QATAR UNIVERSITY

COLLEGE OF ENGINEERING

EVALUATION OF 2D AND 3D TECHNIQUES FOR SENTIMENT VISUALIZATION

BY

WAFA WAHEEDA SYED

A Thesis Submitted to

the Faculty of the College of

Engineering

in Partial Fulfillment

of the Requirements

for the Degree of

Masters of Science in Computing

June 2018

COMMITTEE PAGE

The members of the Committee approve the Thesis of Wafa Waheeda Syed defended on 21/05/2018.

Dr. Khaled Shaban
Thesis/Dissertation Supervisor

Dr. Osama Halabi
Thesis/Dissertation Co-Supervisor

Dr. Noora Fetais
Committee Member

Approved:

Khalifa Al-Khalifa, Dean, College of Engineering

ABSTRACT

SYED, WAFA WAHEEDA, Masters: June : 2018, Masters of Science in Computing

Title: Evaluation of 2D and 3D Techniques for Sentiment Visualization

Supervisor of Thesis: Dr. Khaled Shaban & Dr. Osama Halabi.

With the rise of user generated content on the Internet, sentiment visualization is being highly researched and practiced. Advances in information visualization, such as the use of three-dimensional visualizations need to be applied to sentiment visualization. However, minimal efforts were taken in the literature, to address when two-dimensional (2D) and three-dimensional (3D) visualization techniques can be used for sentiment visualization. In this thesis, we investigate the 2D and 3D visualization techniques based on the visual variables which represent sentiment in sentiment visualization and perform a comparative empirical study. We conduct a task-based evaluation to measure the performance and cognitive load of visualizations where sentiment is represented by different visual variables in both 2D and 3D visualizations. The objective of this work is to find when 2D and 3D visualization techniques can be used for sentiment visualization and which visual variable is comparatively well-suited for visual representation of sentiment in 2D and 3D. We use scatterplot and bar chart in 2D and 3D for case-study. While the results reflect the known fact that 2D has better performance and lower cognitive load, we investigate different scenarios involving the visual representation of sentiment in 2D and 3D visualizations. Additionally, we discuss the trade-offs of using 2D and 3D visualizations for sentiment visualization. We expect this study to help data analysts, sentiment analysis and visualization researchers and developers make an informed decision of when 3D visualization can be used for sentiment visualization.

DEDICATION

To my parents

ACKNOWLEDGMENTS

This thesis would not have been possible without the support and guidance from my supervisors Dr. Khaled & Dr. Osama. Dr. Khaled, for always being open for discussions when I was confused. I hope to be as good a researcher as him. Dr. Osama for introducing me to the world of Human Computer Interaction, and for sharing his passion towards the Visualization and HCI fields with me. I would like to thank my parents and sister for their extensive support with no bounds. I also thank my friends, specially Alaa and Rachael, for their moral support and encouragement. Finally, I thank the participants of the evaluation for their valuable time and feedback.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

This chapter gives a brief overview of the problem addressed in this thesis. The problem statement, motivation, scope and objectives of the thesis are clearly defined in this section. The thesis adopts a quantitative approach to solve the problem, thus the hypothesis is stated and tested. Finally, we provide a short outline of this thesis.

1.1 Overview and Background

A. Information Visualization

Visualization has been well-known and used for ages, as an activity of forming a mental model of something. The practicability, applicability and impact of visualizations grew with the rise of computers. Correspondingly for visualization in scientific computing, a pertinent definition by McCormick et al., in 1987 [1] is that *"Visualization is a method of computing. It transforms the symbolic in the geometric, enabling researchers to observe their simulations and computations. Visualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights."* Since the advancements with web technologies such as the development of browsers, the use of web-based visualizations has increased. Web-based visualizations also facilitate sense making of digital data, either in an explanatory or exploratory manner. The explanatory visualizations often tend to be static and communicate inferences by visually highlighting and representing the inferences from data directly [2]. Whereas, the exploratory visualizations involve data exploration using the interaction features, that are integrated to the visualization charts in order to enable exploration and analysis of the underlying data. The exploratory visualizations are also referred to as interactive visualizations, and they are used to aid analysis by allowing the users to visually explore and make inferences.

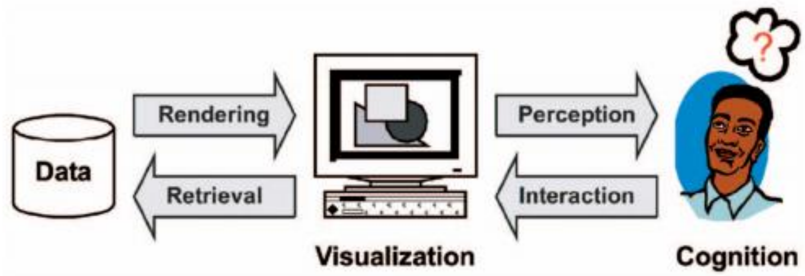

Figure 1: The visualization process [4]

Visualization domain is broadly categorized as scientific and information visualization. Information visualization is a visual representation of data, to facilitate understanding and generate insights [3]. Information visualizations represent data, by mapping data to graphics such as shapes and other graphical properties such as color, size, position, texture and orientation. These graphical properties are also referred as pre-attentive visual elements [4], [5] or visual variables which are used to represent data. Very often, two dimensional (2D) visualizations, involving x and y axes are used for visualizing the data graphically in form of charts or graphs. Similarly, three dimensional (3D) visualizations consisting of x, y and z axes use 3D shapes and graphical properties such as lighting, to visualize data in information visualization. Certainly, visualizations are an integral part of visually driven data analysis and sense-making, as visualizations help users understand the data better. Subsequently, it is inherent that human factors such as perception and cognition are involved in the visualization process, as elaborated in [4] and shown in Figure 1. Here, cognition is the process of human understanding through the interactions made by looking at the visualizations. And visualizations are known to provide cognitive support in communicating insights from the data [4].

B. Sentiment Visualization

With the rise of Internet of Things, social media platforms and other technological advances, there is increasing research and development efforts focused on understanding or making sense of the data obtained. In order to facilitate understanding and decision making, the data is often visualized in form of charts or graphs, as trends, patterns, and outliers in data can be inferred easily.

Twitter, a microblogging social media platform generates around 500 million tweets per day. This data is of high interest to researchers and analysts, for generating insights and making interesting discoveries. In order to understand the overall opinion or sentiment on particular topics discussed on the Internet, the online generated digitized textual data is processed using Natural Language Processing techniques such as Sentiment Analysis. Sentiment Analysis is a process of extracting sentiment from text and it is also known as 'Opinion Mining'. On processing, sentiment polarities such as 'positive' and 'negative' are obtained. 'Neutral' is also used as a sense of sentiment lying between positive and negative [6]. Sometimes the sentiment polarity is also classified across emotions such as 'anger', 'fear', 'disgust', 'sadness', 'surprise', and 'joy' [7]. Visualization of sentiment analyzed data or '*sentiment results'* is known as '*Sentiment Visualization'*, which is a research area belonging to information visualization and visual analytics, these in turn belong to the larger domain of text visualization [8]. Here, the visualization is used to communicate insights or aid analysis.

Sentiment Visualization has emerged as a research domain over the last decade, due to the rise and increasing availability of the user generated content on the internet in form of social media posts. Over the years, sentiment visualizations adopted factors such as story-telling [9] and use of visual metaphor [10] in visualizations to facilitate the communication and exploratory analysis of visualized data.

Advances and trends in information visualization and visual analytics adopt the use of three dimensional (3D) visualizations, however they remain underutilized for sentiment visualization. In the literature for sentiment visualization, it is observed that basic two dimensional (2D) charts such as bar, line, and scatter plots are mostly used [8]. Dashboards with multiple 2D basic charts are also found to be used for providing context, by representing a cumulative view of the sentiment results [5][9][10]. Since the sentiment results are often multi-variate and multi-dimensional, when visualized, the sentiment results are to be sliced and perceived at many levels, from granular-levels of data attributes to aggregated sentiment distributions over the entire data. It is often challenging for researchers, analysts or users to analyze and perceive the different perspectives of the data [6] from individual 2D charts or in a single view using dashboard. Use of 3D for visualizations can facilitate the understanding of such complex data, as three variables can be compared at a time. While recent advances and trends in information visualization and visual analytics have already adopted the use of 3D, they remain underutilized for the sentiment visualization domain.

C. Evaluations in Visualizations

Sentiment visualizations are often exposed to problems such as visual complexity while using 2D or 3D charts, which makes it difficult for users to interpret visualizations. Evaluating visualizations for ease of use is critical, as they require human effort to interpret and understand the visualized data and generate insights. The interpretation, understanding, and insight generation are high level cognitive functions performed by humans [13]. Often, evaluations in visualizations are considered challenging, due to the fact that there are many visualization factors to consider, such as the understanding of data using visualization tasks or the interaction and visualization technique involved.

Since visualizations are created for humans to understand the underlying data and generate insights, they are often evaluated for user performance and experience [14], [15]. Generally, human computer interaction (HCI) evaluation methods such as usability experiments, surveys and questionnaires are used. The nature of these evaluations is either formative or summative. Usability evaluations are widely practiced for measuring the usability metrics such as complexity and learnability of any information visualization system. There are multiple usability evaluation methods that are used for information visualization, depending on the different scenarios and measures [16].

1.2 Problem Statement and Motivation

Besides the existence of 3D visualization techniques for over 100 years [17], to the best of our knowledge, it is not as commonly used as the 2D in sentiment visualization domain. Sentiment visualizations often deal with visualizing the sentiment distribution over different data attributes that are demographic, geographic or temporal [18]–[26].

Visualizing these multi-dimensional sentiment results, over 3D visualizations can aid in generating insights in form of trends, patterns and outliers, as in 3D, the third dimension gives capability to visually represent and spatially compare more variables. Data transparency [27] is also obtained through different rotational scene navigation views.

However, 3D visualizations tend to be more complex when compared to 2D visualizations, as they involve factors such as occlusion, rotational scene navigation and perspective perception of underlying data [17]. Besides that, 2D visualizations also have disadvantages while dealing with complex multidimensional data, as it is not possible to visualize it in a single chart. In addition to using custom visualizations, often multiple 2D traditional charts are aggregated in a single dashboard view, to visualize this complex data. Using many 2D charts in a single view, often overloads the user and there are limitations on the number of graphs to be used [28].

We notice that though 2D and 3D visualizations have their own advantages and disadvantages, they are commonly used in information visualization. In this thesis, we advance the sentiment visualization by adopting 3D visualizations and aim to address the problem of when 2D and 3D visualization techniques can be used for visualizing sentiment results. We measure the performance and cognitive load of 2D and 3D under different conditions, in order to find out when 3D is better than 2D. We expect this study to help data analysts, sentiment analysis and visualization researchers and developers make an informed decision of when 3D visualization can be used for sentiment visualization.

1.3 Scope and Objectives

Sentiment is a multi-class field with polarities such as positive, negative and neutral. In the literature, sentiment is visually represented by visual variables such as color, position, size, shape and texture. According to a recent survey on sentiment visualization by K. Kucher et. al. in [8], color is found to be mostly used for visually encoding sentiment. However, these visual variables behave differently in 2D & 3D visualizations. Thus, by adopting an HCI approach, we aim to address when 2D and 3D visualization techniques can be used for sentiment visualization. We use the traditional and commonly used scatter plots and bar charts in both 2D and 3D. In the scope of this thesis, we limit ourselves to using traditional charts with basic visual elements and no visualization interaction methods in charts, so we can establish a baseline for common usage of traditional charts.

In sum, the objective of this thesis is to study how the sentiment visual variables behave in 2D & 3D visualization environments, as we hope this will enable us to address the research question of when 2D and 3D visualization techniques can be used for sentiment visualization. In this thesis, we aim to perform a comparative usability study to evaluate the user performance and cognitive load while using the sentiment visualizations in both 2D and 3D.

1.4 Hypothesis

This thesis is a comparative and quantitative usability study, as we take an HCI approach to answer when 2D and 3D visualization techniques can be used for sentiment visualization, based on performance and cognitive load metrics. In the usability experiments that we conduct, we have different variables and experimental conditions under which we propose and investigate the following hypothesis:

1. Cognitive Load will be better on 3D charts while using sentiment visual representation as color than position and size
2. Performance will be better when sentiment is visually represented by color in 3D charts than position or size.
3. Scatterplots will have better performance than bar charts when Sentiment is visually represented by size in 3D.

1.5 Outline

The remaining thesis is structured as follows: Chapter 2 briefs upon the related work of the thesis contributions listed above. Chapter 3, we discuss the methodology and experiment design. In chapter 4, we discuss the experimental evaluations and result analysis. Chapter 5 consists of a discussion on the results. Then, we conclude the thesis in chapter 6 and discuss future work.

CHAPTER 2: RELATED WORK

This chapter summarizes the work published in the literature, which is related to the proposed work in this thesis. Here, an overview of the following topics and how they have been addressed in the literature is presented. In addition to sentiment visualization, the use of 2D and 3D visualizations for sentiment visualization, and different usability evaluation methods applied in the literature for evaluating information visualizations and sentiment visualizations are reviewed.

2.1 Sentiment Visualization Techniques

Sentiment visualization has become significant for analysis of textual data which is available and processed in digital form. This digitized textual data is either available offline in computers or online on the internet. There are many sentiment visualization studies in the literature, that address a problem or research gap. In this section, we review the literature for sentiment visualization studies and discuss the data being visualized, type of visualization technique adopted and other common factors. Sentiment visualization studies in literature focus on a variety of topics from domains such as the culture [18], disaster [17][18], e-commerce [21]–[23], health [24]–[26] and politics [25][26]

Sentiment results obtained are multivariate, due to the presence of multiple attributes in data. Based on the domain or topic of research problem addressed by the sentiment visualization study, the data is gathered online or offline. The studies visualize the sentiment distribution over different data attributes of the sentiment analyzed data. For instance in [31], H. Dong et. al., analyze the anomalous information spreading on social media and use Hurricane Sandy as a case-study. They visualize the temporal data attribute in retweeting threads and detect anomalies. E.C. Resende et. al., in [26] studied sentiment over demographic and temporal attributes of tweets about e-cigarettes. In the visualization, the sentiment polarities were distributed over time of tweet and gender of users posting the messages. Color was used to visually represent sentiment as green was used for positive sentiment polarity and red to represent negative sentiment polarity. Here, a traditional visualization chart, connected scatterplot with different shapes such as triangle, circle and square were used as data plots. Though these [26], [31] studies analyze tweets, they studied different data attributes using sentiment visualization. And this depended on the nature and objective of the study.

