

An Interactive Metaphor for 3D Visualization of Large Data Sets Case Study: Diabetes Data

W.D. Hettige



An Interactive Metaphor for 3D Visualization of Large Data Sets Case Study: Diabetes Data

W.D. Hettige
Index No: 13000446

Supervisor: Dr. G.D.S.P. Wimalaratne

December 2017

Submitted in partial fulfillment of the requirements of the
B.Sc. (Hons) in Computer Science Final Year Project (SCS4124)



Declaration

I certify that this dissertation does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any university and to the best of my knowledge and belief, it does not contain any material previously published or written by another person or myself except where due reference is made in the text. I also hereby give consent for my dissertation, if accepted, be made available for photocopying and for interlibrary loans, and for the title and abstract to be made available to outside organizations.

Candidate Name:

.....

Signature of Candidate

Date:

This is to certify that this dissertation is based on the work of

Mr./Ms.

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Supervisor Name:

.....

Signature of Supervisor

Date:

Abstract

In many areas of science and industry, the amount of data is growing fast and often already exceeds the ability to evaluate it. On the other hand, the unprecedented amount of available data bears an enormous potential for supporting decision making. As a consequence, the graphical representation, i.e., the visualization of data, established itself as a useful approach to investigate large datasets. With the development of Web3D technologies, the delivery and visualization of 3D models on the web is now possible and is bound to increase both in the industry and fro the general public.

This thesis mainly explores the role of a primary metaphor in interactive visual analytics. The core concept of this research is to introduce a state machine for 3D visualization. Based on the proposed state machine this thesis attempts to describe the design and implementation of an interactive web application that accesses International Diabetes Federation (IDF) data, in conjunction with geographic location based data, using WebGL based 3D graphics visualize the resulting information. The application was designed with the intention of being usable as a generic web visualization tool, to take IDF information, across a range of topics and present that to the web user in an interactive 3D form.

An experimental trial of 15 participants was carried out and involved each participants using three different forms of web based presentation of IDF data. These forms included tables, 2D and the proposed 3D metaphor. These results ensures that the use of a 3D interactive metaphor in the field of data visualization, in different fields through the reusable nature of the tool, offer the potential to enhance the users interaction with the data.

Keywords: Interactive Data Visualization, 3D metaphor, Web application, WebGL, three.js

Preface

This basis for this research originally stemmed from my passion for developing better methods of interactive data visualization. Some of the concepts has been identified in the literature review has been used in the context and those are justified and referenced accordingly. The sample data set is taken from the International Diabetes Federation which is publically available. I carried out the analytical calculation in the results and evaluation chapter under the supervision of my supervisor.

Acknowledgment

I would like to express my deepest gratitude to my supervisor, Dr. G.D.S.P. Wimalaratne for his valuable and constructive suggestions, kind co-operation and encouragement which helped me to complete this research successfully and also willingness to give his time generously has been very much appreciated. And special honor should go to him for his endurance in listening to myself and his immense knowledge which gave me to decide my own adoptions.

I would like to pay my gratitude to Prof. N.D. Kodikara & Dr. H.A. Caldera for giving me immense support as being my evaluators and giving me various feedbacks which helped me to strengthen my research concept.

I also more generously appreciate the patience, enthusiasm, feedbacks and motivation provided by my friends to achieve my research goals. This thesis also dedicated to my lovely parents who have presented me the greatest opportunity of an education from the scrumptious institution and help through that my whole life. It is a great pleasure for me to acknowledge the assistance and contribution of all the people who helped me to successfully complete my research.

W.D. Hettige

University of Colombo School of Computing,

December 18, 2017

Table of Contents

Declaration	2
Abstract	3
Preface.....	4
Acknowledgment.....	5
List of Tables	8
List of Figures.....	9
List of Acronyms	10
1 Chapter One	11
1.1 Introduction.....	11
1.2 Research Problem.....	13
1.3 Aims and Objectives	13
1.4 Overview of the Methodology	14
1.5 Outline of the Dissertation	15
1.6 Delimitations of Scope	16
1.7 Conclusion	16
2 Chapter 2	17
2.1 Literature Review	17
3 Chapter 3	25
3.1 Design	25
3.2 High Level Design Architecture	30
4 Chapter 4	32
4.1 Implementation.....	32
4.1.1 Technologies.....	32
4.1.2 Hardware Requirements	33
4.1.3 Implementation Process.....	34

5	Chapter 5	40
5.1	Results and Evaluation	40
5.1.1	Experimental Testing	40
5.2	Results	43
5.2.1	Quantitative Results	43
5.2.2	Qualitative Results.....	45
5.2.3	Observational Study	47
5.3	Analysis.....	47
6	Chapter 6	48
6.1	Conclusions and Future Works.....	48
6.1.1	Conclusion	48
6.1.2	Future Works.....	49
	References	50
	Appendix A: Questionnaire.....	54
	Appendix B: Diagrams.....	55

List of Tables

Table 2.1-1 Summary of the Literature Review.....	19
Table 4.1-1 Hardware Requirements for Web3D.....	33
Table 5.2-1 Average speed/accuracy of the task1	43
Table 5.2-2 Average speed/accuracy of the task2	44
Table 5.2-3 Average speed/accuracy of the task3	44
Table 5.2-4 Average user ratings for each of the presentations	45

