. So, the evaluations can be short term to assess current or potential usage and long term to assess the adoption of a visualization in a real usage scenario.

Usability Testing

How participants perform a set of predefined tasks is observed to carry out the usability test. For each session, the evaluators take notes of interesting observed behaviors, suggestions given, comments provided by the participant, lack of understanding, and major problems in interaction. The components of this method are the careful preparation of tasks and feedback material like questionnaires and interview scripts. Its main goal is to improve the design by spotting major flaws and deficiencies in existing prototypes [65]. Nonetheless it can also serve the purpose of eliciting overlooked or missing requirements.

Heidi et al. [64] classified the scenarios into two broad categories called *process* and *visualisation*. The main goal of process group evaluation is to understand the underlying process and the roles played by the visualisations. In contrast, evaluations can focus on the visualisation itself, with the goal to test design decision, explore a design space, benchmark against existing systems, or to discover usability issues. Again Bonneau et al. [16] classified the types of evaluation into three groups:

- Theoretical evaluation - the method is analyzed to see if it follows established graphical design principles,
- Low-level visual evaluation - a psychometric visual user study is performed to evaluate low-level visual effects of the method, and
- Task oriented user study - a cognitive, task-based user study is conducted to assess the efficiency or the usability of the method.

Usually in these evaluations, a part of the visualisation system is tested. In this context, we intend to compare only the concept of Chromatic Aberration against VSUP [35]. To evaluate user experience and performance with the help of a set of questionnaires, we developed an online webpage containing the relevant exercises of visualisation. Instead of asking questions as interviews, we presented everything visually (explained in study design section in chapter 6), so that participants can provide their rating online. For Example: in the system usability test (SUS), we presented 10 questions in 5 scale ratings (1-5) and for Nasa-Tlx we presented 6 questions in 22 scale ratings (1-22).

2.5 Limitations of related works

As stated in the related works section, a plethora of studies have been conducted in these domains, for example: predicting modeling and augmentation of algorithms, time series analyses and comparisons on different diseases and/or on other temporal data, real time predictions from models, measuring chromatic aberration from image distortion, effect of color and light on display devices, uncertainty visualization and decision making, texture analyses and assessments, perceptual textures to represent multi-dimensional dataset, etc. But to our knowledge predictive uncertainty has not been represented with chromatic aberration and so our research evaluates this approach.

Chapter 3

Data Collection, Processing, and Introduction of Models

3.1 Introduction

In this chapter, we discuss the data collection, listing of data properties and snapshot of the collected dataset, data manipulation and pre-processing, model selection and will also provide a brief description of corresponding models, algorithms of model building and training and, generate uncertainty from predicted data, discuss aspects of CA, discuss show our experimental designs and examples of uses of CA in charts.

3.2 Data

Good quality data is an important part in data visualization research. Without having an authentic dataset research cannot be conducted properly and without following a smart data preparation strategy such as cleaning, validating, and consolidating raw data, research cannot succeed in the long run.

3.2.1 Data Collection

Data collection is the process of gathering, measuring, and analyzing accurate information on variables of interest, in an established systematic manner that enables one to conduct research. Due to the global impact of pandemic, different individuals, organizations, or governments are storing data in their own way. After examining different repositories, we found that the most complete data is bundled in ourworldindata.org in the csv format. The following table shows the list of fields/properties of each record where many of them are not relevant to our research. For example: date, location, new_cases, total_cases are some of the useful attributes bolded in the following Table 3.1.

iso_code	continent	**location**
date	**total_cases**	**new_cases**
new_cases_smoothed	**total_deaths**	**new_deaths**
new_deaths_smoothed	total_cases_per_million	new_cases_per_million
new_cases_smoothed_per_million	population_density	new_deaths_per_million
new_deaths_smoothed_per_million	stringency_index	population
new_cases_smoothed_per_million	median_age	aged_65_older
aged_70_older	gdp_per_capita	extreme_poverty
cardiovasc_death_rate	diabetes_prevalence	female_smokers
male_smokers	handwashing_facilities	hospital_beds_per_thousand
life_expectancy	human_development_index	

Table 3.1: COVID Data property list

3.2.2 Sample Data

iso_code	location	date	total_cases	new_cases	total_deaths	new_deaths	new_tests	total_tests	new_vaccinations	population
USA	United States	31/1/21	26249342	112152	449340	1862	943985	309077233	1545397	331002647
USA	United States	1/2/21	26384317	134975	451420	2080	1032022	310109255	1099103	331002647
USA	United States	2/2/21	26499620	115303	454846	3426	1632097	311741352	558458	331002647
USA	United States	3/2/21	26621311	121691	458728	3882	1869029	313610381	1097394	331002647
USA	United States	4/2/21	26745317	124006	462473	3745	1928145	315538526	1325456	331002647
USA	United States	5/2/21	26879739	134422	466138	3665	1829259	317367785	1615502	331002647
USA	United States	6/2/21	26983915	104176	468821	2683	1332964	318700749	2218752	331002647
USA	United States	7/2/21	27073661	89746	470257	1436	828602	319529351	2172973	331002647
USA	United States	8/2/21	27164099	90438	471877	1620	1042164	320571515	1206680	331002647
USA	United States	9/2/21	27259364	95265	474927	3050	1691635	322263150	788573	331002647
USA	United States	10/2/21	27354614	95250	478231	3304	1638478	323901628	1563780	331002647
USA	United States	11/2/21	27460378	105764	481447	3216	1414964	325316592	1620300	331002647
USA	United States	12/2/21	27560048	99670	484374	2927	1411948	326728540	2020288	331002647
USA	United States	13/2/21	27647267	87219	486570	2196	1101360	327829900	2231326	331002647
USA	United States	14/2/21	27712402	65135	487741	1171	663849	328493749	2242472	331002647
USA	United States	17/2/21	27899318	70139	492854	2397	1470698	332013338	1061463	331002647
USA	United States	18/2/21	27969229	69911	495370	2516	1414506	333427844	1455940	331002647
USA	United States	19/2/21	28048511	79282	497994	2624	1353482	334781326	1847276	331002647
USA	United States	20/2/21	28120207	71696	499829	1835	1058252	335839578	1704457	331002647
USA	United States	21/2/21	28177359	57152	501065	1236	683633	336523211	1801134	331002647
USA	United States	22/2/21	28233518	56159	502384	1319	1038714	337561925	1086840	331002647
USA	United States	23/2/21	28305788	72270	504661	2277	1643226	339205151	854609	331002647
USA	United States	24/2/21	28380537	74749	507843	3182	1687287	340892438	1432864	331002647
USA	United States	25/2/21	28458041	77504	510279	2436	1611102	342503540	1809170	331002647
USA	United States	26/2/21	28535390	77349	512357	2078	1475665	343979205	2179947	331002647
USA	United States	27/2/21	28600016	64626	513878	1521	1074248	345053453	2352116	331002647
USA	United States	28/2/21	28651438	51422	514970	1092	633047	345686500	2429823	331002647
USA	United States	1/3/21	28709536	58098	516487	1517	993543	346680043	1663984	331002647
USA	United States	2/3/21	28766634	57098	518430	1943	1611124	348291167	1731614	331002647
USA	United States	3/3/21	28833825	67191	520911	2481	1736631	350027798	1908873	331002647
USA	United States	4/3/21	28901885	68060	522833	1922	1551765	351579563	2032374	331002647
USA	United States	5/3/21	28968304	66419	524636	1803	1430138	353009701	2435246	331002647
USA	United States	6/3/21	29026558	58254	526146	1510	1066184	354075885	2904229	331002647
USA	United States	7/3/21	29067631	41073	526848	702	608815	354684700	2439427	331002647
USA	United States	8/3/21	29112548	44917	527585	737	976094	355660794	1738102	331002647
USA	United States	9/3/21	29170215	57667	529391	1806	1540048	357200842	1602746	331002647

Table 3.2: Screenshot of sample data

In the above Table 3.2, it shows only a snapshot of the whole dataset where there are hundreds of thousands of records for Covid data for more than 237 countries and territories. Though there are numerous fields in the data, we only needed a few of them as listed in the previous section. The dataset is collected as a excel file which includes daily occurances and/or counts of all properties. The total_* fields like total_cases, total_deaths, etc are cumulative and so every day that is updated with the previous day's counts. Data is ordered by date and name of the country correspondingly. If there is no value in a cell for certain date for a country, that cell is empty, so that is needed to be handled during data processing.

3.3 Machine Learning Algorithms

Although we have focused on algorithms in the machine learning domain, it is necessary to briefly introduce the salient algorithms that have been used in our research to process the available data and generate the uncertainties of predictions since uncertainty representation is our prime concern.

3.3.1 Predictive/Forecasting Models

A time series forecasting model comprises a sequence of data points captured, using time as the input parameter. It uses the historical data to develop a numerical metric and predicts values for the next duration, for instance, data for the next few weeks, using that metric.

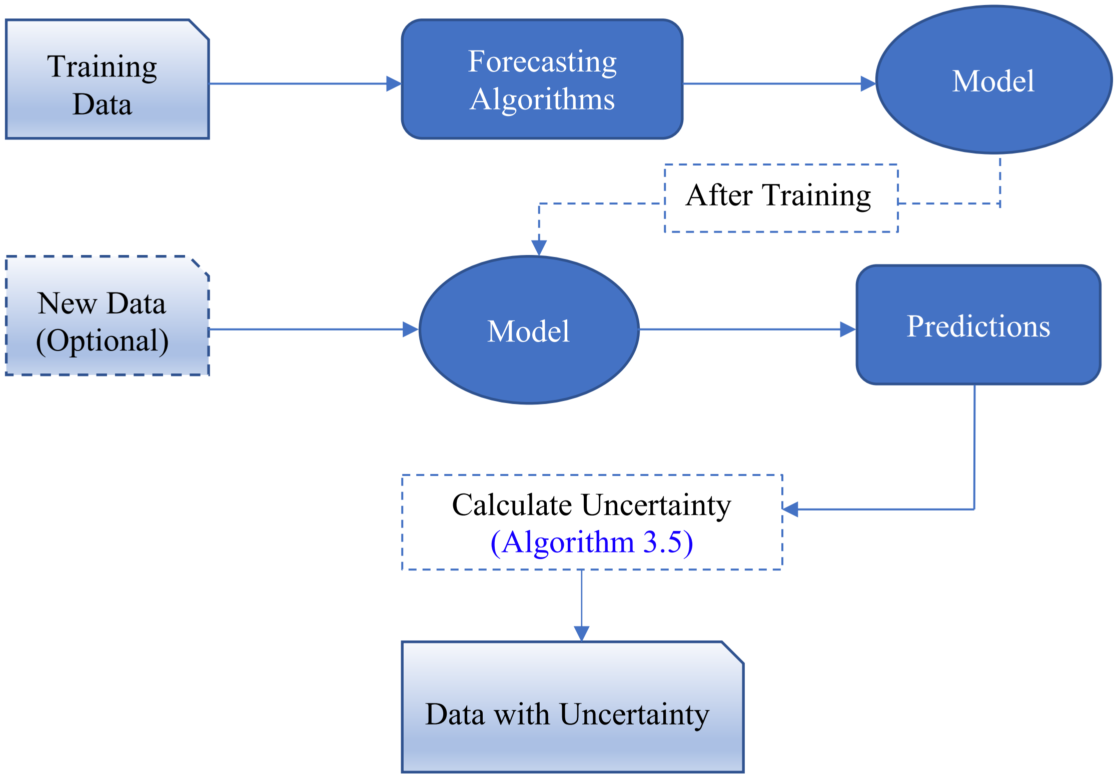


Figure 3.1: Predictive modeling workflow to generate uncertainty

3.3.2 Time Series Analysis vs Forecasting

Sometimes ambiguity arises between time series analysis with time series forecasting when working with temporal data. As per Song et al. [1] and Beneditto el al. [4] in time series analysis, a time series is modeled to determine its components in terms of seasonal patterns, trends, and relation to external factors. In contrast, time series forecasting, Gecili et al. [6] and Brownlee [32] uses the information in a time series (perhaps with additional information) to forecast future values of that series. The COVID-19 dataset is maintained on a global basis, so it is more trustworthy and forecasting models can be considered as suitable for our research to obtain the predicted results and hence generate our required uncertainty data to represent chromatic aberration in visualization area.

3.3.3 Concerns of Forecasting

Time series forecasting is an important area of machine learning. It is important because there are so many prediction problems that involve real life issues with a time component. In forecasting it is very important to understand the goal of the problem and the nature of the available data. For instance, the volume of data, time horizons (short, medium, or long term), and frequency of update an play an important role in forecasting. Sometimes time series data requires cleaning, scaling and even transformation, for example: if there are gaps/missing data, or if there are outliers or corrupt data then those need to be addressed and corrected. Depending on the frequency, a time series can be of yearly (e.g., annual budget), quarterly (e.g., profit), monthly (e.g., cash flow), weekly (e.g., sales quantity), daily (e.g., weather forecast), hourly (e.g., stock market price), minute wise (e.g., calls in a call canter) and even seconds wise (e.g., web traffic). We use the daily forecast mechanism for our research. To compare the results side by side we have created predictions for 200 days from every model and for properties such ‘new_cases’, ‘new_deaths’, 'new_tests', and 'new_vaccinations'.

3.3.4 Example of Forecasting

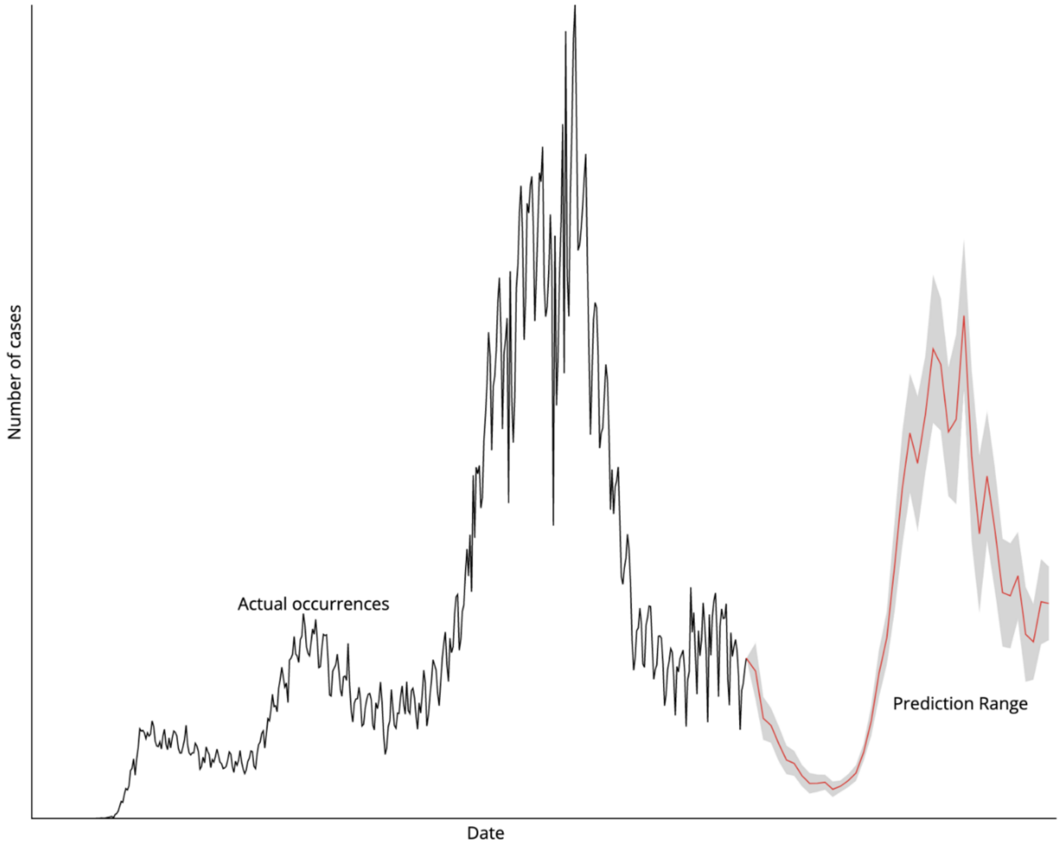


Figure 3.2: Example of daily covid forecasting for 200 days

The above Figure-3.2 shows the daily forecasting of the number of new cases for the United States based on previous statistics. So, the black line on left shows the actual occurrences and the reddish line towards the right shows the predicted number of cases and the greyed background surrounding the predicted line represents the ranges of model prediction, that means the model can predict a value between the lower and upper value for a certain day and that grey area represents the area of uncertainty.

3.4 MLP

The first method we consider is multilayer perceptron (MLP) which is a class of feedforward artificial neural network (ANN) [32]. It is a neural network connecting multiple layers in a directed graph, which means that the signal passes through the nodes only in one direction. It can be used for time series forecasting by taking multiple observations at prior time steps, called lag observations, and using them as input features and predicting one or more-time steps from those observations. The training dataset is therefore a list of samples, where each sample has some number of observations from days prior to the time being forecasted, and the forecast is the next days in the sequence.

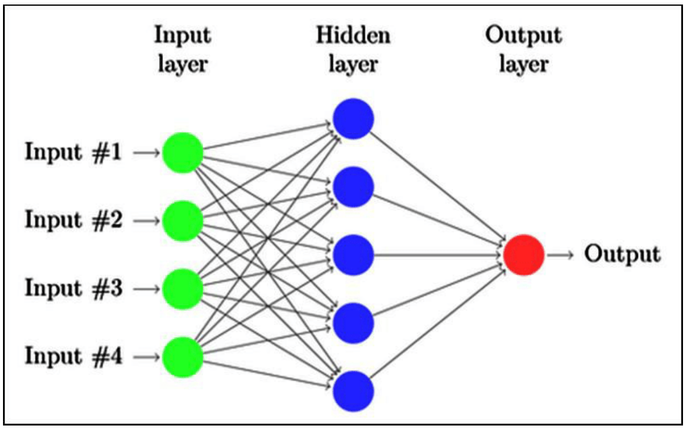


Figure 3.3: Basic Architecture of MLP network [33]

We use the rectified linear activation function on the hidden layer as it performs well and a linear activation function on the output layer because we are predicting a continuous value. We use mean squared error as loss function and the ‘adam’ optimizer for training the network.

The terms shown in the above Figure-3.3 are also used in our code as well as in Algorithm 3.1. So, for better understanding, we briefly introduce some of them as follows:

Model

A machine learning model is a program that can find patterns or make decisions from a previously unseen dataset. A model represents what was learned by a machine learning algorithm.

Keras

Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow. It is also used for distributed training of deep learning models. It provides essential abstractions and building blocks for developing and shipping machine learning solutions with high iteration velocity.

Sequential Model

This is appropriate for a plain stack of layers where each layer has exactly one input tensor and one output tensor. A sequential model is not appropriate when your model has multiple inputs or multiple outputs.

MSE

Mean squared error (MSE) measures the amount of error in model predictions. It assesses the average squared difference between the observed and predicted values. When a model has no error, the MSE equals zero. As model error increases, its value increases. It is also called risk function or loss function.

Dense Layer

A dense layer is a layer that is deeply connected with its preceding layer which means the neurons of the layer are connected to every neuron of its preceding layer. This layer is the most widely used layer in artificial neural networks.

Relu

The Rectified Linear Unit is the most widely used activation function in deep learning models. The function returns 0 if it receives any negative input, but for any positive value it returns that value back.

Adam

Used as an optimization solver for the Neural Network algorithm that is computationally efficient, requires little memory, and is well suited for problems that are large in terms of data or parameters or both. Adam is a popular extension to stochastic gradient descent.

Hidden Layer

A hidden layer in an artificial neural network is a layer in between input layers and output layers. The interior layers are sometimes called “hidden layers” because they are not directly observable from the systems inputs and outputs.

The following steps shows the algorithm used to setup our MLP model:

---

1. Take an instance of 'Sequential' Model from the Keras deep learning library.
2. Add a Dense layer to the model with stating number of inputs (24), number of nodes (500), number of epochs (100) and batch size (100), rectified linear activation function (relu).
3. Add another Dense layer with number of outputs (1), since we predict a continuous value.
4. Compile the model with Mean Square Error (MSE) loss function and 'adam' optimizer.
5. Fit the model with training data set for number of epochs (100) and batch size (100).
6. Make an ensemble of models by following the steps 1 to 5.
7. Get prediction output *yhat* for each time step (day) from all the models of the ensemble. An example of single model output can be derived:

   yhat = model.predict(input)
8. Calculate the ranges (lower bound, mean and upper bound) of each prediction.
9. Calculate uncertainty of the model for each day by using the set of yhats using the uncertainty calculating formula explained in section 3.7

Algorithm 3.1: MLP Model

3.5 CNN

Convolutional Neural Networks are a type of deep neural network developed for computer vision; for instance, two-dimensional image data, although they can be used for one-dimensional data such as sequences of text and time series forecasting. When operating on one-dimensional data, the CNN reads across a sequence of lag observations and learns to extract features that are relevant for making a prediction.

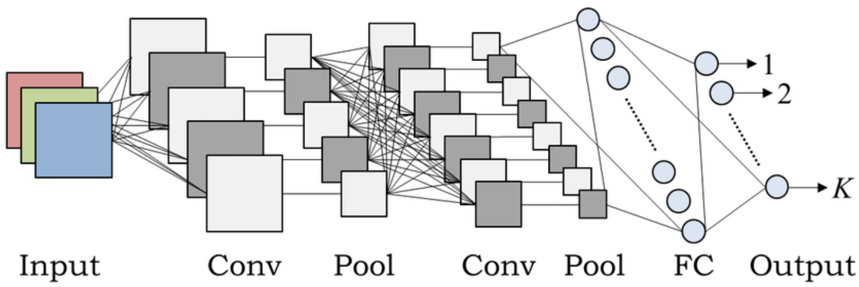


Figure 3.4: Basic Architecture of CNN network [ref. 34]

We define a CNN with two convolutional layers, one max-pooling layer, one flatten layer, and a dense layer from the input sequences. They have a configurable number of filters, kernel-size, pool-size and a rectified linear activation function is used as loss function. The number of filters determines the number of parallel fields on which the weighted inputs are read and projected. A max pooling layer is used after convolutional layers to distill the weighted input features into those that are most salient, reducing the input size by 1/2. The pooled inputs are flattened to generate a long vector before being interpreted and used to make the prediction.

The terms shown in the above Figure-6 are also used in our code as well as in Algorithm 3.2. So, for better understanding, we briefly introduce some terms in the following sub-sections.

Conv1D:

A convolution layer transforms the input in order to extract features from it. This layer creates a convolution kernel that is convolved with the layer input over a single spatial (or temporal) dimension to produce a tensor of outputs.

MaxPooling1D:

Max pooling is a pooling operation that selects the maximum element from the region of the feature map covered by the filter. In other words, it downsamples the input representation by taking the maximum value over a spatial window of size (pool size). Thus, the output after max-pooling layer would be a feature map containing the most prominent features of the previous feature map.

FC

The fully connected layer (FC) operates on a flattened input where each input is connected to all neurons. After feature extraction we need to classify the data into various classes, this can be done using a fully connected (FC) neural network.

Flatten Layer

To flatten the multi-dimensional input tensors into a single dimension this layer is used. For example, if flatten is applied to layer having input shape as (batch_size, 2,2), then the output shape of the layer will be (batch_size, 4)

The following steps shows the algorithm used to setup our CNN model:

---

1. Take an instance of 'Sequential' Model from Keras deep learning library.
2. Add a Conv1D layer to the model defining the number of filters (24), kernel size (500), input shape (100), rectified linear activation function (relu).
3. Add another Conv1D layer with same settings but without input shape.
4. Add another MaxPooling1D layer with pool size of 2.
5. Flatten (reshape) the result of previous step into single dimension before interpreted by the next layer.
6. Add a Dense layer with number of outputs (1), since we predict a continuous value.
7. Compile the model with Mean Square Error (MSE) loss function and 'adam' optimizer.
8. Fit the model with training data set for number of epochs (100) and batch size (100).
9. Create an ensemble of 6 models by following the steps 1 to 8.
10. Get prediction *yhat* for each time step (day) from all the models of the ensemble. An example of single model output can be derived:
    yhat = model.predict(input)
11. Calculate the ranges (lower level, mean and upper level) of each prediction.
12. Calculate uncertainty of the model for each day by using the set of yhats using the uncertainty calculating formula explained in section 3.7.

Algorithm 3.2: CNN Model

3.6 LSTM

The LSTM neural network is a member of RNNs (Recurrent Neural Networks) and it can be used for univariate time series forecasting. It uses an output of the network from a prior step as an input in attempt to automatically learn across sequence data. Having internal memory, LSTM allows it to accumulate internal state as it reads across the steps of a given input sequence.

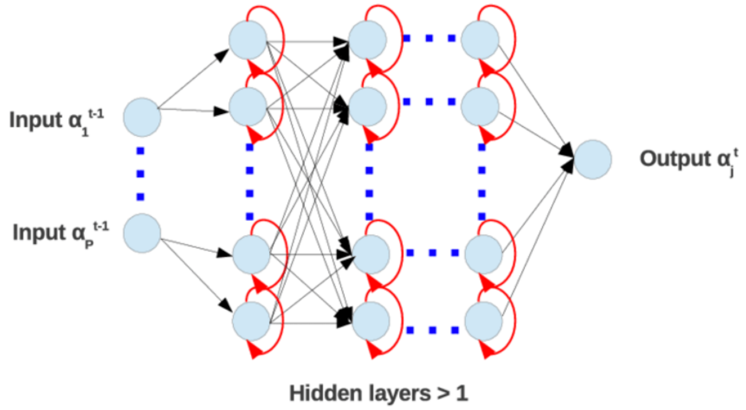


Figure 3.5: Basic Architecture of LSTM network (ref. 55)

For this model, we define a LSTM layer from inputs and subsequently two dense layers. Like other models, rectified linear activation function is used in LSTM layer and in one of dense layer. A simple grid search of model hyperparameters was performed with the predefined configuration.

The terms shown in the above Figure 3.5 are also used in our code as well as in Algorithm 3.3. So, for better understanding, we briefly introduce some of the unknown terms as follows.

LSTM Layer

An LSTM layer learns long-term dependencies between time steps in time series and sequence data. The state of the layer consists of the hidden state (also known as the output state) and the cell state. The hidden state at time step t contains the output of the LSTM layer for this time step.

The following steps shows the algorithm used to setup our LSTM model:

---

1. Take an instance of ‘Sequential’ Model from Keras deep learning library.
2. Add an LSTM layer to the model defining the number of nodes (24), input shape (100), rectified linear activation function (relu).
3. Add a Dense layer for 24 input nodes and ‘relu’ activation function.
4. As we predict single value output, add a Dense output layer of 1 node.
5. Compile the model with Mean Square Error (MSE) loss function and ‘adam’ optimizer.
6. Fit the model with training dataset for number of epochs (100) and batch size (100).
7. Create an ensemble of 6 models by following the steps 1 to 8.

8. Get prediction *yhat* for each time step (day) from all the models of the ensemble. An example of single model output can be derived:
   yhat = model.predict(input)
9. Calculate the ranges (lower level, mean and upper level) of each prediction.
10. Calculate uncertainty of the model for each day by using the set of yhats using the uncertainty calculating formula explained in 3.7.

Algorithm 3.3: LSTM Model

3.7 ARIMA

An Autoregressive Integrated Moving Average (ARIMA) is a statistical analysis model that uses time series data to analyse and understand the data and predict future trends. A statistical model is autoregressive if it predicts future values based on past values. It describes the correlation between data points and considers the difference of the values.

The model has three major components which come from its name – AR (autoregressive term), I (Integrated term) and MA (moving average term). Let us briefly explain each of these components:

- The AR term refers to predicting the next value using the prior values of the dataset. The AR term is defined by the parameter p in ARIMA.
- The Integrated(I) term represents the number of times the differencing operation is performed on series to make it stationary (i.e., data values are replaced by the difference between the data values and the previous values). Tests like ADF can be used to determine whether the series is stationary and help in identifying the d value. Differencing is only needed if the series is non-stationary otherwise, no differencing is needed, and in that case $d=0$.
- MA term is used to define the number of prior/lagged forecast errors used to predict the future values. The parameter q in ARIMA represents the MA term.

3.7.1 Auto ARIMA

Although ARIMA is a powerful model for forecasting time series data, the data preparation and parameter tuning processes is quite time consuming. Before implementing ARIMA, it needs to make the series stationary, and determine the values of p and q as stated earlier. Auto ARIMA makes this complicated task simple for us as it eliminates those time-consuming tasks of optimal p, d, and q parameters by fitting different models and deciding which one is best. Basically, it takes the data and fits many models in a different order before comparing the characteristics. Below are the steps for implementing auto ARIMA, but before going to Algorithm 3.4, we briefly introduce some of the unknown terms for better understanding as follows:

pmdarima

It is a statistical library designed to fill the void in Python's time series analysis capabilities. This includes: cross-validation utilities, built-in time series datasets for prototyping and examples, and time series utilities, such as differencing and inverse differencing.

ADF Test

The Augmented Dickey Fuller test is a common statistical test used to test whether a given time series is stationary or not. It is one of the most widely used statistical test when it comes to analyzing the stationary of a series.

Univariate time series

It refers to a time series that consists of single (scalar) observations recorded sequentially over equal time increments.

The following steps shows the algorithm used to setup our Auto ARIMA model:

---

1. Load data: Collect data from the source repository and load into a data table.
2. Preprocess data: As the prerequisite of the model input is to be univariate, drop other columns from the data table and make sure all empty values replaced with NULL, so that system does not break down during runtime.
3. Build model by using open-source package named *pmdarima*.
4. Fit the model on the univariate series of data generated in step-2 and using parameters *test*='adf', p=3, q=3 to get optimal value of d.

5. Make predictions from the model for 200 days like other models.
6. Calculate series by using the forecasted results in previous step and with the help of panda.Series method.
7. Find the lower and upper bound of the series which will be used to calculate the uncertainties of the prediction.

Algorithm 3.4: ARIMA Model

3.8 Uncertainty Data Generation

Uncertainties are calculated from the ranges of predicted values for every time step (day) during the specified 200 days of forecasting period. That means we have a lower bound, mean, and upper bound of the predictions for each time step. So, the difference between upper and lower limit is the grey area of model prediction. Then we find the maximum difference to set out the domain of the difference. Finally, we divide each difference by the maximum difference and multiply by a scaling factor to keep the maximum result in single digits.

Here is given the steps to find the uncertainties using the machine learning models:

---

1. Read data from filesystem (excel file) to Data-Frame
2. Select Fields for which we need to generate uncertainty data
3. Create Machine Learning model for MLP/CNN/LSTM
4. Split data into training and test set
5. Train model with training set
6. Use model to get predicted or forecasted results
7. Find uncertainties or prediction error from model
8. Continue step 3 to 7 for each field and each model
9. Store uncertainty data as json in filesystem

Algorithm 3.5: Calculate uncertainty (For all models)

3.8.1 Uncertainty Data Scaling

We have shown top-level algorithms in the previous sections to generate uncertainty data from the model predictions. Since the uncertainty values are independent of the display, we need to scale the values to accommodate the displays. The following pseudo code is used to scale the uncertainty data by which it can better fit in display devices.

----

1. Repeat step 2 and 3 for all countries and store them in an array named: all_countries_avg_uncertainties
2. Calculate total uncertainty (county_total_uncertainties) of a country.
3. Calculate average uncertainties of a country as follows: country_avg_uncertainty = county_total_uncertainties/number_of_days
4. Find maximum average uncertainty from all countries: max_uncertainty = find_max_uncertainty(all_countries_avg_uncertainties)
5. country_uncertainty_normal = country_avg_uncertainty / max_uncertainty
6. scaling_factor = 9
7. country_uncertainty = country_uncertainty_normal * scaling_factor

----

Algorithm 3.6: Data scaling

From the algorithm, we see all steps are self-descriptive. Up to step 5, we have calculated the normalized form of uncertainty for every country, that means uncertainties are below or equal to 1 for all countries. So, we have set *scaling_factor = 9* and multiplied it with the country's normal uncertainty to display those smaller values in display in terms of pixels. For example: the countries that have higher uncertainties will be in normal form such as 1, 0.9, 0.8 and after multiplying with scaling_factor those will be 9, 8.1, 6.4 and so on. So, in this way, we could allocate 1 pixel per unit of uncertainty and that helped to visualize the default view in a human recognizable manner and easily understandable the higher uncertainty countries.

3.8.2 Snapshot of uncertainty data

Since the pandemic affected all the countries of the world and there are more than 200 countries, we have only trained the models for top 100 countries which were infected severely. Based on that setup, we have sorted the countries by obtained uncertainties in both ascending and descending orders. Table 3.3 shows the top 10 uncertainty levels for countries and Table 3.4 shows the lowest 10 uncertainty levels of the countries.

Country	Actual Count	Predicted Count	Uncertainty
United States	14,851,118	15,652,300	7.00
India	15,693,425	7,409,636	4.28
Brazil	7,219,982	7,409,636	3.64
Kazakhstan	667,009	651,009	2.43
France	2,088,610	2,307,005	2.15
Peru	432,034	546,901	1.28
Germany	1,700,161	1,599,684	1.21
Spain	1,542,012	1,510,467	1.07
Turkey	3,645,288	3,389,016	1.03
Argentina	2,352,216	2,450,255	1.02

Table 3.3: Top 10 uncertainty countries using MLP model, Correlation (Actual) = 0.87

Country	Actual Count	Predicted Count	Uncertainty
Qatar	36,256	36,796	0.013
Albania	62,292	65,515	0.016
Estonia	90,950	89,900	0.017
Egypt	118,376	124,175	0.019
Moldova	103,270	101,832	0.019
Australia	161,819	147,134	0.021
Algeria	86,238	82,121	0.022
Singapore	178,151	175,400	0.025
North Macedonia	57,447	57,420	0.037
South Korea	277,584	274,766	0.037

Table 3.4: Lowest 10 uncertainty countries using MLP model, Correlation (Actual) = 0.53

Based on the correlations (0.87 and 0.53), we can say the uncertainties are not fully dependent on the number of occurrences. In addition, from the observation of above two tables, it is noticeable that uncertainty is independent of the number of cases (Actual Count). For example: United States has lower number of cases than India but achieved higher uncertainty than India. Again, Kazakhstan and France exhibit similar behavior and if we examine other countries then we will see more examples.

3.8.3 Uncertainty Comparison among Models

Country	MLP	CNN	LSTM	ARIMA
United States	7.00	7.00	3.44	7
India	4.28	0.61	7.00	3.52
Brazil	3.64	0.51	3.24	1.27
Kazakhstan	2.43	0.42	0.35	0.17
France	2.15	0.31	0.81	0.56
Peru	1.28	0.23	0.28	0.22
Germany	1.21	0.19	0.50	0.51
Spain	1.07	0.19	0.67	0.33
Turkey	1.03	0.19	1.21	0.30
Argentina	1.02	0.14	1.08	0.25

Table 3.5: Uncertainty comparisons of models (Based on MLP)

From the above comparison Table 3.5 of four different machine learning models, we notice that the uncertainties greatly vary for each country based on the model. There is no country which has identical uncertainty values for all four models. Though the dataset used in each of the models in similar approach, the variation appears due to their internal mechanism of the model algorithms. Since the model performance is not our goal, we will not examine these results more closely. We will use the uncertainty data that we obtained from model prediction and uncertainty calculation methods to later drive our visualizations.

Chapter 4

Visualization Component Calculations

4.1 Introduction

There are various ways to represent uncertainties in the visualization domain. We have introduced a novel idea named Chromatic Aberration (CA) and in chapters 6 and 7 we evaluate how well it works compared to other existing approaches such as Correll et al. [35]. For this eventual purpose, we have explained the technical background of this method of representation and corresponding algorithms in this chapter.

4.2 Background Architecture

We have seen an example of lateral chromatic aberration in Figure 1.2 (Chapter 1) where all lights with different wavelengths do not focus to the same convergent point because lights having shorter wavelength refract more than the lights with longer wavelength. Inspired by that phenomenon, we can consider a circle that represents the predicted number of new cases for a country in a specific day. But since there is associated uncertainty of the prediction, a single circle will not be sufficient to represent bivariate (number of cases and uncertainty) distribution. So, instead of single circle if we use three different circles with separated RGB color channels, we can then apply lateral shifting from the center of the circle by the amount of uncertainty and blend them together and the resultant outcome would be an approximate representation of CA. The following Figure 4.1 shows such a geometric arrangement on a unit radius circle which is used as a basis of representing CA on a circular shape.

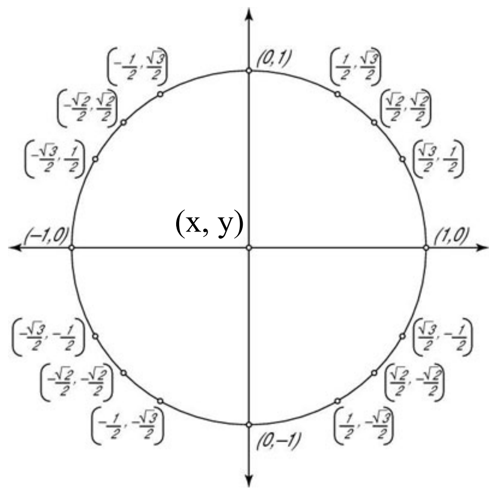


Figure 4.1: Underlying Geometry of CA

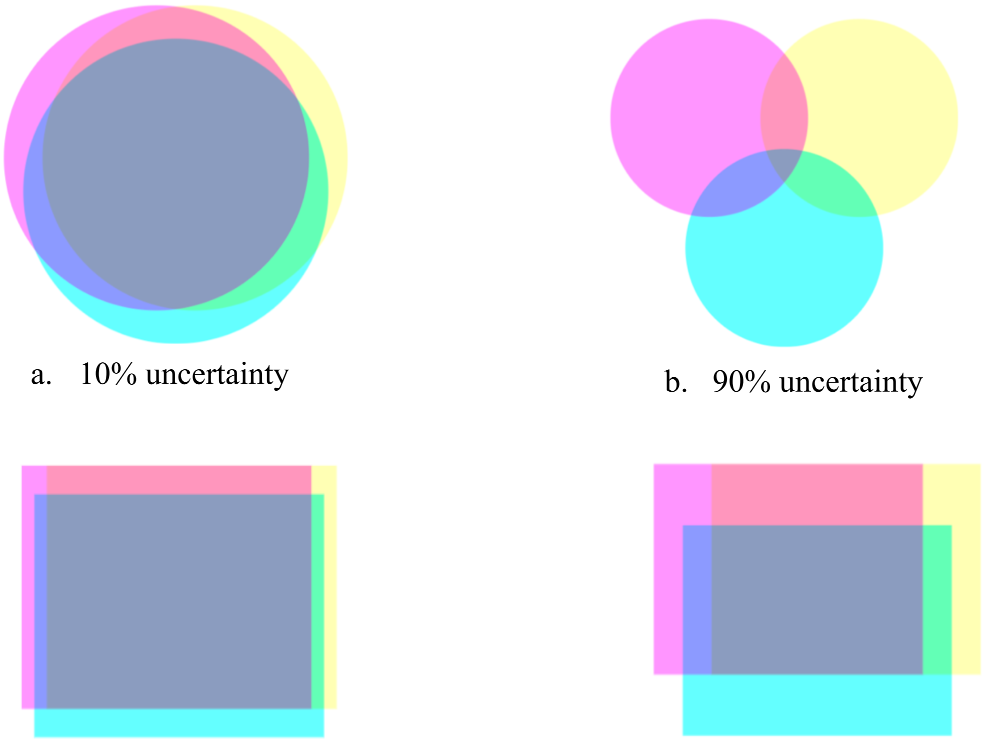
To draw a circle representing aberration as per the above explanation if we draw 3 circles, let's call them 3 chromatic circles, then we can render the technique with the following simple algorithm with the help of Figure 4.1:

- Let's consider the center of the target circle at (x, y).
- Radius (radial offset) of the circle is 'r' represents uncertainty.
- Draw the first chromatic circle with color (R, 255, 255) with a shifted location of (x, y + r) where 'r' denotes red color channel.
- Draw the second chromatic circle with color (255, G, 255) with a shifted location of $(x + r * \frac{+\sqrt{3}}{2}, y + r * \frac{-1}{2})$ where 'G' denotes green color channel.
- Draw the third chromatic circle with color (255, 255, B) with a shifted location of $(x + r * \frac{-\sqrt{3}}{2}, y + r * \frac{+1}{2})$ where 'B' denotes blue color channel.
- Set the standalone css 'mix-blend-mode' to 'darken' to blend all three circles to get the resultant CA appearance.