Based on the commonalty of visual elements and their usage, charts such as bar chart, line chart, pie chart, scatterplot, area chart are known as traditional visualizations. From the sentiment visualization literature, we note that the traditional visualization charts are mostly preferred and are subject to minor modifications. The modifications in traditional visualization charts are often using colors, font attributes like font-weight and font-style, shapes and spatial positioning. For instance, warm and cool colors are used to represent sentiment polarity. In addition, color attributes such as the opacity and brightness are used for different purposes. In [22], the color brightness is used to represent the percentage of reviews in a cluster and in [29], the color opacity is used to represent the reachability of news, where the less opaque color marks the widespread news. Furthermore, this study, uses the vertical displacement (a.k.a. spatial position) of a symbol to mark the polarity of sentiment instead of colors and uses opacity to show reachability of news. The study [32] represents sentiment by using colored lighting to illuminate buildings and regions in pictures, which were obtained from Google street view.

The selection of traditional charts depends on the data type of the attributes visualized, and the tasks to perform in order to obtain different insights to be communicated to users. For chart suggestions, readers can refer to the work of B. Faket et.al., in [33] A. Abela in [34], where different traditional charts are suggested to use based on what insights to facilitate, while communicating the results to users. Moreover, user requirements, platform constraints, and developer biasness also influence the type and complexity of the charts used. For instance, P.K. Novak et. al., in [35] propose an emoji sentiment ranking system for automated sentiment analysis and formalize the results as visualization in form of a sentiment bar. They use horizontal stacked bar chart with customizations such as plotting grey rectangle shaped bar over the bars to map the sentiment range for each emoji. Also, the emoji replace the axis label text in visualization.

Dashboards combine and link multiple visualization charts in a single view. Either traditional charts, custom charts or both are used. TwitInfo by A. Marcus et. al., [36] aggregated and visualized events on Twitter using dashboard with only traditional visualization charts. This system summarized Twitter events over time and location. The dashboard consisted of charts such as: line chart - for displaying the frequency of messages, pie chart - to display overall sentiment and map - to show the sentiment distribution over location. Tweet list - and popular links - are also displayed. The dashboard also allowed users to give user-defined name to the event and displayed event keywords which summarized tweets as text. This allowed the user to interpret the entire summary of an event over time.

In sentiment visualization literature, a few studies propose new custom charts for visualizing the sentiment results. These newly proposed custom charts are usually novel, as they introduce new spatial arrangement of visual elements or use well-known visual metaphors and propose new visual elements. For instance, F. Wanner et. al, in [29] studied the politics domain, using data from Rich Site Summary (RSS) for analyzing and visualizing the sentiment of news articles featuring the 2008 US elections. The contributions of this study lie in the visualization technique as it creates custom visualization, using bars as glyphs and encoding data attributes to position, color, shape and opacity.

Majority of the sentiment visualization studies in literature use 2D visualization charts. There are very few studies like [28][33] which use 3D for sentiment visualization and comparatively, they remain underutilized for sentiment visualization. This could be because of a research gap in literature, on when to choose 3D over 2D for sentiment visualization. In this thesis, we address this gap by adopting an HCI approach to find when 2D and 3D visualization techniques can be used for sentiment visualization.

2.2 Evaluations in Visualizations

A. Evaluations in Sentiment Visualization

If not carefully designed, visualization charts comprising of traditional and custom interactive components, may complicate the ease of use and under-standing. This might lead to reducing the user performance and user experience. Hence the sentiment visualization systems are evaluated using HCI evaluation methods such as heuristic evaluation, interviews, surveys and questionnaire, to find usability problems and assess various aspects of usability. The studies performed usability evaluation for the purpose of testing and inquiry; i.e., user testing methods such as Think-Aloud, and controlled experiments. And user inquiry methods such as questionnaire, survey and interviews were performed for measuring usability. Usability methods which involve improvement by inspection, such as the use heuristic evaluation to find the usability problems, remained underused [18], [38]–[40].

D. Duan et. al., in [14] evaluate the usability, by measuring the user satisfaction of their system using user experiments. They included two other baseline systems in their study and evaluated the usefulness and user satisfaction of the three systems. Tasks were designed to focus on different aspects of the systems and a set of survey questions were used to get the users' ratings. Whereas, in [41] the socialHelix visualization system was evaluated by conducting informal interviews with HCI and sentiment analysis domain experts to evaluate the usability. Study [24] employing a user centered approach while developing visualizations, in the process they undertook usability evaluations to obtain feedback from domain experts. A design study consisting of tasks followed by interviews were conducted iteratively to evaluate usability.

In the sentiment visualization and information visualization literature, many studies that do not evaluate their visualizations. Less than 50% of the information visualization studies were evaluated [16]. This is due to the lack of well-established evaluation and usability guidelines for visualization studies. In [16], H. Lam et. al, identify seven different scenarios for evaluation in information visualizations by an extensive literature review. They based the scenarios based on the objective of the studies and discovered common evaluation scenarios. H. Lam et.al., in [42] further categorize the seven scenarios as process and visualization. Evaluations using user performance, user experience and evaluating visualization algorithms were the scenarios under visualization category.

On the other hand, A. Shamim et.al., in [43], explicitly work on evaluating eleven sentiment visualization techniques through a questionnaire. Their objective was to rank eleven selected sentiment visualization techniques and find important visualization metrics. The results of the study indicated that bar charts to be one of the top-five ranked techniques for sentiment visualization, and the ease-of-use, understanding, user-friendliness, etc. to be important metrics for sentiment visualization.

B. Comparison of 3D & 2D in Literature

For many years, the problem of when 2D or 3D can be used for information visualization has been in existence [44]. So, in this section, we review the existing work for when 2D or 3D can be used in information visualization.

Despite the fact that there are only handful of 3D visualization approaches in the literature for sentiment visualization, to the best of our knowledge there are no efforts made to understand when 2D and 3D visualization techniques can be used for visualizing sentiment results. In [45], the authors study the 2D and 3D techniques for spatial movement data - which consists of two spatial and one temporal variables, which is three-dimensional in nature. Though 2D and 3D visualization techniques have their own advantages and drawbacks, the two techniques are compared using controlled experiments by performing common tasks on 2D & 3D visualizations. This work evaluated the 2D and 3D techniques, which were designed with basic interactions for the spatial movement data. The occlusion aspects which occurred in 2D and 3D were studied, and the results of the experiments showed that it was better to use 3D over 2D for spatial movement data. Additionally, in the work of H. Liao et.al., in [46] studies visual attention of users in 2D maps and 3D geo-browsers, in acquiring spatial knowledge and decision making. The authors perform controlled experiments and use eye-movement data for evaluating the behavior of users while acquiring knowledge in 2D and 3D visualization environments.

However, in [45], [46] the behavior of visual elements such as color, size and shape, which are an integral part of any visualizations, were not studied. Therefore, general inferences about when to use 3D over 2D cannot be made from these studies for different types of visualization tasks and data. In [15], the effectiveness visual variables effect on geographic information visualization 2D visualization were studied, using the eye-tracking. Whereas, for the comparison with 3D on how the visual variables react in 3D was not addressed. Recently, in [47] the visual variables were tested for their guidance and constancy in perception capabilities in 3D. Here the size, shape and hue variables were studied. The results indicated that n 3D hue and shape provided more guidance and? constancy than size. However, it was the vice-versa in 2D, as size was considered to have more visual guidance. From this study we note that the visual variables have different capabilities, behavior and performance in 2D and 3D environments. Additionally, in [48] B.M. Hughes compared 2D and 3D bar charts for chart readability using psychophysical analysis method called constant stimuli, where the participants undertook the evaluation with an overhead projector. However, only chart readability was assessed using accuracy, instead of response time.

In order to understand the behavior of 3D information visualizations, there are studies which compare multiple 3D visualizations using evaluations. For instance, in [49] an empirical study is performed on three different 3D information visualizations, in order to understand the usability factors. Tasks such as search, count and compare were performed on each. From the results of this study, the authors convey that having overview for 3D visualizations is important. These studies do not take into account the 2D information visualizations.

Moreover, studies like the work of S. Dubel et. al., in [44] facilitate 2D and 3D comparison by formalizing the 2D 3D visualization for spatial data in terms of attribute space and reference space presentation characteristics. This work guides researchers dealing with spatial data on when to go for 2D or 3D.

CHAPTER 3: METHODOLOGY

In this chapter, we highlight the approach adopted by the thesis to answer the research question of when 2D and 3D visualization techniques can be used for sentiment visualization. A summative usability evaluation approach is adopted to compare the performance and cognitive load of 2D and 3D visualizations under different experimental conditions. In the following sections, we discuss the sentiment analysis in stages such as data gathering, data preprocessing and data transforming. After the sentiment results data transforming stage, they are ready to be visualized. We discuss the 2D and 3D visualizations created for sentiment results. Additionally, the empirical study – which is the usability experiment designed to compare 2D and 3D visualizations is discussed. Here, the experiment design, participants, materials required, and the experimental procedure adopted are detailed.

3.1 Sentiment Analysis & Visualization

A. Data Gathering

To begin with, we narrow down on a popular use-case for performing sentiment analysis and visualization. We choose the politics domain and select the 2016 US elections as a case-study and decide to study the sentiment on the tweets gathered from the Twitter accounts of election candidates. We use a reviewed dataset from kaggle[1], uploaded by Ben Hammer, the cofounder of kaggle. The dataset is titled “Hillary Clinton and Donald Trump Tweets”, who were the major party candidates in 2016 US elections. This dataset consists of tweets

[1] www.kaggle.com

collected during 5th January 2016 to 27th September 2016 from Hillary and Trump's Twitter accounts. The tweets are associated with the following meta-data: *id, handle, text, is_retweet, original_author, time, in_reply_to_screen_name, in_reply_to_status_id, in_reply_to_user_id, is_quote_status, lang, retweet_count, favorite_count*. The dataset was first subject to preprocessing, where we remove the missing information such as the location from the dataset. Here, as the 2016 US Elections is an event happened in past, we narrowed to study only Trump's tweets as he won the elections. So, using the *handle*, we filter Trump's tweets data for further processing.

B. Data Preprocessing

Subsequently, the tweets which are represented as *text* in preprocessed dataset, are subject to sentiment analysis. Since the scope of this thesis is limited to sentiment visualization, our efforts are minimal on the performance or efficiency of the data processing technique, for instance: the accuracy of sentiment analysis. TextBlob - a python library, which is generally used for performing common Natural Language Processing tasks, is used to perform sentiment analysis. As a result, the sentiment polarity and subjectivity scores for each tweet analyzed are obtained. We only concentrate on the polarity score as it serves as an indication for the sentiment to be either positive, negative or neutral as the polarity scores range from -1 to +1. The negative polarity scores indicate negative sentiment, and the positive polarity scores greater than 0, indicate positive sentiment, while the polarity score of 0 indicates neutral. We process the polarity score of each tweet and store the sentiment as positive, negative or neutral.

C. Data Transformation

In order to study the sentiment of tweets, we select the following variables: *id, text, time, retweet_count* and *favorite_count* and aggregate them over total positive, negative and neutral sentiment for each month. So, we are left with the following variables: *month, tweets, retweet, fav, sentiment* and *polarity* as shown in Table 1. Thereupon, the *tweets, retweet* and *fav* variables represent the aggregated measures: total number of tweets, retweets and favorites for each month, which are distributed over the respective representative sentiment and aggregated polarity scores. We know that, for obtaining insights, such as patterns, outliers and connections from the data, we need to visualize it. The selection of visualization charts often depends on the data type and number of the variables visualized. In statistics, the types of data are broadly categorized as quantitative and categorical. Quantitative data is numerical and can consist of either discrete or continuous values. Whereas the categorical data consists of ordinal or nominal values. In our case, along with the *polarity,* we have the *tweets, retweet* and *fav* count as quantitative data variables. The *month* and *sentiment* are nominal data variables. However, sentiment can also be considered ordinal, as negative, neutral and positive in order of increasing polarity value. Note that, we chose to study Trump's tweets by aggregating the time in months. Additionally, the use of 2D or 3D is governed by the number of variables we want to visualize over the sentiment, as in sentiment visualizations the sentiment is distributed over different data variables, obtained from the sentiment results.

Table 1: Data for Jan, Feb

Month	Tweets	Retweet	Fav	Sentiment	Polarity
Jan	244	535756	1492729	Positive	0.404
Feb	255	778955	2262762	Positive	0.382
[illegible]	[illegible]	23394[illegible]	62253[illegible]	Neutra[illegible]	[illegible]

D. Sentiment Visualization

Furthermore, as mentioned in the scope of this thesis, we do not use custom visualizations or visualization interactions, as we focus on studying different visual representations of sentiment in 2D and 3D visualizations.

According to the recent survey on sentiment visualization [8], color, position and size, are majorly used visual variables to visually represent sentiment in visualizations used in the literature. Thus, the sentiment visualizations generated in both 2D and 3D, visualize sentiment over color, position and size visual variables. The charts visualize sentiment using one visual variable at a time for each set of data attributes. For instance, the charts created, visualize the same data attributes and encode sentiment either by color, size or position.

i) Visualization Design

Based on the questions we want to address using the sentiment visualizations, we select the data variables from sentiment results and visualize. Also, depending on the variables involved, we select the visualization charts to visualize aggregated sentiment results data. As discussed, we are studying the Trump tweets and want to analyze the following:

1. Over the months, what was the overall sentiment distribution of Trump tweets?

2. Did Trump receive less retweet and favorite counts for negative tweets compared to positive tweets?

3. Which months did Trump tweet most positive? And what was the retweet and favorite count?

4. With increasing positive number of tweets, was there an increase in retweet and favorite count?

the variables involved in each of the questions are grouped.