List of Figures

Figure 1.4-1 Overview of the Methodology	15
Figure 2.1-1 Taxonomy for visualization of large data sets.....	22
Figure 2.1-2 A schematic diagram of the visualization process [Ware, 2004]	23
Figure 2.1-3 Reference model for visualization [Card et al., 1999]	24
Figure 3.1-1 State machine for 3D Visualization	27
Figure 3.1-2 3D Globe	29
Figure 3.1-3 GapMinder Animated 2D Geographic Charts.....	30
Figure 3.2-1 Prototype1 of clustering PIMA data	31
Figure 3.2-2 Prototype2 of clustering PIMA data	31
Figure 4.1-1 Reference model adapted for JS web visualization [Card et al., 1999]......	34
Figure 4.1-2 Input JSON format of data.....	35
Figure 4.1-3 Python code snippet of csv to JSON conversion	35
Figure 4.1-4 JSON parser	36
Figure 4.1-5 XMLHttpRequest code snippet.....	37
Figure 4.1-6 Handling JSON.parse() automatically	38
Figure 4.1-7 Main Implementation code snippet.....	39
Figure 5.1-1 Tabular presentation of IDF data	41
Figure 5.1-2 2D presentation of IDF data	42
Figure 5.1-3 3D metaphor	42
Figure 5.2-4 Chart of user response to "Overall rating"	46
Figure 5.2-3 Chart of user response to "User Interaction"	46
Figure 5.2-2 Chart of user response to "How much suitable, the presentation for the task?"	46
Figure 5.2-1 Chart of user response to "How informative was each of the presentation?"	46

List of Acronyms

IVA	Interactive Visual Analytics
IDF	International Diabetes Federation
2D	2 Dimensional
3D	3 Dimensional
IV	Interactive Visualization
VA	Visual Analytics
CSV	Comma Separated Values
HTTP	Hypertext Transfer Protocol
WebGL	Web Graphics Library
CAD	Computer Aided Design
OLAP	Online Analytical Processing

1 Chapter One

1.1 Introduction

Data big or small remains a mystery unless people can find patterns in it. It has become a challenge to utilize both technology and the powerful capabilities of human intuition to extract value and advantage from the information [4]. If the data set is small and simple, no visualization is needed. Usually, by looking at the data in their raw format will do. But many users have also big problems with comparably small data [24], in particular, when the data is complex (heterogeneous, multi-dimensional, highly structured, multimodal, etc.). However large scale multidimensional data contain the useful information and unknown patterns to be uncovered that feed government, organization, and business need. In order to analyze those data and find out familiar patterns, it is essential to analyze the data in every angle. Interactive Visual Analytics (IVA) emerged in the field of data mining. With IVA, the concept of data visualization can be taken a step further by using technology to drill down into charts and graphs for more detail, interactively changing what data you see and how it's processed. Having this idea, people have this in mind whether 3D visualization is the next step for finding patterns inside large data sets.

The pursuit of graphical representation for complex data, as well as its subsequent exploration and understanding, can be supported with interactive visualization software, e.g. Gephi, Processing, PhiloGL, three.js or D3 [11]. VA, however, has requirements that are more complex. It needs to provide a range of visual tools and analytic methods aimed at decision-making and problem-solving, by assisting analysts in visualization, filtering, sorting, recording and sharing of digital data, as well as, viewing, navigation, transformation, manipulation, organization and coordination of insights derived from this data [11]. Metaphors are just that: abstractions. Metaphors describe something unfamiliar in terms of something familiar that may actually be very different. The key to effective

visual metaphors is in their ability to stimulate the formation of mental images of data by reference to previous experience and the familiarity of the adopted visual form [12]. For the VA communication to be effective, the participants of such visual interaction, i.e. an analyst and the recipient of a visual message, need to establish the basis of shared communicative acts and meanings, and they need to agree on a set of common metaphors capable of conveying their personal understanding of reported events and situations [14]. Visual metaphors provide multiple complementary views of information, capable of assisting analysts in perception and reasoning [13].

Grady [15] showed that many conceptual metaphors are constructed on more basic or primary metaphors that arise out of aspects of our human experience. That is, sensory-motor experiences like hunger, itch, movement, strain, warmth become source domains for metaphors.

The Main reason why 3D metaphors are important is that it gives the ability for the user to see how realistic an object is when it is considered from all perspectives. Unlike 2D designs which may need clear instructions on how to deduce design information, 3D designs are almost instruction-less and without any language barriers. It is natural for any human to understand 3D design and experience the virtual reality it creates [6]. 3D clearly plays an important role, the reason for choosing 3D as space is that it will allow constructing a visualization of data without an immediate connection to the three spatial dimensions of our world.

Medical visualization is may be the most prominent application domain that is suitable for visualization. According to the stats provided by the International Diabetes Federation, by 2040 642 million people will suffer from diabetes. Due to the growing unstructured nature of diabetic data from health industry or all other sources, it is necessary to structure and emphasize its size into nominal value with a possible solution. At the individual patient level, clinicians must be able to quickly assess multiple laboratory tests, history, and other

risk factors to judge the risk of disease. Medical officers, researchers etc. analyze these data using the conventional techniques such as graphs, area charts, pie charts etc. which are in 2dimensional space and those techniques limit them on their findings. Information visualization utilizes the high bandwidth processing capabilities of the human visual system to reveal patterns in data that are not evident in non-visual data analysis [9] [10]. Since primary metaphors arise from our lived experience as humans it follows that we can intuitively understand them. By designing VA environments that take advantage of these metaphors we can thus reduce the cognitive load on the user or data analyst. Further, a solid understanding of primary metaphor can assist VA designers in avoiding the creation of displays that add to users' cognitive effort to understand them.

1.2 Research Problem

How a 3D interactive metaphor can offer the potential to enhance the user's interaction with data in finding hidden patterns inside the data by data visualization.