Algorithm 4.1: CA Construction Formula

4.3 Examples of CA in Shapes

By using the above formula explained in section 4.2, a resultant aberration is presented with the uncertainty for the country India (IND) in Figure 4.2 below. To accommodate this properly we normalized the uncertainty data. The center dark-grey area represents the predicted number of new cases, and the color separated edges represent the amount of uncertainty in that prediction.



a. 10% uncertainty

b. 90% uncertainty

c. 7% uncertainty

d. 40% uncertainty

Figure 4.2: Example CA on Bubbles and Rectangles

In the above depiction we have shown four different amounts (in percentage) of uncertainties in two different types (circle and rectangle) of representations. The same formula as explained in previous section (4.2) has been used to draw both the circular and rectangular shapes. We will show more examples in our user study representation and actual application of uncertainty visualization in different charts in the following chapters.

4.4 Other Visualization Experiments - Texture Pattern Generation

In addition to our primary contribution (introduction and evaluation of CA for uncertainty), we also explored a considerable number of experimental designs. We have defined and briefly explained texture generation basics in section 1.2.5 and by which we know how textures can be generated in web using SVG. In some cases, textures are used to emphasize or deemphasize certain parts of the design. Because of the versatility of textures, they can be used or generated in combination with many other design elements, such as typography, lighting, and colors.

There is a subtle difference between patterns and textures. Patterns are visual elements of geometric and mathematical structures that form consistent and repeated graphical shapes on a surface. Visual activity across a surface is more generally a texture when the structure forming the texture is based on irregular and random relationships over given areas. There are various kinds of textures and one of them is visual textures and patterns fall in that category. So, in our perspective, we built textures with the help of SVG patterns where predictions are presented with patterns and the collective outcome for the whole duration will be considered as textures.

As an example, we can consider a streamgraph (defined and briefly explained in section 1.2.4) that emphasizes the prediction of daily of number of new cases for a certain country for a specific duration and that is accomplished by filling the whole shape with flat color. But we can also attempt to represent uncertainty using textures within the streamgraph by slicing it for smaller number of days. For example: Figure 4.3 shows the scenario explained here and underlying mechanism of slicing is explained in the later section.

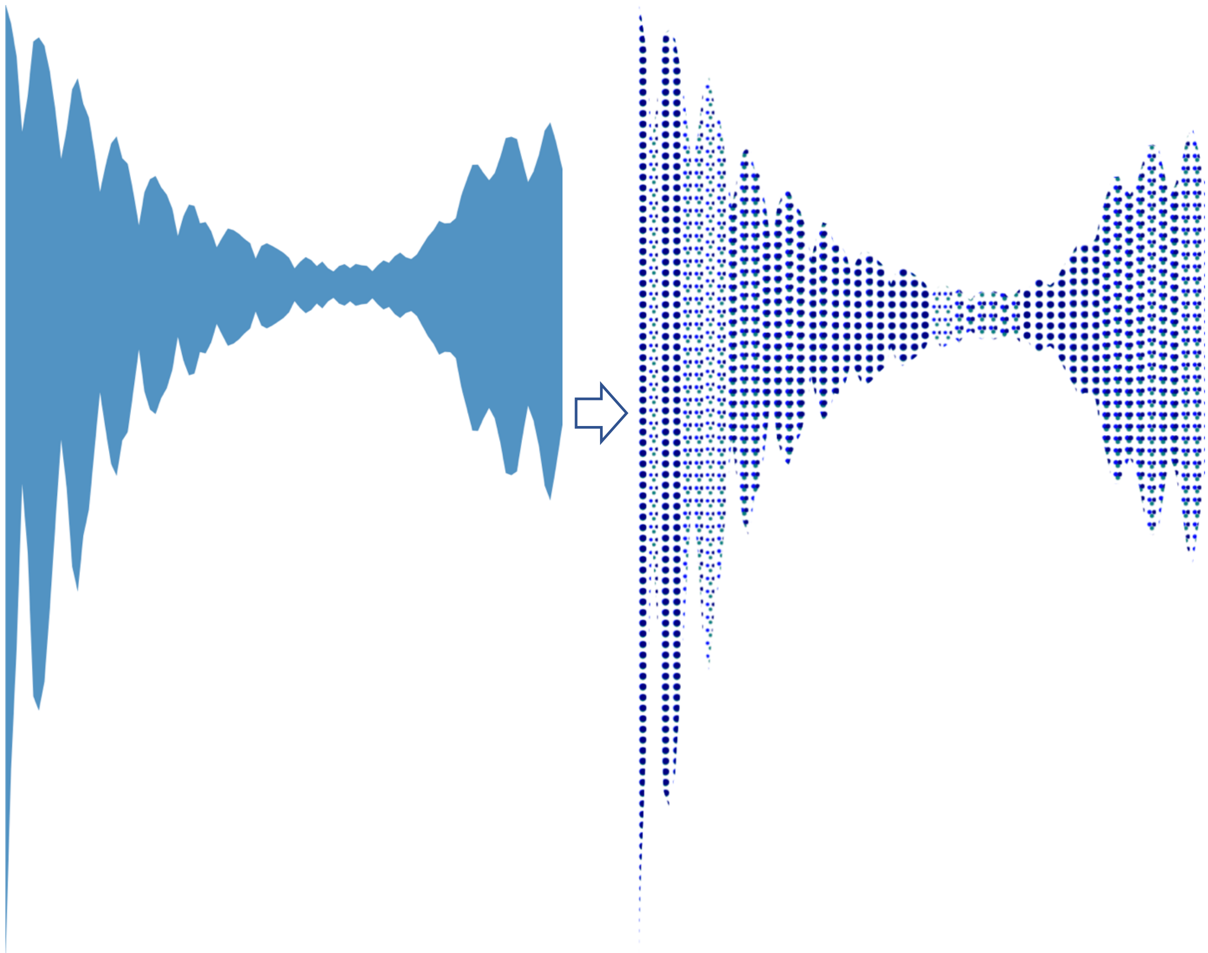


Figure 4.3: Streamgraph Color Filled (left), Texture Filled (right)

4.4.1 Slicing plot

In the above section, the streamgraph is shown as both a color-filled version and a texture-filled version. To better understand how the conversion is done the following Figure 4.4 gives a clear insight. We split the flow in the horizontal direction and make a slice for every 3 days since the horizontal axis represents the time in days. We have also experimented by chopping the graph with other number of days like 2, 4, 5, 6, 7 and so on but we found empirically that 3 days gives best result among all options to pertain the shape and peaks of the curve. This is because if we split it by 2 days then the width of the slice is too small to accommodate the content and if we use higher number of days then the shape of the plot undergoes with distortion and deteriorates the smoothness of the shapes such as peaks.

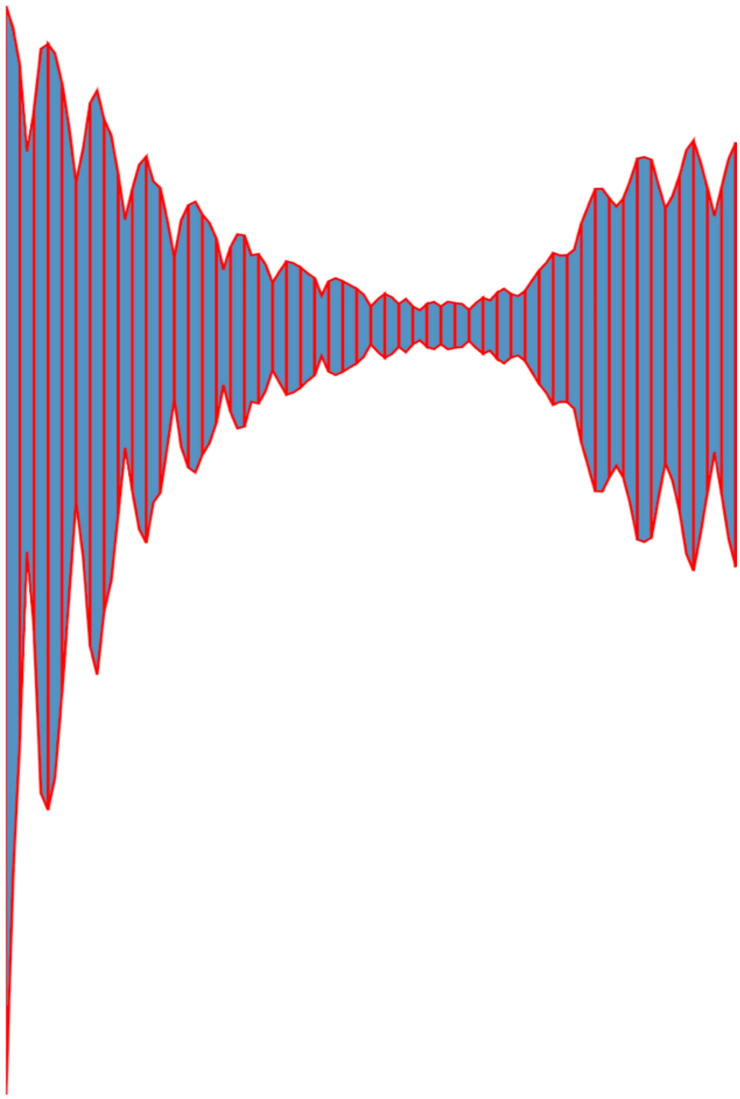


Figure 4.4: Sliced Streamgraph

Since each day of the duration has a different value of prediction and uncertainty, we have averaged the prediction of uncertainties for every three days and presented the corresponding values for the columns. That means, every column in the representation shows different uncertainties where dark green bullets reveal lower uncertainty and light-green bullets represent higher uncertainties in Figure 4.3.

4.4.2 Pattern Generation

Patterns can be generated easily with the help of HTML, CSS and JavaScript as stated in section 1.2.5. We have generated patterns in our experimental designs to apply in textures of various charts. To draw the textures, we have chosen two sets of alternating colors such reddish and bluish colors, used HTML pattern tag with a specific structure of defining id so that it can be used by that id in textures. It needs to define the height, width, shape (circle, rect, etc.), center (cx, cy) and radius(r), attribute (patternUnits for coordinate system). Finally, fill the pattern with a color. The following section shows pattern generation process:

- Repeat steps 2 to 12 for all the countries
- Define two sets of alternating colors to distinguish patterns side by side as follows:
  {0: '#ff0000', 1: '#800000', 2: '#FF00FF'} -> Reddish colors
  {0:'#008080 ', 1: '#0000FF', 2: '#000080'} -> Bluish colors
- Repeat step 3 for three aberration points.
- Repeat step 4 for three alternative colors.
- Repeat steps 6 to 12 for 10 uncertainty scales [0 to 9].
- Define a pattern by using <pattern> tag that defines a graphics object which can be redrawn at repeated x- and y- coordinate intervals to cover an area.
- Set an 'id' of the pattern element. This 'id' is used later during filling texture. We use the following convention to define an id:
  pattern_id = 'pat-' + country + '-' + aber_indx + '-' + rgb_indx + '-' + uncertainty_scale
  where,
  country = name of the country of stream graph
  aber_indx = index of the aberration from [0, 1, 2] from step 3.
  rgb_indx = index of the color channel (0 for red, 1 for green, 2 for blue) from step 4.
  uncertainty = uncertainty in the scale of 0 to 9 (normalized to meet the range) from step 5.
- Set width and height of the pattern.

- Append a shape of the pattern such as 'circle', 'rect', 'ellipse' etc. In our case, 'circle'.
- Set center (cx, cy) and radius (r) of the circle.
- Set attribute 'patternUnits' to 'userSpaceOnUse' that defines the coordinate system for cx, cy, width, and height.
- Fill the pattern by setting a fill color.

Algorithm 4.2: Pattern Generation

4.4.3 Texture Generation

As per explanation in section 1.2.5, textures and patterns are connected where patterns are a smaller component to use in textures with the pattern id. So, textures are considered in a bigger context for instance: html path element of a streamgraph. We can pick the path of a streamgraph and it can be chopped along its temporal direction by number of days such as 3 days (we used in our case to see better results). A new path is then created for each chopped area and define by an id. Each new path is then filled with a pattern id based on the matched parameter for example: uncertainty of that area in our case. Since we have three different circles in each new path according to Algorithm 4.1, we needed to blend them to emphasis the CA representation at the outer edges which refers to uncertainty as given in Algorithm 3.5 in the following section:

---

1. Select the streamgraph container using d3.js.
2. Find the path of the streamgraph by from the value of 'd' property.
3. Divide the upper and lower segments of the path and save in two variables.
4. Determine the number of vertexes (coordinates) in each segment (they would be same).
5. Repeat steps 6 to 11 until all vertexes are traversed.
6. Take three vertexes from upper segment, let's call it p1.
7. Take three vertexes from lower segment, let's call it p2.
8. Build a new path string by joining p1 and p2 with standard rule of using M (moveto), L (lineto) and Z (closepath).
9. Append new 'path' element into the container svg and set 'd' property by the path string.
10. Fill the path with the pattern id (generated by the previous algorithm 4.2) with the following syntax (value of fill attribute 'url(#pattern_id')).
11. Add blend style property 'mix-blend-mode' to 'darken'.

Algorithm 4.3: Texture Generation

Chapter 5

Experimental Designs with Chromatic Aberration & Texture Patterns

5.1 Introduction

As we have shown the underlying mechanisms, backgrounds, and algorithms in previous chapters, in this chapter we present some real-world charts where we have experimented using CA and textures in different ways such as in bubble chart, streamgraphs, parallel coordinate charts, horizontal charts, cell chart with bubbles and squares and a bubble chart in world map view. We will also show the appearance with streamgraphs using different models and corresponding representations with CA. We also briefly discuss some other early experiments with what we call star-fish charts which are generated using multiple streamgraphs for different countries and how their equivalent appearance could be made with CA.

5.2 Web Interface

To visualize different charts, we have developed a web application as in Figure 5.1 with html input controls in the top toolbar and all charts are presented in the main container placed just below the toolbar as follows:

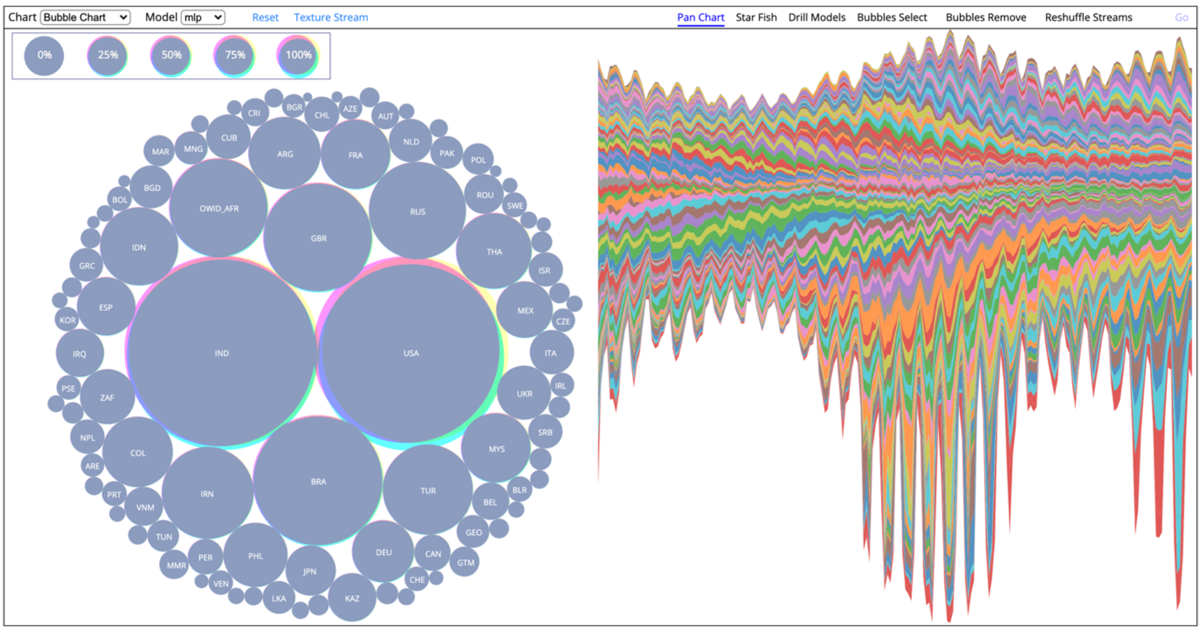


Figure 5.1: Initial Web Interface (Left - Bubble chart, right – Color Streamgraph). The legend gives a perception about the amount of uncertainty as a percentage.

In the following section, we briefly explain the basic functionalities of the input fields in toolbar, utilized for our experimental designs.

Chart dropdown: List of chart names, on selection it will automatically draw the corresponding chart in the main container. Bubble chart, Parallel Coordinates, Horizontal chart, Square Grid Chart, Bubble Grid Chart are available options in the list.

Model dropdown: Names of the predictive models for which we have generated data for finding the uncertainties and presenting as chromatic aberration. MLP, CNN, LSTM and ARIMA are the available options for the list.

Reset: Return to the initial state of the drawing for bubble chart. For this chart it has different type of modes listed in the right side of the toolbar.

Texture Stream: This is a toggle button to switch the stream graph from color-based filling to texture based filling, that means instead of flat color flow it uses bullet like textures to fill the stream but they retain different colors for their own country region. More detail is shown in section 4.5.

The followings are available operational modes of bubble chart:

Pan Chart: Since the bubble chart and stream graph are drawn side by side and they work interactively like filtering the streamgraph with the selection from bubble chart, so sometimes it is necessary to zoom-in/out of the charts and consequently panning the charts in its own space is also advantageous.

Star Fish: changes the drawing mode to interact with mouse events. In this mode user can click on a country bubble to open the corresponding texture stream graph as a wing in a star-fish layout. So, for example, when user select 8-10 countries in each side then the resultant chart will look like starfish. We will show further detail about this layout in later sections.

Drill Models: In this mode when the user selects a country then four stream graphs with Chromatic Aberrated textures are shown in the right panel corresponding to the four predictive models. A detailed explanation will be shown in later section.

Bubbles Select: Select one or more country from the bubble chart and redraw it with the selected countries only. After selection, the 'Go' button will perform a redraw. This helps to compare specific countries.

Bubbles Remove: It is the opposite feature of the bubble select mode. It filters out countries from the bubble chart. In this mode the selected countries are omitted from the chart. After omitting countries on press 'Go' button it redraws with the other countries.

Reshuffle Streams: Redraws the main streamgraph with the selected countries of interest from bubble chart. This allows one to compare streamgraph of one or more countries selectively.

5.3 Filtering

We use data for the top 100 countries based on the total infection rate. As we see from the Figure-12, it is difficult to read the label of the country and difficult to identify the extent of aberration for the smaller circles having lower uncertainties. That's why we implement a filtering option with different perspectives. In the section below we briefly explain them.

5.3.1. Bubble Selection Mode

In this mode, it allows users to select the countries of interest on first click and toggles on the next one. So, when all preferred countries are selected the 'Go' button redraws the bubbles side by side with comparatively bigger sizes.

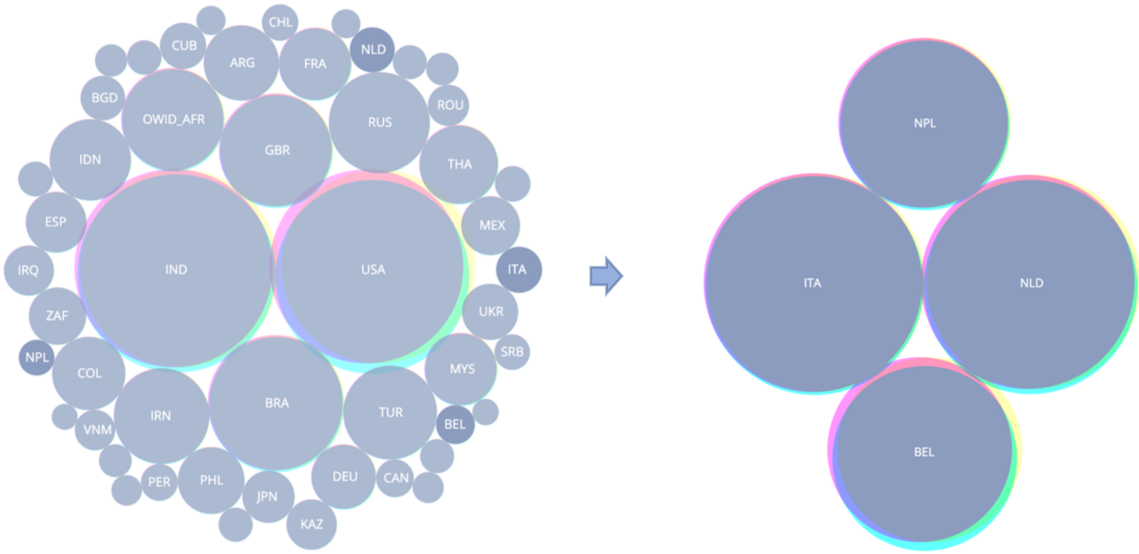


Figure 5.2: Filter by selected countries of interest

5.3.2 Bubble Removal Mode

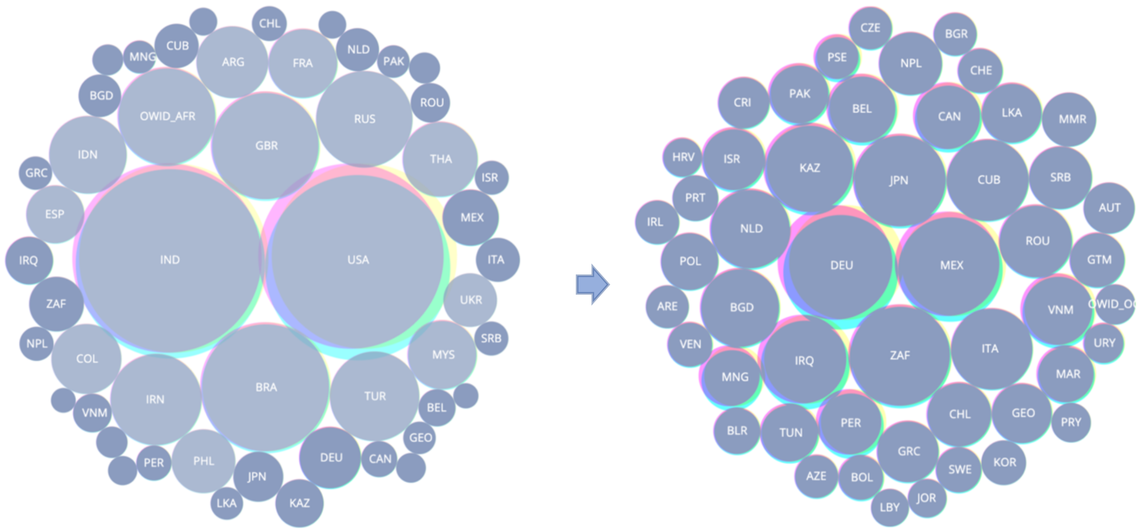


Figure 5.3: Removal of countries of interest. On removing large bubbles, the uncertainty representation noticeably changes among the existing circles.

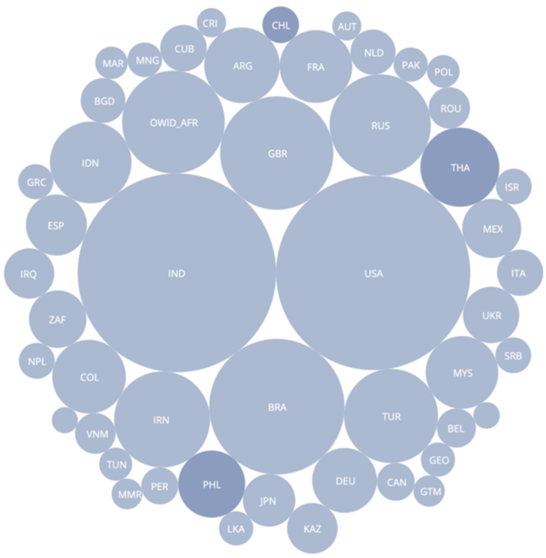
This is the opposite of the previous mode where the user can select the countries to remove from the chart, for instance, removing bigger ones help to find the status of the countries having a smaller size.

5.4 Legend

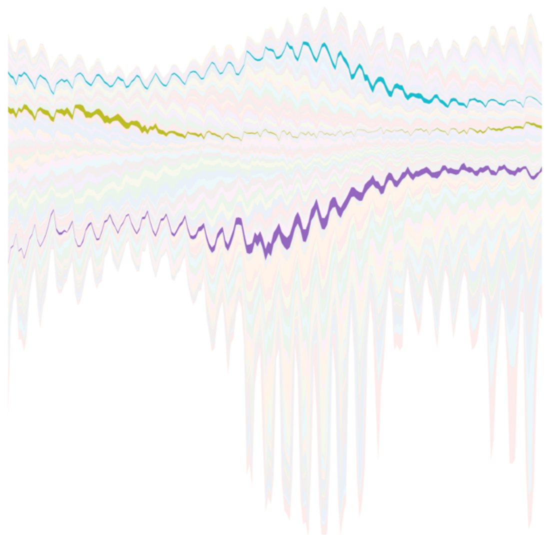
Placed at the top-left corner (Figure-12) just below the toolbar and above the bubble chart with 5 consecutive circles. The circles are drawn for representing 5 different levels of Chromatic Aberration. The circle with 100% uncertainty represents the maximum uncertainty among all the countries drawn in bubble chart.

5.5 Reshuffling Streamgraph

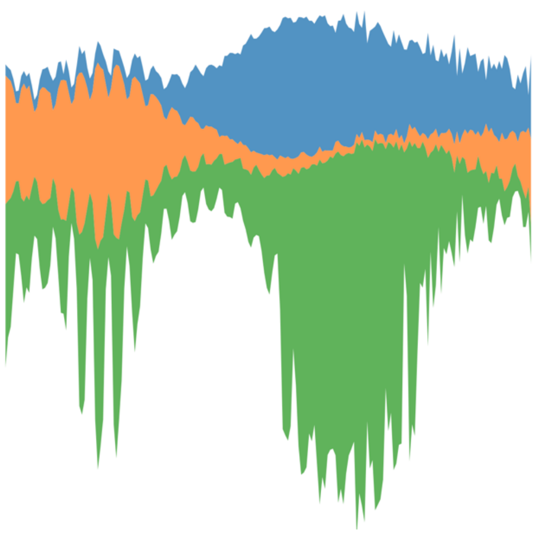
In Figure-12 we found the stream graph with many countries can be difficult to understand, so reshuffling is important to see and compare them side by side with a small number of countries.



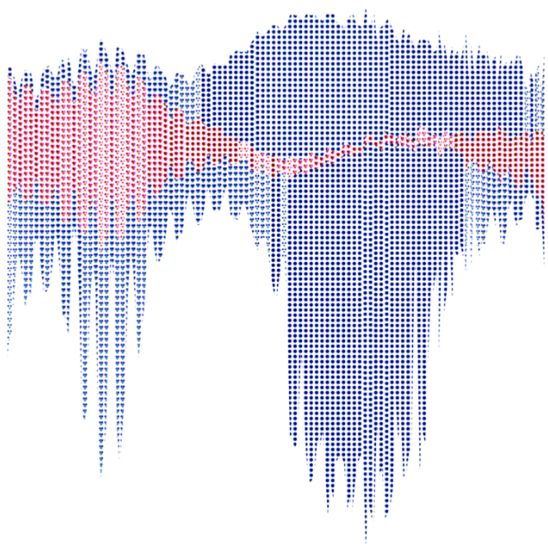
a. Selected countries brightened



b. Highlighted streams



c. Selected Countries Streams



d. Uncertainty texture with streams

Figure-5.4: Reshuffling Main Streamgraph (a => b => c => d)

To serve that purpose, in this mode, a user can choose the countries from the bubble chart. On selecting the countries, the corresponding ones brighten on bubble chart and highlighted the relevant ones in the streamgraph to represent the selection and the rest of the country-streams will be grayed out in both charts (a. and b. in Figure 5.4). Pressing ‘Go’ button confirms the redraw of streamgraph with the selected countries as shown in the Figure 5.4 (c). Then on pressing ‘texture stream’ button it converts the color stream to texture streams as in Figure 5.4 (d).

5.6 Drill-Down All Model Predictions

Since we have explored four different predictive models, we intended to show their predictions side by side altogether for a same country. We have used their predictions data for the same number of days and for same country to draw streamgraphs. Initially we have shown all streamgraphs in color filled version in Figure-16 and then the same graphs are represented with Chromatic Aberration in Figure-17.

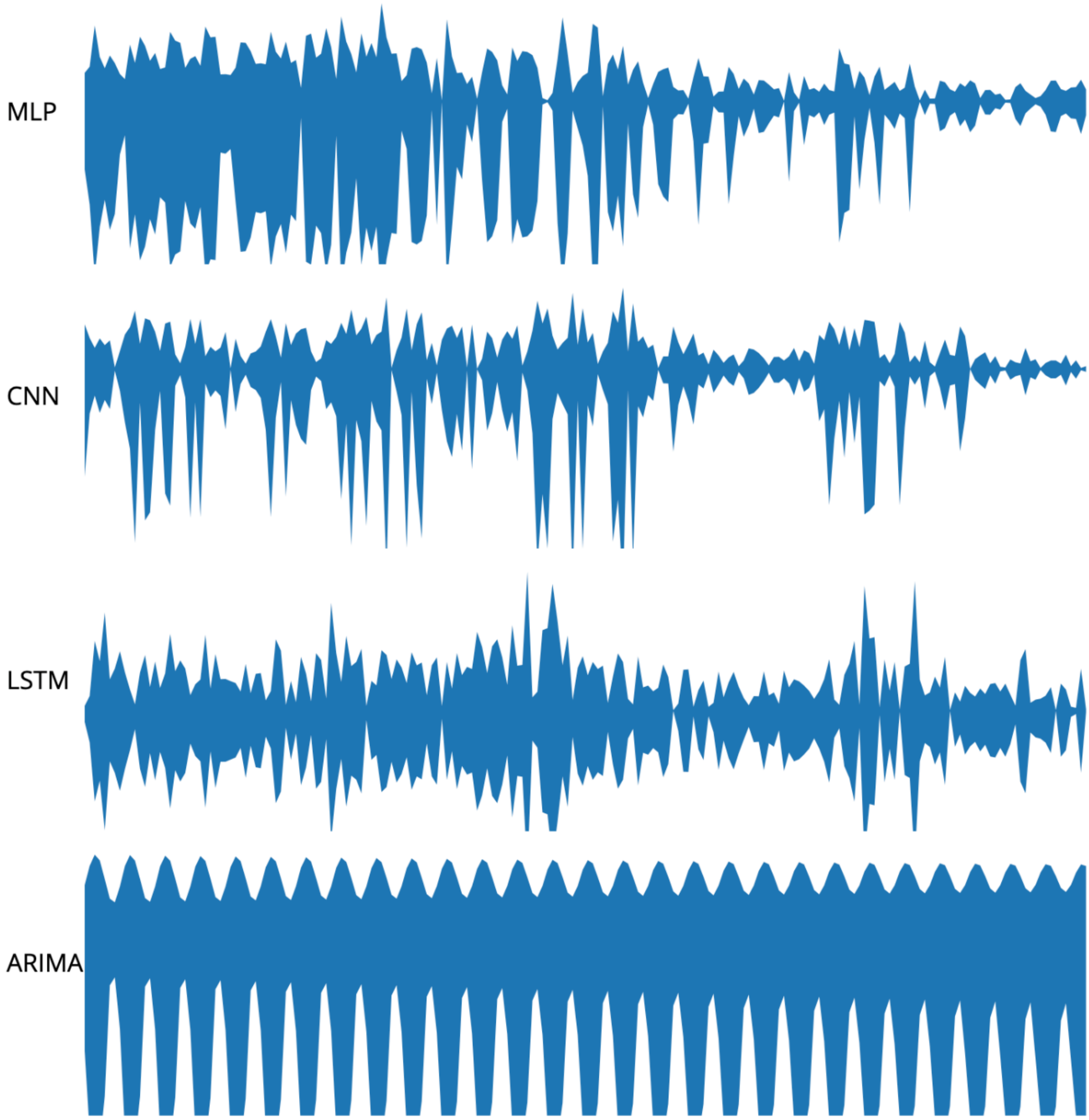


Figure 5.5: Streamgraphs for all models for Brazil

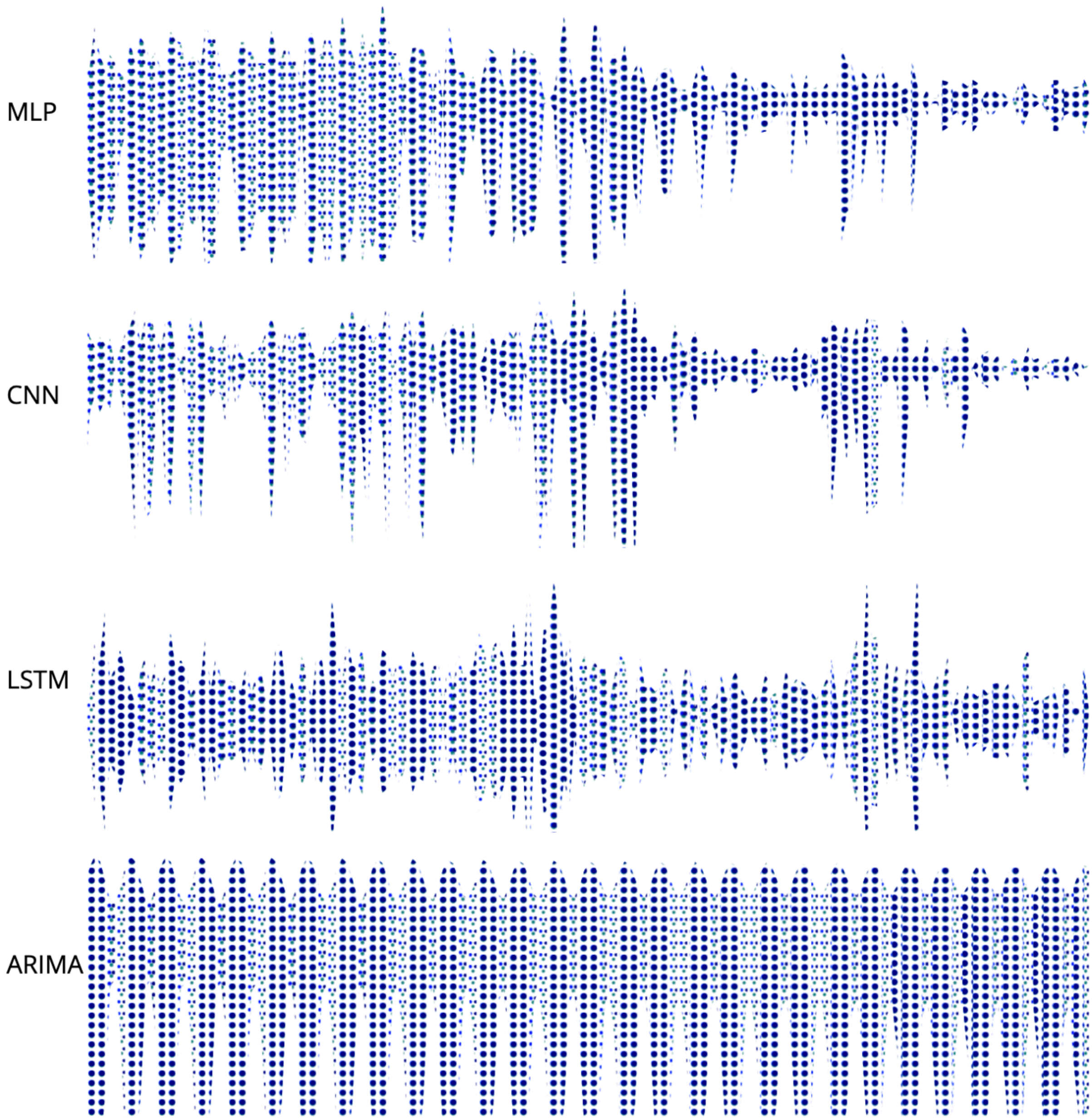


Figure 5.6: Uncertainty presentation on stream by texture

5.7 Star Fish Inspired Design

In this experimental design, a user can draw multiple stream graphs around a bubble chart and the system serves this by dynamically calculating the rotation angle of the stream based on the position of the country cell in respect to center of the map. Then it sets the start point of the stream at the bubble center and the rotation is introduced to avoid overlapping the other countries' streams and by which it attains 'Star Fish' layout presentation. If we call each individual stream as an arm, then the potential benefit of this chart is it allows to draw multiple charts in compact way. If the user needs to explore certain stream more deeply then it allows to do so by panning and zooming the map. To clear the drawn streams user can click reset button and again select more countries to draw another start with other countries.