In addition, to study the sentiment distribution by using visualizations to find answers to the above questions we group the variables as: Month-Tweets, Retweet-Fav, Month-Retweet-Fav and Tweet-Retweet-Fav.

This grouping of variables holds true only for sentiment visualization where sentiment is visually represented as color and size. This is because the color and size visual variables take into account categorical sentiment values as negative, neutral and positive. On the other hand, while sentiment is being represented as position, the sentiment value – polarity score is used. Additionally, to represent sentiment by position, the polarity values are plotted on the y-axis, as doing so will spatially position sentiment and cluster the data values by the sentiment polarity. Here, polarity as we know is a quantitative interval value used to give an indication of the sentiment. So, we have the following grouping of variables to visualize sentiment by position: Month-Polarity, Tweets-Polarity, Retweet-Polarity, Tweets-Polarity-Retweet, Retweet- Polarity-Fav. Note that there are two groups with two variables and two groups with three variables. Thus, in 2D, the groups with two variables are visualized on the x and y axis, while the three variables are visualized as multiple 2D charts with a variable or axis x in common, such that the multiple 2D charts visualized represent the variables in x-y and x-z axis. However, in 3D since three variables are to be used, the groups with two variables were visualized with another variable along the z axis: Month-Tweets-Retweet and Retweet-Fav-Tweets. Note that Retweet and Tweets in this case are the variables represented on the z-axis, however the z-axis does not contribute to the question addressed by the 3D visualization. A summary of the group of variables used for respective sentiment representation in 2D and 3D is presented in Table 2.

Table 2: Summary of Variables and Sentiment Representation used for 2D and 3D

Sentiment Representation	**Group of Variables**	**2D**			**3D**		
		x	**y**	**z**	**x**	**y**	**z**
Color & Size	Month-Tweets	Month	Tweets	-	Month	Tweets	Retweet
Color & Size	Retweet-Fav	Retweet	Fav	-	Retweet	Fav	Tweets
Color & Size	Month-Tweets-Retweet	Month	Tweets	Retweet	Month	Tweets	Retweet
Color & Size	Tweets-Retweet-Fav	Tweets	Retweet	Fav	Tweets	Retweet	Fav
Position	Tweets-Polarity	Tweets	Polarity	-	Tweets	Polarity	Retweet
Position	Retweet-Polarity	Retweet	Polarity	-	Retweet	Polarity	Tweets
Position	Tweets-Polarity-Fav	Tweets	Polarity	Fav	Tweets	Polarity	Fav
Position	Retweet-Polarity-Fav	Retweet	Polarity	Fav	Retweet	Polarity	Fav

Ideally, the bar chart and scatterplots are used to facilitate both overview and comparison of the data underneath. Also, as these charts are commonly used traditional visualization charts, we select bar charts and scatterplots to visualize the sentiment results in 2D and 3D. The bar charts use length which attributes to the visual dimension of size (bar height) to represent the data. And scatterplots use spatial position between the x and y axis to visually represent data. While bar charts are ideally used to represent categorical and numerical value on the x and y axis respectively, scatterplots are used to represent numerical values. Note that the spatial position and size used in explanation above describing the nature of charts and it need not be confused with the visual representation of sentiment in 2D and 3D charts.

For creating the 2D and 3D sentiment visualization charts, we use Tableau software, D3[2] and ThreeJS[3] JavaScript libraries. The ThreeJS library along with D3, is used to create 3D visualizations, as it is built upon WebGL to support 3D graphics for web. In order to find insights from visualizations, we create 2D and 3D charts that visualize sentiment using one visual variable at a time for each group of variables. Figures 2-5 are the 2D scatterplots and bar charts and Figures 6-9 are the 3D scatterplots and bar charts generated.

[2] www.d3js.org

[3] www.threejs.org

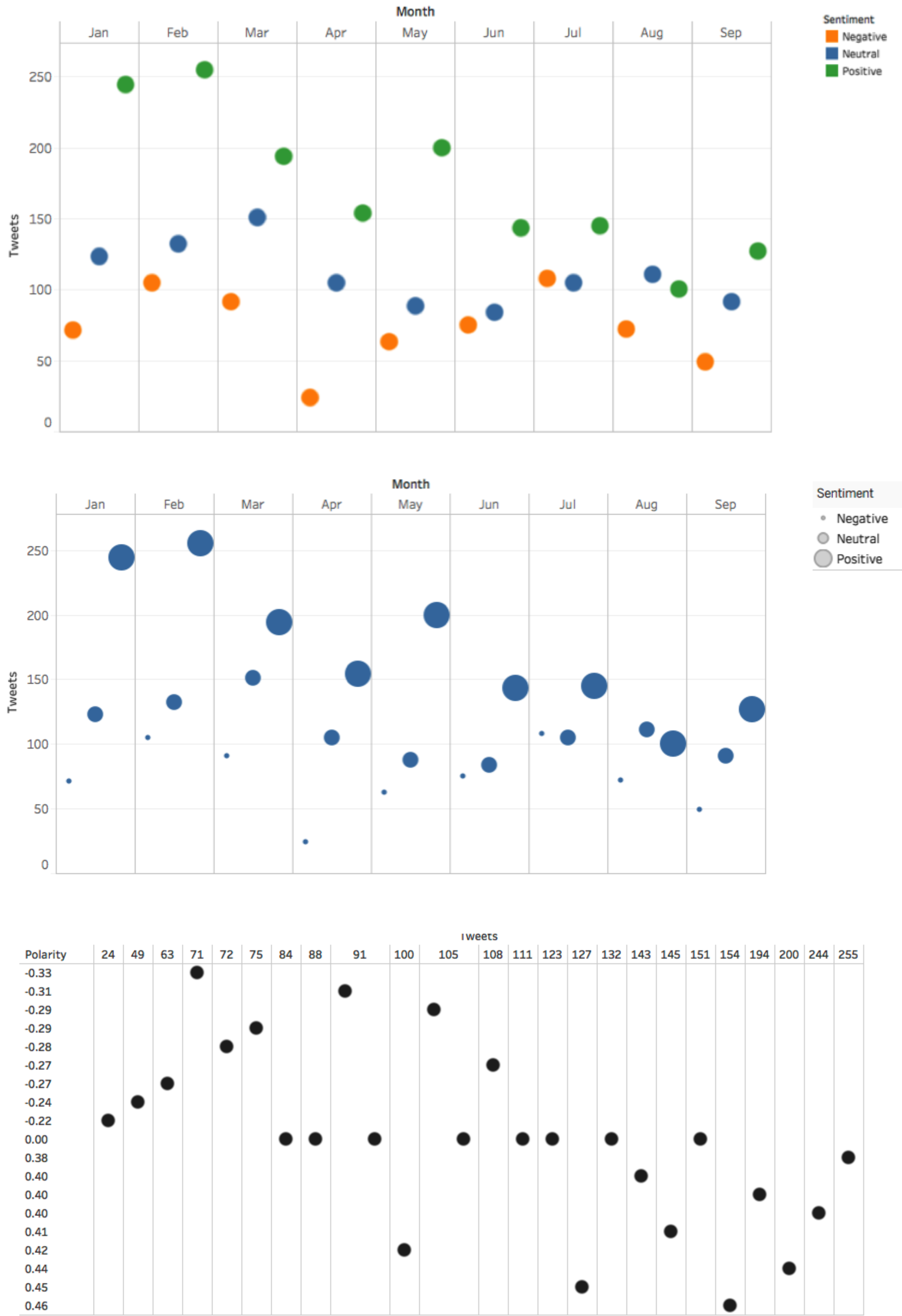

Figure 3: 2D scatterplots for month over no of tweets and polarity over no of tweets of Trump. (From top) Sentiment is represented as color, size and position respectively.

ii) 2D Sentiment Visualization

Taking in to account the visualization task and the data variables to be studied, the 2D scatterplots and bar charts were generated. The other important factor considered was the visual representation of sentiment. First, the sentiment by color, size and position was done for month-tweets and retweet-fav group variables. Sentiment by color was easy to allocate. However, the values were getting overlapped, in order to remove that the data points were grouped by the sentiment dimension.

Also, since the retweet and fav values have data values in a wide range, in order to avoid overlapping, we perform grouping and binning of the visual elements representing data points. Grouping is used for categorical axis, where the data points are visually grouped by values of the sentiment. While, binning is used for quantitative values in axis, such as for retweet and fav variables in order to avoid overlapping and generate clearer charts.

Furthermore, for sentiment by size in 2D, the sentiment data-type was used as ordinal, with negative being the smallest, neutral using medium size and positive sentiment using the largest size. Also, to represent size in bar charts, variable width was applied to bars. Finally, for sentiment by position in 2D, we used the polarity values on y-axis. So, the scatterplots and bar charts using tweets-polarity and retweets-polarity on the x-y axes were generated.

Moreover, since bar charts plot the y-axis as bar height, which adds visual weight on the y-axis, while sentiment being represented as polarity, the data points with neutral sentiment are generated as lines – with no height or visual weight, as they have the polarity value of 0, as shown in in Figure 3.

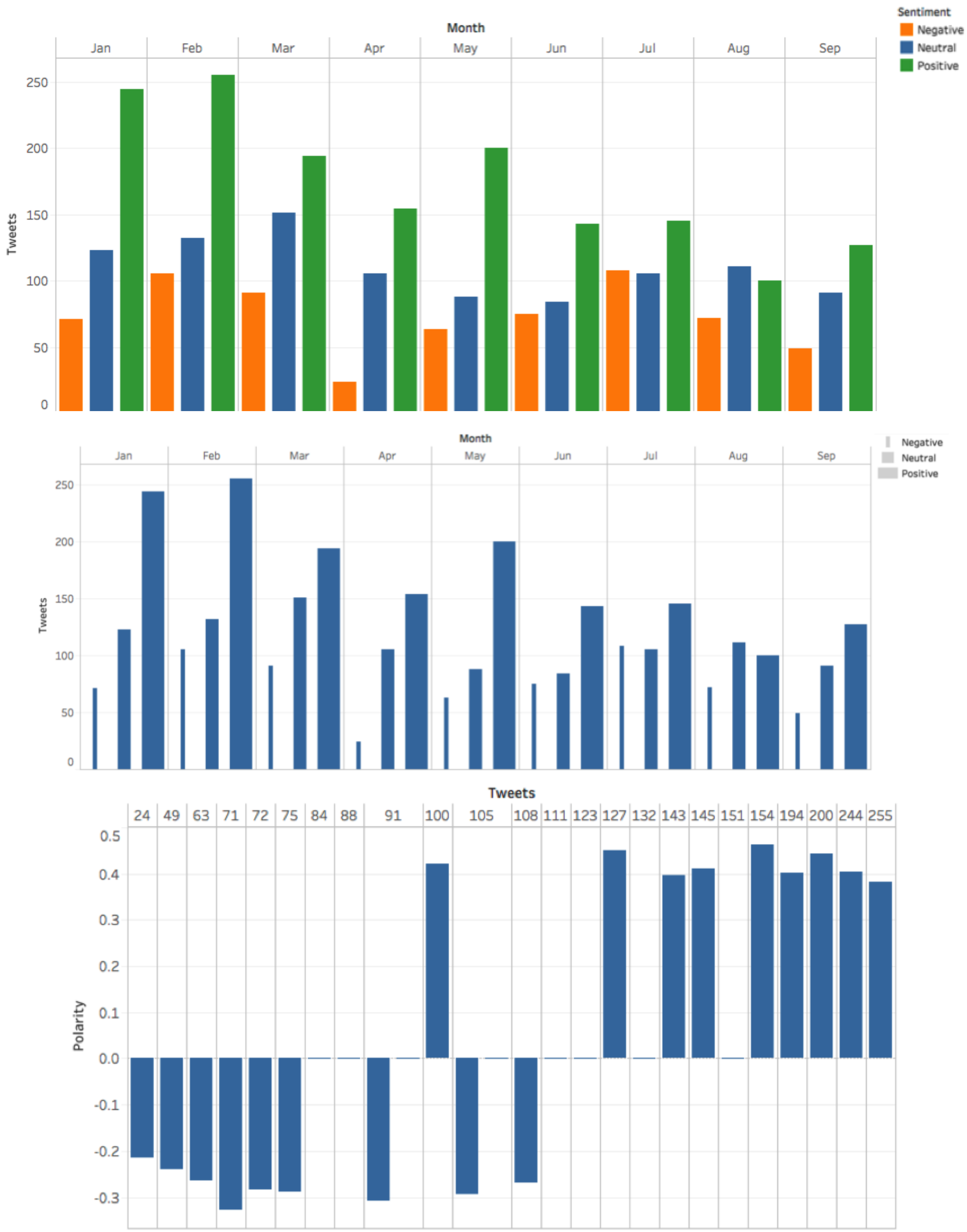

Figure 5: 2D Bar charts where (from top) sentiment is represented as color, size and position respectively

The multiple 2D charts have same variable in the x-axis such that x-y and x-z charts are created, as shown in Figure 4 and 5. For the sentiment by color and size, the following data variables: month-tweets-retweet and tweets-retweet-fav were visualized. However, while sentiment is visually represented by position, the data variables visualized are tweet-polarity-fav and retweet-polarity-fav are used in order to include polarity.

For month-tweet-retweet group of variables, the month-tweet and month-retweet were generated as two separate 2D charts, having same variable along the x axis. Later, the sentiment by color and size 2D charts were grouped vertically along the x-axis to provide a single aggregated view of data. Similar approach was adopted for generating multiple 2D charts for the tweets -retweet- fav with sentiment represented by color, position and size.