1.3 Aims and Objectives

Proving the fact that 3D interactive, online animation further enhanced data engagement highlights the broader principal that entertainment that engages the human senses play a key role in transferring knowledge.

Objectives of this research as follows,

- Literature review on data visualization.
- Evaluation of existing interactive visualizing techniques.
- Investigate how large data sets can be visualized more using a 3D metaphor than conventional techniques.
- Analyze metaphor in every angle to see hidden patterns.
- Develop a 3D metaphor that will support and improve visualization of large data sets.

1.4 Overview of the Methodology

Our own human visual system is not really made for 3D vision. That means our eyes are relatively close to each other and they are looking basically into the same direction. Even though our visual system has been equipped with a number of mechanisms to reconstruct 3D vision in our brain, people simply aren't in a good situation with 3D perception, quite some effort is needed to make 3D work such as interaction. Visualization is in principal expensive because it requires the user's time, so it is really important to think about the cost benefit relationship for any visualization solution. When it comes to data exploration and analysis, people recently started to look into a research direction, which is called Visual Analytics.

Recently, there are some researchers carried out to facilitate interactive visualization [5] [10]. The prevailing techniques suggested by these researchers will be collected and studied to get a deeper understanding in the domain, as an initial step. Next attempt will be to investigate the 3D metaphors found in the existing literature and see how they can be modified to visualize data. Further, the existing 2D models will also be taken into the investigation to understand how they work with data visualization. By analyzing these approaches, prototypes will be designed and developed, based on a state machine which is going to be introduced. Then it will be evaluated and assessed with synthetic and real-world datasets. The developed models might be evaluated based on different factors: scalability, accuracy, and robustness etc. Throughout the process, it will be ensured to maintain core UI principles such as clarity, flexibility, familiarity, and efficiency to facilitate usability aspect. The proposed model will be improved based on the evaluation results. Then as a final step, a new interactive metaphor for 3D visualization of large data sets will be introduced.

The main aim of this study is to aim at combined solutions, where parts of the task are solved in an automated fashion and only those parts of the task are solved interactively, where the benefit outweighs the costs. This whole purpose will be done based on the 3D

metaphor that is going to be introduced. This will be elaborated in the design section [12]. The implementation will be mainly based on web based rendering where the technologies will be used as WebGL. The metaphor will then be evaluated based on a quantitative and qualitative analysis. As the final step a generalized 3D metaphor will be introduced.

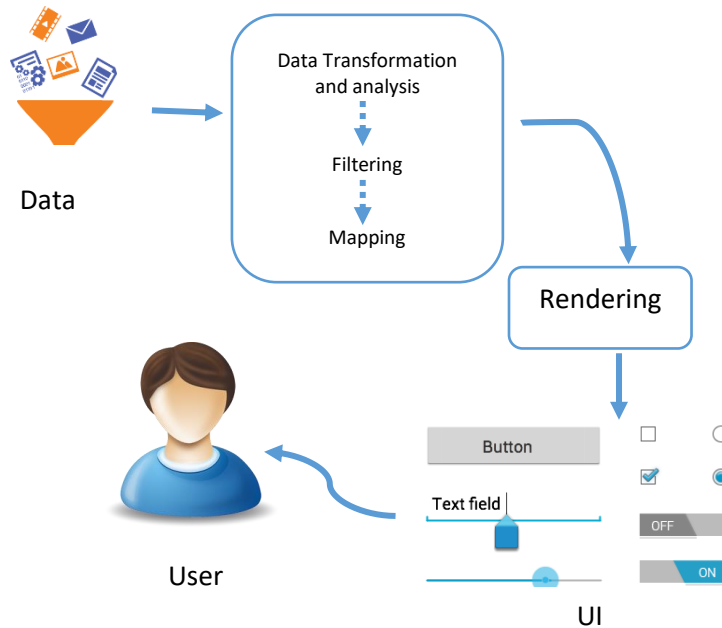


Figure 1.4-1 Overview of the Methodology

1.5 Outline of the Dissertation

This thesis paper is structured as follows. Chapter two explores about the literature review of interactive visualization which terribly describes levels of former work has been done. Chapter three encompasses a detailed annotation of the project design stage. Chapter four demonstrates the implementation of prototype with detailed description of different phases of the software system. In chapter five explicates the evaluation process of system results. The last chapter, chapter six demonstrates the conclusion of thesis and outlines the significance of our contribution and future works.

1.6 Delimitations of Scope

A generalized 3D metaphor for visualization will be introduced which can be used by any other data set. The data will be gathered from publically available IDF data repository and also from the UCI data repository. The research will be conducted within a specified time period, one year.

1.7 Conclusion

This chapter laid the foundations for the dissertation. It introduced the research problem and research questions and hypotheses. Then the research was justified, definitions were presented, the methodology was briefly described and justified, the dissertation was outlined, and the limitations were given. On these foundations, the dissertation can proceed with a detailed description of the research

2 Chapter 2

2.1 Literature Review

Recently, there has been significant interest in approaches that address the visual analysis of large data sets. But most of the time they haven't spoken about the 3D modeling.

Although the traditional approaches have proven their usefulness in many practical applications, they still face significant challenges such as algorithm scalability, increasing data dimensions, and data heterogeneity. To address these challenges visual analytics has been developed in recent years through a proper combination of automated analysis with interactive visualizations [1]. Visual analytics mainly focuses on analytical reasoning using interactive visualizations. Some of the existing theories, models, and frameworks have been identified [1].