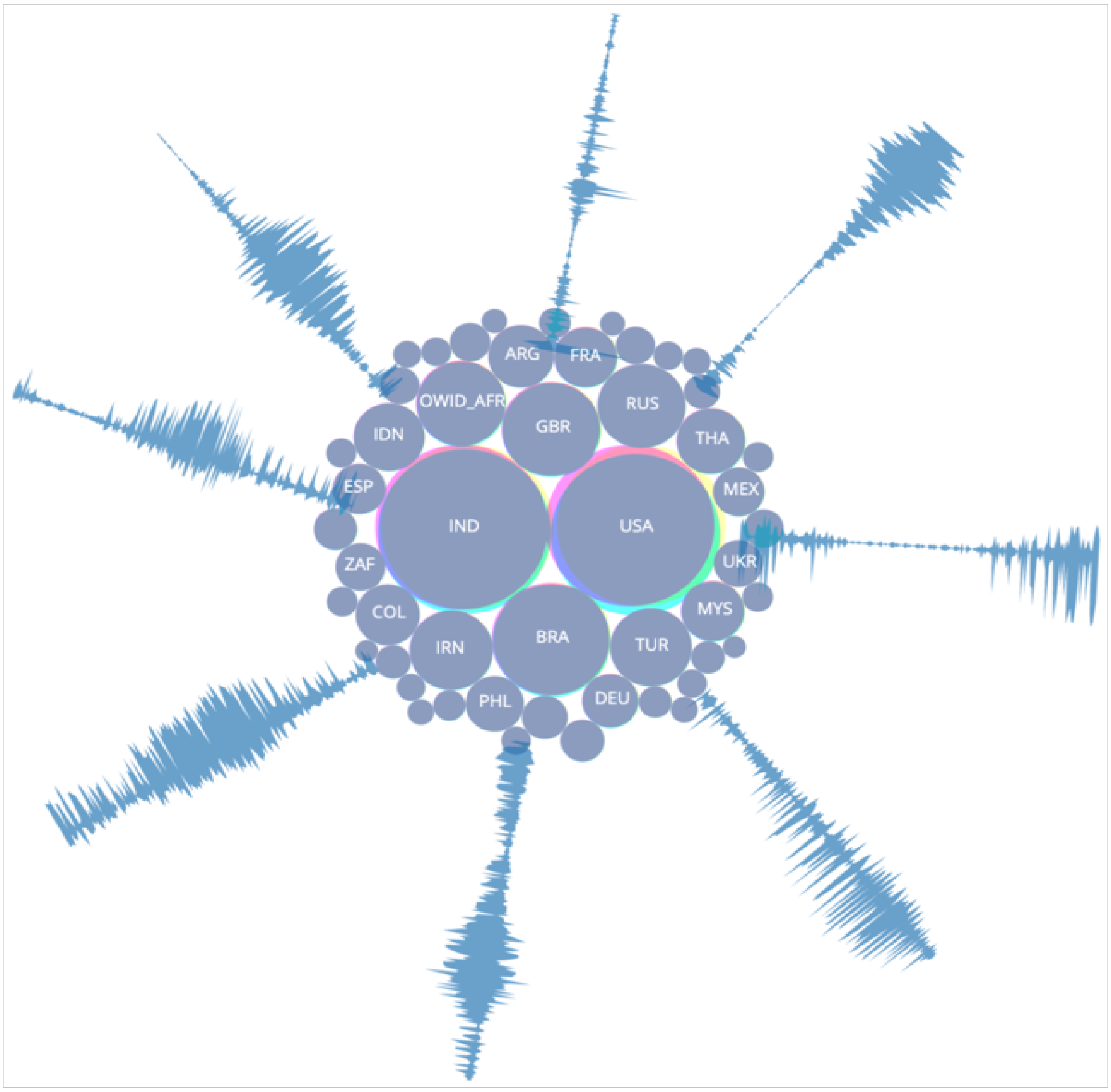


Figure 5.7: Multi Country Stream Graphs filled by color

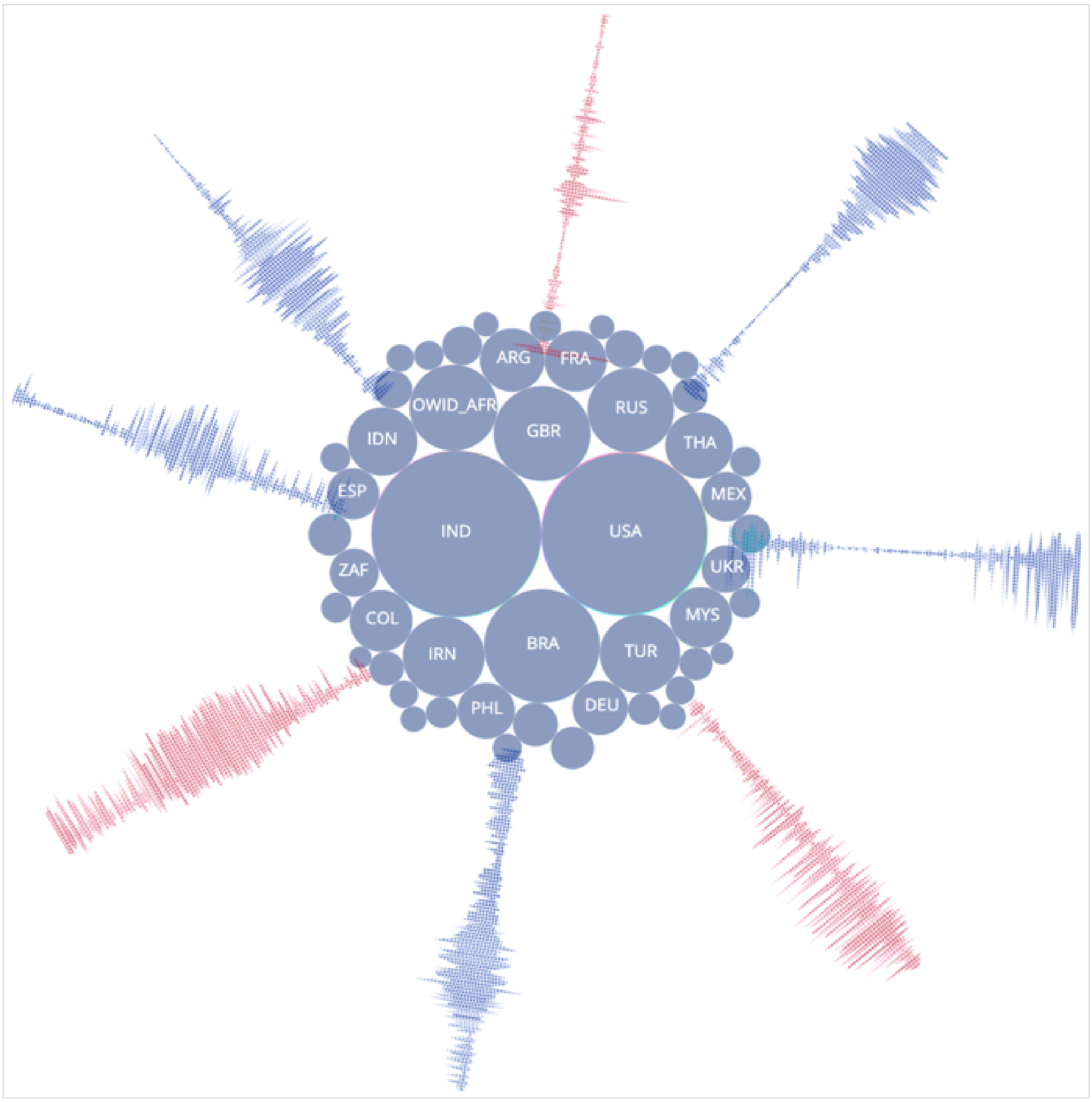


Figure 5.8: Multi Country Stream Graphs filled by Uncertainty Textures

5.8 Parallel Coordinates Chart

Parallel plots or parallel coordinates plots allows one to compare the features of several individual observations (series) on a set of numeric variables. Each horizontal axis represents a variable and often has its own scale. The units can be different, that is the strength of this special kind of plots. The main advantage offered by parallel coordinate is the representation of high dimensional data as a 2-dimensional visualization. Data is represented in the form of a polyline, and it becomes possible to perceive trends shown by data entries from the visualization.

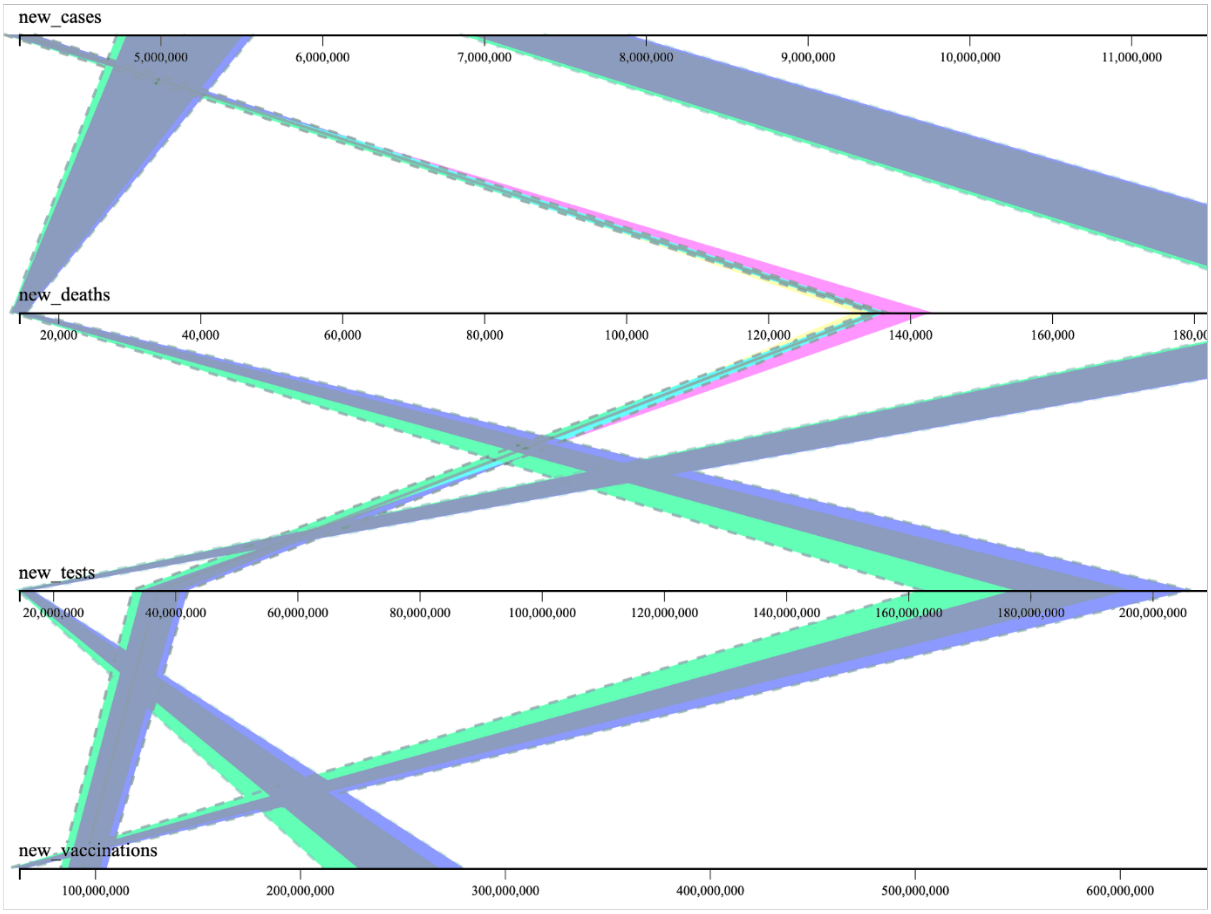


Figure 5.9: Parallel coordinates chart (representing the uncertainty with a similar approach of lateral shifting, so the edge thickness represents uncertainty and overlapping area represents actual counts)

This plot is helpful in our presentation because we have several variables together to visualize one after another and showing the relationships between them. For example, you can compare number of total cases(total_cases) with hospitalized patients (hosp_patients) facilitated by a tooltip showing the country name. Also, it can show the predicted flow (thinner line) along with actual counts (thicker line). The limitation of this chart is frequent overlaps for multiple variables. This particular experimental direction did not proceed much beyond the standard approach.

5.9 Bubble Grid Chart

This chart helps to indicate daily uncertainty presentation for every country as a cell. In this way a user can perceive trends for certain days or a set of consecutive days. In other words, the chart provides a platform that helps you decide which uncertainty requires your attention. This chart gives information to users about daily basis prediction uncertainties for all countries.

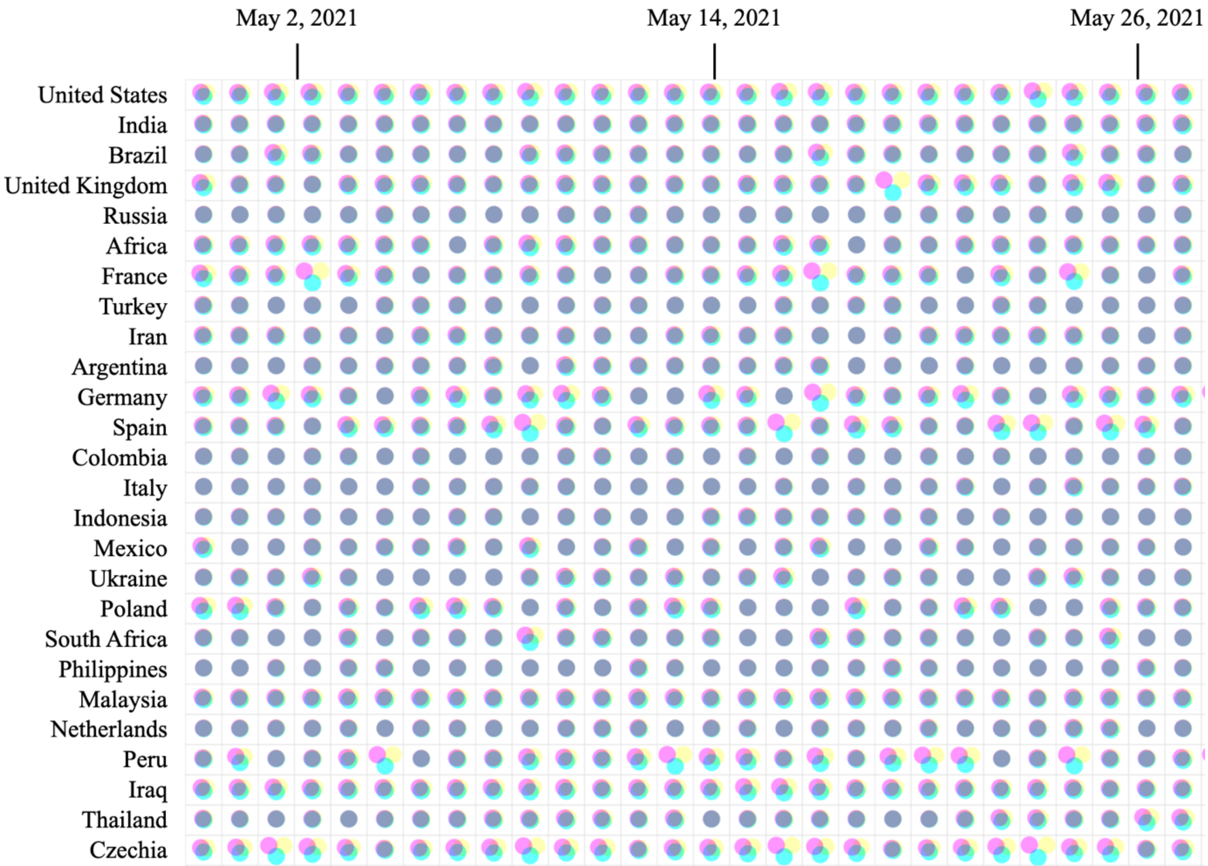


Figure 5.10: Bubble Grid chart with CA textures

5.10 Horizontal Chart

Horizontal charts are small-multiple area charts that allow greater precision for a given vertical space by using colored bands. These charts can also be used with diverging color scales to differentiate positive and negative values.

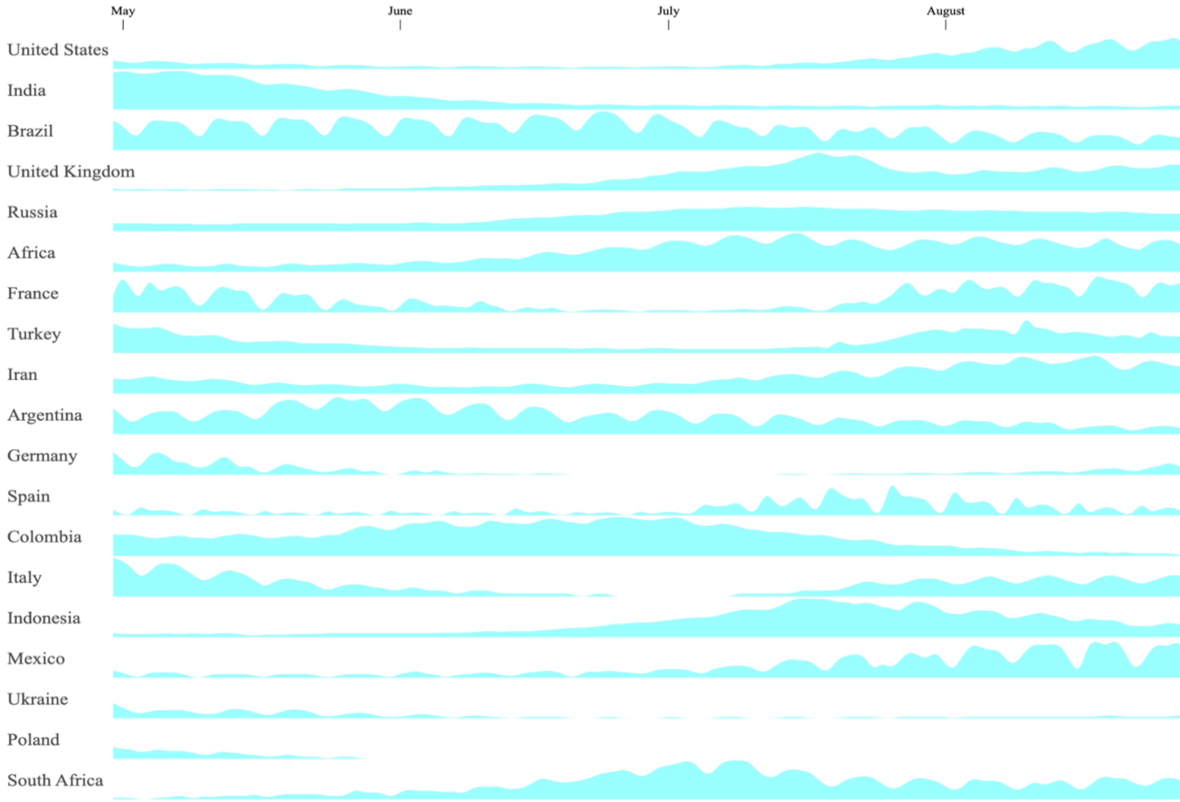


Figure 5.11: Horizontal chart (Color filled)

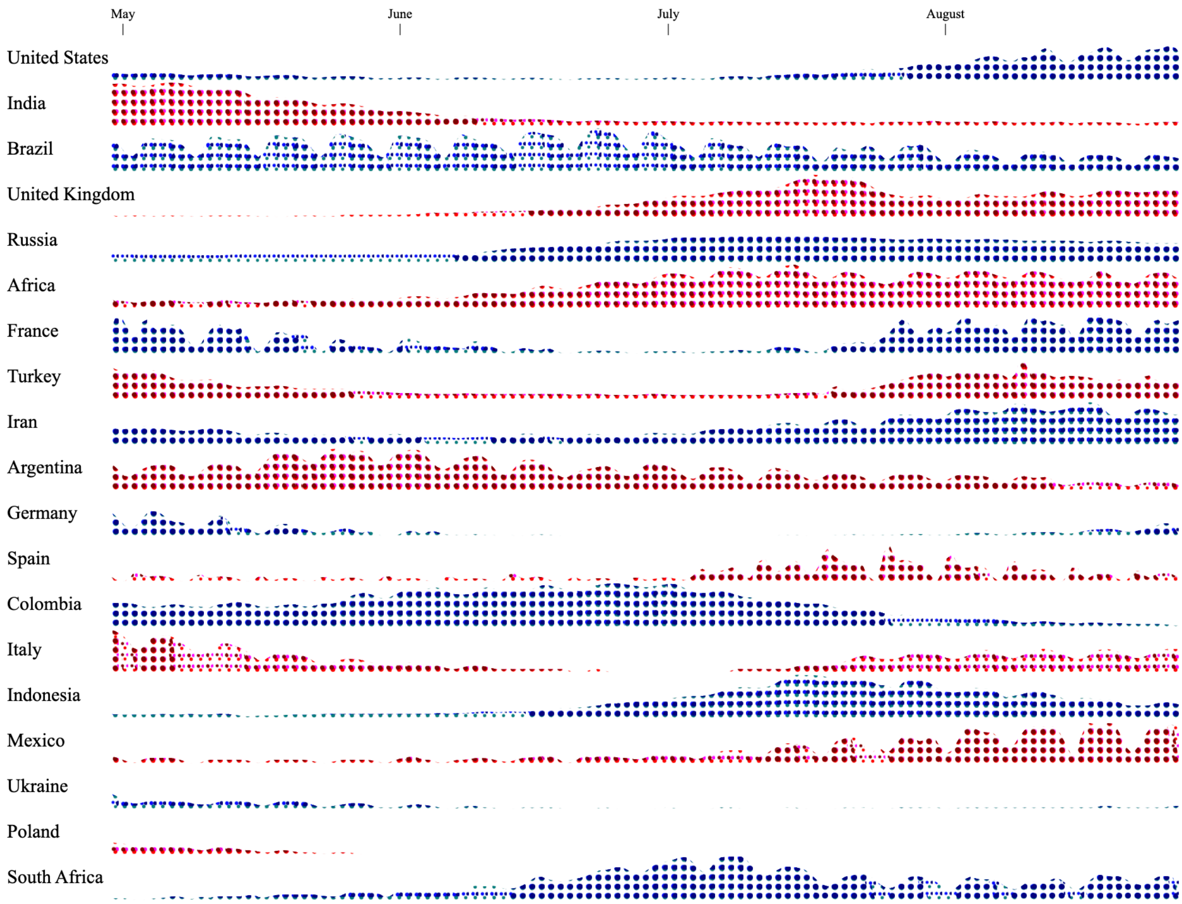


Figure 5.12: Horizontal chart (Uncertainty Texture filled)

5.11 Square Grid Chart

This chart shows information for date vs country, where the horizontal axis represents country and vertical axis represents date. Here we used rectangular shapes whereas the previous example used circular shape. In both case we have used prediction for daily number of new cases.

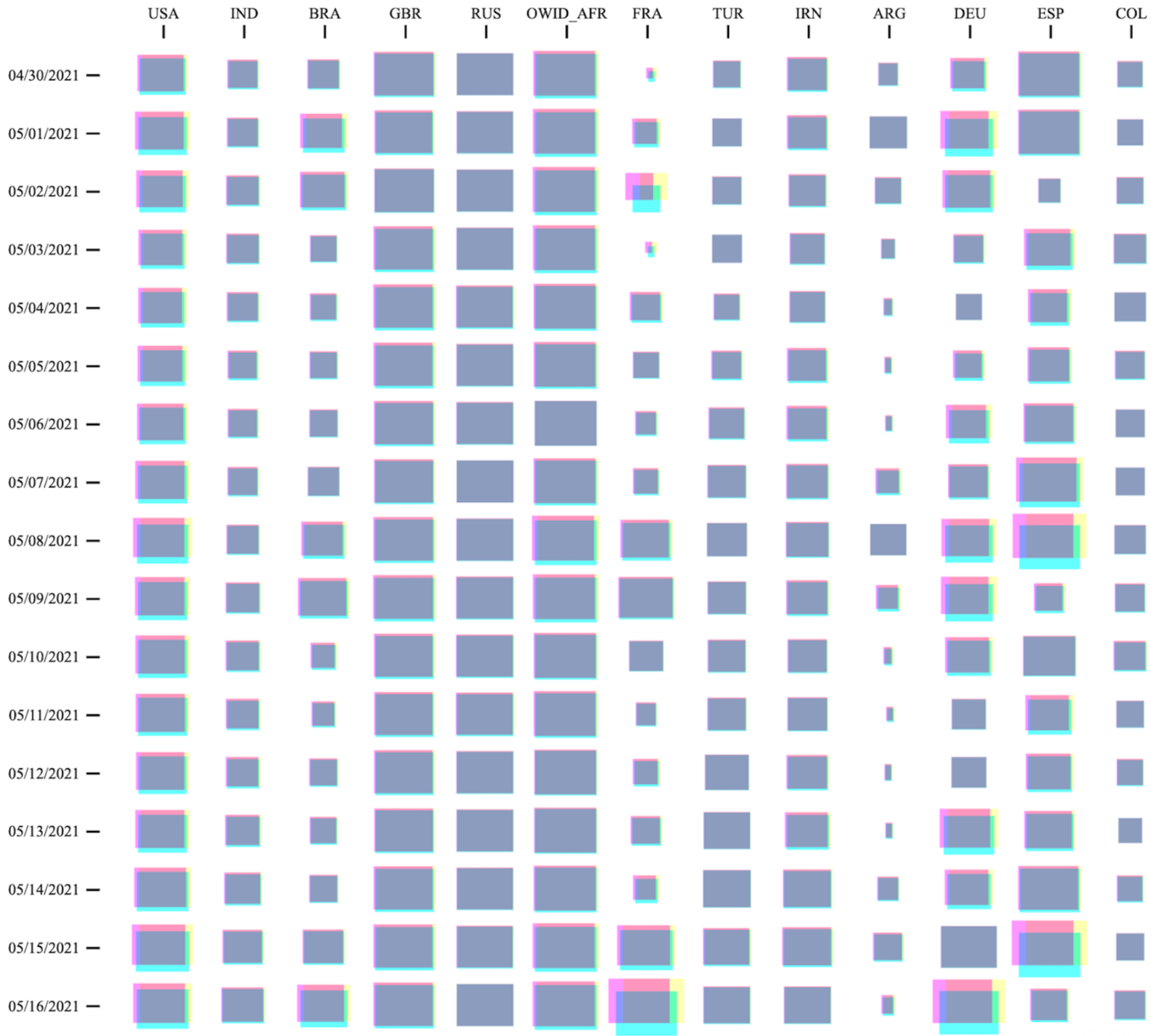


Figure 5.13: Charts of Daily Counts for different countries.

Here, the size of every cell represents the number of new cases whereas the Chromatic Aberration represents the uncertainty of the prediction for that date against the corresponding country. The interesting thing is uncertainty is independent of the predicted count, so there are some smaller cells convey high uncertainty and some bigger cells show lower uncertainty.

5.12 World Map

This is another version of a bubble chart that we experimented with in our designs. In this case all circles are drawn over the respective areas regarding their geographical position on world map.

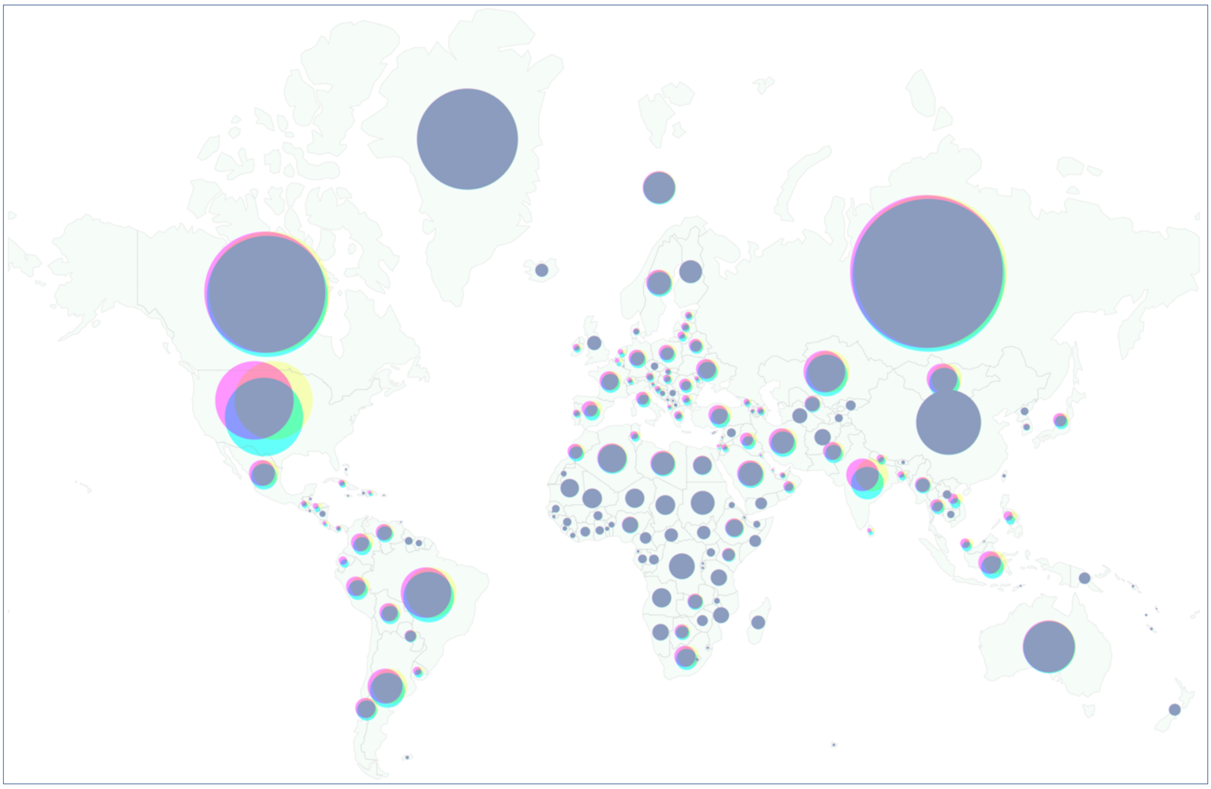


Figure 5.14: Uncertainty in World view

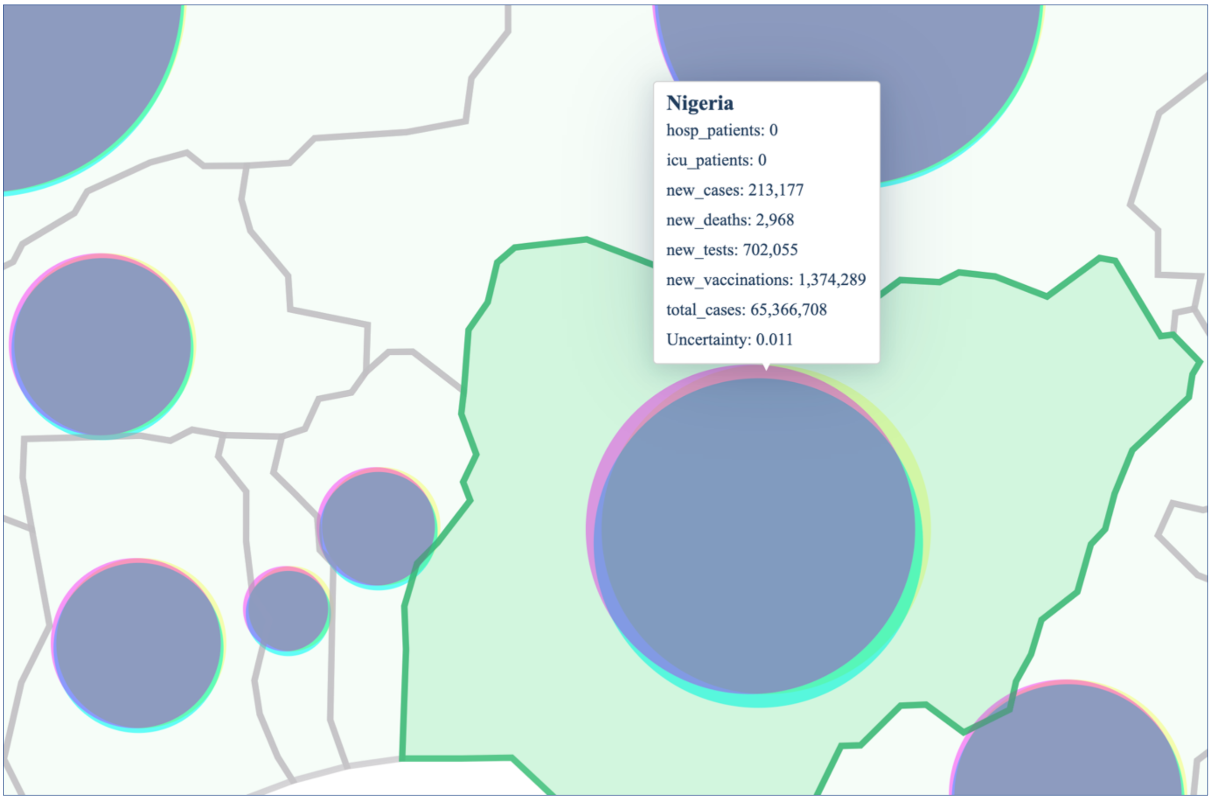


Figure 5.15: Zoomed World Map centering Nigeria

We have added some additional interactive functionalities on the map such panning, zoom in/out, hovering to highlight specific country boundary, show detail in a popup menu regarding the selected (by click) country. In the initial view, the uncertainty presentation is clearly visible only for the countries that have a higher number of counts and uncertainties. So, if the user zooms the map, then the bubbles and its edges are proportionately increased. On clicking the country, it shows the Covid related information along with uncertainty of the country.

5.13 Summary of Experimental Designs

In this chapter we presented a number of experimental designs. While several may be promising enough to refine and expand upon in the future, in order to produce focused research contribution, we converged on Chromatic Aberration and the evaluation of such a method for uncertainty visualization. The following chapters will discuss our design and analysis of a comparative user study with a recently published and prominent alternative.

Chapter 6

Evaluation: User Study Design

6.1 Introduction

Uncertainty visualisation is one of the complex challenges in the visualisation domain and designing a valid user study is also important. The study design usually prepares a particular set of questions that depends on the nature of the research, goal of the research, and the availability of resources, etc. There are various types of user studies such as experimental/interventional studies, descriptive studies, observational studies, within/between subject studies, and so on. We have conducted a within-subject comparative study with the following measures:

- Task time
- Error Rate
- Subjective assessments (NASA-TLX, SUS)

According to lam et al. [64] user performance is predominantly measured in terms of objectively measurable metrics such as time and error rate, yet it is also possible to measure subjective performance such as work quality. The commonly used metrics are task completion time and task accuracy. On the other hand, the goal of evaluating user experience is to understand to what extent the visualization supports the intended tasks as seen from the participants' eyes to provide subjective feedback and to probe for requirements and needs.

6.2 Background and Goal

We have implemented a novel approach of uncertainty visualization and uncertainty data is generated from some existing machine learning predictive models. We then visualise the data itself in web platform in terms of Chromatic Aberration in an interactive fashion. This simulated Chromatic Aberration (CA) artificially separates the Red, Green, and Blue components of colors spatially around visualisation elements, such as squares and circles. The effect is a particular kind of blurriness of color perception. The idea is that the more uncertainty there is in a single predicted datapoint, the more its visual representation will be affected at its outer edge by this artificial chromatic aberration, with the intent of conveying that sense of uncertainty to the viewer through the visual channel.

The purpose of this study is the test whether in fact chromatic aberration can be used successfully to represent uncertainty and determine how accurately viewers can estimate the degree of uncertainty based on a given level of chromatic aberration applied to representative visual elements of predicted data values. This will be determined interactively with users through a web-based visualization system. From our literature review, Correll et al. [35] also visualises uncertainty with an alternate approach called Value-Suppressing Uncertainty Palettes (VSUP). So, the prime goal of this study is to compare of CA with that existing approach VSUP.

6.3 Research Questions

For our research, we have several research questions:

1. How visualising uncertainty with Chromatic Aberration works compared to VSUP in terms of user perception and accuracy?
2. Which representation is more efficient in terms of user response times?
3. How do the two representations compare in terms of user preference?

6.4 Study Material

We have developed a dynamic webpage with the content of study materials to conduct the study session entirely remotely online. It helped to save both participant's and researcher's travelling time to meet in a common place and eliminate the risk of health issues due to pandemic which was still subject to restrictions at the time the study was designed and submitted to ethics. That's why it was mandatory for each participant to have a Computer/Laptop and a fast enough internet connection to share participant's screen and allow for uninterrupted audio conversation.

6.4.1 Technology and Browser Support

The webpage was developed with HTML, CSS, JavaScript, and D3.js for frontend and PHP in backend, deployed in the webspace (web.cs.dal.ca) allocated to the student by the Department of Computer Science, Dalhousie University. Since the webspace has public access over internet, anyone could access the page from anywhere which helped to remain inside the

COVID-19 safety guidelines defined by the Nova Scotia Health department and provincial authority.

We also note that we used CSS color blending to represent Chromatic Aberration which does not work properly in Google Chrome/Safari. It is a well-known issue that they cannot render the blended color properly and when there are large number of cells with color blending in a chart, Chrome often crashes. We found Firefox and Microsoft Edge works without issue and served our purpose, and Firefox/Edge are easy to install on any computer having an internet connection. For this reason, we made either Firefox/Edge mandatory for the participation.

6.4.2 Study Components

As already noted, VSUP is the closest publication that also presented a technique for uncertainty visualisation, but that paper only used a grid-chart representation. In other words, the smallest unit of their representation is a square shape. But in our study, we broadened the test cases somewhat using circles and squares and that's why we created the following core components of our study:

- **CA + Bubble**: Chromatic Aberration is applied on circles in a bubble chart.
- **CA + Grid**: Chromatic Aberration is applied on squares in a grid chart.
- **VSUP + Bubble**: Uncertainties are presented with circular shapes.
- **VSUP + Grid**: Uncertainties are presented with square shapes.

So, first two components use CA and the last two use VSUP representation. In other words, two representations are implemented in four different components with slightly modified approaches.

6.4.3 Counter Balancing

Each component consists of eight questions. The order of the questions is selected randomly which means no two participants would get the questions in same order and the components themselves were presented to the participant in "Balanced Latin Squares" method of counter balancing mechanism proposed [67] to give equal emphasis to each component throughout the study and balance the learning effect. We have explained in detail about the study design and questionnaire in Appendix-E but for reference Figure 6.1 shows an example of balanced-latin-square.

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

A	B	D	C
B	C	A	D
C	D	B	A
D	A	C	B

Figure 6.1: Latin Square (left), Balanced Latin Square (right)

If we consider the four components as A, B, C, and D then the first participant will have the order of the first row, second participant will have the order of second row, etc. This approach ensures, no two consecutive participants will get the same order of components and 8 participants among 32 will get each component first.

Many empirical evaluations of input devices or interaction techniques are comparative. A new device or technique is often compared against alternative devices or techniques. There are two common designs for such experiments *within-subjects design* and *between-subjects design*. We have used the former because we were able to test every component of the system by every participant.

However, counterbalancing conditions using a Latin Square does not fully eliminate the learning effect noted earlier. Note in Figure 6.1 (left), the 4 × 4 Latin square design, component A follows component B for three of the four groups of participants. Thus, there is a tendency for better performance on component B simply because most participants benefited from practice on Component A prior to testing on Component B. This phenomenon is mitigated using a *Balanced Latin Square* Figure 6.1 (right).

Let us consider the following abbreviation of the modules to easier accommodation in tabular representation of task arrangement of user study:

CA+Bubble = CB, CA+Grid = CG, VSUP+Bubble = VB, VSUP+Grid = VG

PID	Modules Order	Questions Order(random)
P1	CB => CG => VG => VB	Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8
P5		Q2, Q3, Q1, Q5, Q4, Q7, Q8, Q6
P9		Q4, Q2, Q3, Q8, Q5, Q6, Q7, Q1
P13		Q1, Q3, Q2, Q4, Q6, Q7, Q5, Q8
P17		Q6, Q4, Q7, Q3, Q8, Q2, Q1, Q5
P21		Q4, Q3, Q2, Q1, Q5, Q7, Q6, Q8
P25		Q3, Q1, Q5, Q4, Q2, Q7, Q8, Q6
P29		Q2, Q5, Q1, Q4, Q6, Q8, Q3, Q7
P2	CG => VB => CB => VG	Q5, Q4, Q8, Q2, Q1, Q7, Q6, Q3
P6		Q1, Q5, Q2, Q3, Q4, Q6, Q8, Q7
P10		Q4, Q2, Q3, Q8, Q5, Q6, Q7, Q1
P14		Q1, Q3, Q2, Q4, Q6, Q7, Q5, Q8
P18		Q6, Q4, Q7, Q3, Q8, Q2, Q1, Q5
P22		Q4, Q3, Q2, Q1, Q5, Q7, Q6, Q8
P26		Q3, Q1, Q5, Q4, Q2, Q7, Q8, Q6
P30		Q2, Q5, Q1, Q4, Q6, Q8, Q3, Q7
P3	VB => VG => CG => CB	Q2, Q1, Q5, Q4, Q3, Q7, Q8, Q6
P7		Q2, Q6, Q1, Q5, Q4, Q7, Q8, Q3
P11		Q4, Q2, Q3, Q8, Q5, Q6, Q1, Q7
P15		Q1, Q3, Q2, Q4, Q6, Q7, Q5, Q8
P19		Q6, Q4, Q7, Q3, Q8, Q2, Q1, Q5
P23		Q4, Q3, Q2, Q1, Q5, Q7, Q6, Q8
P27		Q3, Q1, Q5, Q4, Q2, Q7, Q8, Q6
P31		Q2, Q5, Q1, Q4, Q6, Q8, Q3, Q7
P4	VG => CB => VB => CG	Q7, Q8, Q3, Q4, Q5, Q6, Q1, Q2
P8		Q2, Q3, Q1, Q5, Q4, Q7, Q8, Q6
P12		Q8, Q2, Q3, Q4, Q5, Q6, Q7, Q1
P16		Q1, Q6, Q2, Q4, Q3, Q7, Q5, Q8
P20		Q6, Q4, Q7, Q3, Q8, Q2, Q1, Q5
P24		Q4, Q3, Q2, Q7, Q5, Q1, Q6, Q8
P28		Q3, Q1, Q5, Q4, Q6, Q7, Q8, Q2
P32		Q1, Q5, Q2, Q4, Q6, Q8, Q7, Q3

Table 6.1: Task arrangement of user study

6.5 Recruitment

Since the participants play a central role in any user study, it's important to find the suitable participants for the study based on the attributed research domain.