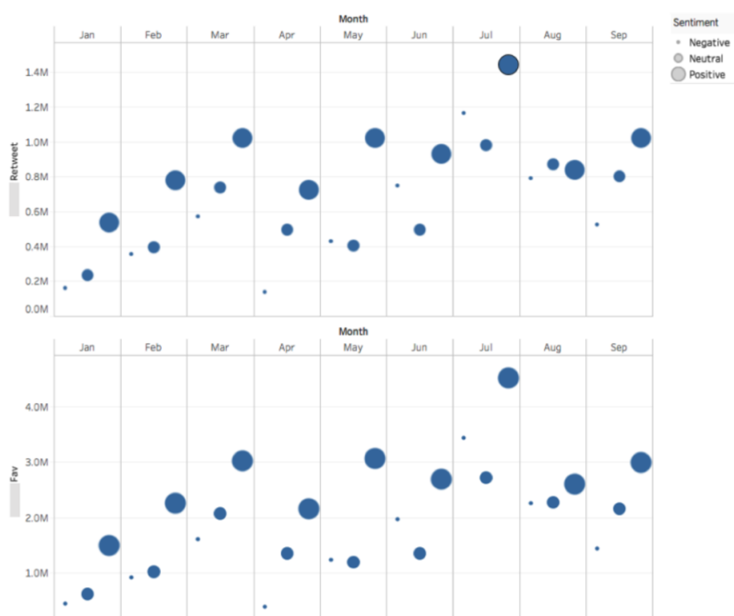

Figure 8: Multiple 2D scatterplots (from top) sentiment represented as size.

For sentiment by position, the common axis has to be the polarity value, as we are studying the sentiment distribution across different variables. So, for the tweets-polarity-fav set of data variables, tweets-polarity and fav-polarity charts are generated with polarity along the y-axis. But for maintaining uniformity with other charts where sentiment is represented as color and size, as shown in Figure 5, we vertically align the two 2D charts in one view. We follow the same approach for generating the 2D charts for retweet-polarity-fav data variables, where sentiment is represented by position.

Furthermore, in 2D charts, we maintained the same look and feel across all 2D charts, such as the colors used for sentiment in charts and the ordinal size used for data points to visually represent sentiment by size. Additionally, we also maintained consistency in the chart dimensions, font type and size and the position of legend.

Figure 9: Multiple 2d bar chart view for (from top) sentiment represented by color

iii) 3D Sentiment Visualization

3D visualizations, have additional attributes such as lighting, shadows, illumination, for 3D graphics or object models. ThreeJS library used to generate 3D charts, offered more control on these 3D properties. 3D visualizations offer additional continuity by visualizing three data variables on x-y-z axis at a time.

For generating 3D scatter plots and bar charts for sentiment visualization, we use the same data variables as in 2D. However, in the case where 2D charts used individually only x and y axis were involved. Familiar with the nature of 3D, we add another variable to visualize on the z-axis. So, we generate 3D scatterplots and bar charts for month-tweets-retweet and retweet-fav-tweet by visually encoding sentiment as color and size. To maintain uniformity with 2D, while representing sentiment as polarity, we visualize polarity on the y-value for the 3D sentiment charts by position. So, 3D scatterplots and bar charts were generated for tweet-polarity-retweet and retweet-polarity-tweet variables, as they are visualized on the x, y and z axis respectively.

We used the same colors as in 2D, for representing sentiment using 3D charts. For the size category, the sentiment was used as an ordinal value, and similar sizes proportionality from 2D were adopted by 3D. In 3D scatterplots – the sphere radius was adjusted, keeping in mind the ordinality of sentiment. This adjustment was done to avoid occlusion as much as possible.

For 3D bar charts, being familiar with the nature of bar charts, the visual height which represented the y-axis value, and the bar width was used to represent sentiment. So the entire volume of bars were representative of the size for sentiment.

On generating 3D charts, the data points were overlapping due to the close values and spatial positioning. So, similar to 2D, binning and grouping were used for removing overlaps across the x, y and z axis. Additionally, in 3D bar charts, to visualize variables along the z-axis, we converted the quantitative variables in to ordinal by grouping.

However, from the 3D charts generated, the 3D bar charts had many occlusion issues. The default initial camera angle is the first view users get of the data on 3D charts. So, while setting the initial viewing camera angle, we noticed more occlusion. We introduced transparency using illumination effect to make the 3D object models in scatterplots and bar charts transparent. This allowed the users to look through the other objects.

Also, as shown in Figure 6 and 7, the axis gridlines for scatterplot and bar chart were purposely made different. As our objective is not to introduce new effects, we do not make modifications that are not being practiced. In order to give an equally good presentation for 3D charts, we referred the 3D bar charts created using the ThreeGraphs[4] online tool and implemented the same axis grid for bar chart.

For 3D scatterplot and bar chart with the sentiment by position, the xz plane acts as a separator. While presenting the 3D charts without viewing control, the users might not see the negative data points. Additionally, as the y axis polarity value is 0, the neutral data points appear to be engraved on the xz axis plane.

In 3D charts, the role of viewing control is very important, as it enables a user to view the data from different rotational angles. In order to be able to study the user behavior, with and

[4] https://threegraphs.com

without viewing control, we initially fix the camera angle such that majority of the data points are visible [50]. Furthermore, in 3D charts, we also maintain consistency in the lighting, initial camera angle and the legend position.

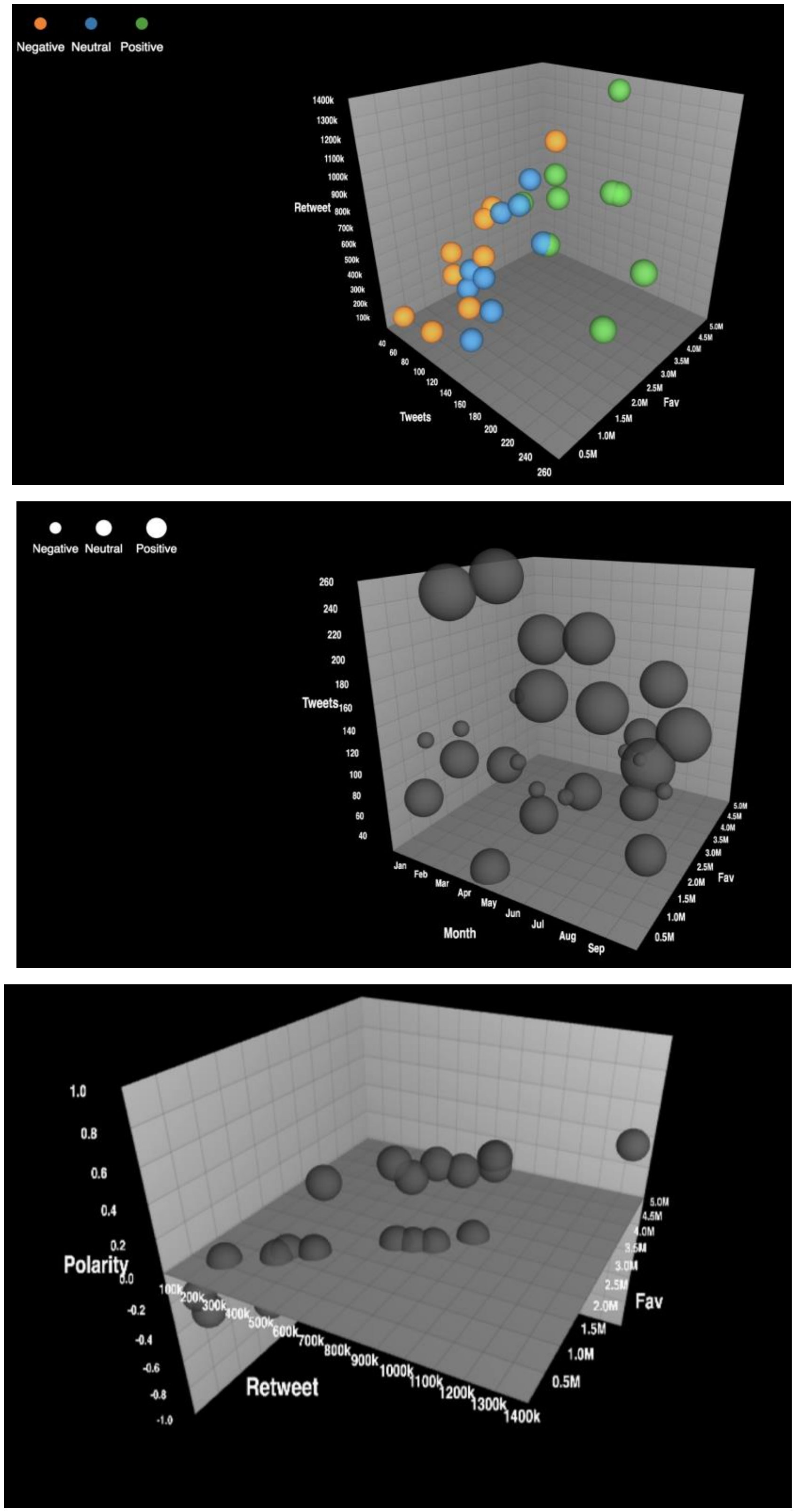

Figure 10: 3D scatterplots (from top) with sentiment visually represented as color, size and position

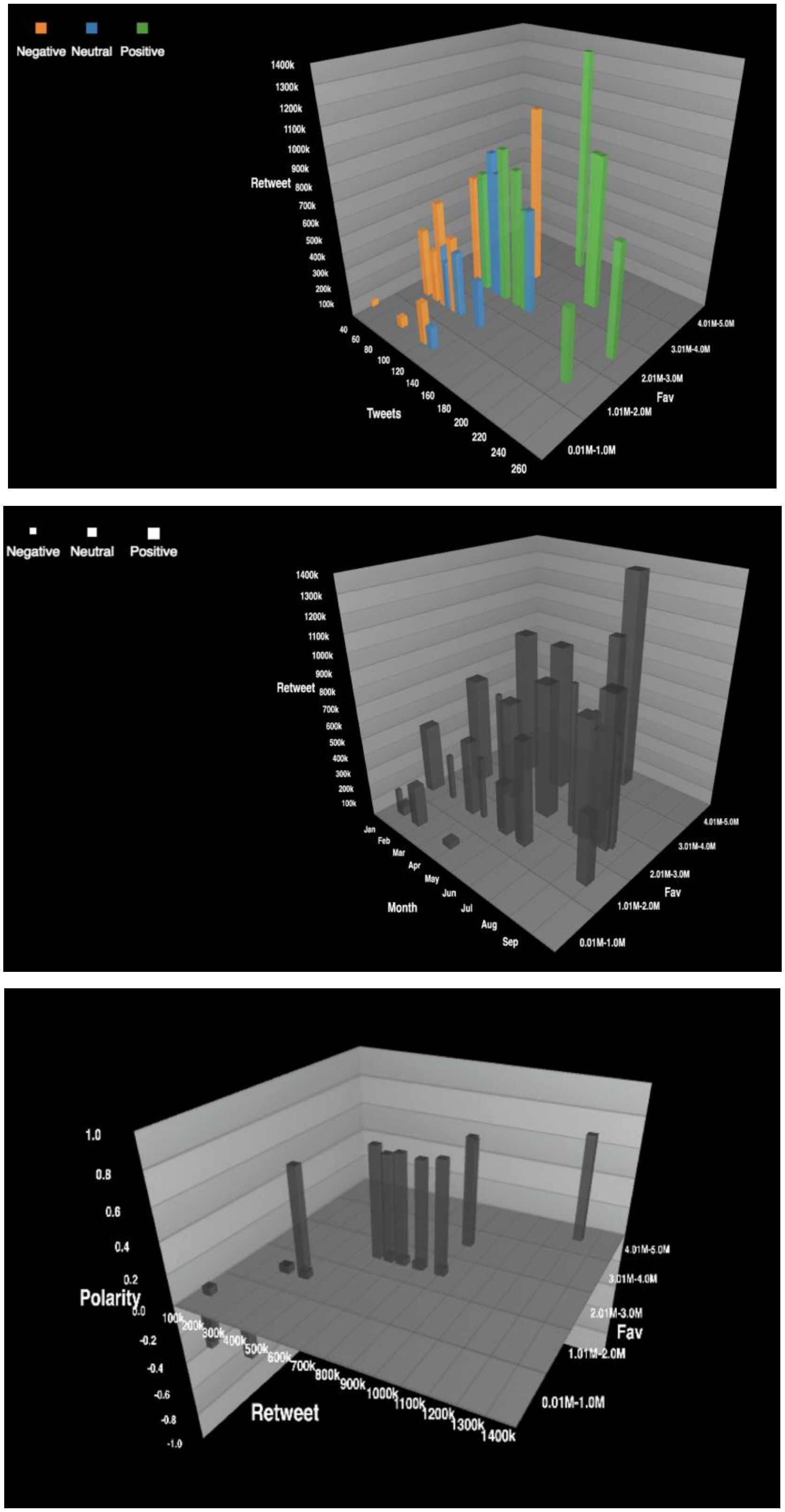

Figure 11: 3D bar charts (from top) with sentiment visually represented by color, position and size

3.2 Empirical Study

In this section, we elaborate on the approach adopted in this thesis, to perform to the empirical study in order to find when 2D and 3D visualization techniques can be used for sentiment visualization. The evaluation method we use is a task-based summative evaluation, where we perform a user testing evaluation and measure the user performance and cognitive load for 2D and 3D visualizations. The experiment design, participants and materials involved in the study are briefed, and finally the procedure for evaluation is discussed.

A. Experiment Design

The main objective of this empirical study is to find when 2D and 3D visualization techniques can be used for sentiment visualization. Also, to test our hypotheses that color is a better representation of sentiment in 3D.

In any visualization, the goal is to understand the data and generate insights in form of connections, trends, patterns and outliers. Though these experiments, we study how the different visual variables, used to encode sentiment in the literature behave in 2D and 3D sentiment visualizations.

In this task-based evaluation, we record and study the user performance and cognitive load of users while performing tasks on 2D and 3D visualizations under different experimental conditions. Since visualization is to facilitate human understanding of data, it differs person to person. Studying only the performance of users for different representation of sentiment variables in 2D and 3D is not enough [51], as it does not allow us to make appropriate conclusions on which visualization technique is better. Thus, we so measure the cognitive load – which is defined as the multidimensional construct of the amount of effort put by an individual user while performing tasks [52]. The different aspects of cognitive load are intrinsic load, extraneous load and germene load. Here, the intrinsic load is due to the element of interaction between the visualization chart being observed and the expertise level of the user. Whereas, the extraneous loads due to the framing of questions or delivery of task instructions. And the germane load is the effort put by the individual user in order to store the knowledge obtained from tasks. It is interesting to note that the extraneous load and germane load are not because of the mental effort, but because of the clarity level in task delivery by the instructor and are in our control [52].