Different data sets have different characteristics and patterns of interest which may require specialized technique to visualize the data. Based on the characteristics of target data visualization techniques can be categorized into four main section, such as graph visualization, text visualization, map visualization and multivariate visualization [2]. When developing a 3D visualization model there are some technical challenges more likely to face by the researcher. Visual scalability, errors, uncertainty, usability and integrated analysis of heterogeneous data are some of them [2].

"How might we design effective interactive systems not only for wrangling individual tables but for performing data integration? And for manipulating text, image, or video data?" [3], this is one of the problems that we are going to address in our research.

The literature reveals that in order to make data more meaningful it is essential to building stories. There is a Chinese proverb that says “One picture is worth ten thousand words”, which is the main motivation of this research [30].

A visualization system to be well formed it should be both intelligent and automatic [8]. On the other hand, the users should be capable of interacting with the system as much as possible and fine tune the display manually. A good visualization system should be able to render a large data set efficiently, render a high-dimensional data set and it should be able to deal with both numerical and categorical values. In order to satisfy these requirements, the visualization system should be interactive and should have an open architecture where it will be easier to integrate with other data analysis components [16].

The capability of a visualization to effectively display large datasets in terms of either the number or the dimension of individual data elements is known as its visual scalability. Eick and Carr identified six factors affecting visual scalability, i.e., human perception, monitor resolution, visual metaphors, interactivity, data structures and algorithms, and computational infrastructure. Besides human perception which is inherently given, the other five factors explain a great diversity of different visualization approaches with respect to visual scalability [23].

This research will provide a 3D metaphor to visualize data. Table 2.1-1 will emphasize some related works and their key findings on this research topic.

Table 2.1-1 Summary of the Literature Review

Reference ID	Sample	Limitations	Key Findings
6	Social network data	<ul style="list-style-type: none"> • Focus on social network data. • Lack of comparison between existing modules. • Not enough examples to show the graphical representation. • Not specifically related to health 	<ul style="list-style-type: none"> • The efficient 3D interactive graphical representation visual OLAP analysis yields a synergy effect among OLAP technologies, 3D graphical presentation methods, and human's cognitive capability. • Cloud computing will address some issues which will arise from building a 3D metaphor such as computational cost and storage resources.
16	Voxel data set (aneurysm-dataset and a human brain dataset)	<ul style="list-style-type: none"> • Need to have sufficient density of points. • Not suitable for datasets with low values, or grayscale/single-valued datasets. 	<ul style="list-style-type: none"> • This technique supports fast and efficient means of accessing and displaying contours and bringing out features generated by several values, or value ranges, which is especially advantageous for datasets with no "inherent" dedicated isosurfaces like colored medical datasets.
17	-	<ul style="list-style-type: none"> • Assumption that the data will never fit into GPU memory. • They have only focused on using GPU-based approaches for single stand-alone workstations. 	<ul style="list-style-type: none"> • They have identified that web-based visualization will play a big role in the future, which will make it necessary to research scalable algorithms for remote visualization and mobile devices. <ul style="list-style-type: none"> • As data sets get larger volume rendering methods will become more complex, so it is better to have data acquisition and pre-processing in order to make visualization more productive and efficient.

18	15GB of uncompressed telecommunication data	This approach is not suitable to visualize higher-order objects in large networks.	<ul style="list-style-type: none"> • Need to have drill-down facilities, drag-and-drop queries in order to explore detailed visualization. • Visualization environment should be flexible in order to experiment on scalable prototypes.
19	Flow field data	<ul style="list-style-type: none"> • The design is not well formed. • Need of powerful resources. 	<ul style="list-style-type: none"> • The environment should integrate modeling and visualization. • Interactive visualization is essential for gaining new insight in the nature of flow.
20	Livestock disease data in food production chains	Proposed visualization strategies are not exclusively designed for IEs.	<ul style="list-style-type: none"> • Compared to the common “flat” data display, immersive environments support a highly realistic representation of the three spatial dimensions and natural means to interact in that space.
3	–	–	<ul style="list-style-type: none"> • Exploratory analysis through visualization is often a critical component for assessing data quality and developing hypotheses. • Interactive tools for data analysis should make technically proficient users more productive while also empowering users with limited programming skills. • Implementation of novel interactive tools to aid analysis will remain an exciting and important topic for computer science research.

Even though there are so many techniques as mentioned in the Table 2.1-1, they are incapable of visualizing a large amount of high dimensional data because:

- Lack of human computer interaction.
- Incapable of dealing with a large amount of data. Today, a data set can easily have millions or even billions of records. How can people visualize both local details and the general overview of such data sets?
- Incapable to dynamically assign data dimensions to visual elements. Most existing methods use only one visual object to visualize one data record. When a record has lots of dimensions, the visual object becomes too complex for human beings.
- Lack of integration with other data mining and knowledge discovery (KDD) tools. The main idea of data visualization is to help in knowledge discovery. Since there are so many well-known KDD techniques available, it would be a great benefit integrating such techniques with visualization.

Utilizing 3D computer graphics to visualize and understand data is by no mean a new field. The concept of 3D visualization has emerged in a number of forms for a range of types of data [31]. The earliest example can be taken as a the first graphical computers in the late 1960's and included Computer Aided Design (CAD) tools in order to visualize architectural, engineering and design features[24].

However, visualization techniques have a different visual scalability with respect to the number of displayed data elements and the number of concurrently shown data dimensions. Most visualization techniques have no inherent limit with regards to the number of displayed data elements and a variety of clutter reduction techniques exist to overcome practical limitations. On the other hand, most techniques are inherently limited with respect to the number of dimensions which can simultaneously be displayed (e.g., scatterplots) [32]. For other techniques like parallel coordinates or scatterplot matrices, the practical limit for the number of dimensions is magnitudes smaller than the one for data elements [33]. As a consequence, an analysis of truly high-dimensional data is challenging.