As we have four components in our study and each component has eight random questions, we decided to hire (4 x 8 = 32) participants to give equal emphasis to every component and questions. The detail procedure and considerations about recruitment are described in the following sub-sections.

6.5.1 Criteria

Given that our application is web-based and online, the population for our study is potentially all over the world including members of the Dalhousie University community. But we require participants to be fluent in English because there are questionnaires which needs to be understood correctly and answered accordingly. They all are at least post-secondary students or professionals who have some degree of computer experience as a user of common computer applications. In summary, here is the checklist of the criteria:

- **Age**: We wanted to eliminate participants of age lower than 17 years and higher than 60 years. Because younger participant might not have sufficient knowledge to understand the scope of the questions and elderly people are more likely to suffer from eyesight issues.
- **Education**: We required the minimum education level to be post-secondary level.
- **Experience**: We do not require any expertise in specific domains, but participants need to have minimum expertise in computer use, such as browsing websites.
- **Vision**: Participant's eyesight must be reasonable to detect objects and pass our color blindness test to participate in the study, explained in section 6.4.
- **Physical Ability**: Participants are not be disabled in a way which prevents them from using keyboard, mouse, browse the web or use computer.
- **Computer**: Participants must own a computer/laptop for the period of study session. Smartphones are not accepted to participate in the study due to the insufficient display size.
- **Internet**: Participants must have a good internet connection to continue the session without interruption with voice and video(screen) sharing.

6.5.2 Hiring Procedure

The internet is relatively cheap and accessible almost all over the world. So, with Covid-19 still an issue, we preferred to use online publicity as a recruitment strategy. We sent a recruitment notice to the Dalhousie University Computer Science undergraduate and graduate mailing list, Dalhousie Computer Science jobs email, physical bulletin boards on campus, in social media like LinkedIn. The recruitment notice outlined the study (process, eligibility criteria, data collection, compensation, and estimated time requirement) and instructions to contact the main researcher. Once a potential participant showed interest with a reply email, the main researcher emailed them with more detailed information and attached a consent form for their perusal and suggested to reply with three potential time slots if they agree with the detail requirements and a consent form content. Participation acceptance was done on a first-come first-serve basis until all places were booked. When participants either became sick or cancelled or did not continue interest up to the sessions, potential wait-list participants were called serially as per their time of participation confirmation.

6.5.3 Scheduling

On confirmation of participation interest by the participant, the main researcher created an MS Teams/Skype event with the agreed time. Participants received a notification in their inbox with the detail of the event including a URL for the event. At the meeting time participants just needed to open the link in either a browser or installed desktop application of the relevant tool to start the session.

The main researcher used a MS Word document as a logbook to manage, track and keep the study process synchronized. The researcher always ensured that two participants participation time could not overlap each other and tried to keep a gap of 30 minutes between two schedules so that unexpected delays could be mitigated by that buffer time.

6.6 Study Procedure

The study session contains several stages such as a color blindness test, modular sessions for each of the core components, module introductions and clarifications before starting a module, and post-session questionnaires after completing core modules. In the following section we explain them briefly.

6.6.1 Start of Event

Since the study was conducted online and schedules were made between two parties (researcher and participant) and an event was created through the online meeting platform or conferencing tool such as MS Teams. The participant just needed to be in front of a computer at the scheduled time of the day and click on the link he or she received in his or her email to meet in the meeting platform. When the participant logged-in to the system, it notifies the researcher that participant is waiting at the lobby, and he needs to be admitted. On approval, the event is instantly started at the online meeting room and both parties will be able to hear each other. Researcher greeted and welcomed the participant and exchanged their formal greetings. If participant faced any technical difficulty to join then the researcher tried to help by possible means.

6.6.2 Briefing

The researcher needed to brief the participant about the steps he or she had to go through and explained how he was going to conduct the session. Participants were also asked if his/her system has a Firefox/Edge browser installed which is mandatory for the study. If not he/she would be requested to install it and the researcher might instruct further if they needed any help regarding the installation of the browser. After confirming the browser is ready to go, the participant was requested to open it and informed him/her that researcher would give two URLs for the session i. for Color Blindness Test and ii. for the Questionnaire about the study.

6.6.3 Color Blindness Test

One of the prime requirements for the selection process is to test for color-blindness of the participants. The participants had to be capable to decern color to provide meaningful input through their participation in the study. To maintain similarity with Correll et al. [35], we presented a set of Ishihara plates [60] attached in Appendix F in a webpage. The URL of the page (Figure 6.1 shows a screenshot of the webform with plate and input fields) is given to the participant through chat (conversation) box of the conference tool and they were requested to fill the input field with what they saw in the image and click next to get next question. This would continue until it ended with all samples identified. We excluded those that misidentified values or who self-reported as having a color vision deficiency.

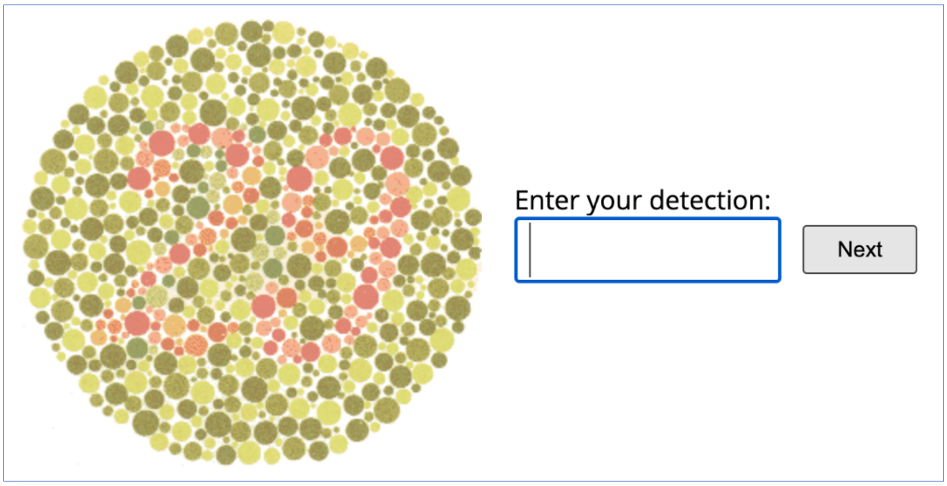


Figure 6.2: Example Color Plate in our portal

6.6.4 Pre-Session Discussion

After the color blindness test was passed successfully, the researcher asked participants about some basic questions which we thought to be relevant with their performance. Because it needs full concentration on the task to understand the question, find relevant values, and uncertainties. For instance, the following information were noted by the main researcher:

- Educational (science, arts, etc.) background
- Professional background (IT, Accountant, etc.)
- Computer skills (Basic, Intermediate, Expert)
- Mathematical and Geometric knowledge
- Visualization and Computer graphics knowledge
- Computer gaming skill
- Measurement knowledge (inch, feet, pixel, etc.)
- Physical Condition (Tired, Sleepy, Hungry, Fresh, etc.)

6.6.5 Overview of the Questionnaire Structure

There were two types of questions in our user study design, and they are as follows:

i. Component Questions
ii. Post Session Questionnaire (PSQ)

As noted previously in 6.2.2, we have four components in our study. So, by Component Questions we refer to the questions relevant to those four core components. On the other hand, PSQ refers to the questions to obtain user feedback from the experience of using the four core components of the system. PSQ includes System Usability Scale (SUS) test questions and NASA-TLX Work-Load Scale test questions.

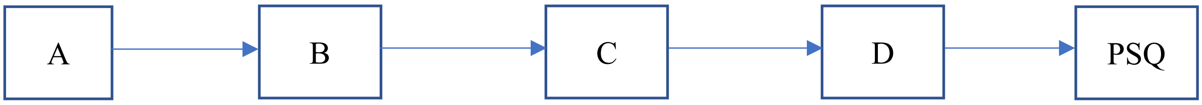


Figure 6.3: Flow of Questionnaires for a participant

If we consider A, B, C, and D as four components of the study then Figure 6.3 shows the flow of the components that come one after another randomly during the session of the participants as discussed in counter-balancing section. It also shows that PSQ appears at the completion of four modules.

At the beginning of every section, the bottom-right part the of the UI will show the Session description. The researcher will describe the features (chart, legend and how question will be asked and what does that mean, etc.). After completion of explanation, the participant is asked to hit 'Start' button as the following screen:

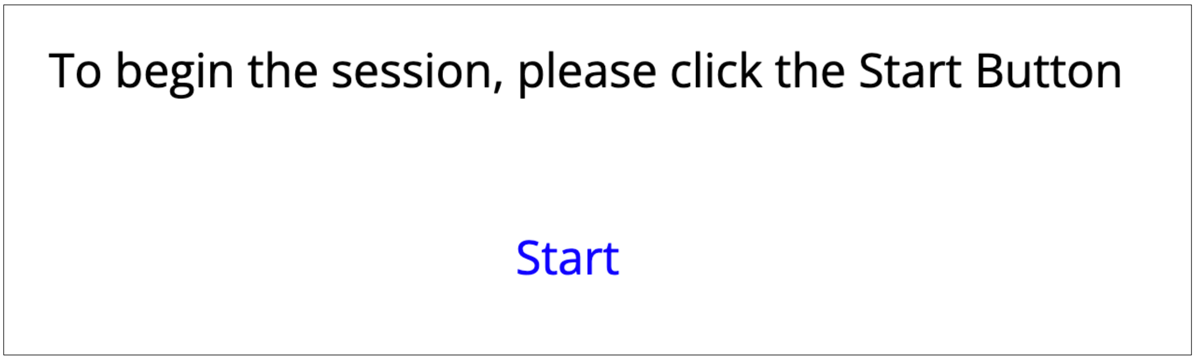


Figure 6.4: Module Start View

Once she or he presses the 'Start' button, the questionnaire will be started immediately and will present one question at a time. Figure 6.4 shows the overall layout of the questions and Figure 6.5 shows an example question.

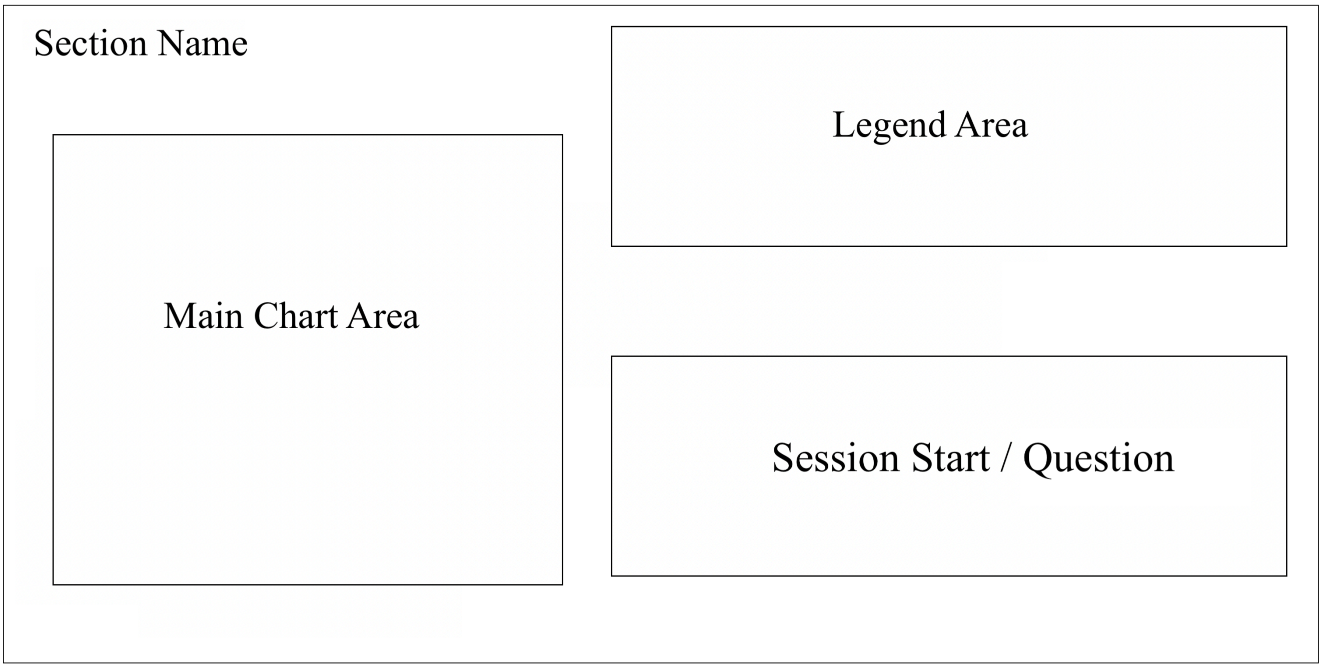


Figure 6.5: Layout of Questionnaire View

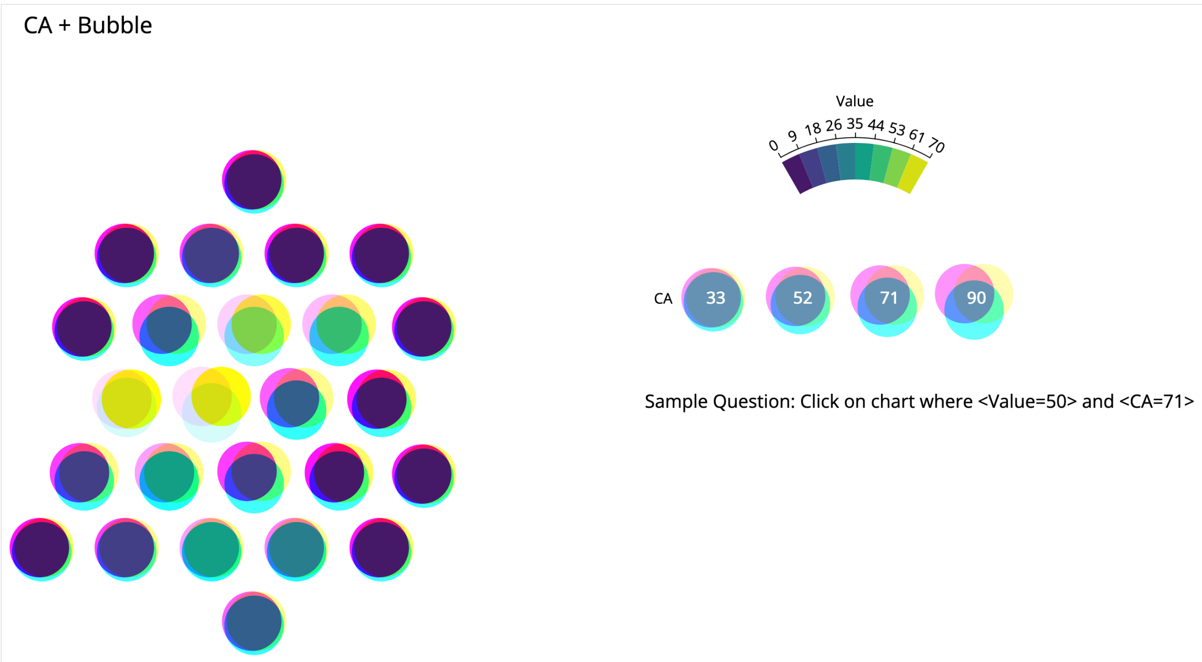


Figure 6.6: Sample Question (number inside CA bubbles denotes the amount of uncertainty represented by those chromatic circles. E.g., CA = 71 refers 71% uncertainty.)

When presented with a question, the user needs to select a cell (bubble or rectangle) from the chart based on the provided Value and Uncertainty/CA combination. An example question is shown in Figure 6.6. After a cell is selected by the user, the next question will appear at the same place until it reaches to eighth question of the section. We will return to the internal format of the component questions in the next section 6.6.6.

Since the bubble chart and the grid chart are two major components of this research and we have four sections with these two components, we present one example with identification procedure for a sample question prior to questionnaire of each section. Examples are given here for the reader of this document but in real application it was described verbally to the participant along with answering more questions if the participant may have. Orders of the questionnaire will be changed by counterbalancing stated above for different session users. So, these are the summary of the next sections:

1. Example of CA + Bubble
2. Questionnaire on CA + Bubble
3. Example of VSUP + Bubble
4. Questionnaire on VSUP + Bubble
5. Example of CA + Grid
6. Questionnaire on CA + Grid
7. Example of VSUP + Grid
8. Questionnaire on VSUP + Grid

Then we ask the following two types of additional questionnaires:

9. Questions on System Usability Scale (SUS)
10. Questions on NASA TLX

6.6.6 Component Questions

We now present a sampling of questions that were presented to the user, with additional explanatory information placed within them. We have not shown all questions here as the complete list can be found in APPENDIX E.

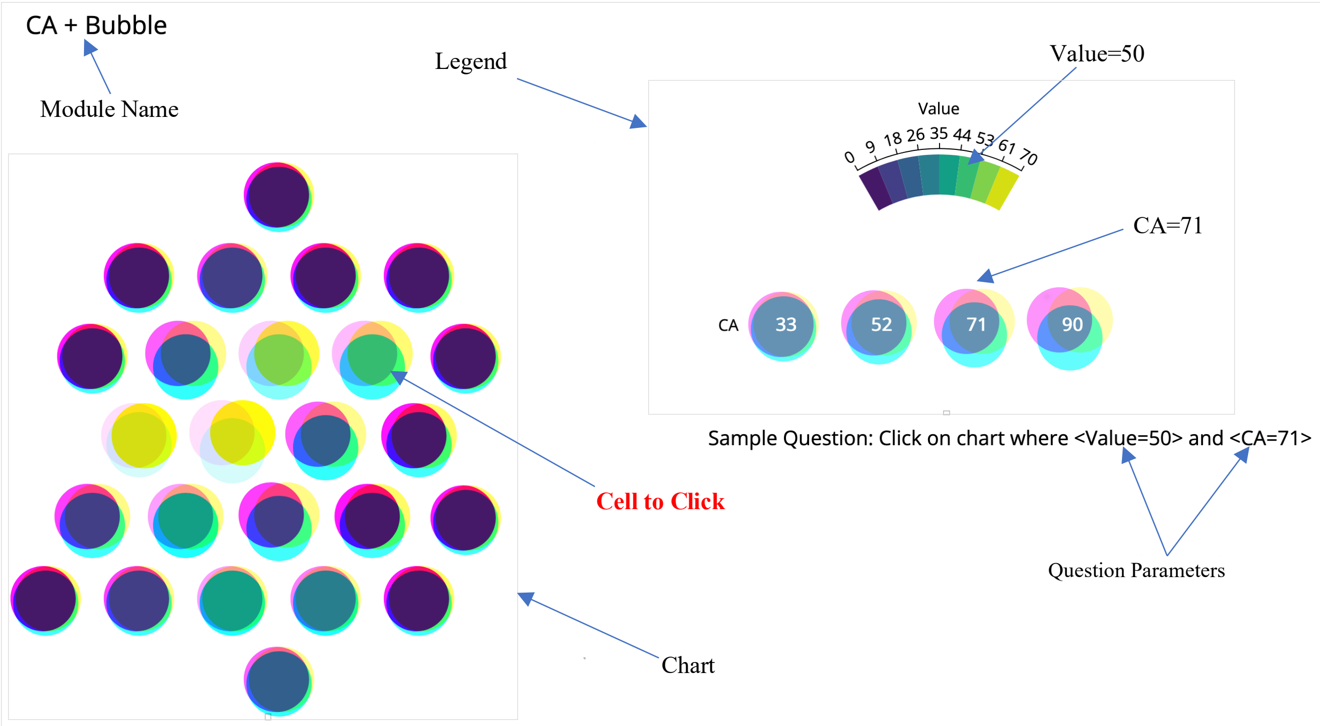


Figure 6.7: Question-Answer Identification Procedure for CA + Bubble

In the (CA+Bubble) example shown in Figure 6.7, we have introduced the various components with arrow indicators such as chart, legend, question parameters, detection of question parameters in the legend and finally, based on the parameter values, the target cell from the chart with the label 'Cell to Click', shown in red.

In this identification task the following aspects need to be considered by the user:

CA = The thickness of the colorful edges of the three overlapping circles

Value = Color of the common(center) portion of the three circles

Based on the above, participants need to answer both the value and the uncertainty simultaneously by clicking on the correct circle. We note that the researcher also explained the mechanism verbally before starting the module.

Figure 6.8 shows a very similar picture, with the only significant difference being that this question is using squares in a grid.

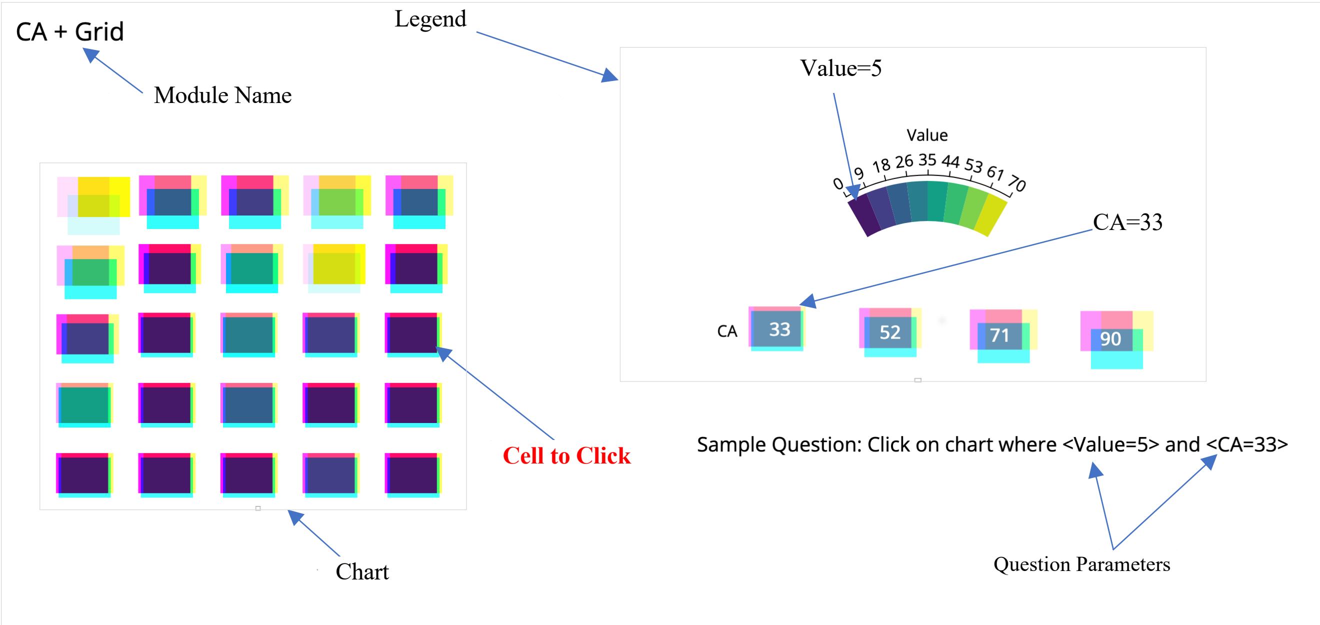


Figure 6.8: Question-Answer Identification Procedure for CA + Grid

In the (VSUP+Bubble) example shown in Figure 6.9, we have likewise introduced the various components with arrow indicators such as chart, legend, question parameters, detection of question parameters in the legend and finally, based on the parameter values, the target cell from the chart with the label ‘Cell to Click’, shown in red.

In this scenario, by using Uncertainty and Value, the user needs to target a single cell from the legend as indicated in Figure 6.9:

Uncertainty = Represents the vertical axis in the legend labeled by ‘Uncertainty’

Value = Represents the horizontal axis on the legend

Based on the above, participants again need to answer both the value and the uncertainty by clicking on the correct circle.

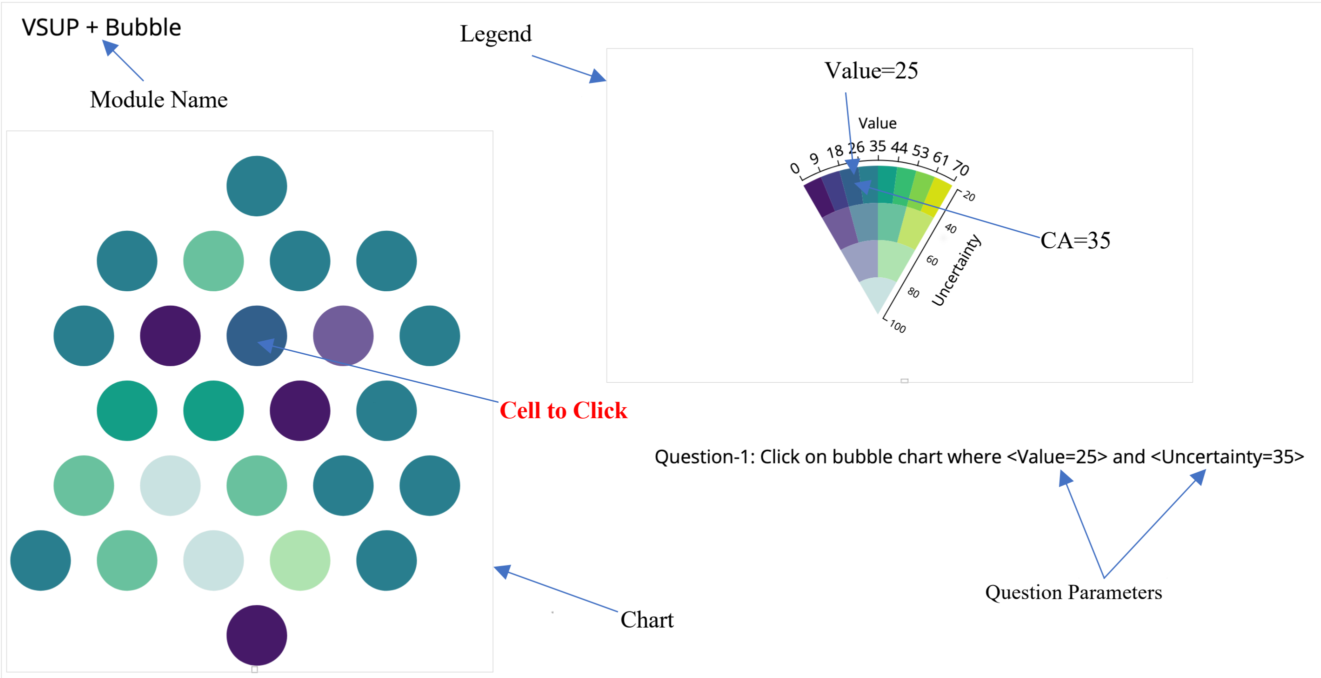


Figure 6.9: Question-Answer Identification Procedure for VSUP + Bubble

Figure 6.10 shows a similar picture to that of Figure 6.9, with only significant difference being that this question is using squares in a grid.

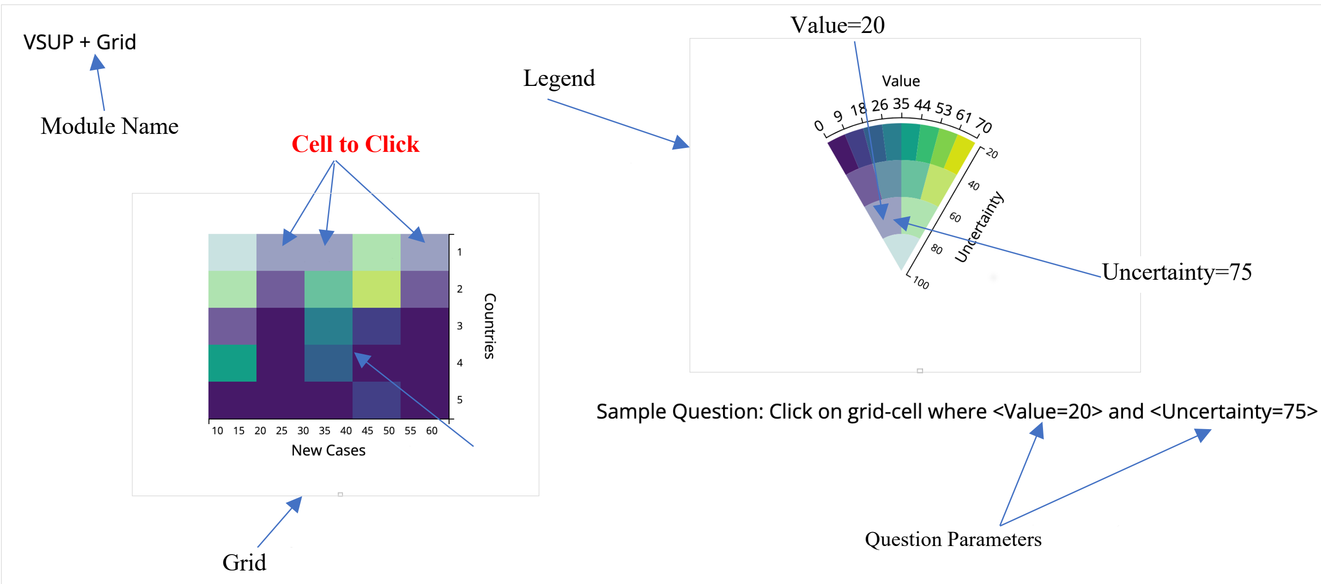


Figure 6.10: Question-Answer Identification Procedure for VSUP + Grid

6.6.7 Example PSQ Questions

We intentionally placed PSQ at the end of core modules so that participants could give their ratings based on their recent experiences gathered from the components. Since we have used our self-developed online web page to conduct the whole session, the system automatically and quantitatively captures the participant's answers and saves in memory from time to time and stores in the server at the end of the study. Figure 6.4 shows System Usability Scale test related question sample for both CA and VSUP.

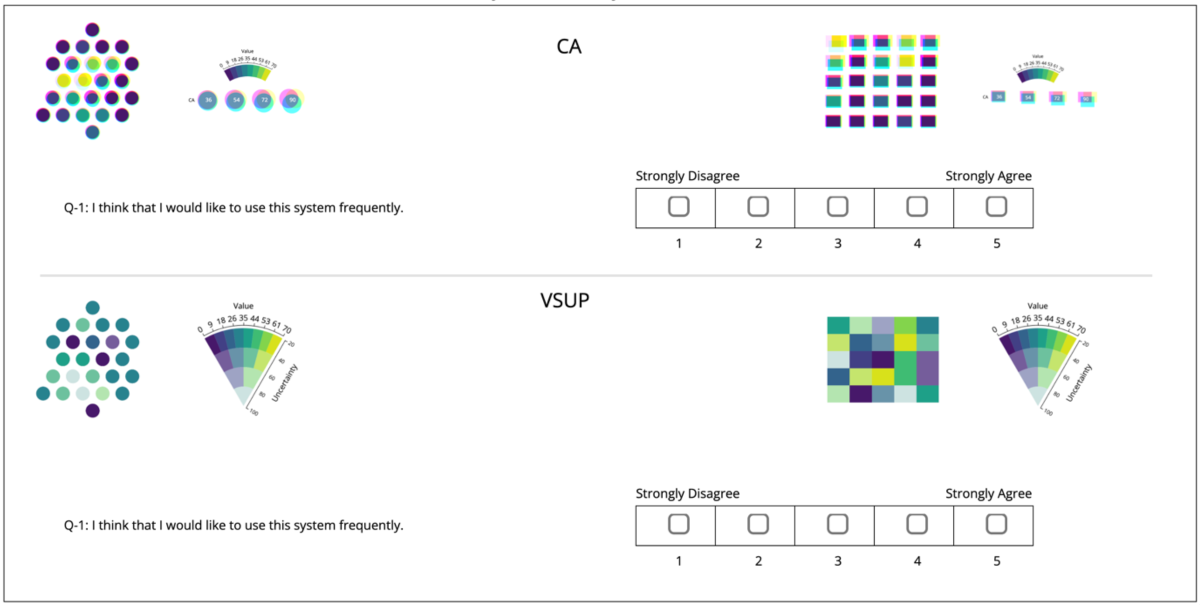


Figure 6.11: System Usability Scale (First Question Example)

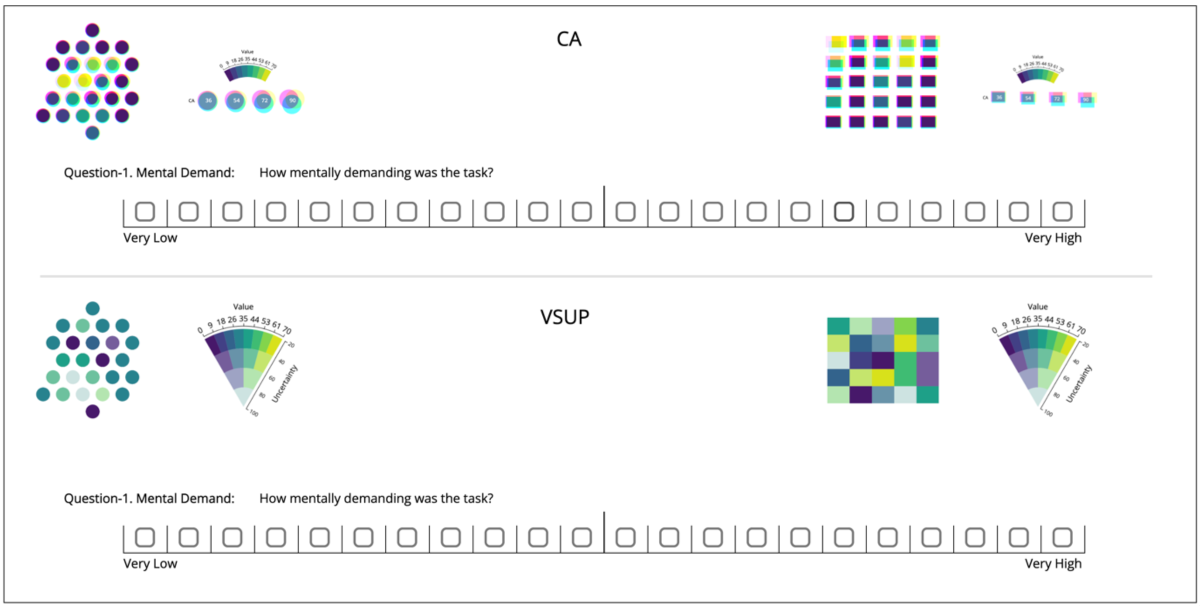


Figure 6.12: NASA-TLX Work-Load Scale (First Question Example)

Since the underlying mechanism is same for both CA+Bubble and CA+Grid, they are grouped together and placed at the top of the UI in the CA section. Similarly, VSUP+Bubble and VSUP+Grid are grouped together for the same reason and placed at the bottom in VSUP section of the UI. For both cases, we have shown the same question and the scale of answer is 1 to 5 where 1 means ‘Strongly Disagree’ and 5 means ‘Strongly Agree’ and the rest of the scales 2, 3, 4 carries in between weights based on their values.

Figure 6.5 shows NASA-TLX Work-Load test related questions where CA and VSUP sections are placed in the top and bottom respectively since they are two groups of four components just like SUS test explained in previous section. But in this case the number of questions is six and the scale range is 1(Very Low) to 22 (Very High). We have not shown all questions here for SUS or NASA-TLX as the complete list can be found in APPENDIX E. We can summarise the questionnaire as follows:

- Each component would appear to the participant conforming counter balancing rule as stated in 6.4.3.
- Every component had eight random ordered questions but the content of the question for every participant were same.
- The Post Session Questionnaire had two types of questions i. SUS and ii. NASA-TLX
- SUS had 10 questions and NASA-TLX has 6 questions.
- Total number of questions = 4 x 8 + 2 (10 + 6) = 64

6.7 Data Collection and Storing

We have developed the webpage, so we have implemented it in such a way that the system can automatically track the status of every answer whether correct or wrong. That means it keeps a record of every question from starting to end in a JSON object. The structure of the sample JSON data is given below:

```
Data = {
  participant-index: 1,
  ca-bubble: {
    Q1: {     // single variable and single target selection
      status: true,  ca: 52, mode: "single-var-one",
      selected: ['Vietnam'], options:
['Vietnam', 'Canada', 'Philippines', 'Kazakhstan', 'Palestine', 'Colombia']
    },
    …
    Q3: {     // single variable and all target selection
      status: true, ca: 90, mode: "single-var-all ",
      selected: ['Mongolia', 'Argentina', 'Russia', 'Peru'], options:
['Argentina', 'Mongolia', 'Peru', 'Russia']
    },
    …
    Q8: {     // double variable with single target selection
      status: true, ca: 71, mode: "double-var ",
      selected: ['Peru'], options: ['Mongolia', 'Peru']
    },
    single-var-one-time: 4.9,  // time required for questions of single variable-single answer
    single-var-all-time: 5.7,  // time required for questions of single variable-all answers
    double-var-time: 6.5  // time required for questions of double variable-single answer
  },
  ca-grid: {... same structure of ca-bubble   ...},
  vsup-bubble: {...  same structure of ca-bubble ...},
  vsup-grid: {... same structure of ca-bubble ....},
  nasa-ca: {1: 14, 2: 13, 3: 14, 4: 15, 5: 15, 6: 15}, // answers of NASA-TLX for CA components
  nasa-vsup: {1: 13, 2: 12, 3: 13, 4: 12, 5: 13, 6: 12},  // answers of NASA-TLX for VSUP components
  sus-ca: {1: 2, 2: 3, 3: 4, 4: 3, 5: 4, 6: 3, 7: 2, 8: 3, 9: 2, 10: 3},  // answers of SUS for CA components
  sus-vsup: {1: 2, 2: 1, 3: 2, 4: 1, 5: 2, 6: 3, 7: 2, 8: 3, 9: 2, 10: 3}  // answers of SUS for VSUP
components
}
}
```

After completion of the entire questionnaire, the generated JSON data is stored on the server with the email address provided by the participant. In the above sample structure, we see for every component it has its own block with the common set of properties for each. The above structure is designed with some self-descriptive properties that it would be helpful later in the results and numerical analysis phase.