Since the study is a task-based evaluation, we measure performance and cognitive load of participants while performing same tasks on 2D and 3D visualizations, which visualize the same dataset. We randomize the order in which participants evaluate 2D and 3D in order to cancel the effect of learnability, as there is a possibility that the users who performed 2D first, learn from the dataset and have shorter response time for 3D. Additionally, we also randomize the order in which sentiment representation is being tested first. Eg: Color-Position-Size or Position-Size-Color etc., as we evaluate the 2D and 3D visualizations under six different experimental conditions. In the evaluation, we record the response time and the correct number of tasks performed by a user, also if they answered correctly the first time. The recorded values depict the user's overall performance.

For measuring the overall cognitive load, there are a variety of ways which could be adopted, such as: subjective assessment – using rating scales and psychophysical measurements such as EEG and eye-tracking [51], [53]. While [54] measures cognitive load by using brain sensing, it also uses subjective assessment from evaluators using NASA-TLX. We adopt the subjective workload assessment technique using the NASA – Task Load Index (NASA – TLX) rating scale to measure the overall workload as the NASA TLX, is task rating scale, where the cognitive workload of tasks is subjectively measured [55]. The NASA TLX approach is a multidimensional rating scale using six different factors: mental demand, physical demand, temporal demand, performance, effort and frustration. Mental demand is the mental effort a user puts in order to perform a task, while the physical demand is the physical activity required to complete the task. Temporal demand is due to time pressure, and the speed of tasks. Performance in NASA TLX is the satisfaction and accomplishment level of users after performing tasks, whereas effort is the level of mental and physical demand they felt while accomplishing the performance level. Frustration is the discouragement, stress, irritation level experienced during the tasks. The first three factors, mental, physical and temporal demand are intrinsic demands experienced by the user while performing tasks, and the remaining three factors - performance, effort and frustration are due to the interaction of users with tasks.

The NASA TLX has two steps – in first, the users are required to select the factor which had more demand from the two factors displayed. This selection is done for 15 combinations of 6 factors. The weights from this step are tallied and they contribute to the overall workload. In the second step, the users are required to rate each of the six factors using the rating scale – consisting of 20 bins, each accounting for a value of 5. The scales are ordered from very low to very high for each of the factors and good to bad for performance. Additionally, the overall workload is obtained as a weighted average, based on the weights and rates generated in NASA TLX [56]. We choose the NASA TLX because it gives us more data and transparency on different aspects of the cognitive load. Also, because we are particularly interested in the mental load, which is one aspect of intrinsic cognitive load, which accounts to the amount of effort a user puts in order to perform the task. And the performance, as it indicates the user's accomplishment and satisfaction level with the tasks. While testing the cognitive load in results and analysis chapter, we consider the overall load to account for cognitive load, as cognitive load is a multidimensional construct [52].

In sum, for recording the cognitive load, the user is asked to weight and rate the difficulty of the tasks performed, after at the end of every set of tasks under the same experimental condition using the NASA TLX. Finally, the user experience and overall satisfaction of users is also captured at the end of the evaluation using a questionnaire post-evaluation.

i) Experimental Conditions:

Given that color, position and size are mostly used visual representations of sentiment in the literature, we test the 2D and 3D visualizations under three conditions each. The sentiment is represented by color, position and size in each of the conditions. So, we have six experimental conditions, which are listed below:

1. 2D sentiment visualization, where sentiment represented by color
2. 2D sentiment visualization, where sentiment represented by position
3. 2D sentiment visualization, where sentiment represented by size
4. 3D sentiment visualization, where sentiment represented by color
5. 3D sentiment visualization, where sentiment represented by position

6. 3D sentiment visualization, where sentiment represented by size

ii) Task Analysis:

The sentiment visualizations are generated to answer a visualization question, discussed in section 3.1 of this chapter. An initial set of tasks are generated for evaluation, keeping in mind the user goals based on the visualized axis variables and visualization tasks the users are most likely to perform on the generated sentiment visualizations. Since the task delivery or instructions given by the instructor during an evaluation, also contribute to the overall cognitive load [56], we conducted 3 pilot studies initially with evaluators from data analysis background, in order to rephrase and finalize the set of tasks to be used for evaluations. In the pilot study, after a brief introduction on the dataset, the sentiment results and the evaluation approach, a set of tasks were performed on 2D and 3D visualizations. We recorded this pilot task-based evaluation session, in order to take qualitative feedback in terms of comments and opinions of the pilot study evaluators. We also discussed with the pilot evaluators, to get their inputs on the tasks, check if they want to add to the evaluation, based on what inferences they would like to obtain from the dataset and visualizations. Later, we incorporate the feedback, rephrase tasks and choose tasks which have only a single correct answer.

As we subject each of the visualizations to visualize sentiment using color, position and size, we group our tasks based on the visual variables encoding sentiment. The color and size visual variables had similar tasks, meanwhile the position visual variable had different tasks, because of different axis variables as mentioned in Table 2.

We also take in to consideration the different answers, which the default views in 3D chart generate, in our sub-condition for 3D without viewing control. We note down the closest possible correct answers which can be given by the evaluators. This is because of the nature of 3D, there is the low perceptual accuracy when compared to 2D [57]. The tasks used in evaluation are included in Appendix A.

B. Participants

In similar evaluation studies using controlled experiments, a participant sample size of 12 subjects in [45] and in other visualization evaluation studies involving eye-tracking techniques [15], [46], a sample size of up to 20 was used. We selected 17 participants as subjects from different age groups, backgrounds and level expertise in data analysis and visualization were recruited.

The participants perform tasks on both 2D and 3D charts but were randomly assigned 2D first and 3D first categories, such that 8 participants evaluated 2D first and the remaining 9 evaluate 3D sentiment visualizations first. We randomize the order to eliminate any effect of learnability on the other.

C. Setup

For the evaluation, an Asus 23-inch desktop computer, with in-built web camera, attached with keyboard and mouse was used. Chrome web browser was used, as the sentiment visualizations evaluated were web-based. The evaluation also required a local http-server as the visualizations were hosted on the local machine. We used the http-server plugin in the Atom editor for firing up a local http-server instance.

In order to measure the response time, number of correctly performed tasks and other performance metrics for tasks performed by user, a screen and voice recording was needed.

We used the loom[5], a free screen and video recording software for chrome. The loom chrome extension required internet connection to generate video recordings of evaluation sessions with screen and voice captures during the evaluations.

For measuring the cognitive load, NASA-TLX, which is a multidimensional rating scale for measuring workload was used. Instead of the paper-pencil questionnaire compiled together by S. Hart in [56], we used iPad Air 2 with an iOS app of NASA-TLX found from the NASA website[6] created by Phil So. The iOS app was installed and used for recording the workload of participants at the end of every experiment condition. The app enabled editing of group id and subject id for different participants, so we modified the subject id and group id before every evaluation. We enter the subject id as a unique id, comprising of name of evaluators and the group id was 2DFirst or 3DFirst, depending on which visualization techniques the evaluator was assigned to perform tasks on, first. The number of trail for each evaluator was automatically calculated by the app. Since we also randomize and record the order of visual variables for sentiment seen by the evaluator during evaluations, the trial number gave a clear indication of which recorded trial is for which the experimental condition. Finally, the results of all evaluators were saved, and the app also had an export option for obtaining the results as a csv file.

[5] https://useloom.com/

[6] https://humansystems.arc.nasa.gov/groups/TLX/

D. Procedure

The evaluation started with a pre-evaluation questionnaire, to record their demographics and also to understand heir background and level of expertise with data analysis and sentiment visualization. As we use the red-green color palette for representing sentiment in our visualizations, we verify if the participants are color-blind. We performed the ishihara color-blind test [58] from a website (colour-blindness.com)[7]. For using 3D with viewing control and perceive data from 3D charts, we assess the spatial abilities of participants using a spatial reasoning test[8].

Then, the evaluators were given a short introduction to some background on sentiment visualization, the dataset we used for sentiment visualization and the evaluation procedure. The evaluators were also being exposed to a brief interactive training on the3D and 2D visualizations. After this, the actual task-based evaluation began. Users were presented with tasks, while recording the screen, to monitor time taken for completing tasks. The evaluators performed tasks on bar charts and scatterplots for different experimental conditions in 2D and 3D. At the end of each experimental condition, NASA-TLX rating was taken, to record the difficulty and overall workload experienced during the group of tasks performed under the same experiment condition.

After finishing all the tasks under the six experimental conditions, the users are presented with a post-evaluation questionnaire to record the overall difficulty of tasks and their overall satisfaction with tasks on 2D and 3D sentiment visualizations. The entire evaluation

---

[7] http://www.colour-blindness.com/colour-blindness-tests/ishihara-colour-test-plates/

[8] https://www.123test.com/spatial-reasoning-test/

lasted for 60-80mins depending on the speed and cognitive capacity of the participant.

In this task-based summative evaluation on 2D and 3D sentiment visualizations, task completion time and task correctness were to be recorded to measure performance and cognitive load of users. The measured metrics were to give an indication of when 2D and 3D visualization techniques are suitable for sentiment visualization.

CHAPTER 4: ANALYSIS & RESULTS

This chapter summarizes the analysis technique and results obtained from the experiments. Here, we discuss our validation of hypothesis, observations and lessons learnt from the experiments.

The data from pre-evaluation questionnaire is processed first. From the results of the color-blind and spatial reasoning tests conducted during the pre-evaluation, we found that none of the participants were color deficient and they had reasonable spatial skills. The results from pre-evaluation questionnaire was analyzed for demographic and background information about the participants. We had 11 female participants and 6 male participants, out of which 70.6% belonged to the 20-30 years age-group, 23.5% were 31-40 years and 5.9% were 41-50 years. These participants were 4 masters and PhD students from data analysis background; 5 researchers and 6 business intelligence professionals and developers with background in virtual reality and information visualization. Note that majority of our participants had data analysis background. Though majority are from data analysis background, 11 of them did not perform sentiment analysis before. However, most of them were familiar with it. We collect this information for our analogy, but regardless of their familiarity all participants were given a brief introduction to sentiment analysis. All our participants have used visualizations before, specifically scatterplots and bar charts, and 11 of them use visualization charts very often. 13 participants were active and inactive Twitter users and the remaining 4 participants are not Twitter users. However, we introduce the Twitter terminology, such as: tweets, retweets and favorites to all participants.

Then, the screen and voice recorded sessions were watched to record the start and end time of tasks, and the answer given by evaluator during the task-based evaluation, conducted after the pre-evaluation questionnaire. 408 samples were generated from 17 participants, who performed tasks on 2D and 3D visualizations under different experiment conditions. The 408 samples were records of tasks, consisting of response time taken to complete the task, along with the entry of, if the task was completed correctly or not, and if it was completed correctly in the first attempt. Additionally, the ratings collected from the NASA-TLX rating scale at the end of each experimental condition for 2D and 3D visualizations, was compiled to generate 612 samples consisting of the ratings for each of the six factors, along with the overall workload, which was measured using the weighted average. The recorded data was ready to be processed to generate results.

4.1 Analysis Technique

Here, for analyzing the data generated from task-based evaluation, we first start by testing our three hypotheses, mentioned in chapter 1. We use descriptive and inferential statistical techniques for analyzing the data obtained from evaluations. Mean and standard deviation are the descriptive statistical techniques used for result analysis. Whereas, the inferential statistical techniques, depend on the number and data type of the independent variables involved in the hypothesis being tested. Since our evaluation study is within the group, we test our hypothesis using Analysis of Variance (ANOVA). The type of ANOVA used also depends on the number and data type of the independent variables in the hypothesis tested. So, we formulate the hypothesis, based on the variables involved.

Table 4: Hypothesis Variables and their Values

V
a
r
i
a
b
l
e
N

We have formulated three hypothesis, by identifying the variables, in order to find which statistical test to use for our study. In the hypothesis 1, we assume that 3D visualizations, the cognitive load will be lesser while using sentiment visual representation as color than position or size. The variables involved here are, chart dimension (d), cognitive load (l) and visual representation of sentiment (v). The chart dimension and visual representation of sentiment, d and v are independent variables, while the cognitive load, l is dependent on d and v. Hypothesis 1 can be symbolically represented using variables as an equation below:

$$l = f(d, v)$$

In hypothesis 2, we state that the 3D visualizations will have better performance when sentiment is visually represented by color in 3D charts than position or size. Here, the performance (p), chart dimension (d) and visual representation of sentiment (v) are the variables. Of which, the performance p is dependent on the chart dimension d and visual representation of sentiment v, as they are the independent variables. Hypothesis 2 can be symbolically represented using variables as in equation below:

$$p = f(d, v)$$

In hypothesis 3, we assume that when sentiment is visually represented by size in 3D, the 3D bar charts will have better performance than 3D scatterplots. Here the performance (p), chart dimension (d), visual representation of sentiment (v) and chart type (T) are the variables involved. The chart dimension d, visual representation of sentiment v and the chart type T, variables are independent and effect the performance p, as it is a dependent variable. Hypothesis 3 is represented as an equation below:

$$p = f(d, v, T)$$

The variables discussed in hypothesis have different values. The summarized list of hypothesis variables and their possible values are found in Table 3.

As we know that the data type of hypothesis variables decides the type of analysis we are going to perform, we breakdown the independent and dependent variables by quantitative, nominal or ordinal data types in Table 4. From the table, we note that there are two nominal independent variables and 1 quantitative dependent variable in hypothesis 1 and 2 and three nominal independent variables and one quantitative dependent variable in hypothesis 3.

Thus, we use two-factor with replication ANOVA, as there are two and more than two independent variables in each hypothesis [59], [60].

Table 6: Hypothesis Variables and their Data Type

Hypothesis	**Independent Variable**	**Data Type**	**Dependent**	**Data Type**
	1.Chart dime			

Note that the performance mentioned in hypotheses is the user's performance measure, which is indicated by the response time nd number of correctly answered tasks under each experiment condition. And the cognitive load is the overall workload obtained from NASA TLX rating scale. The cognitive load is more reflective of the mental demand and performance factor, which is based on the mental effort and satisfaction level of tasks performed under different experimental condition [51]. However, we discuss the hypotheses for performance by considering the response time and cognitive load by considering the overall workload.