The user typically has to pre-select the displayed dimensions, which may become a difficult task without a-priori knowledge or dedicated support [25].

Most commonly used large data sets visualization techniques can be broadly classified into the following three categories. Spatial Layout Visualization, Abstract/Summary visualization, Interactive/Real-Time Visualization. Figure 2.1-1 will represent a taxonomy for large data sets visualization techniques [7].

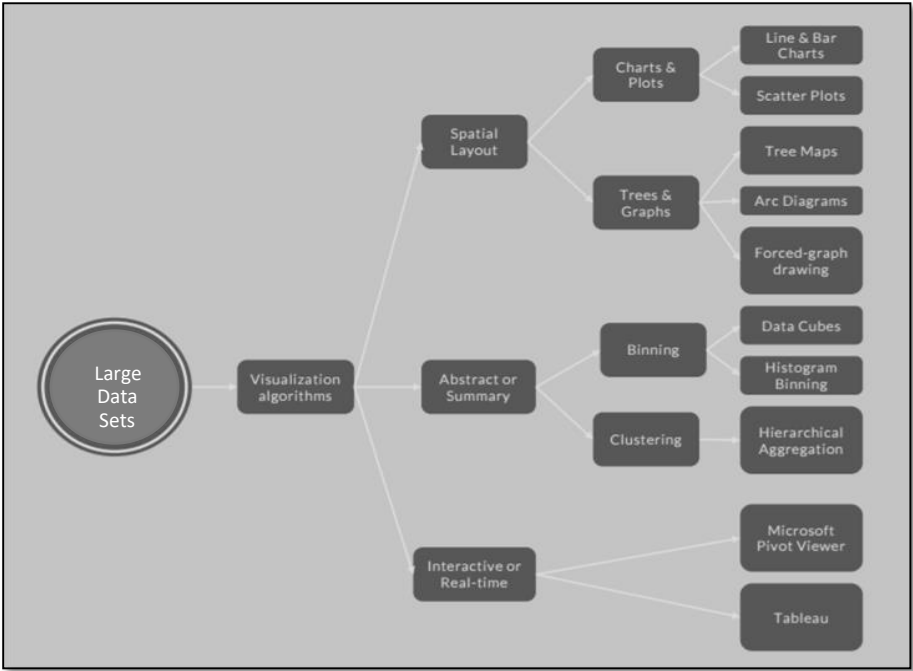


Figure 2.1-1 Taxonomy for visualization of large data sets

IDF data has been selected as the target data to interactively visualize for this project. This data set was chosen as it consists of many levels of geographical detail (from global, to national, to regional data) as well as data based on medical conditions, their spread, mortality rates and medical interventions [34].

Ware [2004] separates the process of data visualization into four: collection and storage of data, pre-processing and transporting the data into human-readable format, displaying the image on screen, and finally the perceiver (see Figure 2.1-2).

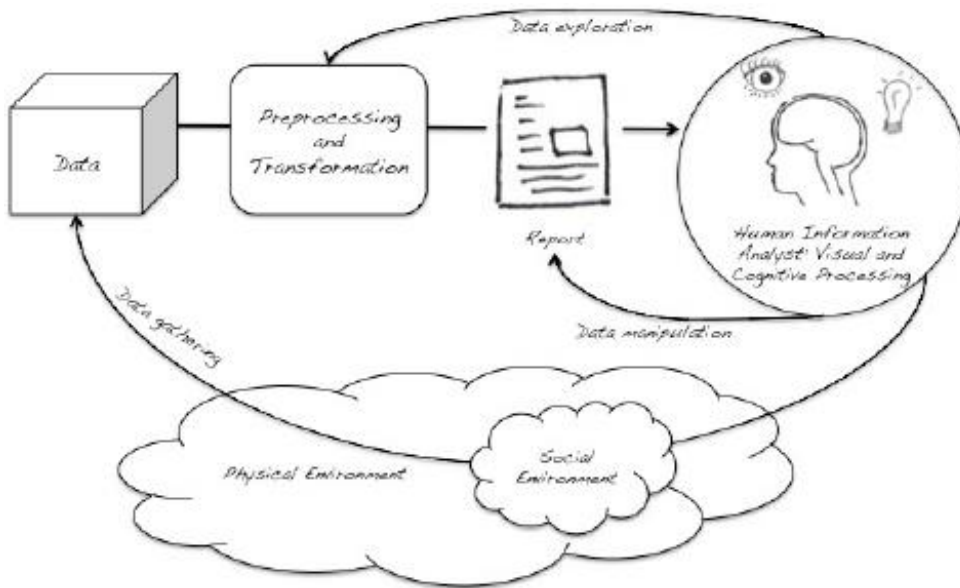


Figure 2.1-2 A schematic diagram of the visualization process [Ware, 2004]

Data is gathered from both physical and social environments. After data is collected, it needs to be pre-processed and transformed into organized canonical data format. At this stage, data entities need to be associated with attribute values. Graphics engine is responsible for mapping the dataset into visual form. Once the visual forms are created, views need to be provided to present transformations (e.g., navigation). Views are interpreted by users through human visual system, visual and cognitive processes.

Although Ware [2004] explains the process of information visualization, it lacks the ability of providing a platform to compare and contrast different information visualization systems. In order to make it simple a new reference model has been introduced.

Card et al. [1999] established a reference model by defining the milestones in the process of information visualization (see Figure 2.1-3). Reference model aims to provide a framework for developers who create new visualizations.