6.8 Session Ending

Once the participant completed the post-session questionnaire, the page immediately informs the participant with following message:

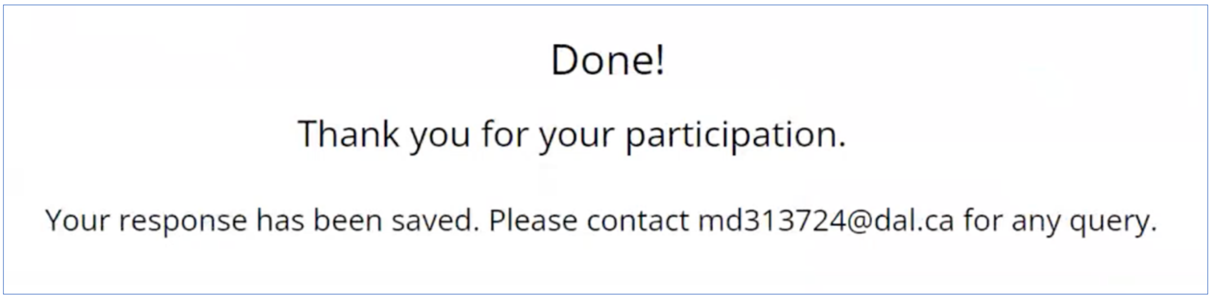


Figure 6.13: Session Ending Greetings

Finally, participants were given whole-hearted thanks for their dedication and participation in the study and immediately sent the promised $10 e-gift card (Amazon) to their email address. A sample of such gift card is attached in APPENDIX G.

Chapter 7

Evaluation: Results and Numerical Analysis

7.1 Introduction

In the previous chapter, we have described the user study design including questionnaire presentation, data collection procedure, data structure, and the data storing mechanism. In this chapter, we discuss and analyze the study generated data with the help of statistical principles which are commonly used for user studies such t-test, and ANNOVA. The goal of the study was to evaluate user performance and user experience of our newly designed approach of uncertainty visualisation and generate quantitative and qualitative (user preferences, SUS, etc.) data. We will use that data in analysis, prepare results, eventually discuss the findings, and finally come to conclusions.

7.1.1 Sample Population Demographics

Age and Gender

The sample population of 32 participants had a distribution of 78.12% male (25/32), 21.88% female (7/32). Given that we did not have any plan to control for gender within the recruitment policy, we have recruited on a first come first join basis. All participants were in the age range of 22-35 years old.

Education

There were 25% CS grad students (8/32), 28.12% CS undergrad students (9/32), 34.37% ICT grad students (11/32), 3% Statistics undergrad students (1/32) and 9.37% telecom professionals (3/32).

Prior experience in visualisation

The following experience with visualization was also noted:

- All CS and ICT students had taken at least one course of visualisation/graphics design in their undergraduate/graduate level and 12 of them had conducted their undergraduate thesis related to visualization or graphics or image processing.
- Telecom professionals also came from a CS background, so they had taken Computer graphics course in their undergraduate degrees.

- All participants had played computer games many times.
- 15 participants have knowledge of animated movies.

7.2 Study Results

We have obtained several kinds of data from the user study such as:

i. Quantitative Questionnaire Results
ii. Time utilization data for each component
iii. SUS data for CA and VSUP
iv. NASA-TLX for CA and VSUP

We analyse all these data in various ways in the following sections which helps to reach conclusions from the study.

7.2.1 Quantitative Questionnaire Results

As we have four core components, we designed the study content for each component individually and collected the log data for each component separately. As we already stated, there were 8 questions for each component and every question carried 1 point. For answering correctly, the participant gains one point and do no lose any points for wrong answers. So, a participant can gain minimum 0 point and maximum 8 points for a component. That point achievement is considered as the user performance of the study and we are going to analyse the user performance on the basis of ANOVA for four components and t-test for two grouped (CA and VSUP) components.

7.2.1.1 One-way repeated measures ANOVA

The user performance results that we received from the study can be summarized in Table 7.1 graphical box plots in Figure 7.2, and the complete raw data is attached in APPENDIX-I.

Groups	N	Mean	Std. Dev.	Variance	Std. Error.
CA + Bubble	32	6.2813	1.301	1.692	0.23
CA + Grid	32	5.5938	1.2916	1.668	0.2283
VSUP + Bubble	32	5.6563	1.4053	1.975	0.2127
VSUP + Grid	32	5.1875	1.2032	1.456	0.2127

Table 7.1: ANOVA Data summary

The results of a one-way ANOVA is considered reliable if the following assumptions hold:

1. the response variable (the dependent variable) is normally distributed.
2. the samples are independent.
3. the variances of populations are equal.

Since the sample are taken from independent interfaces of the questionnaire, requirement 2 fulfilled. Again, as per Keppel's ratio rule of thumb [70], if the ratio of the larger variance to the smaller variance is less than 1.5, then we can assume the variances are approximately equal. So, from Table 7.1, we see that variances are equal which conforms condition (3). Since conditions 2 and 3 are met, we need to ensure data is normally distributed. On this purpose, we conducted Shapiro-Wilk Normality Test and obtained results shown in Table 7.2 which indicates the distributions of the components are approximately in normal distribution which satisfies requirement (1) and we can conduct an ANOVA test. Additionally, we have also showed box-plot (Figure 7.1) and normal distribution graphs in Figure 7.2.

Component	**W**	**P**	**Status**
Ca + Bubble	0.915	0.015	Normal
Ca + Grid	0.932	0.045	Normal
VSUP + Bubble	0.911	0.012	Normal
VSUP + Grid	0.913	0.013	Normal

Table 7.2: Shapiro-Wilk Test of Normality

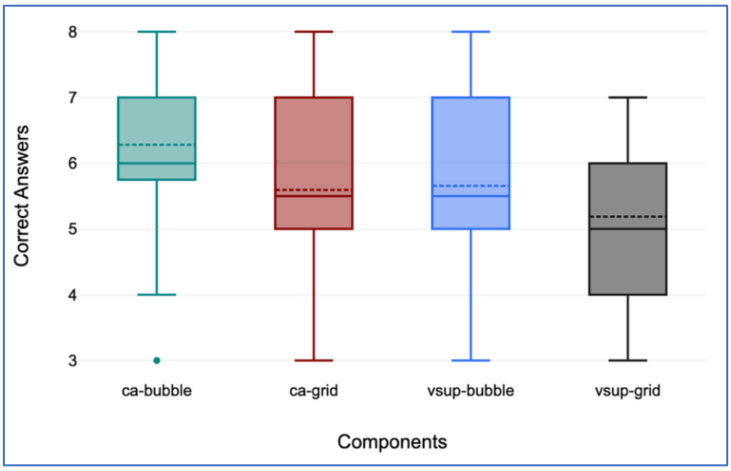


Figure 7.1: Box plot of user performance

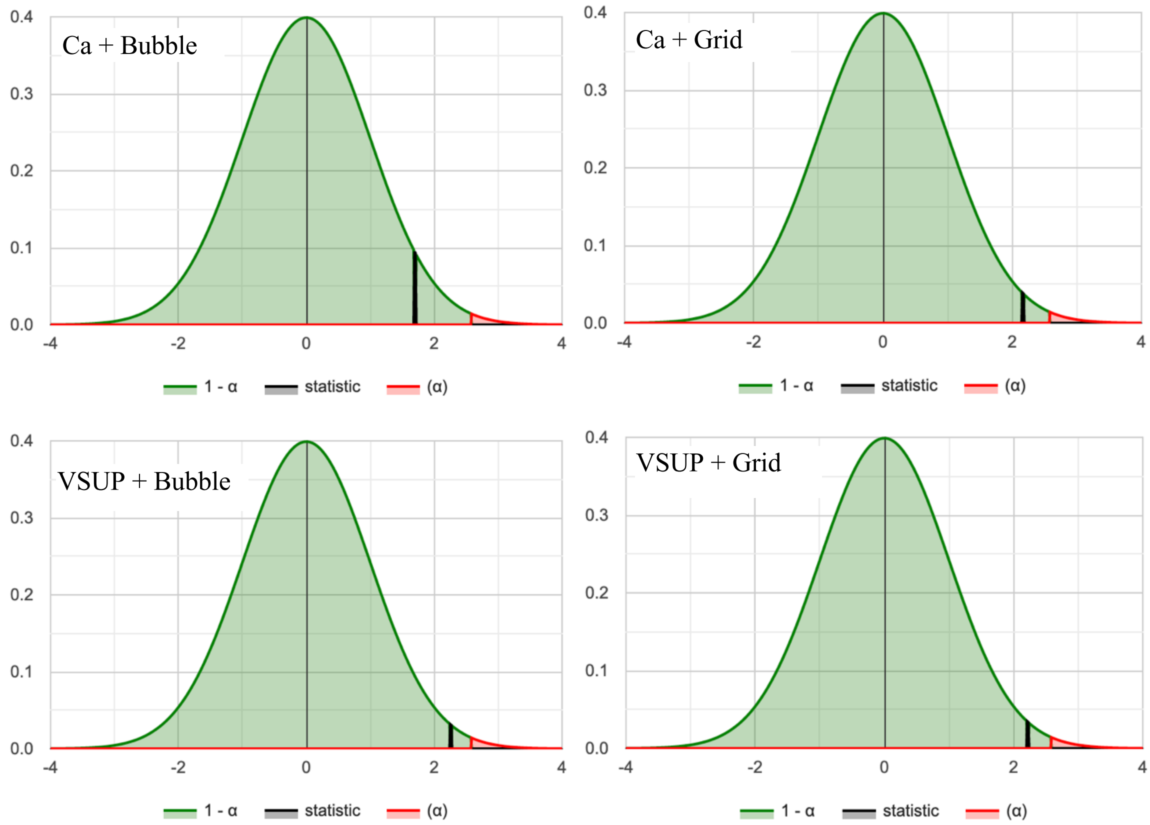


Figure 7.2: Normal Distributions for each component

We obtain the ANOVA summary as in Table 7.3.

Source	**Degrees of Freedom DF**	**Sum of Squares SS**	**Mean Square MS**	**F-Stat**	**P-Value**
Between Groups	3	19.5875	6.5292	3.8499	0.0113
Within Groups	124	210.2851	1.6958		
Total	127	229.8726			

Table 7.3: ANOVA Test Results Summary

So, we briefly point out the findings from the ANOVA test as follows:

(1) Null and Alternative Hypotheses

The following null and alternative hypotheses need to be tested:

Ho: $\mu1 = \mu2 = \mu3 = \mu4$ (Performances were equal for all components)

Ha: Not all means are equal (Performances were not equal for all components)

The above hypotheses will be tested using an F-ratio for a One-Way ANOVA.

(2) Rejection Region

Based on the information provided, the significance level is α=0.05, and the degrees of freedom are $df1$=3 and $df2$=3, therefore, the rejection region for this F-test is $R = \{F: F > 2.678\}$.

(3) Test Statistics

The computed test statistic F equals 3.8499, which is not in the 95% region of acceptance: [-∞: 2.678].

(4) Decision about the null hypothesis

p-value equals 0.0113, [p (x ≤ F) = 0.988735]. It means that the chance of type1 error (rejecting a correct H0) is small: 0.0113 (1.13%). The smaller the p-value the stronger it supports H1. Again, from the sample information we get that F = 3.85 > Fc=2.678, it is then concluded that *the null hypothesis is rejected.*

(5) Conclusion

It is concluded that the null hypothesis Ho *is rejected.* Therefore, there is not enough evidence to claim that all 4-population means are equal, at the $\alpha=0.05$ significance level. In other words, the difference between the averages of some groups is big enough to be statistically significant.

Figure 7.3 summarizes the results of the One-Way ANOVA. And from Table 7.1 we see, CA+Bubble has significantly higher means compared other distributions and CA+Grid has closer mean with VSUP+Bubble, and VSUP+Grid has significantly lower mean among all. So, we can conclude CA has significantly better user results compared to VSUP.

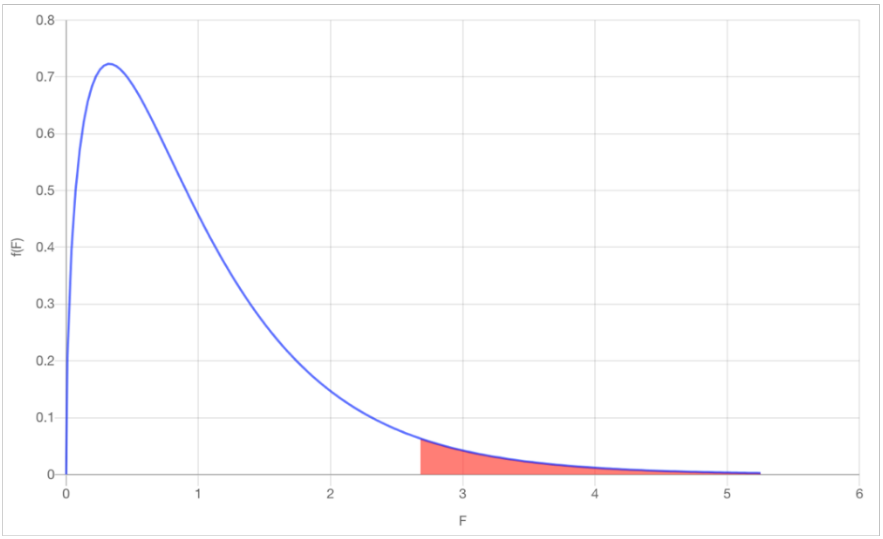


Figure 7.3: ANOVA Results: F=3.85, p-value=0.0113, Ho rejected.

7.2.1.2 Paired t-test

We have generated the CA and VSUP data from the four components performance data by grouping the two pairs (CA+Bubble with CA+Grid and VSUP+Bubble with VSUP+Grid).

Now the statistical summary of CA and VSUP data are shown in the following Table 7.4.

Group	CA	VSUP
Mean	5.938	5.422
SD	1.105	1.078
SEM	0.195	0.191
N	32	32

Table 7.4: Summary of CA vs VSUP performance

We present test result of Shapiro-Wilk normality test for significance level of 0.005 in the following table 7.6 where we see both distributions do not differ significantly from a normal distribution. We also show the normal distribution graphs in Figure 7.5.

Group	CA	VSUP
Skewness	-0.4622	0.07107
Kurtosis	-0.8658	-0.8737
p-value	.017	0.017
W	0.916	0.956

Table 7.5: Shapiro-Wilk Normality Test

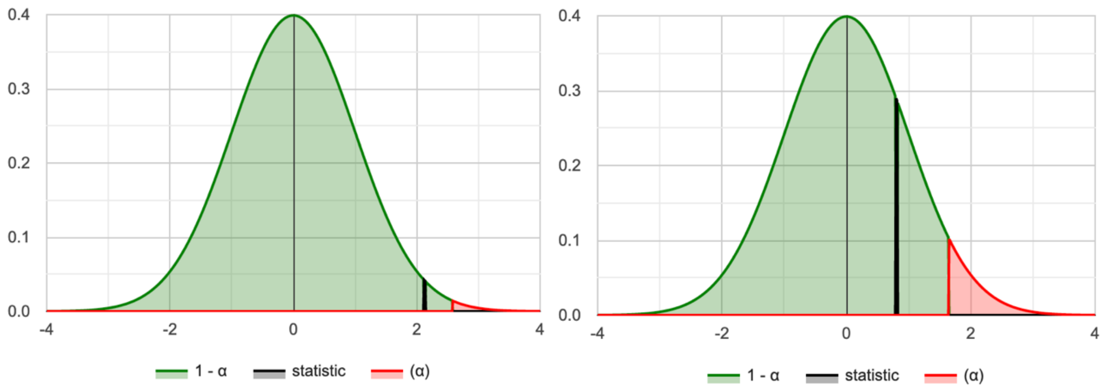


Figure 7.4: Shapiro-Wilk Normal Distribution CA (left), VSUP (right)

The following steps show the paired t-test results for the given data and draws conclusion from the test:

(1) Null and Alternative Hypotheses

The following null and alternative hypotheses need to be tested using paired t-test: Ho: $\mu_D = (\mu 1 - \mu 2) >= 0$ (performance of CA is higher or equal to performance of VSUP) Ha: $\mu_D = (\mu 1 - \mu 2) < 0$ (performance of CA is less than performance of VSUP) This corresponds to a left-tailed test, for which a t-test for two paired samples be used.

(2) Rejection Region

Based on the information provided, the significance level is α=0.05, and the critical value for a left-tailed test is $t_c = -1.696$.

The rejection region for this left-tailed test is $R=\{t : t < -1.696\}$

(3) Test Statistics

The computed t-statistic = 3.61

(4) Decision about the null hypothesis

Since it is observed that $t = 3.61 \geq t_c = -1.696$, it is then concluded that *the null hypothesis is not rejected.* Using the P-value approach: The p-value is $p = 0.9995$, and since $p = 0.9995 \geq 0.05$, it is concluded that the null hypothesis is not rejected.

(5) Conclusion

It is concluded that the null hypothesis Ho *is not rejected.*

Confidence Interval: The 95% confidence interval is $0.224 < \mu_D < 0.807$.

We can visualize the paired T-test scenario graphically as follows in Figure 7.5:

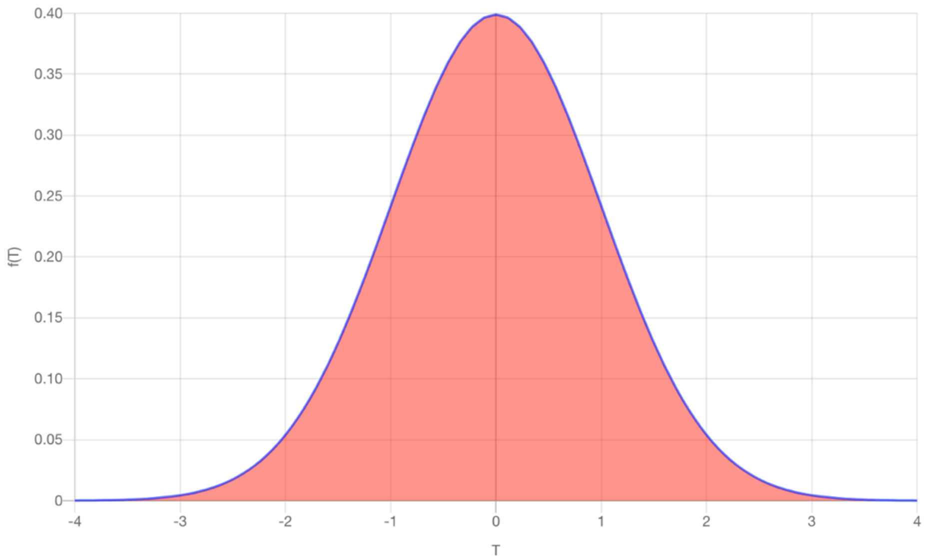


Figure 7.5: Paired t-test gaussian plot with p-value=0.9995 for CA vs VSUP performance.

Finally, based on above statistical test results, analysis and hypothesize conclusion, we can say that performance of CA quantitatively surpassed the performance of VSUP.

7.2.2 Time Utilization Results

Our automated system tracked effective response time for every component separately. The statistical summary of the timing data is represented in the following table 7.6

Group	CA	VSUP
Mean	8.675	9.647
SD	2.320	3.123
SEM	0.410	0.552
N	32	32

Table 7.6: Summary of CA vs VSUP timing

The Shapiro-Wilk tests on both distributions showed that they met the normality test with the following results:

For CA = $W(32) = 0.959$, $p = 0.254$

For VSUP = $W(32) = 0.977$, $p = 0.716$

The following steps show the paired t-test results for the given time data and draws conclusion from the test:

(1) Null and Alternative Hypotheses

The following null and alternative hypotheses need to be tested:

Ho: $\mu_D = (\mu 1 - \mu 2) <= 0$ (CA response was equal or faster than VSUP response)

Ha: $\mu_D = (\mu 1 - \mu 2) > 0$ (CA response was slower than VSUP response)

This corresponds to a right-tailed test, for which a t-test for two paired samples are used.

(2) Rejection Region

Based on the information provided, the significance level is $\alpha = 0.05$, and the critical value for a right-tailed test is $t_c = 1.696$.

The rejection region for this right-tailed test is $R=\{t : t > 1.696\}$

(3) Test Statistics

The computed t-statistic is equal to -2.656

(4) Decision about the null hypothesis

Since it is observed that $t = -2.656 \le t_c = 1.696$, it is then concluded that *the null hypothesis is not rejected.*

Using the P-value approach: The p-value is $p = 0.9938$, and since $p = 0.9938 \ge 0.05$, it is concluded that the null hypothesis is not rejected.

(5) Conclusion

It is concluded that the null hypothesis Ho *is not rejected.*

The 95% confidence interval is $-1.718 < \mu_D < -0.226$.

We can visualize the paired T-test scenario graphically as shown in Figure 7.6.

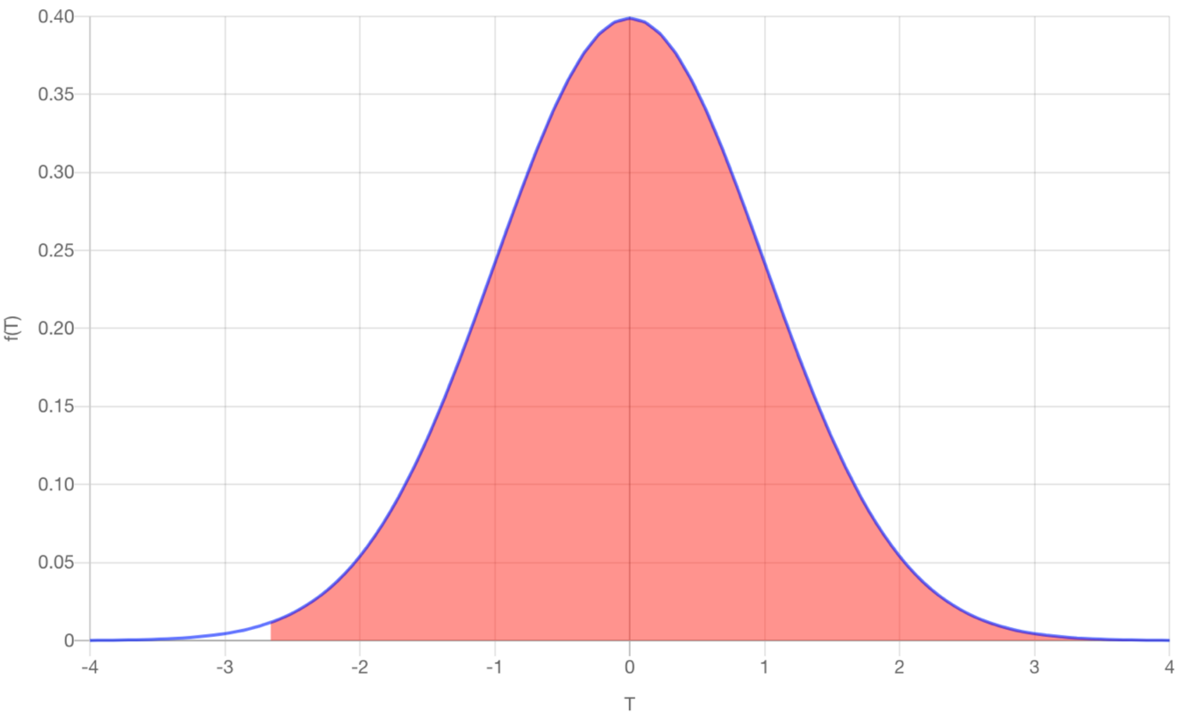


Figure 7.6: Paired t-test gaussian plot with p-value=0.9938 for CA vs VSUP timing.

Finally, based on above statistical test results, analysis and hypothesize conclusion, we can essentially say that user performance in CA method was faster than VSUP method.

7.2.3 SUS Results

The SUS provides a quick tool for measuring the usability of various kinds of systems based on user experience. It consists of a 10-item questionnaire with five scale response from participants starting from Strongly agree to Strongly disagree. Collectively its use is in classifying the ease of use of the system being tested. We will interpret the results by normalizing the scores to produce a percentile ranking. By convention of SUS scoring, based on Sauro [69], we converted SUS results to SUS scores by the following rules:

- For odd items: subtract one from the user response.
- For even-numbered items: subtract the user responses from 5
- This scales all values from 0 to 4 (with four being the most positive response).
- Add up the converted responses for each user and multiply that total by 2.5. This converts the range of possible values to a range from 0 to 100 instead of 0 to 40.

The statistical overview of the scores is given below in Table 7.7.

Group	CA	VSUP
Mean	60.078	61.094
SD	16.307	14.227
SEM	2.883	2.515
N	32	32

Table 7.7: SUS scores summary of CA vs VSUP

The Shapiro-Wilk tests on both distributions showed that they do not meet normality test with the following results:

For CA = W(32) = 0.913, p = 0.013

For VSUP = W(32) = 0.889, p = 0.003

The following steps show the Kruskal-Wallis Test results, which is non-parametric alternative to the paired t-test since the distributions are not normal. The purpose of the test is to assess whether or not the samples come from populations with the same population median.

(1) Null and Alternative Hypotheses

The following null and alternative hypotheses need to be tested:

Ho: The samples come from populations with equal medians.

Ha: The samples come from populations with medians that are not all equal.

The above hypotheses will be tested using the Kruskal-Wallis test.

(2) Rejection Region

Based on the information provided, the significance level is α=0.05, and the number of degrees of freedom is $df = 2 – 1 = 1$. Therefore, the rejection region for this Chi-Square test is $R = \{\chi2: \chi2 > 3.841\}$.

(3) Test Statistics

The computed H statistic is = 0.146

(4) Decision about the null hypothesis

Since it is observed that $\chi^2=0.146 \leq \chi_c2 = 3.841$, it is then concluded that the null hypothesis is not rejected. The p-value is $p = 0.702$, and since $p = 0.702 \geq 0.05$, it is concluded that the null hypothesis is not rejected.

(5) Conclusion

It is concluded that the null hypothesis *Ho* is not rejected.

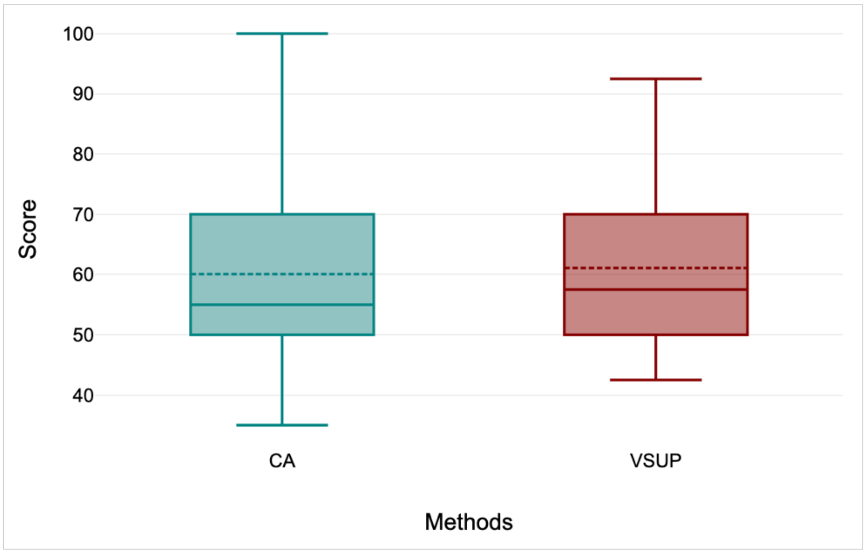


Figure 7.7: SUS rating plots for visualization methods

Finally, although the scores of the methods are slightly varying according to Figure 7.7, the differences ($\chi 2$ = 0.146, p = 0.702, df = 1) were not statistically significant as per Kruskal-Wallis test at $\alpha = 0.05$.

7.2.4 NASA-TLX Results

TLX stands for Task Load Index and is a measure of perceived workload. Just like SUS data, we have collected Nasa-TLX test data from our online system. A TLX method increments of high, medium, and low estimates for each point result in 21 gradations on the scales. To score, we subtract 1 from the given rating in the range of 1-21, and multiply by 5. For example, if user gives a rating 5, the score would be 20: (5-1) x 5.

Methods	NASA-TLX	Shapiro-Wilk Normality Test (α = 0.05)		
		Test Statistic (W)	**p-value**	**Status**
CA	Mental Demand	0.906	0.009	Not normal
	Physical Demand	0.914	0.014	Not normal
	Temporal Demand	0.948	0.128	Normal
	Performance	0.932	0.044	Not normal
	Effort	0.942	0.085	Not normal
	Mental Frustration	0.916	0.017	Not normal
VSUP	Mental Demand	0.863	0.001	Not normal
	Physical Demand	0.903	0.007	Not normal
	Temporal Demand	0.938	0.067	Not normal
	Performance	0.887	0.003	Not normal
	Effort	0.901	0.006	Not normal
	Mental Frustration	0.877	0.002	Not normal

Table 7.8: Normality test results of NASA-TLX score

Since almost the datasets didn't follow a normal distribution, we used the Kruskal-Wallis non-parametric test to evaluate the differences across the two methods of uncertainty representations (CA and VSUP) on NASA-TLX ratings. The following null and alternative hypotheses need to be tested with Kruskal-Wallis test.

Ho: The samples come from populations with equal medians
Ha: The samples come from populations with medians that are not all equal

The following table shows the summary of such test results of Kruskal-Wallis test at the α = 0.05 significance level:

NASA-TLX	X2	P	df	H	Conclusion
Mental Demand	0.19	0.6626	1	0.19	Not Rejected
Physical Demand	0.062	0.8038	1	0.062	Not Rejected
Temporal Demand	0.018	0.8932	1	0.018	Not Rejected
Performance	3.61	0.0574	1	3.61	Not Rejected
Effort	0.062	0.8038	1	0.062	Not Rejected
Mental Frustration	0.173	0.6772	1	0.173	Not Rejected

Table 7.9: Kruskal-Wallis test results of NASA-TLX

No statistically significant differences were found between the learning conditions on: mental demand ($\chi_2 = 0.19$, $p = 0.6626$, $df = 1$), physical demand ($\chi_2 = 0.62$, $p = 0.8038$, $df = 1$), temporal demand ($\chi_2 = 0.018$, $p = 0.8932$, $df = 1$), performance ($\chi_2 = 3.61$, $p = 0.0574$, $df = 1$), effort ($\chi_2 = 0.62$, $p = 0.8038$, $df = 1$), and mental frustration ($\chi_2 = 0.61$, $p = 0.6772$, $df = 1$) for the significance level $\alpha = 0.05$.

7.3 User Comments:

Although participants did not offer many informative comments, we note a few comments that were made during the experiment. Participants (4, 21) commented that "*CA representation is deterministically difficult*" but we also noted that in these cases the comment was the opposite of their performance given that they performed better in CA than VSUP. It is interesting, nonetheless. Some other participants (19, 24) made a more nuanced comment, stating that "*CA representation is complex but gives more confidence to find target*". Another comment that was commonly expressed by participants (14, 25, 31) is that "*Colors are very close in VSUP which made them puzzled to select target*".

7.4 Summary of the results

We obtained two types of results from the study i. Quantitative and ii. Qualitative. Based on a statistical analysis in the former sections, we can summarise as that the Quantitative results were better for CA than the VSUP approach whereas subjective results were not significantly different from each other. In other words, user performance and speed of target identification was significantly better in CA than VSUP although user preference was more or less similar.

Chapter 8

Limitations

There are several limitations of this work that we wish to highlight. Although we did not have a prerequisite for participants to be university students, based on the responses we received we obtained all participants from universities (undergraduate and graduate students). This may imply that results may not generalize to significantly younger or older demographics. Though we believe the VSUP study also had participants with similar educational levels.

Secondly, in the CA representation one needs to be careful so that chromatic objects and adjacent objects do not overlap. An additional level of care must be applied in case of implementing zooming based on the zoom-scale of the visualization in order to keep them consistent.

Thirdly, although our approximation of CA is computationally inexpensive and accessible for use in widely used web-based visualizations (i.e. d3), if one were to implement a more complex CA rendering method then further study with participants may be required. In particular, we note that in real world chromatic aberration the chromatic blurring appears continuously from inner edge to outer edge.

Chapter 9

Conclusions and Future Work

In this thesis, we propose a novel approach for uncertainty visualisation, namely Chromatic Aberration. We conducted a within subject comparative user study with VSUP and our system to assess user performance accuracy/error rate, task completion time, and subjective assessment with NASA-TLX and SUS. From numerical analysis and evaluation of the results, we see user performance and perception is both statistically improved and faster compared to VSUP whereas in the subjective assessment do not vary significantly.

Again, by analyzing the incorrect answers from the user study, we notice that in both representations the correctness of uncertainty detection is higher than the value detection. Most probably, this happens due to the number of values (8) is higher than the number of levels of uncertainties (4). Similarly, comparing CA and VSUP representations for the same, we notice that in VSUP suffers from more erroneous detection compared to the CA representation.

Although we stated the absence of blurriness in our CA representation as limitation, however, our simplified implementation allows us to reduce the aberration to both double and/or single parameter, which facilitates chromatic aberration tuning with regards to the amount of represented uncertainty. It also allows one to implement the approach relatively easily using standard d3 and SVG operations. However, additional research could be conducted that examine more sophisticated effects. In addition, further research could be conducted with more levels of uncertainties than were tested in both in Correll et al. [35] and the present work, for instance 8-levels instead of 4-levels. The role of CA might also be explored in animated visualizations. And finally, other future work may refine and expand upon some of our other experimental designs such as the starfish streamgraph layout briefly discussed.

References:

[1] Song, Xin; Xiao, Jun PhD; Deng, Jiang PhD; Kang, et al. Time series analysis of influenza incidence in Chinese provinces from 2004 to 2011. Received March 1, 2016, Accepted May 20, 2016, Medicine: June 2016 - Volume 95 - Issue 26 - p e3929

[2] Muhammad Ali, Dost Muhammad Khan, et al. Forecasting COVID-19 in Pakistan, received: August 17, 2020; Accepted: November 10, 2020; Published: November 30, 2020.

[3] COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population. Appl. Sci. 2020, 10(11), 3880; Received: 5 May 2020 / Revised: 23 May 2020 / Accepted: 29 May 2020 / Published: 3 June 2020

[4] Christophorus Beneditto, Aditya Satrio et al. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET, https://doi.org/10.1016/j.procs.2021.01.036

[5] Leo J, Luhanga E, Michael K. Machine Learning Model for Imbalanced Cholera Dataset in Tanzania. The Scientific World Journal. 2019 Jul; 2019: p. 1–12.

[6] Emrah Gecili, Assem Ziady, Rhonda D. Szczesniak. Forecasting COVID-19 confirmed cases, deaths and recoveries: Revisiting established time series modeling through novel applications for the USA and Italy. Received: June 30, 2020; Accepted: December 5, 2020; Published: January 7, 2021.

[7] Sathler C, Luciano J. Predictive modeling of dengue fever epidemics: A Neural Network Approach. 2017. Data Science for Drug Discovery, Health and Translational Medicine. December 10, 2017. I590.

[8] Miranda GHB, Baetens JM, Bossuyt N, Bruno OM, Baets BD. Real-time prediction of influenza outbreaks in Belgium. Epidemics. 2019 Sep; 28: p. 100341.

[9] Muktevi Srivenkatesh, Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver Disease. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-2, December 2019.

[10] K. Koh, B. Kim & J. Seo. 2014. Effect of lateral chromatic aberration for chart reading in information visualization on display devices. Advanced Visual Interfaces. Como, Italy, 289-292.

[11] H. S. Yoo. 2007. Color illusions on liquid crystal displays and design guidelines for information visualization. Master of Science, Virginia Tech.

[12] T. Boult & W. Wolberg. 1992. Correcting chromatic aberrations using image warping. CVPR, Champaign, IL, 684–87.

[13] M. K. Johnson & H. Farid. 2006. Exposing digital forgeries through chromatic aberration. Multimedia and security, Geneva, Switzerland, 48-55.

[14] M. Hullin, E. Eisemann H.P. Seidel & S. Lee. 2011. Physically-based real-time lens flare rendering. ACM SIGGRAPH, Vancouver, 108:1–108:9.

[15] S. Lee, E. Eisemann & H.P. Seidel. 2010. Real-time lens blur effects and focus control. ACM SIGGRAPH, Los Angeles, 1-7.

[16] Bonneau et al. Overview and State-of-the-Art of Uncertainty Visualization, The University of Grenoble, France e-mail: Georges-Pierre.Bonneau@ujf-grenoble. fr. ISBN: 978-1-4471-6496-8

[17] Simon Barthelme, Pascal Mamassian. Evaluation of Objective Uncertainty in the Visual System. Received June 8, 2009; Accepted August 12, 2009; Published September 11, 2009.

[18] Henning Griethe et al. The Visualization of Uncertain Data: Methods and Problems. Computer Graphics, 18051 Rostock, Germany. January 2006, 2,988 reads, 36 publications, 262 citations.

[19] Deitrick, S., Edsall, R.: The influence of uncertainty visualization on decision making: An empirical evaluation. In: Progress in Spatial Data Handling, pp. 719–738. Springer Berlin Heidelberg (2006).

[20] Lundstr¨om, C., Ljung, P., Persson, A., Ynnerman, A.: Uncertainty visualization in medical volume rendering using probabilistic animation. IEEE Transactions on Visualization and Computer Graphics 13(6), 1648–1655 (2007).

[21] Pang, A., Wittenbrink, C., Lodha., S.: Approaches to uncertainty visualization. The Visual Computer 13(8), 370–390 (1997).

[22] Rudolf Netzel and Daniel Weiskopf, Texture Based Flow VisualizaTion. November 2013, Computing in Science and Engineering 15(6): 96-102,

[23] Jesus J. Caban, Alark Joshi, and Penny Rheingans. Texture-based feature tracking for effective time-varying data visualization, IEEE Transactions on Visualization and Computer Graphics (Volume: 13, Issue: 6, Nov.-Dec. 2007). Page(s): 1472 – 1479.

[24] Sven Bachthaler, Daniel Weiskopf. Animation of Orthogonal Texture Patterns for Vector Field Visualization. IEEE Transactions on Visualization and Computer Graphics (Volume: 14, Issue: 4, July-Aug. 2008), Page(s): 741 – 755.

[25] Jin Huang, Zherong Pan, Guoning Chen, Wei Chen, Hujun Bao. Image-Space Texture-Based Output-Coherent Surface Flow Visualization. IEEE Transactions on Visualization and Computer Graphics (Volume: 19, Issue: 9, Sept. 2013). Page(s): 1476 – 1487.

[26] Andrea Kratz, Daniel Baum, and Ingrid Hotz. Anisotropic Sampling of Planar and Two-Manifold Domains for Texture Generation and Glyph Distribution. IEEE Transactions

on Visualization and Computer Graphics (Volume: 19, Issue: 11, Nov. 2013). Page(s): 1782 – 1794.

[27] D. Weiskopf. On the role of color in the perception of motion in animated visualizations. Conference: 10-15 Oct. 2004, Austin, TX, USA. IEEE Visualization 2004. ISBN: 0-7803-8788-0.

[28] C.G. Healey; J.T. Enns. Building perceptual textures to visualize multidimensional datasets. 18-23 Oct. Research Triangle Park, NC, USA. 1998. Proceedings Visualization '98 (Cat. No.98CB36276). ISBN: 0-8186-9176-X.

[29] R.P. Botchen; D. Weiskopf; T. Ertl. Texture-based visualization of uncertainty in flow fields. VIS 05. IEEE Visualization, Minneapolis, MN, USA. 23-28 Oct. 2005. ISBN: 0-7803-9462-3.

[30] Aasim Kamal · Parashar Dhakal, et al. Recent advances and challenges in uncertainty visualization: a survey. May 2021, Journal of Visualization 24(5):1-30.

[31] Galit Shmueli, Kenneth C. Lichtendahl Jr. Practical Time Series Forecasting with R: A Hands-On Guide [2nd Edition] (Practical Analytics) Paperback – July 19, 2016. Page 18-19. ISBN-13 978-0997847918

[32] Jason Brownlee. Deep Learning Models for Univariate Time Series Forecasting. https://machinelearningmastery.com/how-to-develop-deep-learning-models-for-univariate-time-series-forecasting.