4.2 Results

First, we report results obtained from inferential statistical techniques and then discuss results from descriptive techniques. Using the two-factor ANOVA with replication, we tested our three hypotheses for statistical significance using the Excel Analysis ToolPak.

In our three hypotheses, we maintained the significant level to be 0.05, which is 5%, as this is recommended for sample sizes less than 30. Generally, in hypothesis testing, the null-hypothesis states that there is no significant difference between the groups or attributes in group and the alternate hypothesis is that there is significant difference. In order to prove our hypothesis, our objective is to reject the null-hypothesis. When the p-value is lesser than the value of 0.05 which is the significance level, the null hypothesis gets rejected.

Before performing the hypotheses testing using two-factor ANOVA, the results needed to be re-arranged, such that the first column represented the group with only group values at the start of the sample, and the other columns were the group variables being tested for statistical significance. The results of the tests were to be interpreted in a following way: the p-value for sample is the value for significance level in groups and the p-value for columns is the value for the significance level in group variables which are being tested.

The results of the two-factor ANOVA inferential statistical tests on our three hypotheses is summarized in Table 5. For hypothesis 1, we consider the overall workload of 17 participants while performing the tasks under different conditions for 2D and 3D. In hypothesis 2, we look at performance in terms of the response time taken to complete the tasks. And in hypothesis 3, the response time for tasks performed on 3D scatterplot and bar chart under different conditions was considered.

Table 7: Test Results Table for Hypotheses

Hypothesis	F	P-value	F crit
1	0.660	0.519	3.091
2	1.567	0.214	3.091
3	0.507	0.604	3.091

For all three hypotheses, the p-value was greater than 0.05, which depicts that there is no significant difference and we fail to reject the null hypothesis. Thus, our hypotheses do not hold true. We apply the descriptive statistics, for each of our hypothesis to verify if they hold true or not.

i) Cognitive Load for 2D and 3D Sentiment Visualizations

In hypothesis 1, we formulate that the cognitive load will be lesser on 3D visualizations using color to represent sentiment. For verifying our hypothesis, we find the mean overall workload for 2D and 3D charts while sentiment is visually represented by either color, position or size. We plot the mean overall workload in a column chart and standard error (SE) as error bars the chart for 2D and 3D tasks performed separately under each experimental condition. From the column chart in Figure 8, we notice that the mean overall workload obtained from NASA TLX was lesser on 3D charts using sentiment visual representation as color over position and size. The mean overall workload for 3D charts where sentiment is visually represented by color, position and size was 35.82, 46.29 and 41.29 respectively with a SE of 4.91, 5.37, 4.66. This proves our hypothesis that the cognitive load will be lesser on 3D charts with color than over 3D charts visually representing sentiment by position and size.

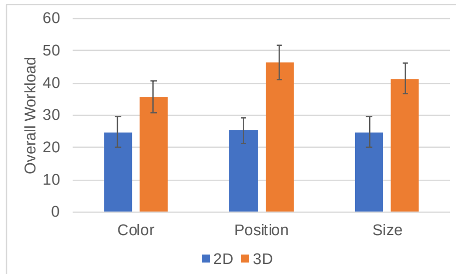

Figure 12: Mean overall workload for 2d and 3d visualizations, while sentiment is visually represented as color, position and size

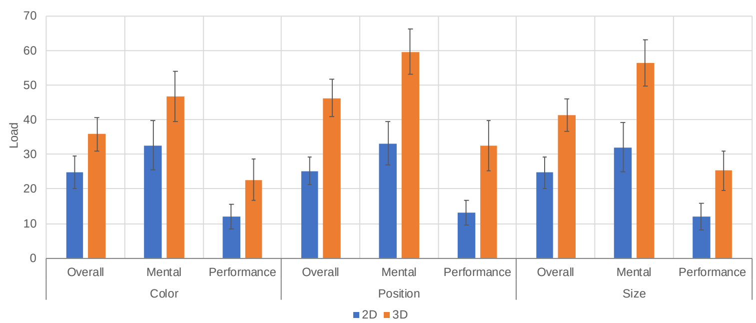

Figure 13: Comparison of mean overall workload, mental demand and performance rating in different tasks for 2D and 3D visualizations, while sentiment is visually represented as color, position and size

On the other hand, it is interesting to note that the 2D charts had almost the mean overall workload of 24.79, 25.23 and 24.71 for sentiment visually represented by color, position and size, with a SE of 4.78, 4.02 and 4.67 respectively. Since the NASA-TLX overall workload was considered as cognitive load while testing our hypothesis, it is interesting to compare the NASA-TLX mental demand and performance ratings for 2D and 3D visualizations under different experimental conditions. The chart in Figure 9 visually summarizes the comparison of mean overall workload, mental demand and performance measures from NASA-TLX for different tasks performed using 2D and 3D visualizations, while sentiment was visually represented as either color, size or position. From the chart, it is evident that there was high mental demand for tasks performed using 3D visualizations where sentiment is visually being represented as position. The sentiment representation by color in 3D visualization seemed to have low mental demand, which highly contributes to the cognitive load. This confirms our hypothesis 1, that color is the best visual representation for sentiment while using 3D visualizations. While for 2D visualizations, the mental demand seemed to be lowest for sentiment represented by color and the next for size. Sentiment represented by position in 2D had the highest mental demand.

ii) Performance in 2D and 3D Sentiment Visualizations

In hypothesis 2, we hypothesize that the performance is better when color is used to visually represent sentiment in 3D visualizations. Note that here, we only discuss 3D sentiment visualizations under different sentiment visual representations. In terms of the response time, better performance translates to the lesser response time. We aggregate the response time for all tasks performed by the participants and group them by 2D and 3D, under which we sub-group by color, position and size sentiment representation condition. The lowest mean response time for 3D visualizations, from Figure 10, is for sentiment representation by color. This proves our hypothesis 2 that while using 3D visualizations, users can perform faster, i.e., have better performance.

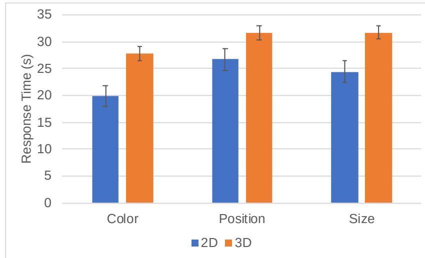

Figure 16: Mean response time in seconds for 2D and 3D visualizations, while sentiment is visually represented as color, position and size

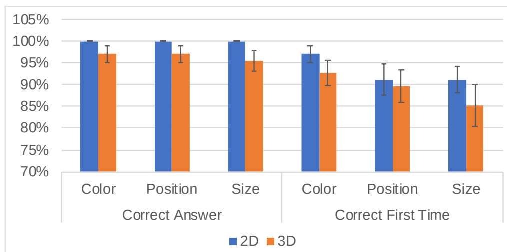

Figure 17: Percent of correctly answered tasks and tasks correct the first time in 2D and 3D visualizations, where sentiment is visually represented by color, position and size

From the chart, 3D visualizations had 111.18, 126.29 and 126.88 as mean response time in seconds for sentiment visually represented by color, position and size respectively, with SE of 14.43, 17.71 and 10.42. Note that the SE here is greater than 10%. For 2D visualizations the mean response time is always lesser than 3D visualizations. The mean response time of 79.47, 106.71 and 97.47 in seconds has been noted with the SE of 9.70, 12.90, and 8.25 for sentiment visually represented by color, position and size respectively. Here, we note that the response time has been lowest for sentiment representation by color in 2D visualizations as well. This testifies that our hypothesis holds true for 2D visualizations as well as it does for 3D visualization.

We also find the percentage for the correctly performed tasks and tasks performed correctly in first attempt in 2D and 3D visualizations. The results are plotted in a chart shown in Figure 11. Here, the 3D visualizations using color to represent sentiment visually, have better performance in terms of the percent of tasks performed correctly in first attempt. This holds true in the case, in the 2D visualizations, as there is better performance for sentiment represented by color.

As per the ISO/IEC 9126-4 Metrics, calculating effectiveness is recommended for measuring usability. We measure the effectiveness of 2D and 3D sentiment visualizations under different conditions, where sentiment is visually represented by color, position and size. Effectiveness is the completion rate of correctly finished tasks. As per our experiment design, we have had 17 participants who perform 4 tasks under six experimental conditions for both 2D and 3D sentiment visualizations. So, 408 total tasks performed by all participants under different experiment conditions for 2D and 3D visualizations. Thus, 68 tasks were performed under each experimental condition on 2D and 3D separately. All 68 tasks were performed correctly for 2D visualizations under the three experimental conditions where sentiment is visually represented by color, position and size. However, for 3D visualizations, 66 tasks were completed correctly when sentiment was represented by both color and position. And 65 correctly performed tasks when sentiment was represented by size. From the column chart in Figure 12, we notice that there is 100% completion rate for 2D. In 3D visualizations, there was 96% successful completion rate for tasks performed under sentiment by size condition and it was the least for 3D visualizations.

iii) Performance of Scatterplots and Bar charts for Sentiment Visualization

In hypothesis 3, we state that when sentiment is represented by size, scatterplots will have better performance than bar charts in 3D visualization. In order to test this, we aggregate the response time for all tasks performed under 2D and 3D – grouped by the type of chart used to perform task – i.e., bar chart and scatterplot. We also sub-group by the tasks performed under different sentiment representations - color, position and size. After performing descriptive statistics on the data, we plot a chart for 3D scatterplots and 3D bar charts as shown in Figure 13. Here, we study the 3D visualizations closely. The mean response time of 59.82, 72.71 and 53.41 in seconds for 3D scatterplot with sentiment visually represented as color, position and size was noted respectively with SE of 9.61, 12.96 and 4.56. For 3D bar charts, the mean response time in seconds was 51.35, 53.59 and 73.47 for color, position and size sentiment representation respectively with a SE of 6.95, 7.83 and 8.87.

From the chart in Figure 11, we note that the 3D scatter plots have lower mean response time i.e., better performance, when sentiment is being represented as size. This proves our hypothesis that 3D scatterplots are better while sentiment is represented as size, than 3D bar charts. However, it is interesting to note that in the other two cases of sentiment representation by color and position, the 3D bar charts perform better. This could be because of the nature of bar charts in 3D, as they use visual height to represent data points on y-axis, using variable width for representing sentiment ordinally led to confusions, as differentiating volume of bars in 3D can be harder.

On the other hand, for 2D scatterplot and bar chart visualizations, have lower mean response time when compared to 3D scatterplot and bar chart visualizations from Figure 12. The highest performance i.e. low mean response time for 3D visualizations was for sentiment represented by color in 3D bar charts and the next best was using 3D scatterplot, where sentiment represented by size.

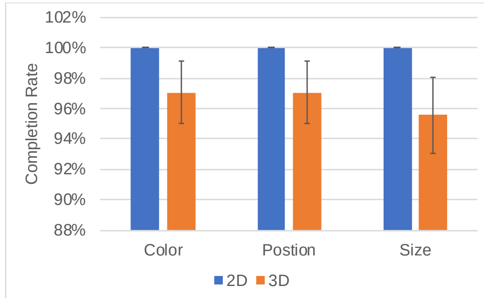

Figure 20: Successful completion rate (effectiveness) for 2D & 3D visualizations for color, position and size visual representations of sentiment

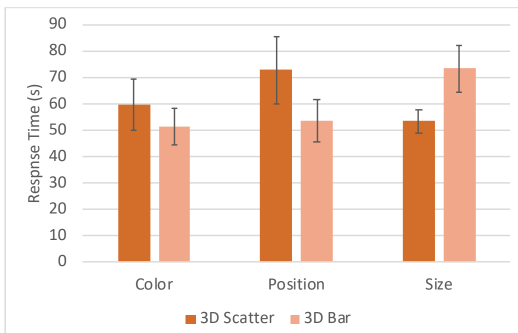

Figure 21: Mean response time in seconds for 3d scatterplots and bar charts for color, position and size visual representations of sentiment

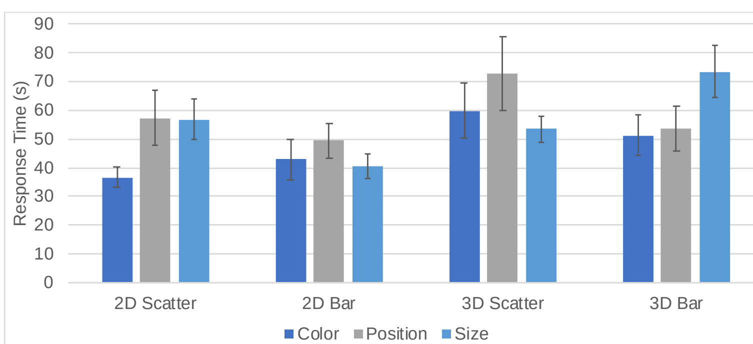

Figure 22: Mean response time in seconds for 2d & 3d scatterplots and bar charts for color, position and size visual representations of sentiment

iv) Subjective Evaluation

From the post-evaluation questionnaire, subjective feedback on overall difficulty and satisfaction level of using scatterplots and bar charts in both 2D and 3D was recorded using a 5-point Likert scale. The points in Likert Scale were from 1-5 and ranged from easy to hard and very low to very high respectively for difficulty level and satisfaction level.

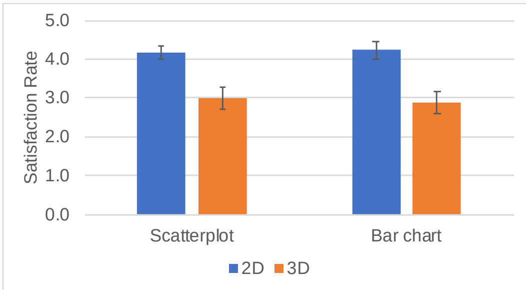

Figure 23: Mean satisfaction level for 2d & 3d scatterplots and bar charts

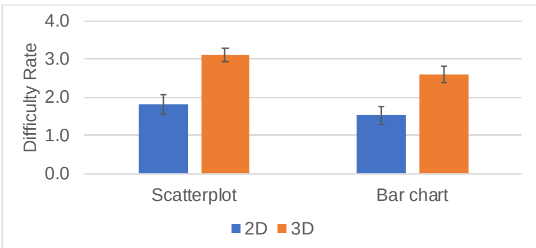

Figure 24: Mean difficulty level for 2d & 3d scatterplots and bar charts

The participants were almost equally satisfied with 2D scatterplots and bar charts, as the mean of 4.18 and 4.24 was observed respectively as shown in Figure 15. 2D bar charts were slightly more satisfying as they were more clear, easy to read and understand. On the other hand, 3D scatterplots got better mean satisfaction score of 3.00 than 3D bar charts, which is 2.88.