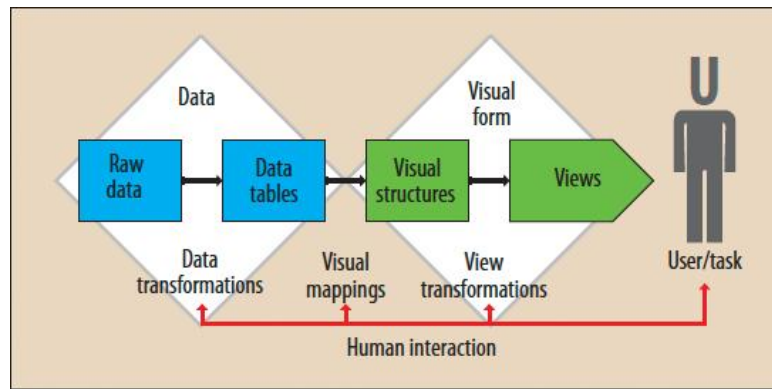


Figure 2.1-3 Reference model for visualization [Card et al., 1999]

According to the Card et al.'s model, raw data needs to be transformed into structured data using data tables. Data tables contain the metadata. Further transformations can be performed on metadata to derive additional characteristics of the data. Visual mappings constitute to the core of the reference model by converting data tables into visual structures. Visual structures are abstract, representing the derived characteristics of the data, which undergoes to the visual mappings. At this stage human vision can process the visual structures, whereas data tables were representing the mathematical relations. Once the visual structures are created and graphical views are acquired by view transformations, human interaction can alter the views (e.g., colour, size and orientation).

According to the Card et al.'s model (Figure 2.1-3), human interaction can take part at any stage of the visualization process. User selects the relative raw-data in the data transformations stage. During the visual mappings, user determines the mapping between the data variables and abstract structures (e.g., axes). User can interactively control the presentation of data, for instance, by distorting at the view transformations stage.

3 Chapter 3

3.1 Design

Data visualization is the presentation of data in a pictorial or graphical format. The main goal of is to provide the user a qualitative understanding of the embedded information in a natural and direct way.

In this approach visualization process is divided into three key stages:

1. Rendering data stage
2. Backward transformation stage
3. Knowledge extraction stage

Rendering data stage: In order to display different values differently, the data should be rendered. In this stage, we map each value in a particular dimension to a member in that visual element domain.

Let's take D as a set of n dimensions in a data set and E as the space of m visual elements which consist of visual objects and their features (To provide an effective visualization system we choose n to be equal to m, this is considered as the association step). These values can be mapped using a transition function as follows.

$$D = \{d_1, d_2, \dots, d_n\}$$

$$E = \{e_1, e_2, \dots, e_n\}$$

$$F_i : d_j \rightarrow e_j \text{ (where } j = 1, 2, 3, \dots, n)$$

Human eyes are not capable of identifying very small visual differences, and also they are not capable of handling a display with overwhelmingly rich visual features. So the transformation functions should be chosen precisely in order to get rid of mentioned constraints. The most appropriate method would be to use a clustering approach where we will be able to assign members of visual elements more reasonably to data values.

Backward transformation stage: The viewer should have the knowledge to reverse the transformed display and restore the original picture in their mind. This process is done by the humans, that's why we have mentioned earlier to map one visual element to only one data dimension where it would be much easier to understand.

Knowledge extraction stage: In our approach, we are not going to provide any solutions for a given data set by analyzing it. We are mainly focusing on building an interactive 3D metaphor where it would be a framework for users who want to visualize large sets of data. Based on the visualization module the users can interact with the system and identifies patterns in the data set and they would be able to come up with strong statements regarding the data set.

As represented in the state diagram of Figure 3.1-1, the visualizations can be updated when the user manually selects an update visualization button on the visualization. Following is a sequence of actions that can occur in response to such update selection. A user views the visualization. The user applies filters and drill down the data set to see the updated visualization.