[33] Aayush Agrawal, Building Neural Network from scratch. https://towardsdatascience.com/building-neural-network-from-scratch-9c88535bf8e9

[34] Taiwo Oladipupo Ayodele. Types of Machine Learning Algorithms. University of Portmouth. United Kingdom.

[35] Michael Correll, Dominik Moritz, Jeffrey Heer. Value-Suppressing Uncertainty Palettes. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. April 2018. Paper No.: 642 Pages 1–11.

[36] Jessica Hullman. Why Authors Don't Visualize Uncertainty. IEEE Transactions on Visualization and Computer Graphic. Jan. 2020, pp. 130-139, vol. 26.

[37] Shunan Guo, Fan Du, Sana Malik, et al. Visualizing Uncertainty and Alternatives in Event Sequence Predictions. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. May 2019. Paper No.: 573. Pages 1–12.

[38] Michelle Korporaal, Ian T. Ruginski, and Sara Irina Fabrikant. Effects of Uncertainty Visualization on Map-Based Decision Making Under Time Pressure. Human-Media Interaction, a section of the journal Frontiers in Computer Science. Received: 22 May 2020. doi: 10.3389/fcomp.2020.00032.

[39] Max Schneider, Michelle McDowell et al. Effective uncertainty visualization for aftershock forecast maps. Natural Hazards and Earth System Sciences. Discussion started: 3 September 2021. https://doi.org/10.5194/nhess-2021-237.

[40] Ken Brodlie, Rodolfo Allendes Osorio, and Adriano Lopes. 2012. A review of uncertainty in data visualization. In Expanding the frontiers of visual analytics and visualization. Springer, 81–109. DOI: http://dx.doi.org/10.1007/978-1-4471-2804-5_6

[41] Michael Correll and Michael Gleicher. 2014. Error bars considered harmful: Exploring alternate encodings for mean and error. IEEE Transactions on Visualization and Computer Graphics 20, 12 (2014), 2142–2151. DOI: • http://dx.doi.org/10.1109/TVCG.2014.2346298

[42] Miriam Greis, Passant El Agroudy, et al. 2016. Decision-Making under Uncertainty: How the Amount of Presented Uncertainty Influences User Behavior. In Proceedings of the 9th Nordic Conference on Human-Computer Interaction. ACM, 52. DOI: http://dx.doi.org/10.1145/2971485.2971535

[43] Lydia R Lucchesi and Christopher K Wikle. 2017. Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. Stat (2017). DOI:http://dx.doi.org/10.1002/sta4.150

[44] Alan M MacEachren, Robert E Roth, James O'Brien, Bonan Li, Derek Swingley, and Mark Gahegan. 2012. Visual semiotics & uncertainty visualization: An empirical study. IEEE Transactions on Visualization and Computer Graphics 18, 12 (2012), 2496–2505. DOI: http://dx.doi.org/10.1109/TVCG.2012.279

[45] Maria Riveiro. 2007. Evaluation of uncertainty visualization techniques for information fusion. In 10th International Conference on Information Fusion. IEEE, 1–8. DOI: http://dx.doi.org/10.1109/ICIF.2007.4408049

[46] Judi Thomson, Elizabeth Hetzler, Alan MacEachren, Mark Gahegan, and Misha Pavel. 2005. A typology for visualizing uncertainty. In Electronic Imaging 2005. International Society for Optics and Photonics, 146–157.

[47] N. Boukhelifa, M.-E. Perrin, S. Huron, and J. Eagan. How data workers cope with uncertainty: A task characterisation study. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pages 3645–3656. ACM, 2017.

[48] 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, 25(1):903–913, 2019.

[49] R. Finger and A. M. Bisantz. Utilizing graphical formats to convey uncertainty in a decision-making task. Theoretical Issues in Ergonomics Science, 3(1):1–25, 2002.

[50] J. Hullman, P. Resnick, and E. Adar. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. PloS one, 10(11):e0142444, 2015.

[51] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson. When (ish) is my bus?: User-centered visualizations of uncertainty in everyday, mobile predictive systems. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 5092–5103. ACM, 2016.

[52] M. Fernandes, L.Walls, S. Munson, J. Hullman, and M. Kay. Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, page 144. ACM, 2018.

[53] M. Skeels, B. Lee, G. Smith, and G. G. Robertson. Revealing uncertainty for information visualization. Information Visualization, 9(1):70– 81, 2010.

[54] C. M. Wittenbrink, A. T. Pang, and S. K. Lodha. Glyphs for visualizing uncertainty in vector fields. IEEE transactions on Visualization and Computer Graphics, 2(3):266–279, 1996.

[55] Z. Wang et al. Model identification of reduced order fluid dynamics systems using deep learning. International Journal for Numerical Methods in Fluids. Int. J. Numer. Meth. Fluids 2014; 00:1-18.

[56] Alan Conrad Bovik et al. Multichannel Texture Analysis Using Localized Spatial Filters. received April 18, 1988; revised June 15, 1989. Recommended for acceptance by W. E. L. Grimson. Department of Electrical and Computer Engineering, University of Texas, Austin, TX 78712.

[57] Ying Tang, Huamin Qu et al. Natural Textures for Weather Data Visualization. August 2006. Information Visualization, 2006. IV 2006. Tenth International Conference.

[58] Olga Scrivner, Vinita Chakilam, et al. Topic Analysis through Streamgraph via Shiny Application: a Social Collaborative Approach Proceedings of the 51st Hawaii International Conference on System Sciences | 2018. ISBN: 978-0-9981331-1-9

[59] Stéfan van der Walt and Nathaniel Smith. 2015. Mpl colormaps. https://bids.github.io/colormap/, (2015).

[60] LeGrand H Hardy, Gertrude Rand, and M Catherine Rittler. 1945. Tests for the detection and analysis of color-blindness. I. The Ishihara test: an evaluation. JOSA 35, 4 (1945), 268–275.

[61] Brooke, J. (1986). SUS: a "quick and dirty" usability scale. In P. W. Jordan; B. Thomas; B. A. Weerdmeester; A. L. McClelland (eds.). Usability Evaluation in Industry. London: Taylor and Francis.

[62] NASA (1986). Nasa Task Load Index (TLX) v. 1.0 Manual.

[63] Shapiro, S. S.; Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". Biometrika. 52 (3–4): 591–611.

[64] Heidi Lam, Enrico Bertini, et al. Empirical Studies in Information Visualization: Seven Scenarios. IEEE Electronic Library (IEL) Journals. 06 December 2011. Page(s): 1520 – 1536. ISSN: 1077-2626. IEEE Electronic Library (IEL) Journals

[65] S. Greenberg and B. Buxton, “Usability Evaluation Considered Harmful (Some of the Time),” Proc. Conf. Human Factors in Computing Systems (CHI), pp. 217-224, 2008.

[66] J. McGrath, “Methodology Matters: Doing Research in the Behavioral and Social Sciences,” Readings in Human-Computer Interaction: Toward the Year 2000, Morgan Kaufmann, 1994.

[67] I. Scott MacKenzie, Within-subjects vs. Between-subjects Designs: Which to Use? https://www.yorku.ca/mack/RN-Counterbalancing.html

[68] Shapiro, S. S.; Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". Biometrika. 52 (3–4): 591–611.

[69] Jeff Sauro, PhD. Measuring Usability with the System Usability Scale (SUS). https://measuringu.com/sus. February 3, 2011.

[70] Keppel, Geoffrey: Design and Analysis, 3rd Edition. Prentice-Hall.

DALHOUSIE
UNIVERSITY

RESEARCH ETHICS BOARDS

APPLICATION FORM

Prospective Research

This form should only be used if new data will be collected. For research involving only secondary use of existing information (such as health records, student records, survey data or biological materials), use the *REB Application Form – Secondary Use of Information for Research.*
This form should be completed using the *Guidance for Submitting an Application for Research Ethics Review*.

SECTION 1. ADMINISTRATIVE INFORMATION [File No: 2022-6028 office only]

Indicate the preferred Research Ethics Board to review this research:
[ ] Health Sciences OR [X] Social Sciences and Humanities

Project Title: Visualizing Uncertainty with Chromatic Aberration

1.1 Research team information				
Lead researcher (at Dalhousie)	Name	Md Rashidul Islam		
	Email (@dal)	md313724@dal.ca	Phone	902-448-3533
	Banner #	B00870359	Academic Unit	Comp Sci
Co-investigator names, affiliations, and email addresses	Dr. Stephen Brooks, sbrooks@cs.dal.ca			
Contact person for this submission (if not lead researcher)	Name			
	Email		Phone	
Study start date	March 12, 2022	Study end date	March 25, 2022	

1.2 For student submissions (including medical residents and postdoctoral fellows)			
Degree program	Master of Computer Science		
Supervisor name and department	Dr. Stephen Brooks		
Supervisor Email (@dal)	sbrooks@cs.dal.ca	Phone	902-494-2512
Department/unit ethics review (if applicable). **Undergraduate minimal risk research only**.			
Attestation: [ ] I am responsible for the unit-level research ethics review of this project and it has been approved. Authorizing name: Date:			

1.3 Other reviews		
Other ethics review (if any) for this research	Where?	
	Status?	
Scholarly/scientific peer review (if any)		
Is this a variation on, or extension of, a previously approved Dal REB submission?	[X] No [ ] Yes Dal REB file #________________	
If yes, describe which components of the current submission are the same as the previously approved submission (list section numbers), and which components are different from the previously approved submission (list section numbers). You may also use highlighting to clearly indicate revised text.		

1.4 Funding		[x] Not Applicable
Funding (list on consent form)	Agency	
	Award Number	
	Institution where funds are/will be held	[ ] Dalhousie University [ ] Other: ______________________
	Was a Dal release of funds agreement issued for this award?	[ ] Yes Date of RoF Agreement: ____________

1.5 Attestation(s). The appropriate boxes *must* be checked for the submission to be accepted by the REB

[X] I am the **lead researcher** (at Dalhousie) named in section 1.1. I agree to conduct this research following the principles of the Tri-Council Policy Statement *Ethical Conduct for Research Involving Humans* (TCPS) and consistent with the University *Policy on the Ethical Conduct of Research Involving Humans*.
I have completed the TCPS Course on Research Ethics (CORE) online tutorial.
[X] Yes [ ] No

For Supervisors (of student / learner research projects):
[X] I am the **supervisor** named in section 1.2. I have reviewed this submission, including the scholarly merit of the research, and believe it is sound and appropriate. I take responsibility for ensuring this research is conducted following the principles of the TCPS and University Policy.
I have completed the TCPS Course on Research Ethics (CORE) online tutorial.
[X] Yes [ ] No

SECTION 2. PROJECT DESCRIPTION

2.1 Lay summary

2.1.1 In **plain language**, describe the rationale, purpose, study population and methods to be used. Include a summary of background information or literature to contextualize the study. What new knowledge, or public or scientific benefit is anticipated? [maximum 500 words]

Visualization is a way of representing digital information to the user as a collection of shapes or lines such as circles, rectangles and curves. Each of the visual shapes represents some aspect of the data. For example, a circle's size might represent the population of a country. However, some data has uncertainty and in some cases, we may want to also incorporate the uncertainty into the visual elements in the charts. For example, we may want to make a circle blurry if the data it represents is uncertain to some degree. In this work we are introducing a new technique to visualize such information uncertainties in computer display called Chromatic Aberration (CA).

In our visualization, the CA for a visual element (such as a circle) will be created with Red, Green, and Blue versions of that circle. But the position of the Red, Green, and Blue versions will be separated from each other, where the amount of separation is determined by the amount of uncertainty in the data. The resultant circle would have an outer edge which will look like a colourful blur. The thickness of that outer edge is made proportional to the amount of uncertainty. The prime concern of the study is to detect how well the participants could perceive the level of uncertainty based on the thickness of the colour-blurred edges.

One common source of uncertainty comes when attempting to predict the future. Guesses about future data values always includes some amount of uncertainty. Common examples of forecasting future events include weather prediction, traffic congestion, and outbreaks of transmissible diseases. This is also true of COVID-19 data forecasting. We have used four computational methods for prediction to estimate future pandemic data values. Moreover, the predictions from these models are represented by visualizations that incorporated uncertainty.

To assess our new Chromatic Aberration approach for visualizing uncertainty, we have designed a study to investigate whether our new technique (CA) can be used successfully to represent uncertainty and determine how accurately viewers can detect those levels of uncertainty in the charts. In particular, we will compare CA with an existing approach called VSUP [Correll et al., 2018], which relies solely on a customized colour palette for representing uncertainty. The comparative evaluation will be conducted interactively with users through our online website.

The potential new knowledge will a novel method of data visualization, which may be applicable to a wider variety of applications. The aim of the study will be to produce a journal paper that will report the suitability of chromatic aberration for this purpose.

[ ] This is a pilot study.
[X] This is a fully developed study.

2.1.2 Phased review. If a phased review is being requested, describe why this is appropriate for this study, and which phase(s) are included for approval in this application. Refer to the guidance document before requesting a phased review.
[x] Not applicable

2.2 Research question

State the research question(s) or research objective(s).

The focus of the research is to calculate uncertainty from the forecasted results of machine learning predictive models and then represent these uncertainties in visualization in terms of chromatic aberration. In particular, we will conduct a comparative evaluation of visual uncertainty representations: our proposed chromatic aberration method and Value-Suppressing Uncertainty Palettes (VSUP) [Correll et al., 2018].

2.3 Recruitment

2.3.1 Identify the study population. Describe and justify any inclusion / exclusion criteria. Also describe how many participants are needed and how this was determined.

The population for our study will include members of the Dalhousie University community but may extend beyond to other universities and to the general public. We also require participants to be fluent in English because there will be questionnaires and interviews.

One of the prime criterions for the selection process is to test for color-blindness of the participants. The participants must be capable to decern color in order to provide meaningful data for the study. As in Correll et al. [2018] we will "present participants with a set of Ishihara plates [Hardy 1945], and exclude those that misidentified values or who self-reported as having a color vision deficiency". See Appendix F.

The study population will be at least post-secondary students or professionals who have some degree of computer experience as a user of common computer applications. In particular, they must have some knowledge of how to use the internet because the study will be conducted online. The study program will be deployed on a server and participants need to make sure they have internet connection with their computer or laptop, and they can access and use it through the freely available Firefox browser.

We aim to recruit 32 participants. We have four sections in the survey. Each component has 8 questions. We have used counter balancing among four sections as well as 8 questions of each section. The order of the sections are presented using a balanced-latin-square approach and questions will come up randomly within each section. To ensure equal priority of the components and to make the study fair, we decided to select (4x8=32) participants.

2.3.2 Describe recruitment plans and append recruitment instruments. Describe who will be doing the recruitment and what actions they will take, including any screening procedures.

Recruitment will be conducted by the primary researcher under the supervision of the supervisor. Due to COVID restrictions imposed by the provincial authority and for the sake of respecting health priority of Dal community, we decided to contact with the participants through email and digital messaging boards. Participants will initially be recruited through Dalhousie's digital message boards, including Notice Digest (notice.digest@dal.ca), the Computer Science Mailing List (cs.all@dal.ca) and the Dal Students emails (dalstudent@dal.ca) and physical bulletin boards on campus. If necessary, further recruits will be sought from similar message boards at other Canadian universities as well as message boards used in the data visualization community.

When potential participants respond to the recruitment notice, we will email them the inclusion criteria (English fluency, some experience with computers, full color vision) to assure that they meet the inclusion criteria. The screening email is given in Appendix C.

2.3.3 If you require permission, cooperation, or participation from a community, organization or company to recruit your participants, describe the agreement obtained from the relevant group(s). Attach correspondence indicating their cooperation and/or support (required). Describe any other community consent or support needed to conduct this research. (If the research involves Indigenous communities complete section 2.11).

[x] Not applicable

2.4 Informed consent process

2.4.1 Describe the informed consent process:

A) How, when and by whom will the study information be conveyed to prospective participants? How will the researcher ensure prospective participants are fully informed?

Prospective participants will receive a copy of the consent letter (Appendix B) and second email (Appendix C) after they indicated an interest in participating in the study by going through "Initial Email or Poster in Bulletin Board on Campus" (Appendix C). They will be instructed to read the consent letter before giving their consent. The email will also indicate to prospective participants that they can ask clarifying questions regarding the study.

B) Describe how consent will be documented (e.g. written signature, audio-recorded, etc.).

Consent will be recorded by based on the participants email responses after going through the consent letter and the second detailed email from the study investigator.

Participants who opt not to provide consent and not to participate will not be able to participate in the survey later.

[X] Append copies of all consent information that will be used (e.g. written consent document, oral consent script, assent document/script, etc.).

Note: If the research will involve third party consent (with or without participant assent), and/or ongoing consent, ensure these are described above.

2.4.2 Discuss how participants will be given the opportunity to withdraw their participation (and/or their data) and any time (or content) limitations on this. If participants will not have opportunity to withdraw their participation and/or their data explain why.

Participants are informed in the consent form and that they can withdraw from the study at any time.

Participants can opt to withdraw their data from the study up to 1 week after the interview because after that we will use the data in de-identified form (with numerical analysis) for our report. If a participant opts to withdraw in time from the study, their questionnaire, survey response, all recorded audio and video (screenshare only) will also be deleted permanently.

2.4.3 If an alteration/exception to the requirement to seek prior informed consent is sought, address the criteria in TCPS article 3.7A. If the alteration involves deception or nondisclosure, also complete section 2.4.4.

[X] Not applicable

2.4.4 Describe and justify any use of deception or nondisclosure and explain how participants will be debriefed.

[X] Not applicable

2.5 Methods, data collection and analysis

2.5.1

A) Where will the research be conducted?

Research will be conducted remotely using Microsoft Teams or Skype based on the participant's convenience. So, the participants will be able to participate in the study from their homes.

B) What will participants be asked to do?

We have developed a web application and for this reason participants do not need to install any specialized software in his/her own machine for this study other than a browser (Firefox) and a communicating medium (software), for example: Skype. We will provide a URL and then participants will be asked to navigate to the application. From there they will directly access the visualizations.

They will be asked to share their screen using Skype or MS Teams to show the running application. They will then be given short interactive tasks that will give them instruction on how to interact with the methods of visually representing uncertainty. The same tasks will be done when evaluating the competing uncertainty visualization approaches. Images of the screenshots are given in Appendix E along with the Questionnaire and examples of our implementations in different kind of charts. Each section of the questionnaire will therefore be proceeded by a short explanation session. The participants need to observe and understand the features to answer questions correctly, and we will brief them to understand the contents whenever needed.

More specifically, the general study design will have four sections of accuracy comparison between Chromatic Aberration (CA) with prior work VSUP and named them as:

- CA + Bubble
- VSUP + Bubble
- CA + Grid
- VSUP + Grid

C) What data will be collected using what research instruments? (Note that privacy and confidentiality of data will be covered in section 2.6)

The following data will be collected:

- Answers to the questionnaire questions (Appendix E) will be collected online using our self-developed web application. So, user will go through the questionnaire through browser and answer the question and our system will automatically track the response and save as JSON in file system with their email address.
- Video (screenshare only) and audio recording of screenshare session will be collected using Microsoft Teams or Skype.
- Timing information will also be recorded to facilitate a comparison of the time requirements of each competing visualization approach.

It will be a within subject experiment to "control for the variation in the interpersonal differences" [Correll 2018]. Moreover, we will use sequence counter balancing to counter act any learning effects within each subject.

D) How much of the participant's time will participation in the study require?

Approximately 1 hour will be required. The participant should go through the presentation in dynamic web application for up to 30 minutes and 30 minutes is anticipated for the completion of questionnaire section which is the main component of the survey.

[x] Append copies of all research instruments (questionnaires, focus group questions, standardized measures, etc)
[ ] This is a clinical trial (physical or mental health intervention) – ensure section 2.12 is completed

2.5.2 Briefly describe the data analysis plan. Indicate how the proposed data analyses address the study's primary objectives or research questions.

There will be several aspects to data collection and analysis. First of all, we will record the audio of participants while they perform tasks; we anticipate that participants could express frustration or describe their intensions while performing the tasks. We will use this information to help interpret the log data, described next. Second, the timings of the required tasks will be recorded automatically by our system. Third, there will a color vision test and tasks related questionnaires [Appendix E] to fill out by participants. Fourth, after all the tasks are finished, they will answer a questionnaire about the experience of using our application for two selection techniques. The questions will include the System Usability Scale (SUS) [Brooks 1986] and the NASA-TLX [NASA 1986] standardized questionnaires.

The timing of finishing tasks and user estimates of values and uncertainty will be objective quantitative measures of performance; responses to rating questionnaires will give us subjective quantitative data such as the degree of user satisfaction with the interface and confidence about finishing tasks. The data will also give us information that will allow us to further refine our interface in future work.

Post-session questionnaires (Appendix E) will provide us with additional feedback, which might not be apparent in the previous data and will help us to understand the preference of a participant's choices more comprehensively in terms of these visualization techniques. The scoring method will be straightforward where each question will carry 1 point in every section.

Participant responses will be compared and anonymized. Positive averaged scores for the approach will support the hypothesis that chromatic aberration is more useful for uncertainty visualization over alternatives. More specifically, we will use the Shapiro-Wilk normality test [Shapiro & Wilk 1965] to determine if the responses followed a normal distribution. For comparisons we will use standard t-tests.

Participants' feedback is also requested in written form. This feedback, in addition to comments made by the user during the screenshare, will be used as suggestions for future work.

2.5.3 Describe any compensation that will be given to participants and how this will be handled for participants who do not complete the study. Discuss any expenses participants are likely to incur and whether/how these will be reimbursed.

Every participant will receive compensation of $10 (Walmart/Amazon E-Gift card) from the researcher after the study. The compensation will be given even if the participant does not finish the study. The gift-card will be sent to their email and there won't be any other expenses in the study.

Since the gift-card will be provided through email, there will be automatic history in mailbox and hence no need to sign of participant payment receipt.

2.6 Privacy and confidentiality

2.6.1

A) Describe who will have knowledge of participants' identities.

Only the researcher will know the relationship between participant's name and unique participant IDs.

B) Describe the level of identifiability of the study data (anonymous, anonymized, de-identified/coded, identifying) (see TCPS Chapter 5A – types of information for definitions).

Data from this study will be associated to participants IDs (Coded Information).

C) Specify which members of the research team (or others) will have access to participants' data and for what purpose.

Project supervisor might have access on participants information for validation and justify their achievement with their qualification. In other words, to justify how much the educational background or knowledge level helps to answer the questionnaire properly.

D) Describe measures to ensure privacy and confidentiality of study documents and participant data during the data collection and analysis phase. [Note that plans for long term storage will be covered in 2.6.2]

- Address: handling of documents/data during data collection; transportation or transfer of documents/data; storage of documents/data (during the study).
- If a key-code will be maintained, describe how it will be kept secure.
- For electronic data, describe electronic data security measures, including file encryption and/or password protection as applicable.
- For hard copy documents, describe physical security measures (specify location).

We will use our self-developed web application for the questionnaire. The questionnaire will include multiple choice questions and identification questions based on provided parameters. No personal information will be asked from the participants other than email to send the gift-card and computer skill/profession to evaluate our study performance based on their qualification. We will store the audio and video (screenshare only) and securely at Dalhousie university and will keep until research work is evaluated and nobody would be able to access the data other than researcher. The researcher will be responsible to keep the data strictly secret and will not share or disclose it to anyone. After evaluation, researcher will permanently erase all data (audio, response, screenshare video) relevant to participation.

[ ] This research involves personal health records (ensure section 2.13 is completed)
2.6.2 Describe plans for data retention and long-term storage (i.e. how long data will be retained, in what form and where). Will the data eventually be destroyed or irreversibly anonymized? If so, what procedures will be used for this? Discuss any plans for future use of the data or materials beyond the study currently being reviewed. Survey responses will be stored on the Dalhousie Servers (secure web-space allocated for the researcher by Dalhousie University) through our online web application automatically. Recorded audio and video from the screenshare will be stored on the same secure server of dal-space. The response data will be stored with users email address initially and after completion of study, it will be added to the report in de-identified form and the researcher will destroy the files after successful completion of the research. Only the researcher will have access to the collected data in this study and there is no plan to use the collected data beyond the study. [ ] This research will be deposited in a data repository (ensure section 2.14 is completed)
2.6.3 Describe if/how participant confidentiality will be protected when research results are reported: A) For quantitative results - In what form will study data be disseminated? [ ] Only aggregate data will be presented [X] Individual de-identified, anonymized or anonymous data will be presented [ ] Other. If "other", briefly describe dissemination plans with regard to identifiability of data. [ ] Not applicable, only qualitative data will be presented B) For qualitative results - Will identifiable data be used in research presentations/publications? If participants will be quoted, address consent for this and indicate whether quotes will be identifiable or attributed. [ ] Not applicable, only quantitative data will be presented Participants are given the option to allow/disallow the researchers to use quotes when disseminating results in the consent form. These quotes would be collected from the written questions in the questionnaire (Appendix E) and from the recorded audio capture during the user study. We will use participant quotes anonymously.
2.6.4 Address any limits on confidentiality, such as a legal duty to report abuse or neglect of a child or adult in need of protection, and how these will be handled. Ensure these are clear in the consent documents. (See the guidance document for more information on legal duties and professional codes of ethics). [X] Not applicable

2.6.5 Will any information that may reasonably be expected to identify an individual (alone or in combination with other available information) be accessible outside Canada? And/or, will you be using any electronic tool (e.g. survey company, software, data repository) to help you collect, manage, store, share, or analyze personally identifiable data that makes the data accessible from outside Canada? [] No [X] Yes. If yes, refer to the University *Policy for the Protection of Personal Information from Access Outside Canada*, and describe how you comply with the policy (such as securing participant consent and/or securing approval from the Vice President Research and Innovation). Explained in Consent Form, Page: 18. 5th and 6th paragraphs.

2.7 Risk and benefit analysis
2.7.1 Discuss what risks or discomforts are anticipated for participants, how likely risks are and how risks will be mitigated. Address any particular ethical vulnerability of your study population. Risks to privacy from use of identifying information should be addressed. If applicable, address third party or community risk. (If the research involves Indigenous communities also complete section 2.11) The use of publicly available data surrounding Covid 19 may cause some degree of discomfort to some participants, given that the data is representative of a pandemic which is of concern to all. In addition, it is possible the use of simulated chromatic aberration may cause some minor eye strain. Beyond the above noted concerns, there are no anticipated physical, mental, economic, or social risks associated with participation beyond those associated with everyday computer use. There may be some minor discomforts for participants in that they will be using a new software application for the first time if someone didn't have the similar experience. We do not anticipate that this will exceed the usual levels of ambiguity or confusion commonly experienced when someone uses new software for the first time.
2.7.2 Identify any direct benefits of participation to participants (other than compensation), and any indirect benefits of the study (e.g. contribution to new knowledge). Participating in the study might benefit participants in terms of knowledge which will help them to participate in paid surveys in future or conduct and contribute their own survey if ever required. But there will not be any direct benefit.

2.8 Provision of results to participants and dissemination plans.

2.8.1 The TCPS encourages researchers to share study results with participants in appropriate formats. Describe your plans to share study results with participants and discuss the process and format. Participants are given a chance to add their e-mail address to receive the results of this study when it has been accepted for publication. Those participants that provided their email addresses for this purpose will receive a summary of the findings after the results are published but nobody will know other participants information.
2.8.2 If applicable, describe how participants will be informed of any material incidental findings – a discovery about a participant made in the course of research (screening or data collection) that is outside the objectives of the study, that has implications for participant welfare (health, psychological or social). See TCPS Article 3.4 for more information. [X] Not applicable
2.8.3 Describe plans for dissemination of the research findings (e.g. conference presentations, journal articles, public lectures etc.). Results from this study will be used for the lead researcher's MCS thesis paper and possibly for publication in Computer Science journals or conferences or in final thesis defense.

2.9 Research Team
2.9.1 Describe the role and duties of all research team members (including students, RA's and supervisors) in relation to the overall study. Dr. Brooks is a faculty member (Professor) in Computer Science. He will provide guidance during the study trials and will take part in the analysis after the event since he has previous expertise on guiding students earlier on visualization related research, conducting user studies and numerical analysis of the study results. Investigator Rashidul Islam has developed the study design under the direction of Dr. Brooks. This study is an integral part of his MCS Thesis component. Since the primary researcher is new to this kind of research and user study, he will discuss the process with the supervisor on an ongoing basis. For example: during study design the researcher tried a variety of alternative approaches such as noise, blurriness, transparency, etc. to compare with his novel technique of Chromatic Aberration (CA). During these phases supervisor discussed the researcher's idea and discussed the pros and cons together. After significant discussion we decided to compare our method with [Correll 2018].

2.9.2 Briefly identify any previous experience or special qualifications represented on the team relevant to the proposed study (e.g., professional or clinical expertise, research methods, experience with the study population, statistics expertise, etc.).

Dr. Brooks has previous experience in designing and executing user studies in the field of data visualization. Investigator Rashidul Islam is novice in conducting such user studies.

2.10 Conflict of interest

Describe whether any dual role or conflict of interest exists for any member of the research team in relation to potential study participants (e.g. TA, fellow student, teaching or clinical relationship), and/or study sponsors, and how this will be handled.
[X] Not applicable

2.11 Research involving Indigenous peoples
Consult TCPS Articles 9.1 and 9.2 in determining whether this section is applicable to your research.

[X] Not applicable – go to 2.12

2.11.1 If the proposed research is expected to involve people who are Indigenous, describe the plan for community engagement (per TCPS Articles 9.1 and 9.2). If community engagement is not sought, explain why the research does not require it, referencing TCPS article 9.2.

2.11.2 State whether ethical approval has been or will be sought from Mi'kmaw Ethics Watch and if not, why the research does not fall under their purview. If the research falls under the purview of other Indigenous ethics groups, state whether ethical approval has been or will be sought.

2.11.3 Describe plans for returning results to the community and any intellectual property rights agreements negotiated with the community with regard to data ownership (see also 2.11.4 if applicable). Append applicable research agreements.

2.11.4 Does this research incorporate OCAP (Ownership, Control, Access, and Possession) principles as described in TCPS Article 9.8?
[ ] Yes. Explain how.
[ ] No. Explain why not.

2.12 Clinical trials
[X] Not applicable – go to 2.13

2.12.1 Will the proposed clinical trial be registered? [ ] No. Explain why not. [ ] Yes. Indicate where it was/will be registered and provide the registration number.
2.12.2 If a novel intervention or treatment is being examined, describe standard treatment or intervention, to indicate a situation of clinical equipoise exists (TCPS Chapter 11). If placebo is used with a control group rather than standard treatment, please justify.
2.12.3 Clearly identify the known effects of any product or device under investigation, approved uses, safety information and possible contraindications. Indicate how the proposed study use differs from approved uses. [ ] Not applicable
2.12.4 Discuss any plans for blinding/randomization.
2.12.5 What plans are in place for safety monitoring and reporting of new information to participants, the REB, other team members, sponsors, and the clinical trial registry (refer to TCPS Articles 11.6, 11.7, 11.8)? These should address plans for removing participants for safety reasons, and early stopping/unblinding/amendment of the trial. What risks may arise for participants through early trial closure, and how will these be addressed? Are there any options for continued access to interventions shown to be beneficial?

2.13 Use of personal health information [X] Not applicable – go to 2.14
2.13.1 Research using health information may be subject to Nova Scotia's *Personal Health Information Act*. Describe the personal health information (definition explained in the guidance document) required and the information sources, and explain why the research cannot reasonably be accomplished without the use of that information. Describe how the personal health information will be used, and in the most de-identified form possible.
2.13.2 Will there be any linking of separate health data sets as part of this research? [ ] No [ ] Yes If yes:

A) Why is the linkage necessary? B) Describe how the linkage will be conducted (it is helpful to append a flow diagram) C) Does that linkage increase the identifiability of the participants?
2.13.3 Describe reasonably foreseeable risks to privacy due to the use of personal health information and how these will be mitigated.

2.14 Data Repositories [X] Not applicable
2.14.1 Identify and describe the data repository in which the research data will be deposited. What is its focus, who are its target users, who can access deposited data and under what circumstances? For how long will the data be kept in the repository?
2.14.2 Describe the data set to be released to the repository. If there is personal and/or sensitive information in the data, describe how you will prepare the data for submission to the repository and mitigate risks to privacy. Identify all fields that will be included in the final data set (include as an appendix).
2.14.3 Is agreeing to have one's data deposited a requirement for participation in the study? If yes, provide a justification. If no, indicate how participants can opt in or out.

SECTION 3. APPENDICES

Appendices Checklist. Append all relevant material to this application in the order they will be used. This may include:
[X] Reference list
[ ] Permission or support/cooperation letters (e.g. Indigenous Band Council, School Board, Director of a long-term care facility, anyone whose permission you need to conduct recruit participants or conduct research)
[ ] Research agreements (required for research involving Indigenous communities)
[X] Recruitment documents (posters, oral scripts, online postings, invitations to participate, etc.)
[ ] Screening documents
[X] Consent/assent documents or scripts
[X] Research instruments (questionnaires, interview or focus group questions, etc.)
[ ] Debriefing and/or study results templates
[ ] List of data fields included in data repository
[ ] Confidentiality agreements

Consent Form Templates
Sample consent forms are provided on the Research Ethics website and may be used in conjunction with the information in the *Guidance* document to help you develop your consent form.

DALHOUSIE UNIVERSITY

CONSENT FORM

Project title: Visualizing Uncertainty with Chromatic Aberration

Lead researcher: Md Rashidul Islam, Dalhousie University, md313724@dal.ca, +1(902)4483533

Other researchers
Dr. Stephen Brooks, sbrooks@cs.dal.ca

Funding provided by: NIL

Introduction

We invite you to take part in a research study being conducted by Rashidul Islam, who is an MCS (Master of Computer Science) student at Dalhousie University. Choosing whether or not to take part in this research is entirely your choice. There will be no impact to you if you decide not to participate in the research. The information below tells you about what is involved in the research, what you will be asked to do and about any benefit, risk, inconvenience or discomfort that you might experience.

You should discuss any questions you have about this study with Rashidul Islam or Dr. Stephen Brooks. Please ask as many questions as you have and we will be happy to answer your questions. If you have questions later, please contact Rashidul Islam.

Purpose and Outline of the Research Study
We have a new technique for displaying uncertain information in a visualization. The new visualization approach has been developed on a website that contains the new design. There are several possible ways to visualize uncertain data. So, it is important to assess how effective his new technique is compared to other possible techniques. To evaluate the new visualization technique the study will compare this new method with an existing approach that visualizes uncertainty with a special color palette. The two visualization approaches will be shown to users in an interactive website. The users will then be asked some question about the visualizations, and their feedback will be recorded.

Who Can Take Part in the Research Study
Anyone can participate in this study who has basic knowledge for recognizing simple shapes such as circles, rectangles, ellipse, partial filling of circles etc. They also need to have access to a computer browser; for instance: Firefox, have good internet connection, and have a microphone connected to the computer to communicate with researcher. In addition, participants must have color vision and not be impaired by color blindness.

What You Will Be Asked to Do
If you decide to participate in this research, you will be asked to navigate to a web application through your computer browser (Firefox). You will be asked to connect to a meeting with an

audio connection and screen sharing with Skype or Teams. Your audio and shared screen will be recorded for the study evaluation. You will be recommended to close all other applications besides the navigated application and the communication software itself. You have to complete the survey. If you have any questions or need clarification, then researcher can explain as he will be available to you for the entire duration. The length of the session would be approximately 1 hour.

Possible Benefits, Risks and Discomforts

Benefits: As a user you will be able interact with new types of visualizations. Other than that, there will not be any direct benefit.

Risks: Looking at images that contain colors that are blurry may produce some eye strain. Also, the data used in the examples are country level Covid 19 statistics which may cause some concern for some participants. Beyond these noted potential minor issues, no significant risks are anticipated with this study beyond being bored or fatigued or confused by using a new piece of software just like what you may feel for using any other new software. To reduce these discomforts, we will offer you breaks between activities whenever you need.

Your audio and screen share will be recorded and will be stored in secure space in our Dalhousie University server using your email address. So, until the study is evaluated, it will remain identifiable only by the researcher but no one else will have access to it. At the end of the research evaluation all data will be securely erased by researcher.

The researcher will use their Dalhousie University credentials for the Microsoft Teams meeting, which will ensure that the Teams meeting recordings are securely stored in Canada. However, during the live Teams meeting, audio and video content is routed through the United States, and therefore may be subject to monitoring without notice, under the provisions of the US Patriot Act while the meeting is in progress. After the meeting is complete, meeting recordings made by Dalhousie are stored in Canada and are inaccessible to external authorities.

However, we can alternatively use Skype if you prefer. Although it is a widely used conferencing tool, we do not have detailed internal knowledge about their data routing. So, similar to a Teams meeting, it is possible that Skype users can also be subject to hidden monitoring.

Compensation / Reimbursement
Every participant will receive compensation of $10 (Walmart/Amazon E-Gift card) and it will be given from the researcher to the participant’s email after the study. The compensation will be given even if the participant does not finish the study.

How your information will be protected:
Before starting the study we will inform you that your screenshare, questionnaire/answers, conversation will be recorded for future use and this information will be stored by the research team and only they will know about your participation information.

The information that we will take from you will remain highly confidential and secure. Only the research team at Dalhousie university who work with us have access to this information and all of us have an obligation to keep all these study information protected from any kind of unauthorized access. Your identity information (name and contact information) also be securely stored separately in an encoded approach. Instead of using your name or contact information, we will create a new ID number by encrypting your base information and which will be used as your participation number. In addition to that, all information will be kept secure in an encrypted file on the researcher's password-protected computer, and we will not maintain any paper/printed documents.

Since the study is a core component of our thesis research, we will explain and share our findings in the researcher's thesis report and thesis's defense. We expect you allow us to use your quotes in our report if needed. But the report will not include any individual information but group results. This means that nobody will be able to identify a single participant's information from our reports.

If You Decide to Stop Participating
You are fully free to leave the study at any time. If you want to stop participating during the study, you can also decide whether you want to allow us to use or remove any of the information that you have given us up to that point. If you want not to keep your participation in the study after completing the study and want us to remove your data, then you can decide for up to 1 week. After that time, we will remove your data but your responses will be anonymized in an analysis report and so there will be no way to trace your data individually.

How to Obtain Results
At the end of the study, we will provide you with a short description of group results but not individual results. You can obtain these results by letting researcher know during participation.

Questions
You are always welcome to reach out us with whatever questions or concerns you may have about your participation in this research study. Please contact at any time to Md Rashidul Islam at +1 (902) 448 3533, md313724@dal.ca or Dr. Stephen Brooks at sbrooks@dal.ca with your questions, suggestions, comments, or concerns about the research study.

If you have any ethical concerns about your participation in this research, you may also contact Research Ethics, Dalhousie University at (902) 494-3423, or email: ethics@dal.ca (and reference REB file # 20XX-XXXX).

Signature
Signatures will not be required as part of this informed consent process. Your email reply with assent will be considered as consent.

APPENDIX C – First Email

Hello Everyone,

Research Title: Visualizing Uncertainty with Chromatic Aberration.

I am writing to let you know that we are going to conduct a research study for my thesis component of my Master of Computer Science, at Dalhousie University. For this purpose, we are recruiting participants to take part in a user study. The primary goal of the user study is to evaluate of our newly designed technique for uncertainty visualization in a web-based application through user feedback. In the study we will compare it with an existing solution published in an international journal.

The survey will be conducted completely online as we have a publicly accessible web application, and all questions will be asked and answered online. So, you will be able to participate from your own computer at home. You must have a computer with internet connection suitable for audio-video conferencing (screen share only) using Skype or MS Teams. For this you must be willing to share your screen with the researcher and allow it to be recorded. Please be assured that your responses will be kept confidential and that the results will be reported in thesis and research paper only in anonymous form.