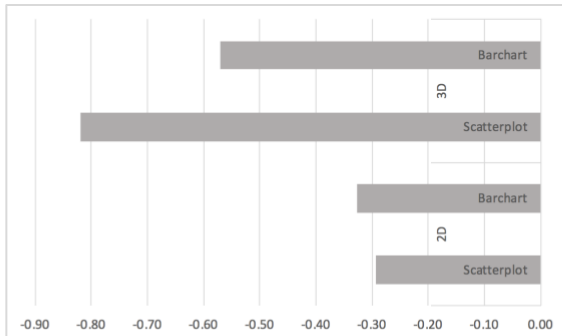

Figure 25: Percent correlation between satisfaction and difficulty level for 2d and 3d sentiment visualizations

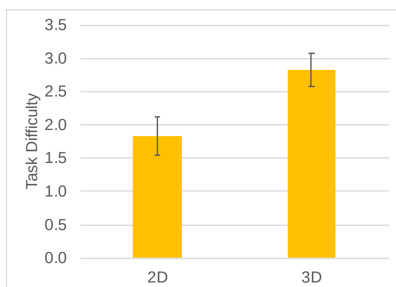

Figure 26: Mean task difficulty for 2d and 3d sentiment visualizations

Meanwhile, the difficulty level for 2D and 3D charts was lesser for 2D scatterplot and bar charts, compared to 3D as shown in Figure 16. For 2D, the bar charts had slightly less mean difficulty of 1.53, where the scatterplots had 1.53. In 3D, the scatterplots had more mean difficulty score of 3.12 than bar charts with 2.59. However, while 3D scatterplots have better satisfaction, seem to have more difficulty. So, we perform correlation on the mean satisfaction and mean difficulty level scores. From Figure 17, the negative correlation between satisfaction and difficulty levels is highest for 3D scatterplots and it is lowest for 2D scatterplots. This means that for 2D scatterplot, the satisfaction and difficulty level scores are more correlated than for 3D scatterplots. We assume, this could be because of the task difficulty and clarity in 2D and 3D. In the post-evaluation questionnaire, the users also rate the overall task difficulty in 2D visualizations and in 3D visualizations using a 5-point Likert scale, ranging from very easy to very hard. And answer if the tasks were clear for 2D and 3D visualizations or not. So, from Figure 18, we see that the mean task difficulty score is more difficult for 3D which is 2.82 than 2D, which is 1.82. And there was 94% task clarity for 2D, while only 76% was for 3D sentiment visualizations.

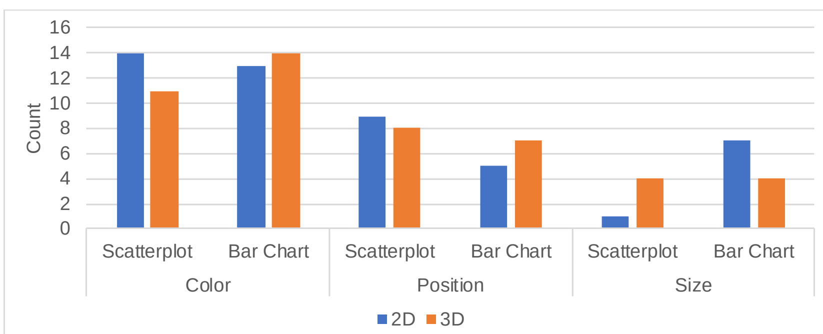

Figure 27: Overall best visual representation for sentiment in 2d and 3d bar chart and scatterplot

The post-evaluation questionnaire also gathered some feedback on the overall best sentiment visual representation. And overall best sentiment representations for 2D and 3D bar charts and Scatterplots. Here, the participants were told to select up to two choices for each chart type. 13 out of 17 participants, said the overall best representation for sentiment was color and the remaining 4 said it was position. While for 2D and 3D bar chart and scatterplots, the overall best sentiment representations are as shown in Figure 19. Even through the subjective feedback, we got color to be the overall best sentiment visual representation for all charts in 2D and 3D.

CHAPTER 5: DISCUSSION

Our experiments were with the 17 participants who undertook similar tasks under different conditions in both 2D and 3D visualizations. However, we randomize on which participant performs 2D or 3D first, to cancel the learnability factor obtained by participants from the previous set of tasks in either 2D or 3D first. Also, the tasks were designed in a uniform manner for 2D and 3D sentiment visualizations to avoid bias in terms of their difficulty level. So, the tasks designed for 2D and 3D visualizations, dealt with similar data variables, as they had similar task questions and a single correct answer for each task.

As mentioned in chapter 4, each participant performed 24 tasks. 12 tasks each on 2D and 3D visualizations, under each experiment condition. Here, the same 12 unique task questions were asked for both 2D and 3D visualizations, for each participant. However, the question if 12 tasks and the time spent on 3D visualizations by each participant is too short for understanding the nature of 3D visualizations exists and remains.

We know that from Figure 10, 2D sentiment visualizations have had better performance in terms of the mean response time in seconds, than the 3D sentiment visualizations. However, due to the nature of 3D, the response time, does not include the interaction time spent with viewing control. For 3D visualizations, the participants perform tasks with and without viewing control, as we study the effect of viewing control on 3D visualizations where sentiment is represented by color, position and size.

From Figure 20, we can find the mean response time taken for tasks with and without using viewing control in 3D visualizations. From the chart, 3D with viewing control has no impact on response time, while sentiment is represented by color. Whereas, for sentiment represented as position, we see the maximum impact of having increased response time for 3D visualizations with viewing control.

The performance and cognitive load of 2D sentiment visualizations were clearly better than the 3D sentiment visualizations. However, the 2D and 3D visualizations have their own trade-offs. While for 2D, there is shorter response time and high accuracy in terms of the number of correctly performed tasks from Figure 10 & 11, along with greater satisfaction levels from Figure15. However, the sentiment distribution over only two variables can be compared at a time. For visualizing more variables, we use multiple 2D charts, with common x axis. We let the participants perform tasks on single 2D charts and multiple 2D charts and the response time taken by them under different sentiment representations such as color, position and size are shown in Figure 21. We notice that there is more mean response time for tasks performed on 2D multiple charts over 2D single charts.

The advantages of using 3D is that we can visualize sentiment distribution over three variables at a time, in order to study relations and interpret from the sentiment distribution over three data variables. Using 3D, it becomes easier to separate coincident points. However, there are disadvantages like longer response time and low accuracy. Additionally, interpreting size in 3D is hard, as shown in Figures 10 & 11, the mean response time, mean number of correctly answered tasks and tasks answered correct first time is bad for 3D visualization using sentiment by size. Given that there is high occlusion and distortion due to perceptual views account for the low performance of 3D charts, methods such as slicing, including transparency and multiple views can be applied to reduce those effects.

As color, position and size are visual variables which have pre-attentive processing, we investigate on which visual variable is most effective overall. From Figures 8 & 9, we note that 3D visualizations while being represented by color have lower overall workload and mental workload for sentiment representation by color and then size. 3D visualizations using position had the highest cognitive load. From Figures 10 & 11, we note that the 3D visualizations have better performance in terms of response time and accuracy for sentiment represented by color and then position. 3D visualizations using size to represent sentiment had the lowest performance. While these are the quantitative results, we look at the subjective evaluation results in Figure 19, the 3D visualizations using color for sentiment representation had the highest overall satisfaction level and then followed by sentiment representation using position.

While comparing the satisfaction level of charts to the performance, there is a slight disagreement. As pointed by N. Neilson in [61], the users prefer fast-easy but are not satisfied.

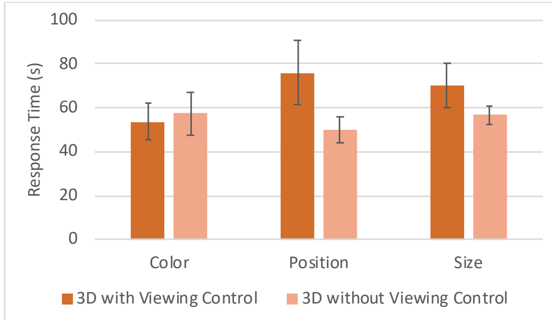

Figure 33: Mean response time in seconds for 3d visualizations with and without viewing control, for sentiment represented by color, position and size

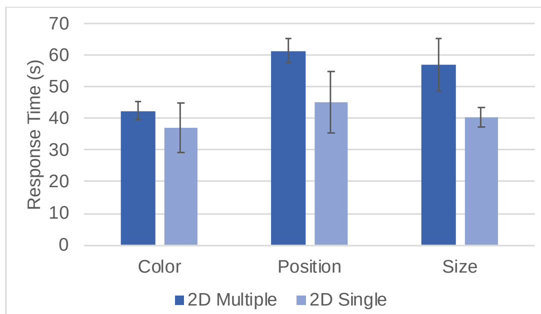

Figure 34: Mean response time in seconds for 2d single and multiple visualizations for sentiment represented by color, position and size

Thus, for sentiment representation in 2D and 3D visualizations, while considering performance, color is the best representation in 2D and 3D. While considering cognitive load, similar levels of overall workload was observed in 2D and color had the lowest cognitive load in 3D. This clearly states that for sentiment visualizations where sentiment is being studied across different data variables, color is the best representation.

While most mistakes were fixed during the pilot study, we discovered two fixes which were not done while performing the evaluations. Though we maintain uniformity while generating the 2D charts and 3D charts for sentiment visualization, after performing some experiments, we learnt that the 3D charts had legend located on top-left corner, while it was placed closer on the right in 2D charts. While trying to maintain uniformity within 2D charts and 3D charts, we missed on the legend position while concentrating on making the 2D and 3D visualization environments homogenous. We believe that this might have slight effect on the cognitive load of participants while performing tasks.

Also, we did not record the interaction time with the viewing controls in 3D sentiment visualizations. From the post-evaluation questionnaire, we discovered that 3 out of 17 participants had difficulties with rotational viewing controls in 3D visualization. We believe that this had some effect on the overall performance of 3D, as the users perform tasks with and without viewing control.

CHAPTER 6: CONCLUSION & FUTURE WORK

In this thesis, we advance the sentiment visualization techniques by adopting 3D for sentiment visualization. Also, we addressed the question of when 2D and 3D visualization techniques can be used for sentiment visualization using a comparative usability evaluation on 2D and 3D sentiment visualizations. In this empirical study, the user-testing approach was adopted, and task-based evaluation was performed. Performance and cognitive load were measured for 2D and 3D sentiment visualizations under different conditions, where sentiment was visually represented by color, position and size. We recorded the 2D and 3D sentiment visualization's task completion time and task correctness to measure performance under color, position and size visual representations of sentiment. At the end of experimental condition, the cognitive load was recorded using the NASA-TLX rating scale. The user experience and overall satisfaction was captured at the end of the evaluation using a questionnaire post-evaluation. The recorded data, helped us measure performance and cognitive load metrics which addressed the question of when 2D and 3D visualization techniques can be used for sentiment visualization.

From the evaluation results, there was lower cognitive-load and better performance for 2D sentiment visualization than 3D sentiment visualization. However, we investigated further on when 3D can be used under different conditions of visual representation by sentiment, by color, position and size. 3D visualizations while being represented by color have lower overall workload and mental workload while the position had the highest cognitive load. There was better performance in terms of response time and accuracy for 3D sentiment visualization, where sentiment was also represented by color. However, 3D visualizations using size to represent sentiment had the worst performance. From the subjective feedback obtained through post-evaluation questionnaire, the 3D visualizations using color for sentiment representation had the highest overall satisfaction level and then followed by sentiment representation using position. Thus, it is clear that the ideal scenario for using 3D visualizations is when sentiment is represented by color. There are trade-offs for choosing between 2D or 3D visualization techniques for sentiment visualization. But with the work of this thesis, we addressed the research gap and provide a guideline for data analysts, sentiment analysis and visualization researchers and developers, make an informed decision of when 3D visualization can be used for sentiment visualization.

In future, we plan on conducting the same study by including more visual variables such as shape and texture. Also, we want to apply more than one visual variable at a time, by creating different combination of pairs of visual variables, which can be used together to represent sentiment. We also want to perform the same study using the psychophysical eye-tracking and EEG techniques for measuring cognitive load of 2D and 3D sentiment visualizations. Furthermore, we want to compare the subjective workload assessment using NASA-TLX and psychophysical approaches using the eye-tracking and EEG to more accurately address the question of when 2D and 3D visualizations can be used for sentiment visualizations.

BIBLIOGRAPHY

[1] McCormick BH, "Visualization in scientific computing," *ACM SIGBIO Newsl.*, vol. 10, no. 1, pp. 15–21, 1988.

[2] N. Iliinsky and J. Steele, *Designing data visualizations: representing informational relationships*. O'Reilly Media, 2011.

[3] R. Mazza, *Introduction to information visualization*. Springer Science & Business Media, 2009.

[4] M. Tory and T. Moller, "Human factors in visualization research," *IEEE Trans. Vis. Comput. Graph.*, vol. 10, no. 1, pp. 72–84, 2004.

[5] C. Ware, *Information visualization: perception for design*. Elsevier, 2012.

[6] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. trends Inf. Retr.*, 2008.

[7] Y. Zhao, B. Qin, T. Liu, and D. Tang, "Social sentiment sensor: a visualization system for topic detection and topic sentiment analysis on microblog," *Multimed. Tools Appl.*, Aug. 2014.

[8] K. Kucher, C. Paradis, and A. Kerren, "The State of the Art in Sentiment Visualization," in *Computer Graphics Forum*, 2017.

[9] W. Cui, H. Qu, H. Zhou, W. Zhang, and S. Skiena, "Watch the story unfold with textwheel: Visualization of large-scale news streams," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 2, p. 20, 2012.