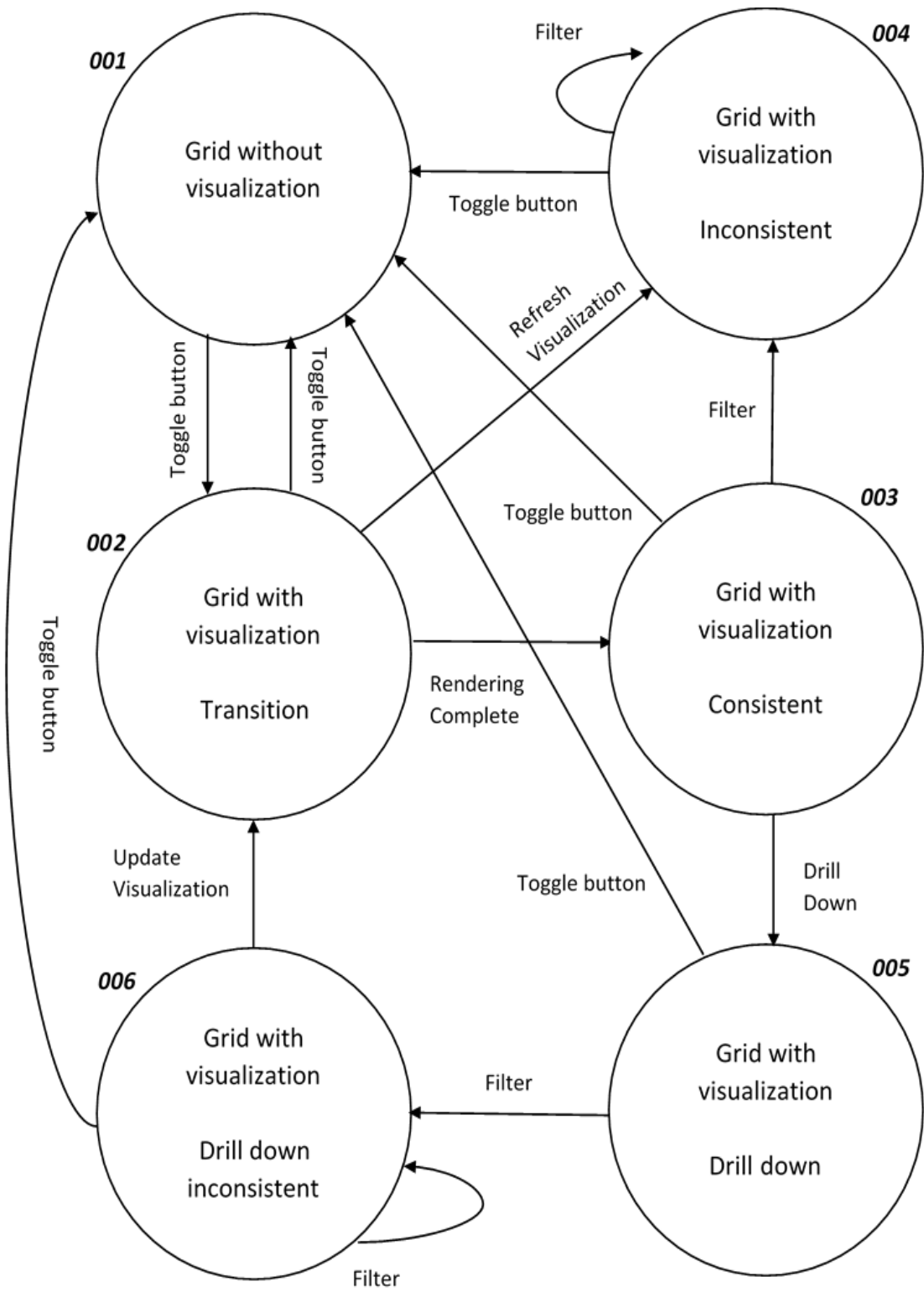


Figure 3.1-1 State machine for 3D Visualization

A first state 001 is where the user interface shows a grid without a visualization, in other words, only the grid is being used by the user. A second state 002 is where the user interface shows a grid with a transition to a visualization, in other words, the visualization is being rendered for the data on the grid. A third state 003 is where the user interface shows a grid with a transition to a visualization, and the visualization is consistent with the grid data. In other words, the visualization has been rendered in combination with the grid and is consistent with the grid data. A fourth state 004 is where the grid is presented, but the visualization is inconsistent with the grid data. In other words, the visualization has become invalid after one or more filters were applied on the dataset. A fifth state 005 is where the grid is presented with visualization drill down, that is, drill down has been performed on the visualization. A sixth state 006 is where the grid is presented, but the visualization drill down is inconsistent, since one or more filters have been applied after drill down.

The 3D metaphor was designed to be a generic 3D visualizer which provides web based 3D visualization of the data set, and handled interactions from the users. The entire metaphor was constructed for a web based platform utilizing JavaScript as its underlying programming language.

The 3D metaphor interface was mainly based on the principals of creating a realistic immersive 3D space and then inserting data into that space so as to allow the users to experience to be immersive within the space, and also allowing them to interact with the metaphor and understand the data has been presented. The interaction has been achieved by implementing a 3D spherical globe with real world texture (c) where the user will get a life like experience. The data has been presented using colored cylindrical tubes, where the metaphor showcased numerical data values in specific geographic locations.



Figure 3.1-2 3D Globe

Mainly the proposed metaphor consist of two parts, convert the data into a readable format (JSON) and present that data using an interactive metaphor (3D globe). In the metaphor each of the data point consist of the location (longitude and latitude) and a “value”. The “value” describe the data point to be shown in the metaphor. For an example if we take people with diabetes the “value” field would be the number of people with the diabetes in each country. By using a generic “value” variable the proposed 3D metaphor would be able to access the different data sets using the same metaphor.

The 3D metaphor was implemented using JavaScript and WebGL as the 3D development platform. The data that has been arranged in csv file was converted to a JSON format using python. The entire system allows the users to interact with the IDF data in an interactive 3D form.

Mainly the 3D nature of the metaphor was designed in order to let the user to see individual locations as 3D cylinder objects extending out from the surface of the planet to

height based on the magnitude of the value (for example areas with high diabetes will consist tall cylinders where, low areas will consist of short cylinders). Comparing this with the one shown in Figure 3.1-3 demonstrates one of the key advantages of the 3D interface. Assume that both systems utilize colors in the same way, but the 3D metaphor allows the user to see the neighbor values as well when there are tall cylinders where as it is other way around in the 2D design where the large circle covers the nearby areas so that the user will not be able to find out hidden patterns. The novelty of the proposed design is that it would allow the viewer to locate high values are located near low values and can highlight areas of opportunity.