This study consists of a single session and will be completed approximately in 1 hour. Compensation is $10 e-gift card for participation in the study.

Please contact me if you are interested in participating or if you need any further information about the user study.

Best Regards,

Md Rashidul Islam
md313724@dal.ca

APPENDIX D – Second Email

Good day Everyone,

We are recruiting participants to take part in the research study of Master of Computer Science, Dalhousie University. The user study aims to get user feedback on the visualization of uncertainty in a web-based application. A potential benefit could be that A potential benefit could be that you will interact with new types of visualizations.

This study consists of a single session and will be conducted completely online to ensure participants and researcher safety and respect the imposed special measures during the COVID-19 pandemic. We created a dynamic web application that allows testing and evaluating various features related to our visualizations. After an initial privacy check, the participant will be requested to browse the application from their own computer and share their screen with the researcher while using it. With the help of screen and audio sharing, participants will be given an introduction to the system by the researcher, answering any questions they might have.

The web application will present different methods of visualizing uncertainty in data. Participants will be accessing the application from their own computer, and they will be able ask any question that arises to the researcher will remain connected online with the participant.

After reviewing and interacting with the application, the participant is provided with a questionnaire which asks for feedback on each proposed method of visualizing uncertainty in data. A series of statements are provided about each visualization feature with multiple choice questions and the participant decides by choosing one of them to represent what extent they agree with the statement. For each visualization, the participant is also provided an opportunity to provide written feedback to the researchers. After completing these sections, the participant is provided an opportunity to provide a general but holistic written feedback at the end of the questionnaire module.

The following requirements are necessary for participation to qualify as participant in the study:

- You must have a computer with internet connection suitable to audio-video conferencing and must be willing to share your screen with the researcher and allow to record it.
- You must be able to install required software such as MS-Teams or Skype for conversation.
- You must be able browse the application and use it.
- You must have full color vision without color blindness.

The length of the session would be approximately 1 hour. Compensation is $10 e-gift card for participation in the study.

Thank you for your consideration. If you agree to participate, please contact the main researcher at md313724@dal.ca for a list of potential time slots to schedule the session.

Thank you,
Md Rashidul Islam, MCS Student
+1 (902) 884 3533
Dalhousie University
6299 South St, Halifax, NS B3H 4R2

EXAMPLES AND QUESTIONNAIRE

Project title: Visualizing Uncertainty with Chromatic Aberration

Lead researcher: Md Rashidul Islam, Dalhousie University, md313724@dal.ca, +1(902)4483533

Other researchers
Dr. Stephen Brooks, sbrooks@cs.dal.ca

Funding provided by: NIL

E.1 Questionnaire Setup and Arrangement:

The existing evaluation of uncertainty representation named VSUP used grid-chart method with a custom color set. We will be comparing VSUP with Chromatic Aberration (CA)using both a grid-chart and bubble-chart. So, the questionnaire arrangement is made with the following sections:

- A: CA + Bubble
- B: CA + Grid
- C: VSUP + Bubble
- D: VSUP + Grid

To make the comparison fair, we have grouped our uncertainties to 4 levels since VSUP also uses four levels of uncertainties. In our case, we have quantized our CA data and made four equidistant values of [33, 52, 71, 90] to draw the aberration in both circles and rectangles. In addition, to fill the circles and rectangles of CA, we have used the eight standard VSUP colors to make the evaluation consistent.

We have also implemented counterbalancing in the questionnaire presentation. That means every four users will see the questionnaire in one of the following orders:

A	B	D	C
B	C	A	D
C	D	B	A
D	A	C	B

Figure E.1: Balanced Latin Squares

Every section consists of eight questions, but the order of the questions is randomly chosen by the system. So, the number of questions and the content of the questions will remain the same but in a different order for different participants.

So, at the first place when participant will be navigated to the given URL of our online application, they will see the following screen to provide their email address:

Figure E.2: Questionnaire Email Screen

After providing the email address, the user will see one of the four sections of the questionnaire. The layout of the questionnaire design will be as follows:

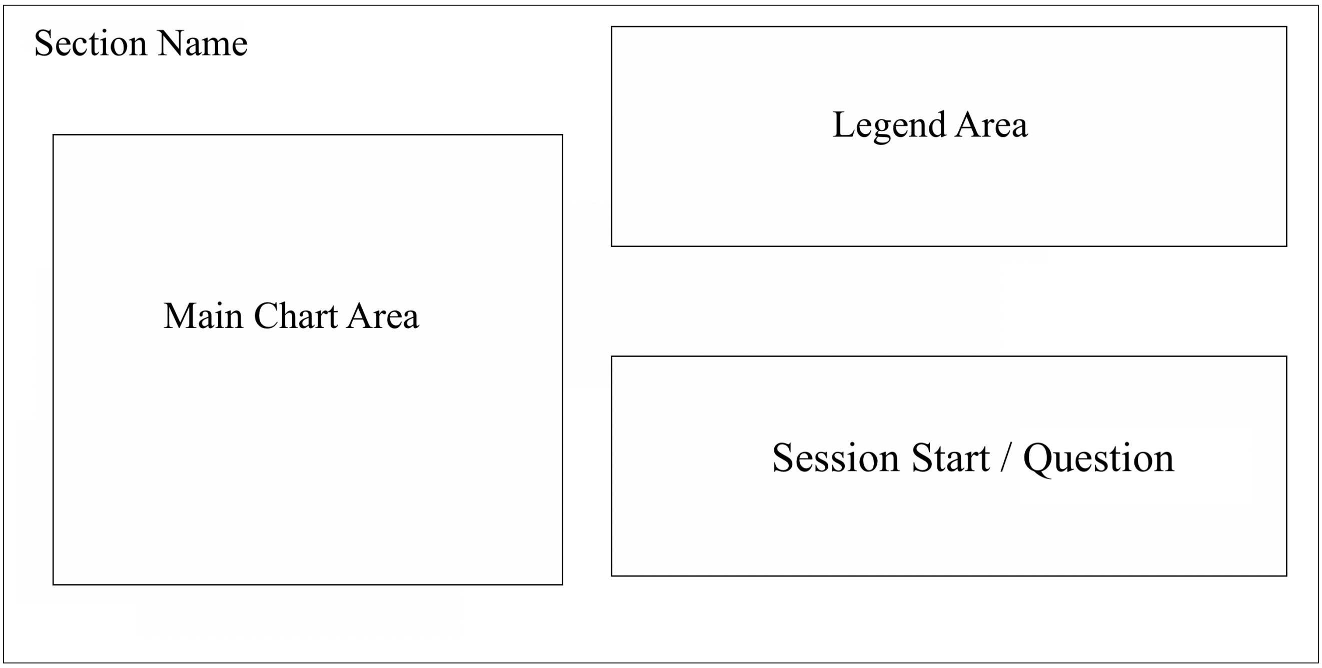


Figure E.3: Layout of Questionnaire View

At the beginning of every section, the bottom-right part the of the UI will show the Session description. The researcher will describe the features (chart, legend and how question will be asked and what does that mean, etc.). After completion of explanation, the participant is asked to hit 'Start' button as the following screen:

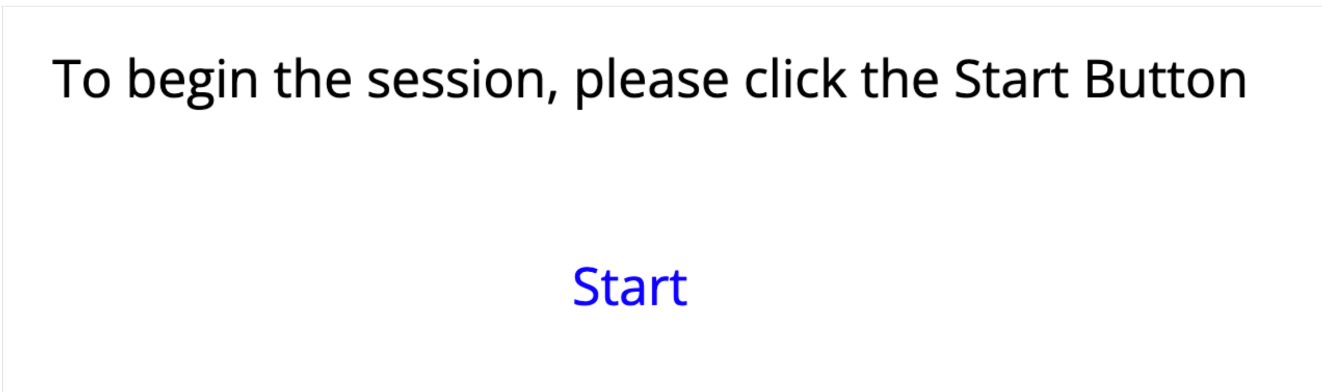


Figure E.4: Module Starter View

Once she or he presses the 'Start' button, the questionnaire will be started immediately and will present one question at a time. For example:

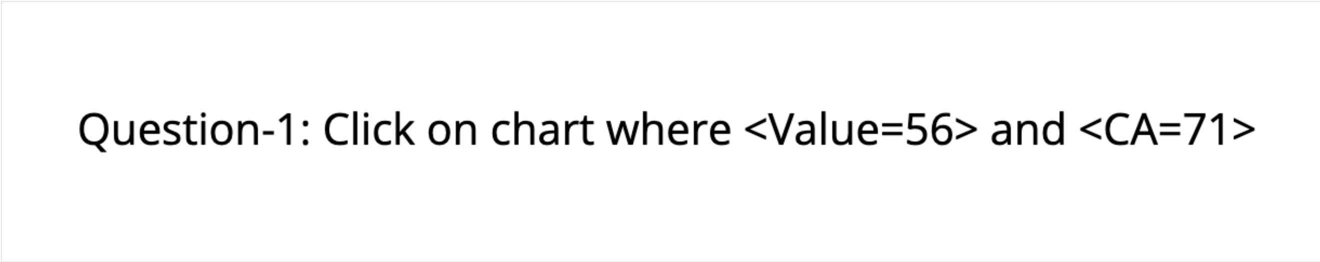


Figure E.5: Sample Question

The user then needs to select a cell (bubble or rectangle) from the chart based on the provided Value and Uncertainty/CA combination. After a cell is selected by the user, the next question will appear at the same place until it reaches to eighth question of the section.

Since the bubble chart and the grid chart are two major components of this research and we have four sections with these two components, we present one example with identification procedure for a sample question prior to questionnaire of each section. Examples are given here for the reader of this document but in real application it will be described verbally to the participant along with answering more questions if the participant may have. Orders of the questionnaire will be changed by counterbalancing stated above for different session users. So, these are the summary of the next sections:

11. Example of CA + Bubble
12. Questionnaire on CA + Bubble
13. Example of VSUP + Bubble
14. Questionnaire on VSUP + Bubble
15. Example of CA + Grid
16. Questionnaire on CA + Grid
17. Example of VSUP + Grid
18. Questionnaire on VSUP + Grid

Then we ask the following two types of additional questionnaires:

19. Questions on System Usability Scale (SUS)
20. Questions on NASA TLX

E.2 Example of CA + Bubble:

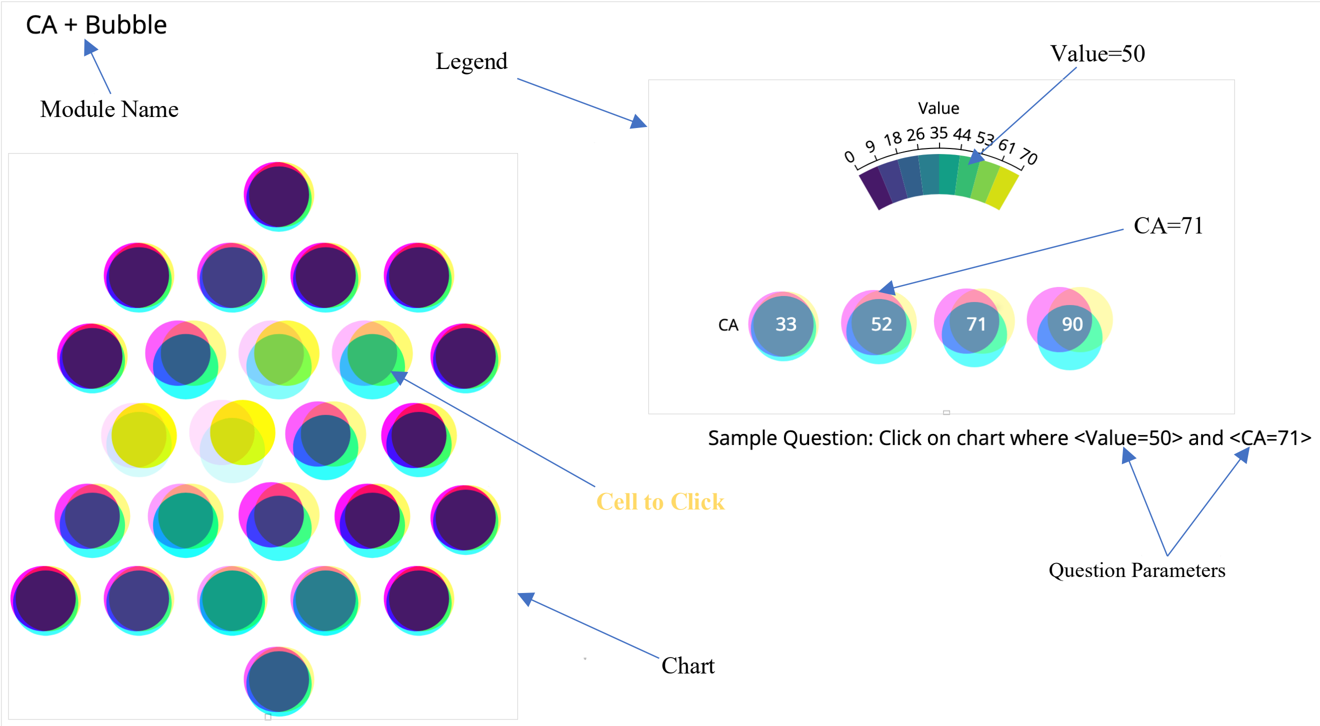


Figure E.6: Question-Answer Identification Procedure

Description:
In this example, we have introduced the different components with arrow indicators such as Chart, Legend, question parameters. Detection of question parameters in legend and finally based on the parameter values finding the target cell from the chart with the label 'Cell to Click'.

In identification the following rules are needed to be used:
CA = The thickness of the colorful edges of the three overlapping circles
Value = Color of the common(center) portion of the three circles.

Based on the above instruction participant need to answer the questions of this model in next section. Researcher will also explain the mechanism verbally before starting the module.

E.3 Questionnaire on CA + Bubble

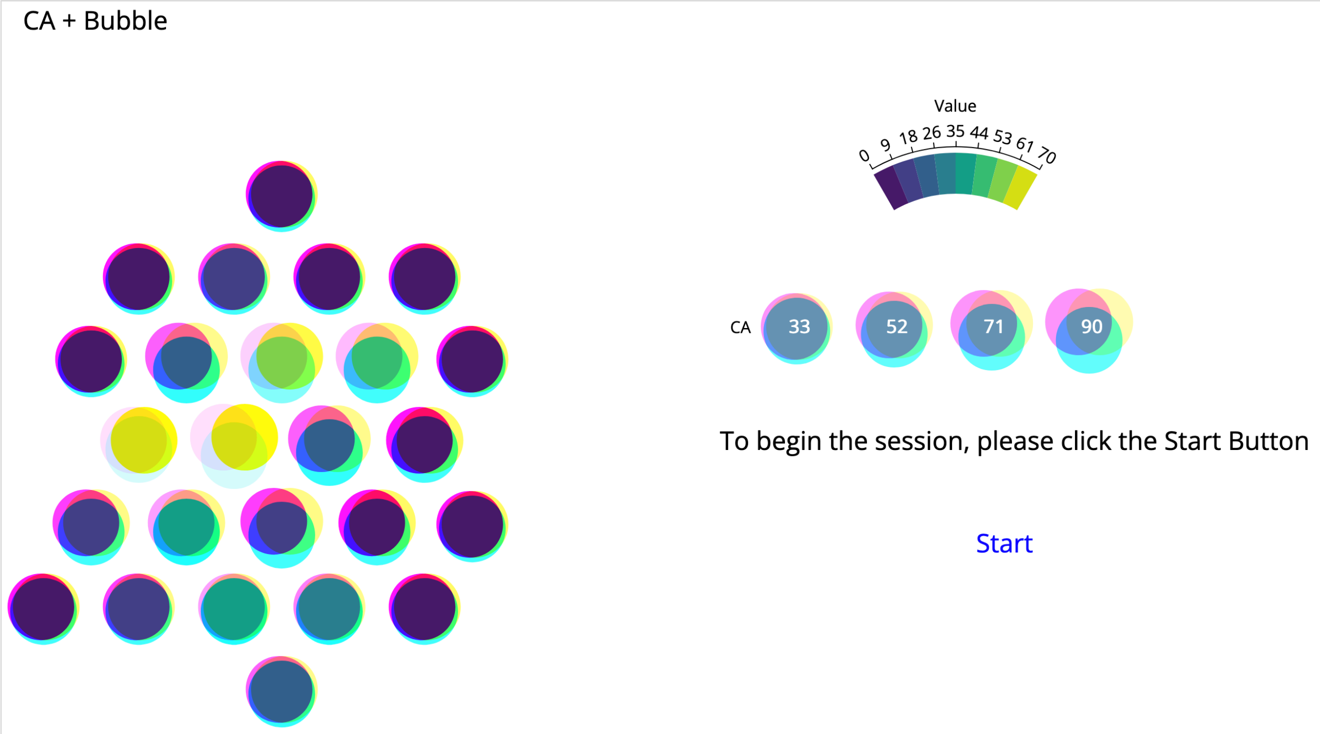


Figure E.7: CA + Bubble Questionnaire UI

Questions:
On pressing ‘Start’ button it will start to show the questions one by one as follows (orders of the questions will be changed by counterbalancing for different session users)

Question-1: Click on <a bubble> in chart where <CA=90>

Question-2: Click on <a bubble> in chart where <CA=54>

Question-3: Click on <a bubble> in chart where <CA=36>

Question-4: Click on <all bubbles> in chart where <CA=72>

Click when finished: Done

Question-5: Click on <all bubbles> in chart where <CA=90>

Click when finished: Done

Question-6: Click on <a bubble> in chart where <Value=32> and <CA=54>

Question-7: Click on <a bubble> in chart where <Value=56> and <CA=90>

Question-8: Click on <a bubble> in chart where <Value=16> and <CA=90>

Figure E.8: Questions on CA + Bubble

E.4 Example of VSUP + Bubble :

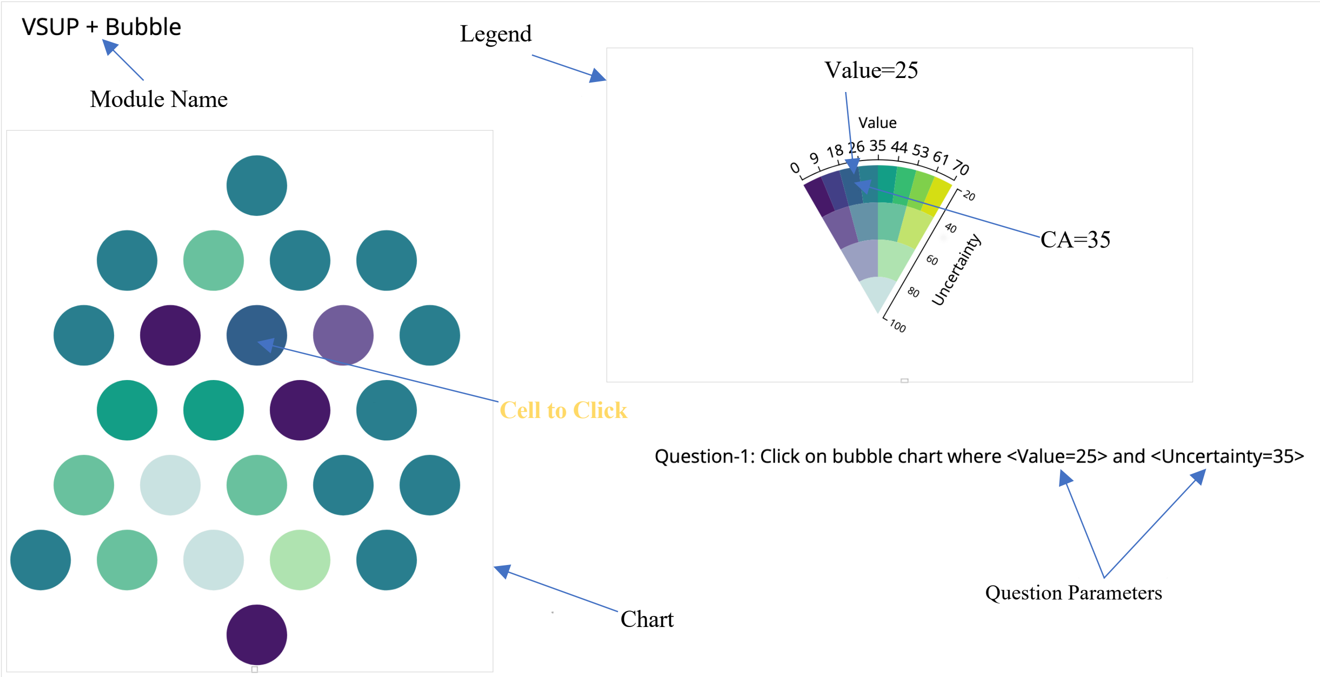


Figure E.9: Question-Answer Identification Procedure

Description:
In this example, we have introduced the different components with arrow indicators such as Chart, Legend, question parameters. Detection of question parameters in legend and finally based on the parameter values finding the target cell from the bubble chart with the label 'Cell to Click'.

In identification the following rules are needed to be used:
Uncertainty = Represents the vertical axis in the legend labeled by 'Uncertainty'
Value = Represents the horizontal axis on the legend.

In this scenario, by using Uncertainty and Value, we get single cell from the legend as indicated above.

Based on the above instruction participant need to answer the questions of this model in next section. Researcher will also explain the mechanism verbally before starting the module.

E.5 Questionnaire on VSUP + Bubble :

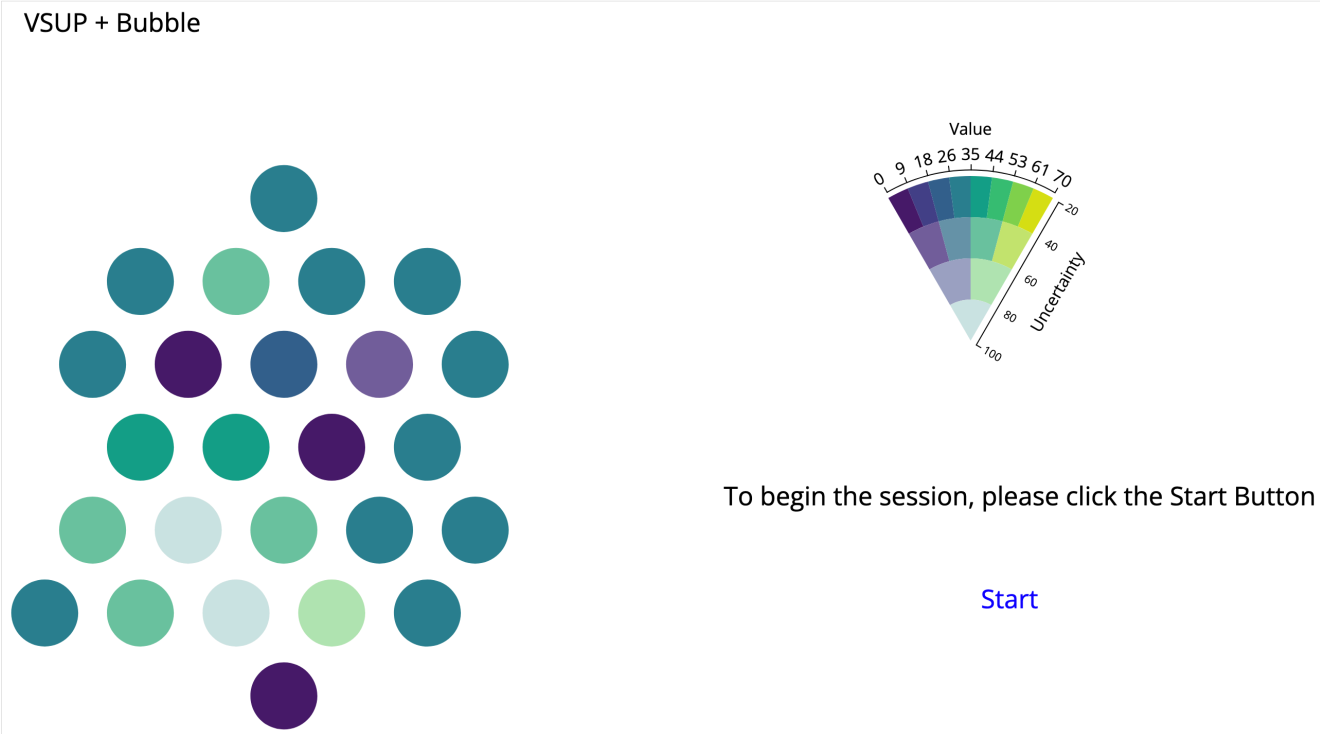


Figure E.10 : VSUP + Bubble Questionnaire UI

Questions :
On pressing ‘Start’ button it will start to show the questions one by one as follows (orders of the questions will be changed by counterbalancing for different session users)

Question-1: Click on <a bubble> in chart where <Uncertainty=78>

Question-2: Click on <a bubble> in chart where <Uncertainty=56>

Question-3: Click on <a bubble> in chart where <Uncertainty=37>

Question-4: Click on <all bubbles> in chart where <Uncertainty=33>
Click when finished: Done

Question-5: Click on <all bubbles> in chart where <Uncertainty=23>
Click when finished: Done

Question-6: Click on <a bubble> in chart where <Value=8> and <Uncertainty=38>

Question-7: Click on <a bubble> in chart where <Value=51> and <Uncertainty=43>

Question-8: Click on <a bubble> in chart where <Value=34> and <Uncertainty=89>

Figure E.11: Questions on VSUP + Bubble

E.6 Example of CA + Grid:

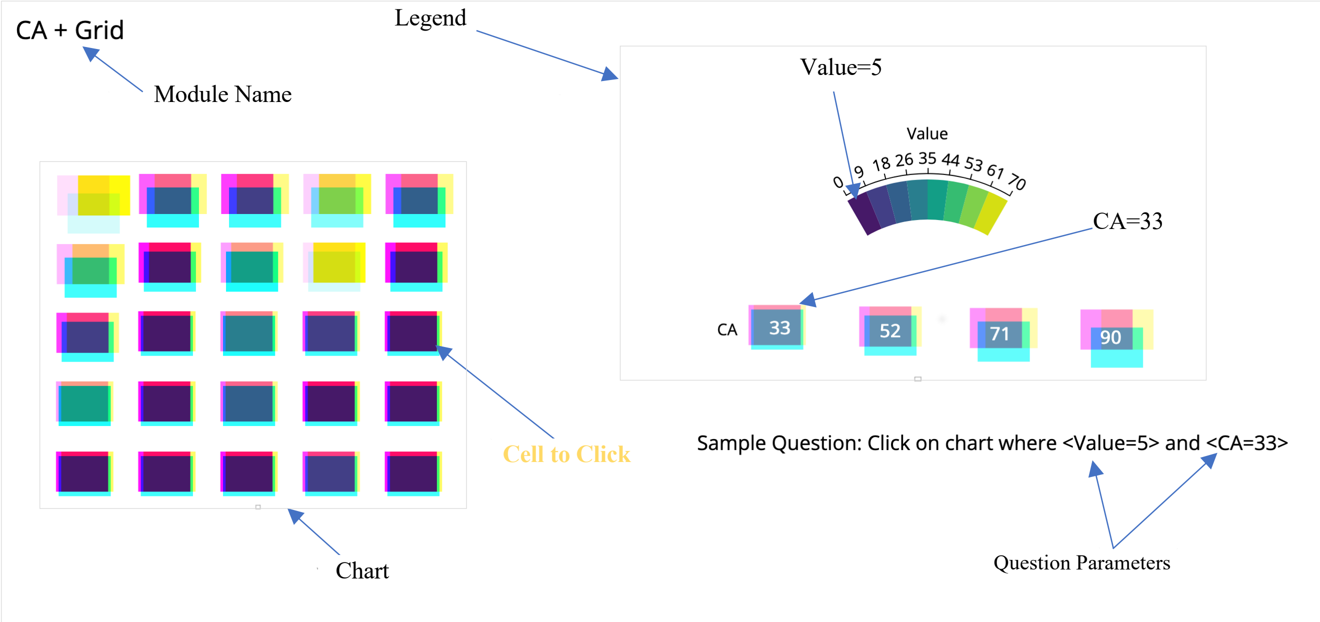


Figure E.12: Question-Answer Identification Procedure

Description:
In this example, we have introduced the different components with arrow indicators such as Chart, Legend, question parameters. Detection of question parameters in legend and finally based on the parameter values finding the target cell from the chart with label 'Cell to Click'.

In identification the following rules are needed to be used:
CA = The thickness of the colorful edges of the three overlapping rectangles
Value = Color of the common(center) portion of three rectangles.

Based on the above instruction participant need to answer the questions of this model in next section. Researcher will also explain the mechanism verbally before starting the module.

E.7 Questionnaire on CA + Grid

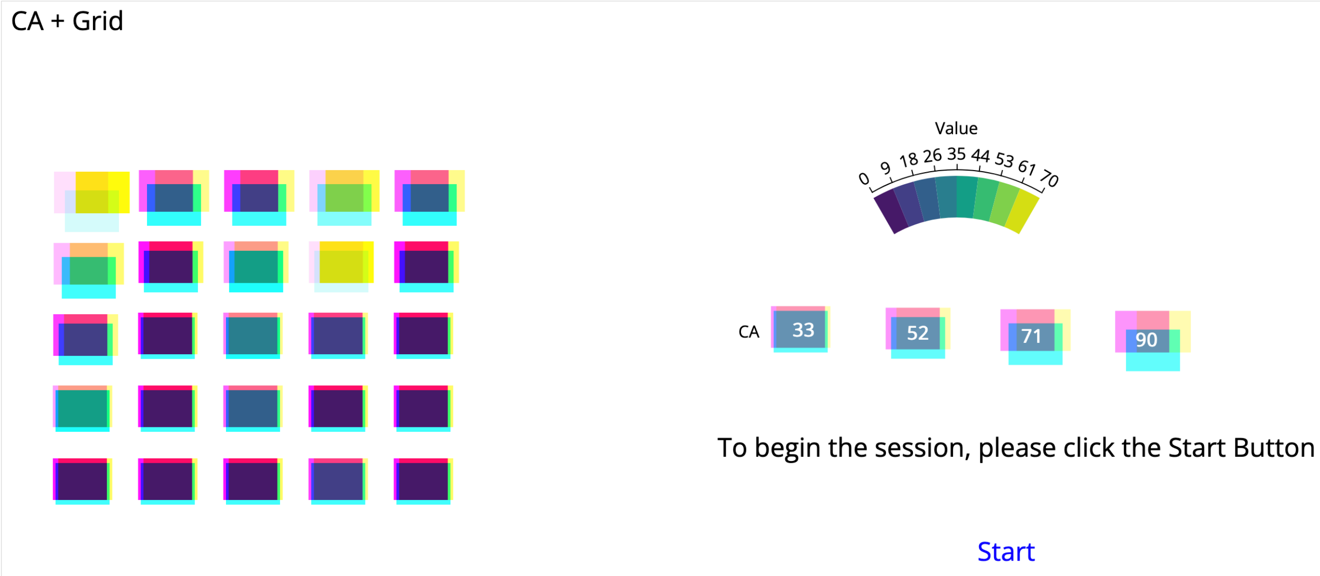


Figure E.13: CA + Grid Questionnaire UI

Questions:
On pressing 'Start' button it will start to show the questions one by one as follows (orders of the questions will be changed by counterbalancing for different session users)

Question-1: Click on <a square> in chart where <CA=90>

Question-2: Click on <a square> in chart where <CA=54>

Question-3: Click on <a square> in chart where <CA=72>

Question-4: Click on <all squares> in chart where <CA=36>

Click when finished: Done

Question-5: Click on <all squares> in chart where <CA=90>

Click when finished: Done

Question-6: Click on <a square> in chart where <Value=40> and <CA=72>

Question-7: Click on <a square> in chart where <Value=64> and <CA=72>

Question-8: Click on <a square> in chart where <Value=8> and <CA=72>

Figure E.14: Questions on CA + Grid

E.8 Example of VSUP + Grid

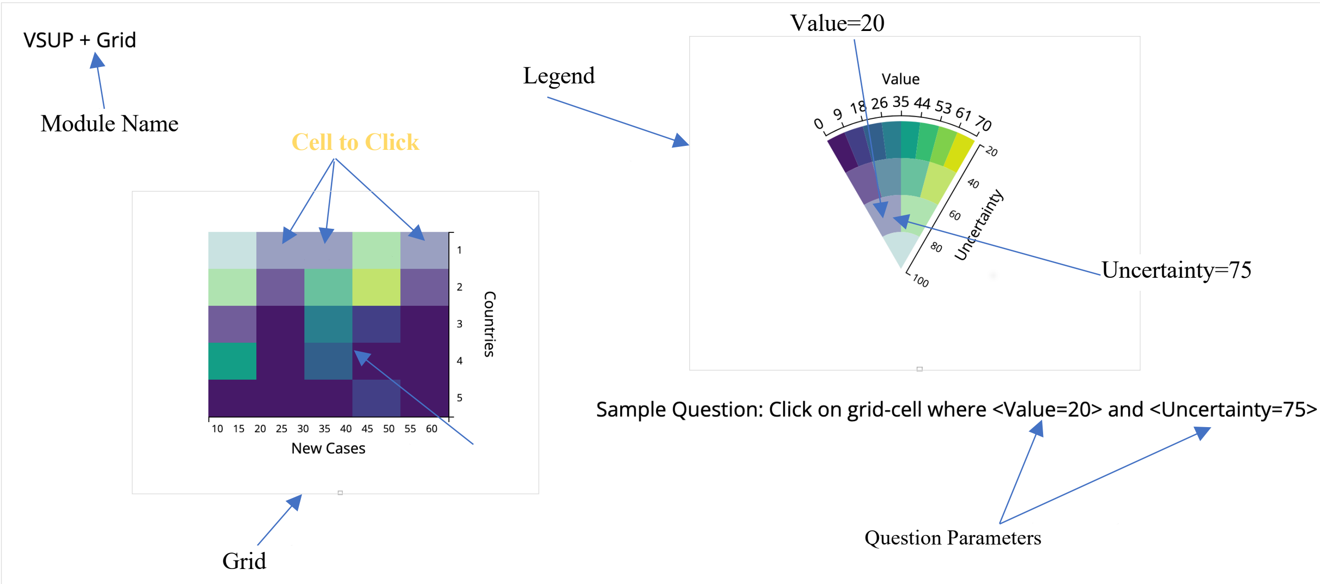


Figure E.15: Question-Answer Identification Procedure

Description:
In this example, we have introduced the different components with arrow indicators such as Grid, Legend, question parameters. Detection of question parameters in the legend and finally based on the parameter values selecting the target cell from the grid with label ‘Cell to Click’.

In identification the following rules are needed to be used:
Uncertainty = Represents the vertical axis in the legend labeled by ‘Uncertainty’
Value = Represents the horizontal axis on the legend.

In this scenario, by using Uncertainty and Value, we get single cell from the legend as indicated above. Here we found three cells in grid with the target legend cell, so clicking on one of them will be considered as correct answer.

Based on the above instruction participant need to answer the questions of this model in next section. Researcher will also explain the mechanism verbally before starting the module.

E.9 Questionnaire on VSUP + Grid

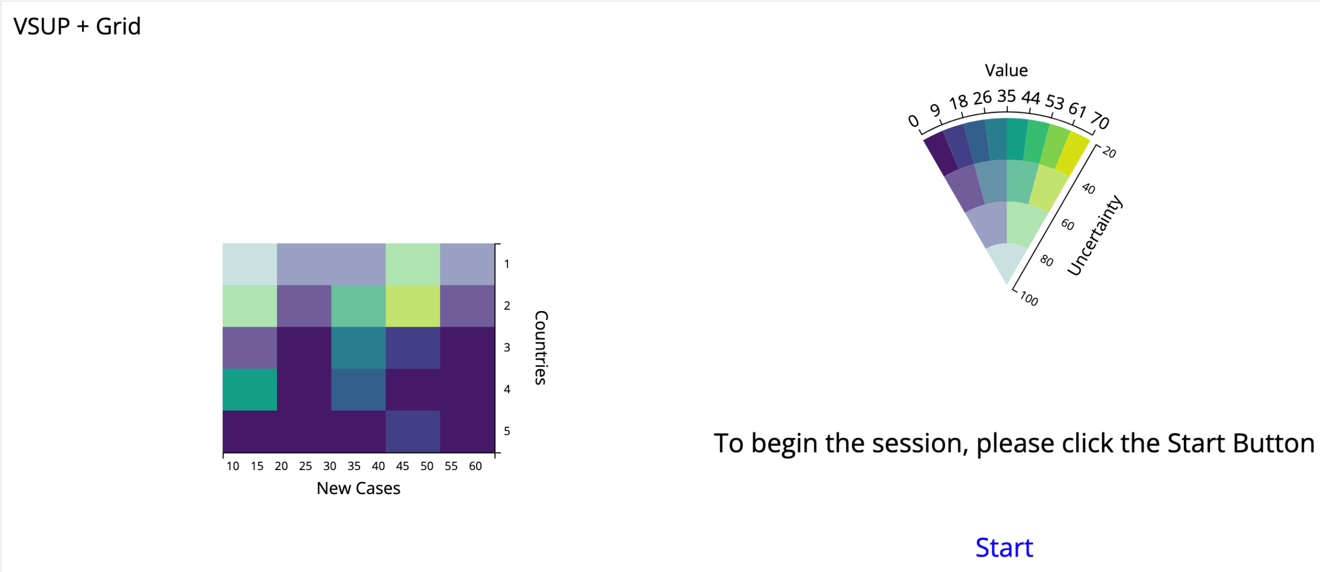


Figure E.16: VSUP + Grid Questionnaire UI

Questions:
On pressing 'Start' button it will start to show the questions one by one as follows (orders of the questions will be changed by counterbalancing for different session users)

Question-1: Click on <a square> in chart where <Uncertainty=78>

Question-2: Click on <a square> in chart where <Uncertainty=56>

Question-3: Click on <a square> in chart where <Uncertainty=89>

Question-4: Click on <all squares> in chart where <Uncertainty=43>
Click when finished: Done

Question-5: Click on <all squares> in chart where <Uncertainty=38>
Click when finished: Done

Question-6: Click on <a square> in chart where <Value=27> and <Uncertainty=37>

Question-7: Click on <a square> in chart where <Value=36> and <Uncertainty=23>

Question-8: Click on <a square> in chart where <Value=19> and <Uncertainty=33>

Figure E.17: Questions on VSUP + Grid

E.10 Questions on System Usability Scale (SUS):

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently.	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated.	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly.	1	2	3	4	5
8. I found the system very cumbersome to use.	1	2	3	4	5
9. I felt very confident using the system.	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system.	1	2	3	4	5

E.11 Questions on NASA TLX:

1. **Mental** — How mentally demanding was the task?