[10] J. Zhao, N. Cao, Z. Wen, Y. Song, L. Yu Ru, and C. Collins, "# FluxFlow: Visual

analysis of anomalous information spreading on social media," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1773–1782, 2014.

[11] O. Hoeber, L. Hoeber, M. El Meseery, K. Odoh, and R. Gopi, "Visual Twitter Analytics (Vista) Temporally changing sentiment and the discovery of emergent themes within sport event tweets," *Online Inf. Rev.*, vol. 40, no. 1, pp. 25–41, Feb. 2016.

[12] J. Ministr, J. Rácek, and D. Toth, "Visualization of the discussion content from the internet," in *IDIMT 2012 - ICT Support for Complex Systems, 20th Interdisciplinary Information Management Talks*, 2012, vol. 38, pp. 297–303.

[13] R. Patterson *et al.*, "A human cognition framework for information visualization," *Comput. Graph.*, vol. 42, pp. 42–58, 2014.

[14] D. Duan, W. Qian, S. Pan, L. Shi, and C. Lin, "VISA," in *Proceedings of the 5th International Symposium on Visual Information Communication and Interaction - VINCI '12*, 2012, pp. 22–28.

[15] S. Garlandini and S. I. Fabrikant, "Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization," in *International Conference on Spatial Information Theory*, 2009, pp. 195–211.

[16] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Seven guiding scenarios for information visualization evaluation," 2011.

[17] R. Brath, "3d infovis is here to stay: Deal with it," in *3DVis (3DVis), 2014 IEEE VIS International Workshop*, 2014, pp. 25–31.

[18] A. Capodieci, D. D'Aprile, G. Elia, F. Grippa, and L. Mainetti, "Visualizing cultural digital resources using social network analysis," in *IC3K 2015 - Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2015, vol. 1, pp. 186–194.

[19] K. Misue and K. Taguchi, "Development of Emotion-weather Maps," *Procedia Comput. Sci.*, vol. 60, no. 1, pp. 292–301, 2015.

[20] Y. Lu, X. Hu, F. Wang, and S. Kumar, "Visualizing social media sentiment in disaster scenarios," *Proc. 24th …*, 2015.

[21] Y. Wu, F. Wei, S. Liu, and N. Au, "OpinionSeer: interactive visualization of hotel customer feedback," *IEEE Trans. …*, 2010.

[22] D. Oelke *et al.*, "Visual opinion analysis of customer feedback data," in *VAST 09 - IEEE Symposium on Visual Analytics Science and Technology, Proceedings*, 2009, pp. 187–194.

[23] S. Fang, L. Miao, and E. Lin, "Visualization and clustering of online book reviews," in *IVAPP 2014 - Proceedings of the 5th International Conference on Information Visualization Theory and Applications*, 2014, pp. 187–194.

[24] B. Kwon, S. Kim, S. Lee, J. Choo, J. Huh, and J. . Yi, "VisOHC: Designing visual analytics for online health communities," *EEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 71–80, 2016.

[25] B. Chee, K. G. Karahalios, and B. Schatz, "Social Visualizations of Health Messages," in *Proceedings of the Hawaii International Conference on System*

Sciences, 2009, pp. 1–10.

[26] E. C. Resende and A. Culotta, "A demographic and sentiment analysis of e-cigarette messages on Twitter," *Comput. Heal. Sci. Work. 6th ACM Conf. Bioinformatics, Comput. Biol. Heal. Informatics*, 2015.

[27] J. Rekimoto and M. Green, "The information cube: Using transparency in 3d information visualization," *Proc. Third Annu. Work. Inf. Technol. Syst.*, pp. 125–132, 1993.

[28] E. Tufte and P. Graves-Morris, *The visual display of quantitative information.; 1983*. 2014.

[29] F. Wanner, C. Rohrdantz, F. Mansmann, D. Oelke, and D. A. Keim, "Visual sentiment analysis of RSS news feeds featuring the US presidential election in 2008," *CEUR Workshop Proc.*, vol. 443, no. February, 2009.

[30] V. Dehiya and K. Mueller, "Analyzing Hillary Clinton's Emails," in *Presented at IEEE Visualization*, 2016.

[31] H. Dong, M. Halem, and S. Zhou, "Social media data analytics applied to hurricane sandy," in *Social computing (SocialCom), 2013 International conference*, 2013, pp. 963–966.

[32] A. Bulbul and R. Dahyot, "Social media based 3D visual popularity," *Comput. Graph.*, vol. 63, pp. 28–36, 2017.

[33] B. Saket, A. Endert, and C. Demiralp, "Task Based Effectiveness of Basic Visualizations," Sep. 2017.

[34] A Abela, “Choosing a good chart - The Extreme Presentation(tm) Method,” 2006. [Online]. Available: http://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html. [Accessed: 01-Oct-2017].

[35] P. Kralj Novak, J. Smailović, B. Sluban, and I. Mozetič, “Sentiment of Emojis.,” *PLoS One*, vol. 10, no. 12, p. e0144296, Jan. 2015.

[36] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller, “TwitInfo: Aggregating and Visualizing Microblogs for Event Exploration,” *CHI ’11 Proc. 2011 Annu. Conf. Hum. factors Comput. Syst.*, pp. 227–236, 2011.

[37] S. Gobron, J. Ahn, G. Paltoglou, M. Thelwall, and D. Thalmann, “From sentence to emotion: a real-time three-dimensional graphics metaphor of emotions extracted from text,” *Vis. Comput.*, vol. 26, no. 6–8, pp. 505–519, Jun. 2010.

[38] S. M. Mohammad, “From once upon a time to happily ever after: Tracking emotions in mail and books,” *Decis. Support Syst.*, vol. 53, no. 4, pp. 730–741, Nov. 2012.

[39] A. Brew, D. Greene, and D. Archambault, “Deriving insights from national happiness indices,” in *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference*, 2011, pp. 53–60.

[40] A. Molla, Y. Biadgie, and K.-A. Sohn, “Network-Based Visualization of Opinion Mining and Sentiment Analysis on Twitter,” in *IT Convergence and Security (ICITCS), 2014 International Conference on IEEE*, 2014, pp. 1–4.

[41] N. Cao, L. Lu, Y.-R. Lin, F. Wang, and Z. Wen, “SocialHelix: visual analysis of sentiment divergence in social media,” *J. Vis.*, vol. 18, no. 2, pp. 221–235, May 2015.

[42] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, S. Carpendale Empirical, and H. Lam Enrico Bertini Petra Isenberg Catherine Plaisant Sheelagh Carpendale, “Empirical Studies in Information Visualization: Seven Scenarios,” *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 9, pp. 1520–1536, 2012.

[43] A. Shamim, V. Balakrishnan, and M. Tahir, “Evaluation of opinion visualization techniques,” *Inf. Vis.*, vol. 14, no. 4, pp. 339–358, 2015.

[44] S. Dübel, M. Röhlig, H. Schumann, and M. Trapp, “2D and 3D presentation of spatial data: a systematic review,” in *IEEE VIS International Workshop*, 2014, pp. 11–18.

[45] F. Amini, S. Rufiange, Z. Hossain, Q. Ventura, P. Irani, and M. McGuffin, “The impact of interactivity on comprehending 2D and 3D visualizations of movement data,” *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 1, pp. 122–135, 2015.

[46] H. Liao, W. Dong, C. Peng, and H. Liu, “Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers,” *Cartogr. Geogr. Inf. Sci.*, vol. 44, no. 6, pp. 474–490, Nov. 2017.

[47] B. Liu, W. Dong, and L. Meng, “Using Eye Tracking to Explore the Guidance and Constancy of Visual Variables in 3D Visualization,” *ISPRS Int. J. Geo-Information*, vol. 6, no. 9, p. 274, 2017.

[48] B. M. Hughes, “Just Noticeable Differences in 2D and 3D Bar Charts: A Psychophysical Analysis of Chart Readability,” *Percept. Mot. Skills*, vol. 92, no. 2, pp. 495–503, Apr. 2001.

[49] U. Wiss and D. A. Carr, “An empirical study of task support in 3d information visualizations,” in *Information Visualization, 1999. Proceedings. 1999 IEEE International Conference*, 1999, pp. 392–399.

[50] M. S. T. Carpendale, D. J. Cowperthwaite, and F. D. Fracchia, “Extending distortion viewing from 2D to 3D,” *IEEE Comput. Graph. Appl.*, vol. 17, no. 4, pp. 42–51, 1997.

[51] W. Huang, P. Eades, and S.-H. Hong, “Measuring Effectiveness of Graph Visualizations: A Cognitive Load Perspective,” *Inf. Vis.*, vol. 8, no. 3, pp. 139–152, Sep. 2009.

[52] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, “Cognitive Load Measurement as a Means to Advance Cognitive Load Theory,” *Educ. Psychol.*, vol. 38, no. 1, pp. 63–71, Mar. 2003.

[53] X. Chen and R. Jin, “Statistical modeling for visualization evaluation through data fusion,” *Appl. Ergon.*, vol. 65, pp. 551–561, Nov. 2017.

[54] E. Peck, B. Yuksel, A. Ottley, R. . R. Jacob, and R. Chang, “Using fNIRS brain sensing to evaluate information visualization interfaces,” 2013.

[55] S. G. Hart, “Nasa-Task Load Index (NASA-TLX); 20 Years Later,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 50, no. 9, pp. 904–908, Oct. 2006.

[56] S. Hart and L. Staveland, *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*, vol. 53. Elsevier, 1988.

[57] A. Russell and Bielewicz J.S, “2D vs. 3D Graphing Styles: A Comparison of the Accuracy of Reader Perceptions for 2D vs. 3D Graphs.” 33rd Annual Western Pennsylvania, Undergraduate Psychology Conference, 2005.

[58] J. Birch, “Efficiency of the Ishihara test for identifying red-green colour deficiency,” *Ophthalmic Physiol. Opt.*, vol. 17, no. 5, pp. 403–408, Sep. 1997.

[59] W. R. Rice, “Analyzing tables of statistical tests,” *Evol. JSTOR*, vol. 43, no. 1, pp. 223–225, 1989.

[60] J. Lazar, J. Feng, and Hochheiser, *Research methods in human-computer interaction*. India: John Wiley & Son Ltd, 2010.

[61] Jakob Neilson, “User Satisfaction vs. Performance Metrics,” 2012. [Online]. Available: https://www.nngroup.com/articles/satisfaction-vs-performance-metrics/. [Accessed: 15-Apr-2018].

APPENDIX

Tasks List

1. Month-Tweets

Q1: Which month had max # of --- tweets?

1.Positive: Feb

2. Negative: Feb, Jul

3. Neutral: Mar

Q2. What is the min # of Tweets with --- tweets?

Positive: Aug

Negative: Jun

Neutral: Apr

Q3. What is the max # of Tweets with --- sentiment?

1.Positive: 240-260

2.Negative: 100-120

3.Neutral: 140-160

Q4. What is the min # of Tweets with --- sentiment?

1.Positive: 100

2.Negative: 20-40

3.Neutral: 80-100

2. Retweet – Fav

Q1: Highest RT & Fav value for --- tweets?

Positive: 1.4M, 4.5M

Negative: 1.2M,3.5M

Neutral: 0.9M,2.75M

Q2. Lowesr RT & Fav value for --- tweets?

1.Positive: 0.5M, 1.5M

2.Negative: 0.2M,0.75M

3.Neutral: 0.1M,0.4M

Q3. Is the highest RT-Fav Value for --- more than --- tweets?

Positive - Negatvie: YES

Negative - Neutral: YES

Neutral - Positive: NO

Q4. Is the lowest RT-Fav Value for --- more than --- tweets?

Positive - Negatvie: YES

Negative - Neutral: NO

Neutral - Positive: NO

3. Month – Retweet - Fav

Q1: Which months had most --- RT & Fav?

Positive: JUL

Negative: JUL

Neutral: JUL

Q2. Which month the --- tweets had the Lowest RT & Fav?

Positive: JAN

Negative: APR

Neutral: JAN

4. Tweets -Retweet -Fav

Q1. How many # of Tweets with --- sentiment had MAX RT & Fav?

Positive: 140

Negative: 115

Neutral: 110

Q2. How many # of Tweets with --- sentiment had MIN RT & Fav?

Positive: 240

Negative: 25

Neutral: 125

5. Polarity- Tweets

Q1. Which Sentiment had max tweets

Positive: 255

Q2. Which Sentiment had min tweets

Negative: 20

Q3: What is MAX # of tweets for --- sentiment

Positive: 255

Negative: 100-120

Neutral: 140-160

Q4: What is MIN # of tweets for --- sentiment

Positive: 100

Negative: 20

Neutral: 80

6. Polarity – Retweet

Q1. which --- Sentiment had Max/Highest Retweet & how many RT

Positive: 1.4M

Q2 which --- Sentiment had Min/lowest Retweet & how many RT

2.Negative: 0.1M – 0.20M

Q3: What was the HIGHEST RT the --- tweets have?

1. Positive: 1.4M

2. Negative: 1.2M

3. Neutral: 1.0M

Q4: What was the LOWEST RT the --- tweets have?

1.Positive: 0.5M

2. Negative: 0.1M

3.Neutral: 0.1M

7. Polarity - Tweets – Retweet

Q1. what senitment had max T & max RT?

1.Positive:258 1.4M

Q2. what senitment had min T & min RT?

1.Positive:100 0.5M

2.Negative:20 0.1M

3.Neutral:80 0.2M

Q3: What is the max T & RT val for --- tweets?

1.Positive:258 1.4M

2.Negative:115 1.2M

3.Neutral:150 1.0M

Q4: What is the min T & RT value for --- tweets?

1.Positive:100 0.5M

2.Negative:20 0.1M

3.Neutral:80 0.2M

8. Polarity -Retweet- Fav

Q1. what sentiment had max RT & max Fav?

Positive:1.4M 4.5M

Q2 what sentiment had min RT & min Fav?

1. Negative:0.1M 0.4M

Q3: What was the max RT & Fav value for --- tweets?

Positive:1.4M 4.5M

Negative:1.2M 3.4M

Neutral:1.0M 2.7M

Q4: What was the min RT & Fav value for --- tweets?

Positive:0.5M 1.4M

Negative:0.1M 0.4M

Neutral:0.2M 0.6M