Very Low — Very High

2. **Physical Demand** — How physically demanding was the task?

Very Low — Very High

3. **Temporal** — How hurried or rushed was the pace of the task?

Very Low — Very High

4. **Performance** — How successful were you in accomplishing what you were asked to do?

Very Low — Very High

5. **Effort** — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

6. **Frustration** — How insecure, discourages, irritated, stressed, and annoyed were you?

Very Low — Very High

APPENDIX F – Ishihara Color Blindness Test Plates

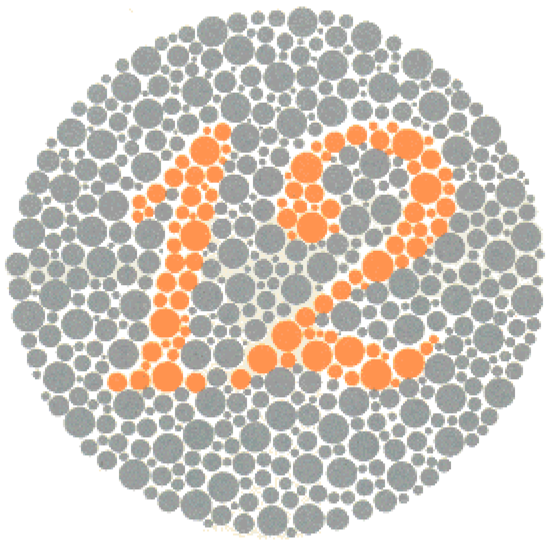


Plate-1

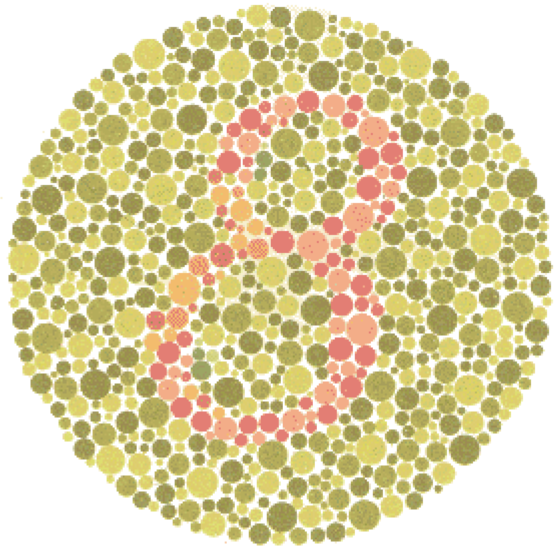


Plate-2

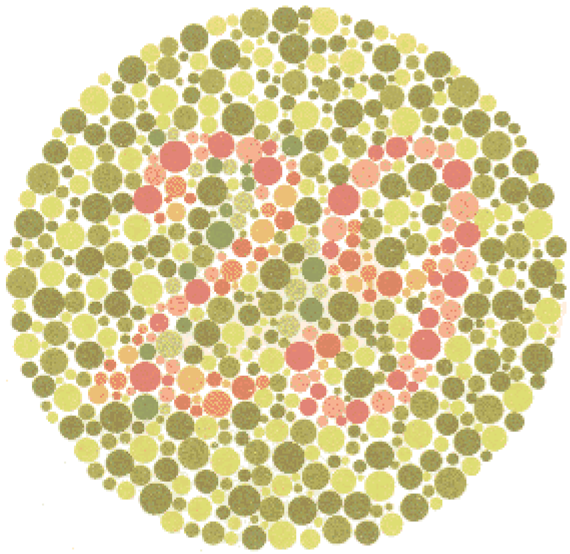


Plate-3

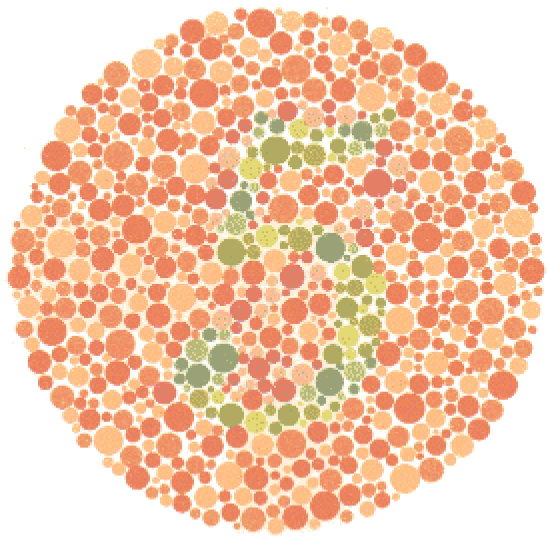


Plate-4

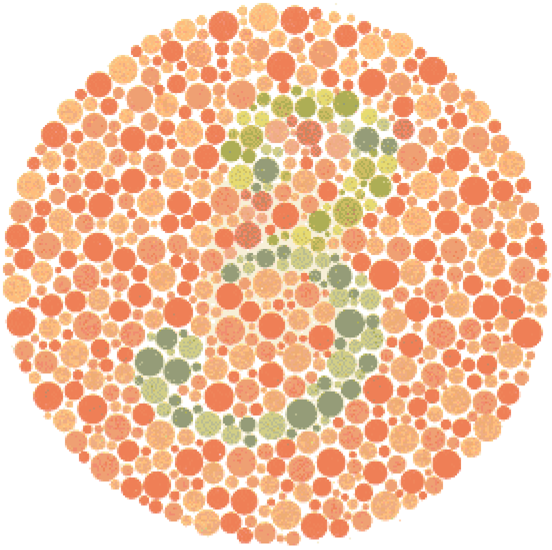


Plate-5

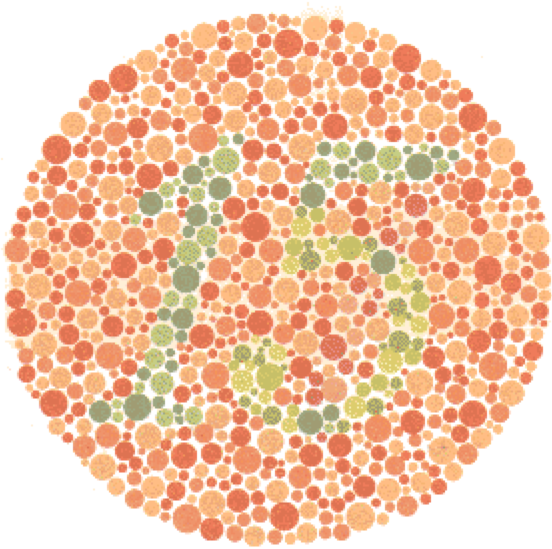


Plate-6

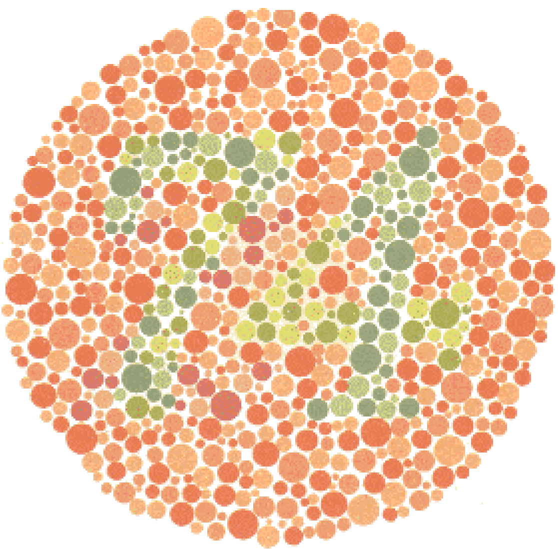


Plate-7

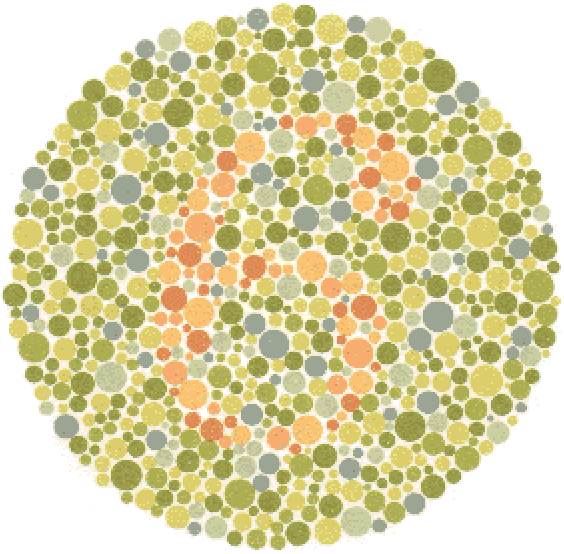


Plate-8

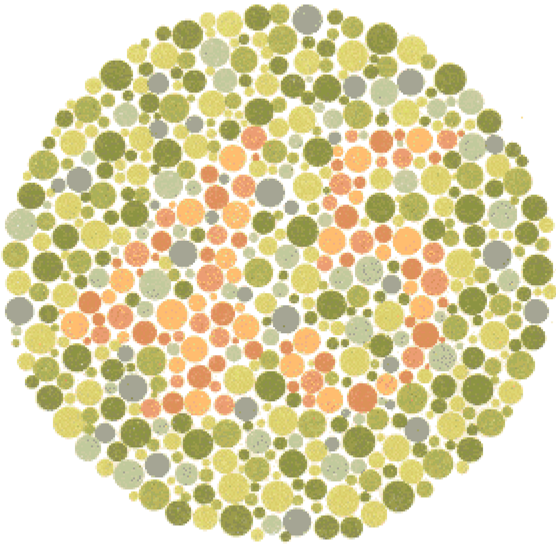


Plate-9

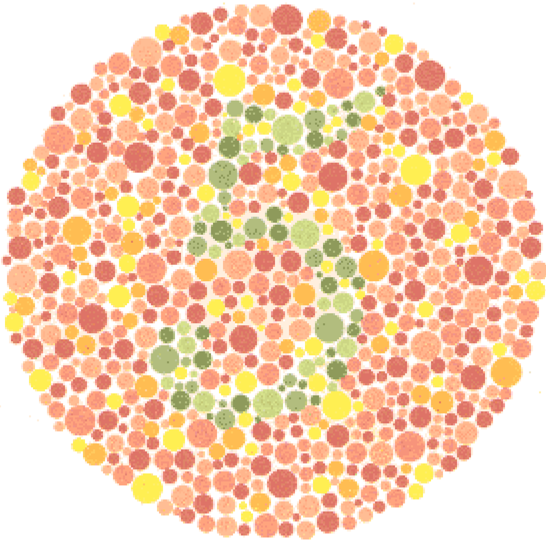


Plate-10

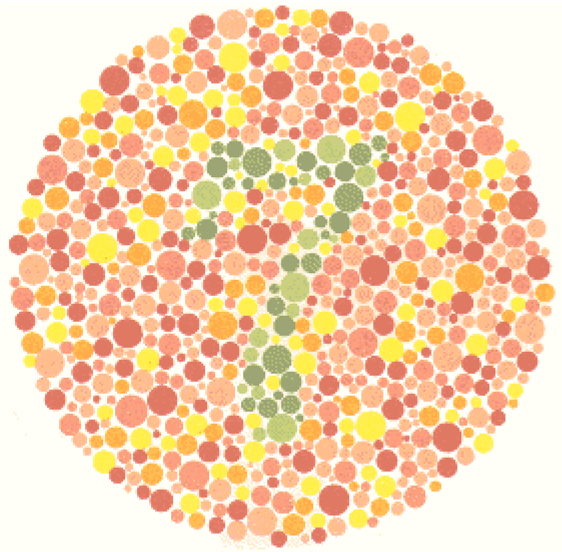


Plate-11

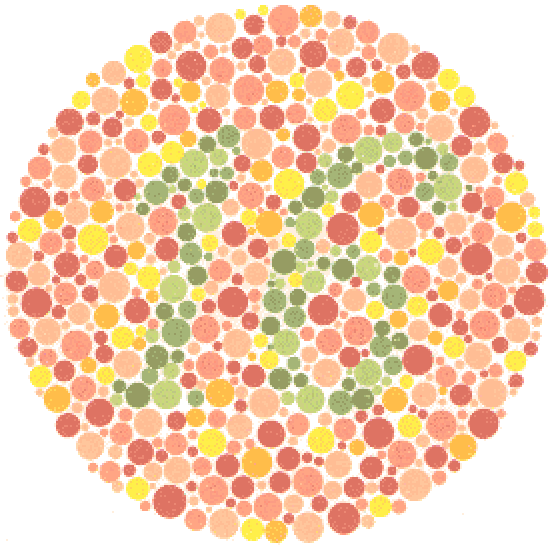


Plate-12

APPENDIX F – Ishihara Color Blindness Test Plates

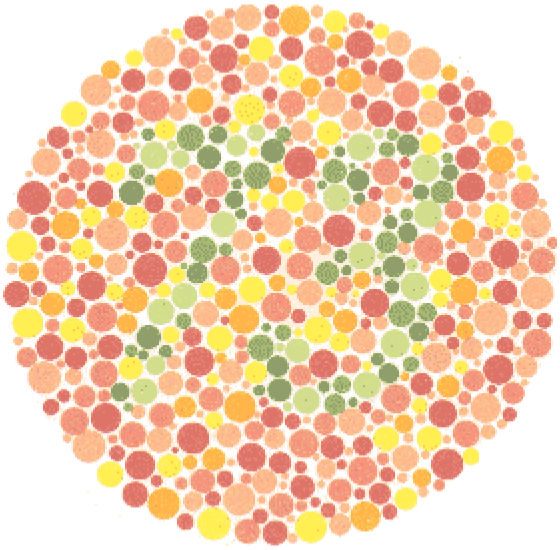


Plate-13

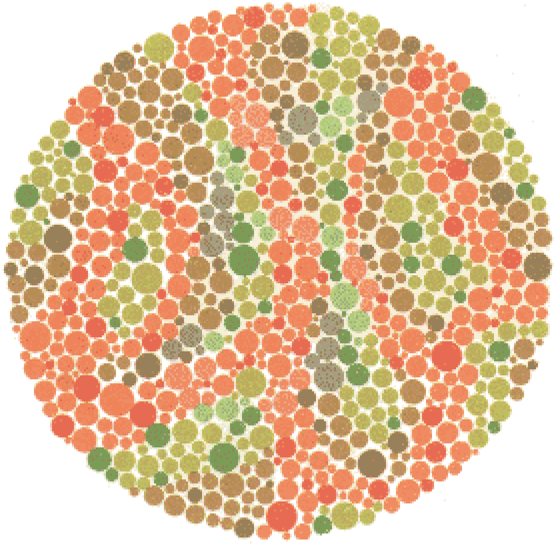


Plate-14

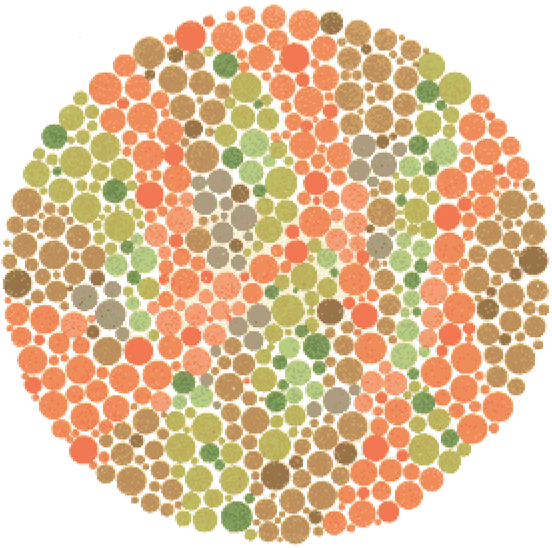


Plate-15

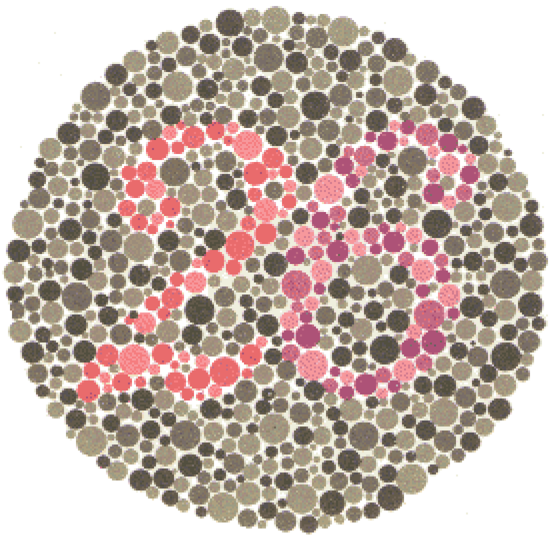


Plate-16

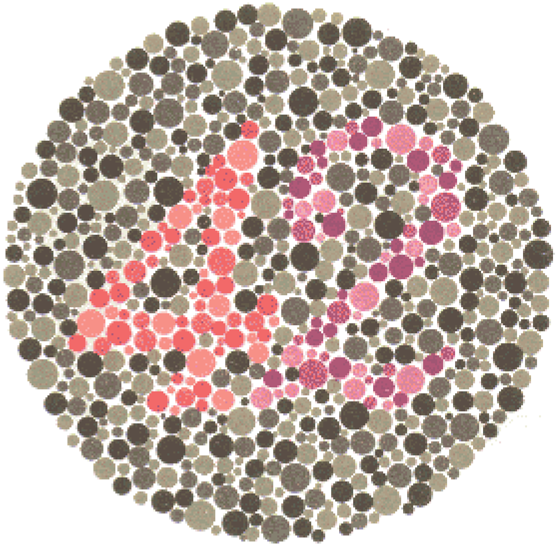


Plate-17

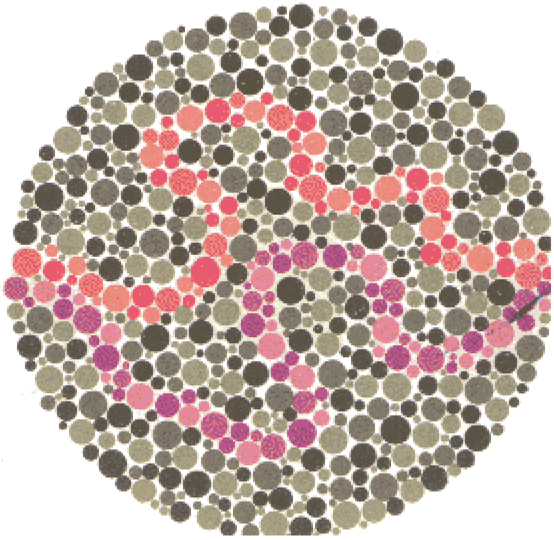


Plate-18

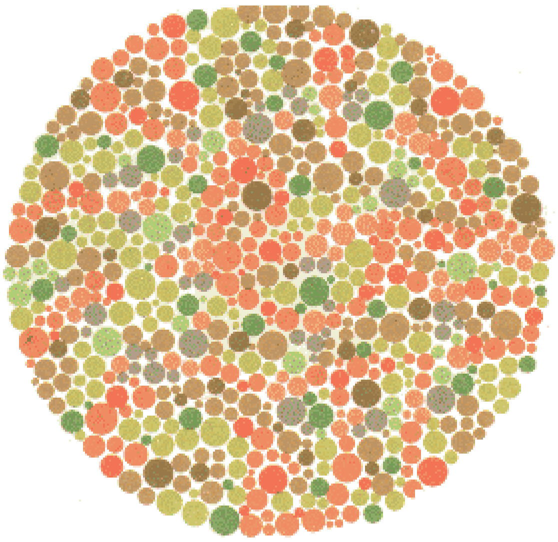


Plate-19

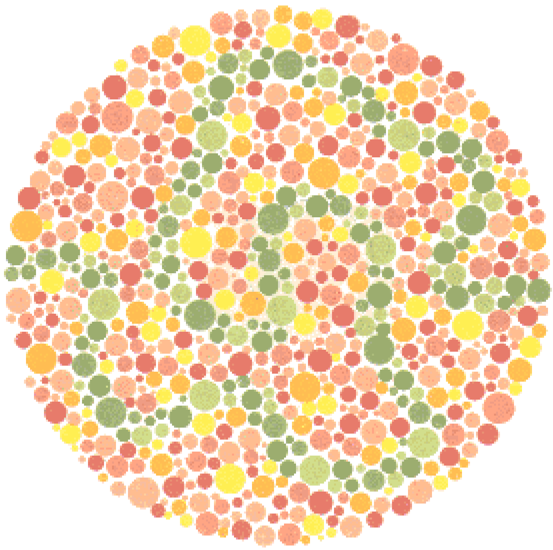


Plate-20

APPENDIX F – Ishihara Color Blindness Test Plates

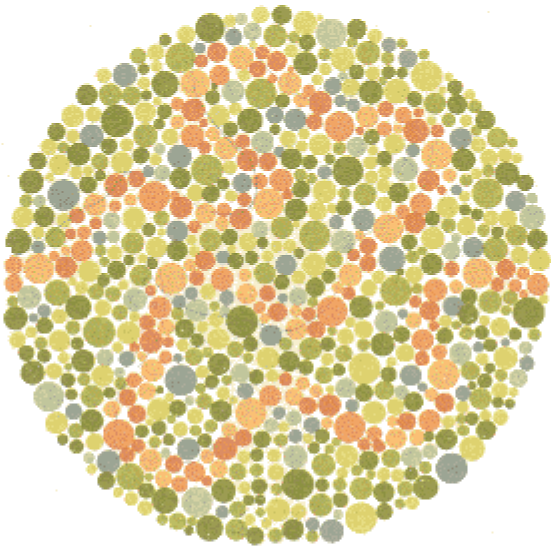


Plate-21

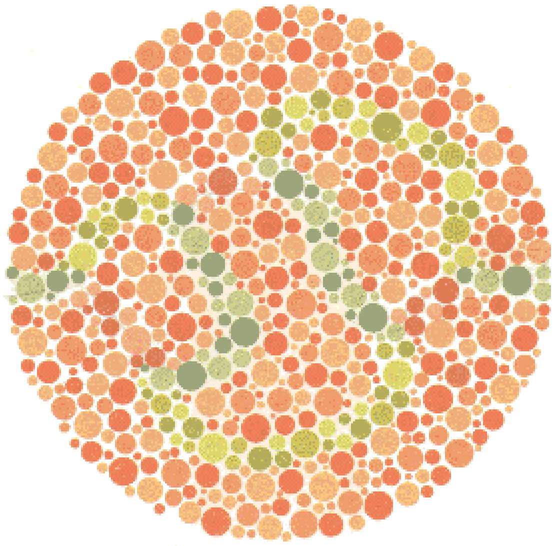


Plate-22

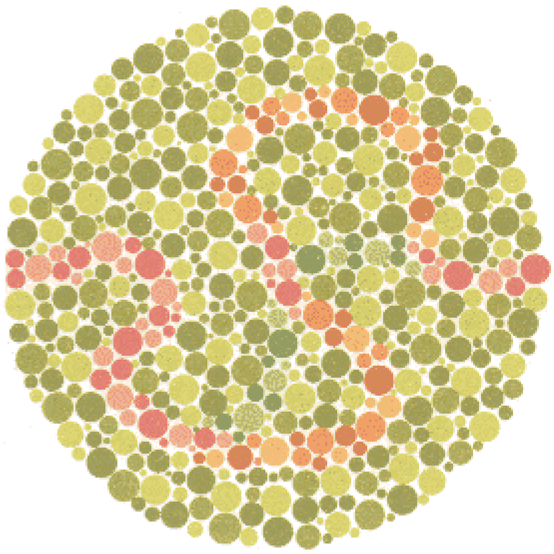


Plate-23

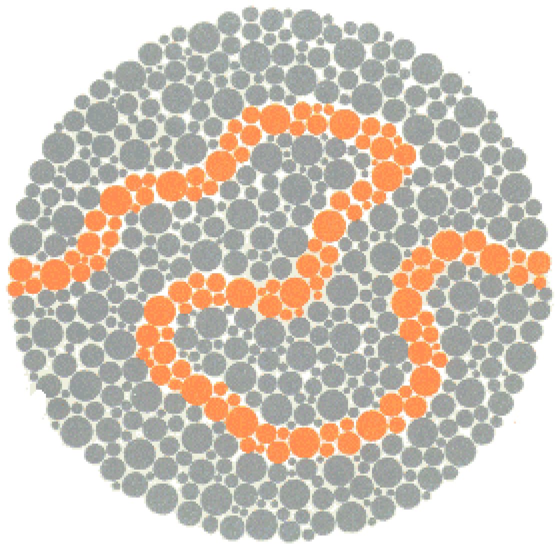


Plate-2

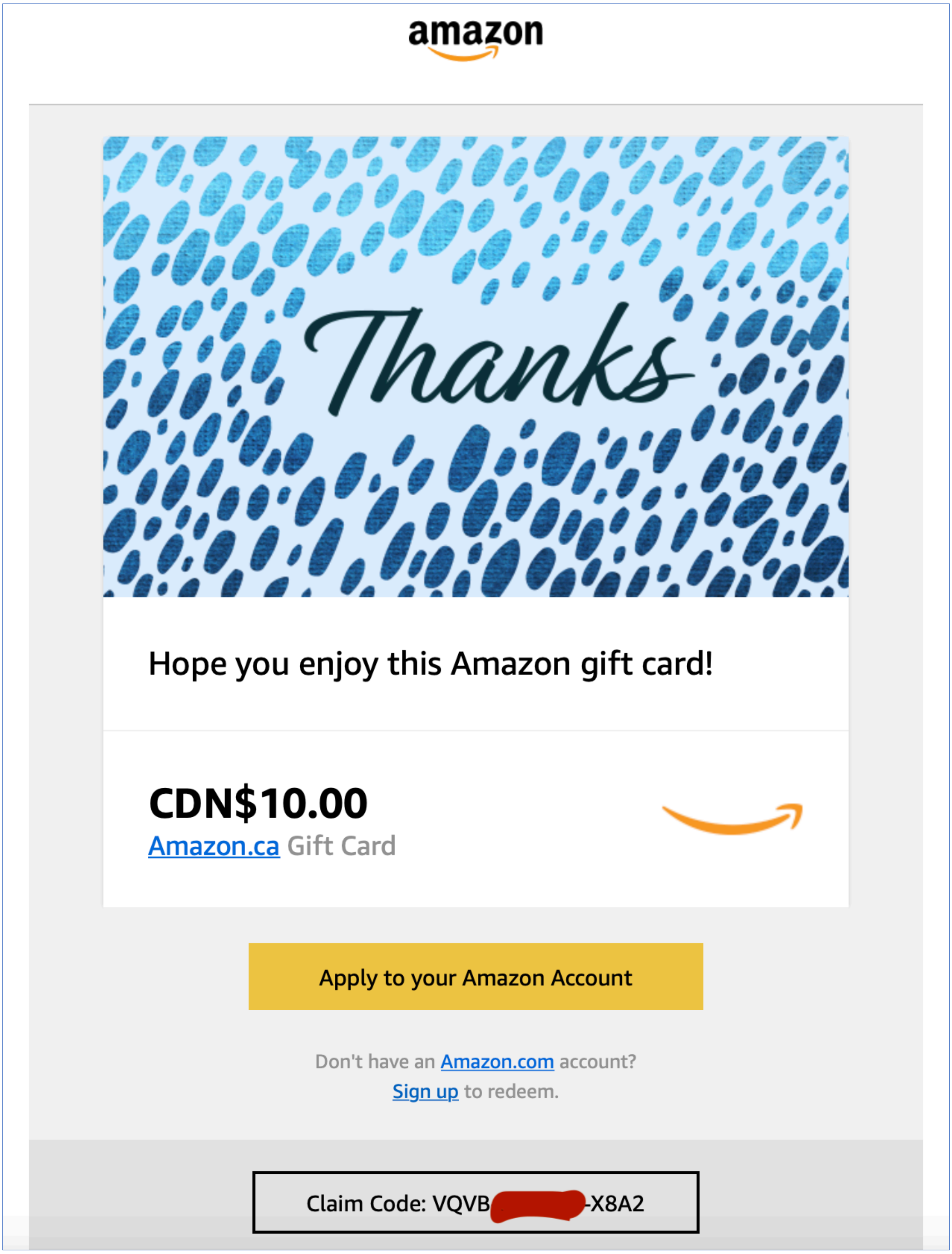


Figure G.1: Amazon gift-card (Claim Code redacted since it is sent to the participant)

APPENDIX H – REB Approval Letter

DALHOUSIE UNIVERSITY
Research Services

Social Sciences & Humanities Research Ethics Board
Letter of Approval

March 11, 2022
Rashidul Islam MD
Computer Science\Computer Science

Dear Rashidul Islam,

REB #: 2022-6028
Project Title: Visualizing Uncertainty with Chromatic Aberration

Effective Date: March 11, 2022
Expiry Date: March 11, 2023

The Social Sciences & Humanities Research Ethics Board has reviewed your application for research involving humans and found the proposed research to be in accordance with the Tri-Council Policy Statement on *Ethical Conduct for Research Involving Humans.* This approval will be in effect for 12 months as indicated above. This approval is subject to the conditions listed below which constitute your on-going responsibilities with respect to the ethical conduct of this research.

Effective March 16, 2020: Notwithstanding this approval, any research conducted during the COVID-19 public health emergency must comply with federal and provincial public health advice as well as directives from Dalhousie University (and/or other facilities or jurisdictions where the research will occur) regarding preventing the spread of COVID-19.

Sincerely,

Dr. Karen Foster, Chair

PID refers to Participant Id based on Session Order among the participants.

Study results of four components (all questions, out of 8):

PID	CA+Bubble	Ca+Grid	VSUP+Bubble	VSUP+Grid
1	6	4	4	5
2	7	5	5	6
3	7	7	5	5
4	8	6	8	5
5	7	5	5	6
6	6	5	7	5
7	5	3	4	3
8	7	5	7	4
9	6	8	5	5
10	4	5	4	5
11	8	6	7	7
12	6	7	5	4
13	3	4	5	4
14	7	5	4	4
15	6	6	7	6
16	6	4	7	5
17	8	6	7	7
18	6	4	3	5
19	7	8	6	7
20	5	7	5	5
21	6	7	5	6
22	5	4	4	3
23	8	7	8	7
24	7	7	6	6
25	7	7	7	6
26	5	5	6	4
27	6	5	5	5
28	8	6	7	4
29	5	4	3	6
30	7	6	7	6
31	8	6	7	7
32	4	5	6	3
Avg	6.3	5.6	5.7	5.2

Table I.1: Questionnaire Raw Scores of four components

Study Results: CA vs VSUP (averaged from four components)

PID	CA	VSUP
1	5.0	4.5
2	6.0	5.5
3	7.0	5.0
4	7.0	6.5
5	6.0	5.5
6	5.5	6.0
7	4.0	3.5
8	6.0	5.5
9	7.0	5.0
10	4.5	4.5
11	7.0	7.0
12	6.5	4.5
13	3.5	4.5
14	6.0	4.0
15	6.0	6.5
16	5.0	6.0
17	7.0	7.0
18	5.0	4.0
19	7.5	6.5
20	6.0	5.0
21	6.5	5.5
22	4.5	3.5
23	7.5	7.5
24	7.0	6.0
25	7.0	6.5
26	5.0	5.0
27	5.5	5.0
28	7.0	5.5
29	4.5	4.5
30	6.5	6.5
31	7.0	7.0
32	4.5	4.5
Avg	5.9	5.5

Table I.2: Study Raw Scores of CA vs VSUP

CA Results for SUS:

PID	Q#1	Q#2	Q#3	Q#4	Q#5	Q#6	Q#7	Q#8	Q#9	Q#10
1	4	1	5	2	5	2	5	1	5	2
2	3	3	3	3	4	3	3	3	3	3
3	3	4	5	4	4	4	2	4	5	3
4	5	4	4	4	4	3	4	4	4	1
5	4	2	4	2	5	2	5	1	5	2
6	5	1	5	1	5	1	5	1	5	1
7	3	3	4	3	5	4	4	4	4	4
8	2	4	2	4	3	2	2	4	3	4
9	3	4	4	4	4	3	5	3	4	2
10	5	1	5	5	5	5	5	1	5	5
11	4	3	4	5	4	5	4	3	4	4
12	3	5	4	2	2	4	2	3	5	5
13	3	2	3	4	4	3	4	3	4	5
14	5	1	4	1	4	1	5	1	5	2
15	4	3	2	2	2	4	5	2	4	2
16	5	5	5	5	5	5	5	5	5	5
17	2	2	4	4	3	4	2	4	4	3
18	5	2	4	5	3	5	4	5	5	4
19	5	1	5	3	5	1	5	1	5	1
20	2	2	4	3	4	3	4	5	5	4
21	4	2	4	4	3	4	3	2	4	2
22	4	2	5	4	5	2	5	2	5	3
23	2	3	3	4	5	5	4	3	4	2
24	2	4	2	3	4	4	2	4	3	2
25	5	1	5	5	5	1	5	1	4	5
26	4	2	4	2	4	3	4	4	3	4
27	4	3	2	4	4	4	5	5	2	4
28	4	4	5	2	5	4	3	3	3	3
29	2	4	3	4	4	2	4	1	2	5
30	4	4	1	5	3	2	5	5	5	3
31	4	2	5	2	5	4	4	5	5	2
32	4	3	3	2	4	1	1	2	3	5
Avg	3.7	2.7	3.8	3.3	4.1	3.1	3.9	3	4.1	3.2

Table I.3: SUS Raw scores for CA

VSUP Results for SUS:

PID	Q#1	Q#2	Q#3	Q#4	Q#5	Q#6	Q#7	Q#8	Q#9	Q#10
1	3	2	2	2	5	4	3	4	2	4
2	3	3	3	3	4	3	3	3	3	3
3	5	1	3	2	4	1	4	2	2	3
4	4	2	4	2	2	2	4	1	2	1
5	4	2	4	2	5	1	2	1	5	2
6	5	1	5	1	5	2	4	2	5	1
7	4	3	4	5	4	5	4	4	5	4
8	5	2	5	2	5	2	5	2	5	1
9	4	4	3	4	4	3	5	3	2	4
10	5	5	1	5	5	5	5	1	5	5
11	5	4	5	5	4	4	5	3	5	5
12	4	2	3	2	4	2	4	2	4	2
13	2	3	3	2	4	2	4	3	5	5
14	3	4	2	4	3	2	4	2	3	3
15	2	2	5	4	4	4	1	5	5	2
16	5	5	5	5	5	5	5	5	3	5
17	4	1	2	1	4	5	5	1	5	5
18	3	5	4	4	5	5	4	5	5	5
19	5	2	5	3	5	1	5	1	5	1
20	4	4	1	2	4	3	2	2	4	2
21	2	1	5	5	5	5	4	3	5	4
22	2	4	3	3	3	3	3	3	4	4
23	1	2	1	3	5	2	3	4	3	1
24	4	2	4	1	4	2	4	2	5	2
25	2	3	5	5	5	1	5	1	4	5
26	5	4	5	4	5	3	4	3	4	4
27	4	2	3	3	5	3	4	5	2	5
28	5	3	4	3	4	4	5	2	4	5
29	4	4	4	3	5	3	2	3	3	5
30	4	3	5	1	4	3	4	1	3	4
31	1	2	4	5	4	2	3	2	3	4
32	5	1	5	4	5	1	1	2	3	5
Avg	3.7	2.8	3.7	3.1	4.3	2.9	3.8	2.6	3.8	3.5

Table I.4: SUS Raw scores for VSUP

CA Results for NASA-TLX:

PID	Q#1	Q#2	Q#3	Q#4	Q#5	Q#6
1	3	2	1	4	8	3
2	13	9	9	12	14	14
3	19	1	18	4	17	17
4	17	0	13	16	15	1
5	11	1	3	20	3	3
6	19	10	15	17	13	6
7	16	15	18	18	17	19
8	16	13	6	8	17	14
9	16	6	12	19	16	6
10	21	21	0	21	21	21
11	16	3	10	18	15	5
12	17	11	12	7	10	11
13	11	14	11	12	7	6
14	4	2	4	18	3	1
15	4	7	18	17	19	16
16	21	21	21	21	21	21
17	15	4	15	5	14	3
18	20	21	20	21	21	20
19	19	0	1	20	1	0
20	10	7	11	12	16	7
21	9	18	2	16	13	18
22	11	8	7	14	16	10
23	17	5	12	6	19	16
24	10	4	5	10	11	2
25	19	11	11	10	4	3
26	11	11	12	12	12	9
27	17	0	1	19	17	17
28	16	7	17	15	15	14
29	12	7	8	12	6	1
30	4	5	13	7	10	11
31	15	18	15	14	9	17
32	20	3	13	13	5	8
Avg	14	8.3	10.4	13.7	12.7	10

Table I.5: NASA-TLX Raw Scores for CA

VSUP Results for NASA-TLX:

PID	Q#1	Q#2	Q#3	Q#4	Q#5	Q#6
1	19	18	19	17	10	18
2	13	9	9	12	14	15
3	3	1	3	16	17	4
4	1	0	6	11	8	1
5	11	1	3	20	3	3
6	20	10	15	18	14	5
7	13	17	9	18	19	7
8	4	3	16	18	5	6
9	20	8	14	17	19	7
10	0	0	0	21	21	21
11	14	2	9	20	18	2
12	13	11	13	12	15	8
13	13	14	10	12	7	2
14	12	6	7	10	9	8
15	17	13	2	13	17	5
16	21	21	21	21	21	21
17	14	20	21	19	20	8
18	20	21	20	21	21	20
19	19	0	1	20	0	0
20	1	1	3	18	6	3
21	4	6	17	20	19	13
22	14	8	10	9	10	15
23	19	6	10	14	21	20
24	1	1	2	17	1	1
25	19	11	11	10	4	3
26	10	11	12	13	14	8
27	15	0	1	19	15	19
28	19	1	20	20	20	20
29	19	11	6	9	8	6
30	17	15	18	21	21	10
31	4	8	9	19	4	8
32	18	2	15	16	5	4
Avg	12.7	8	10.4	16.3	12.7	9.1

Table I.6: NASA-TLX Raw Scores for VSUP

Total time(minutes) utilized: for 8 questions per component

PID	CA+Bubble	CA+Grid	VSUP+Bubble	VSUP+Grid
1	5.6	4.5	4.3	6
2	3.7	4.7	5.1	4
3	4.1	4.2	4.7	4.8
4	5.2	5.8	5	6
5	4.2	3.7	2.8	4
6	1.6	4.2	2.8	2.3
7	1.6	3.3	6.1	2.4
8	4.8	2.5	3.4	3
9	4.9	5.7	4.7	6.5
10	4.5	6.5	5.4	3.8
11	4.2	5.7	6.2	5.5
12	3.9	2.6	2.6	7
13	6.8	4.3	4.2	5.9
14	1.3	3.1	1.8	5.4
15	3.2	5.1	7.6	3.9
16	3.9	2.9	3.7	7.7
17	5.6	5.2	6.6	6
18	1.8	3.4	2	1.4
19	7.8	6.1	8.7	4.8
20	2.3	5.6	4.3	2.7
21	4.7	3.7	4.3	6.2
22	2.5	6.2	3.3	1.9
23	5.9	4.8	6.5	7.1
24	4.4	4.4	3.8	5.3
25	6	5.5	8.8	7.5
26	2.5	3.5	2.4	3
27	3.8	4.1	6.1	3.9
28	4.5	6	6.3	8.3
29	3	1.6	1.7	2.3
30	4.8	5	5.7	7.3
31	4.3	6.2	5.1	5.7
32	5.7	4.4	5.4	5.7
Avg	4.1	4.5	4.7	4.9

Table I.7: Time Utilization for Full Questionnaire