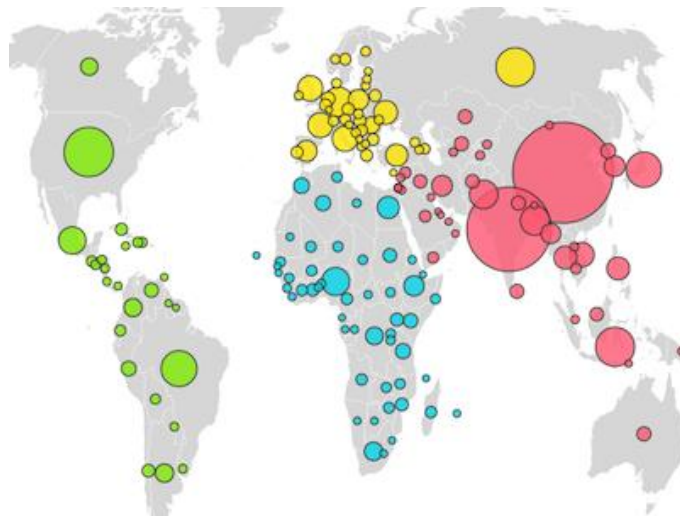


Figure 3.1-3 GapMinder Animated 2D Geographic Charts

3.2 High Level Design Architecture

A different kind of high level structure has been implemented using the same technologies in order to find out the usability of the 3D interactive metaphor. Pima Indians Diabetes data set has been used as the sample data set in order to clarify the patterns using the metaphor. This prototype can be only used by users with a prior knowledge and have a clear understanding on the domain. FIG shows the implementation of the complex architecture with different clusters. As mentioned this prototype is built to give the idea that the 3D visualization can be extended in a way that it would end up providing very valuable findings.

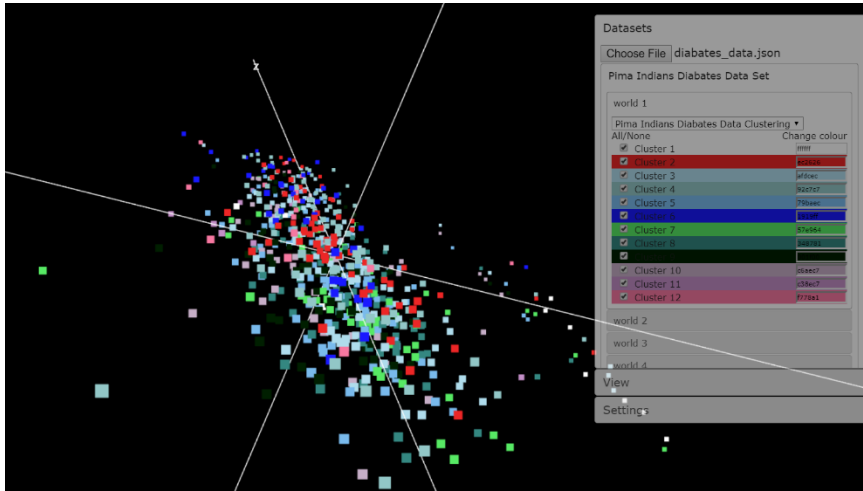


Figure 3.2-1 Prototype1 of clustering PIMA data

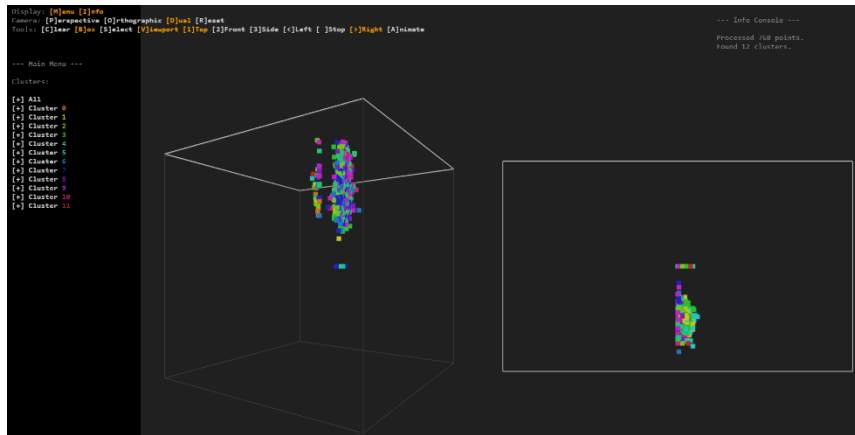


Figure 3.2-2 Prototype2 of clustering PIMA data

4 Chapter 4

4.1 Implementation

4.1.1 Technologies

In recent years, 3D graphics began to play a major role to enhance the multimedia web experience. The rise of Web Graphics Library (WebGL) has allowed lower level access to graphics hardware of ever increasing power.

WebGL is a JavaScript API used to render 2D and 3D graphics to the screen in a browser. Programming directly in the WebGL API can be very complicated and messy, this is exactly why some really useful libraries have recently come about. One such library is Three.js.

Perhaps the most famous library/API for web-based 3D graphics is Three.js [22]. Three.js is a lightweight 3D library that hides a lot of the WebGL complexities and makes it very simple to get started with 3D programming on the web.

In order to interact view and interact with the 3D content the browsers have to install some third party plugins, by installing such browsers it will damage the cross-platform compatibility and contributes to the inability of 3D to work on all browsers and operating systems. This is no longer a problem with WebGL since it is supported by all the browsers where the developers can access graphics hardware acceleration via the browser. It is not required to install any third party plugins which will enhance the cross-platform compatibility.

4.1.2 Hardware Requirements

In order to use advanced 3D graphics, the newest and the higher-end of the host computer is, the better the experience would be. Your computer hardware needs to have a minimum of 2 GB system memory and a video graphics card that supports WebGL. It is recommended that you have at least 4 GB of system memory.

Your hardware should have a nonmobile graphics card with at least 512 MB of video memory. For the best performance, it is recommended that you have a graphics card with at least 1 GB of video memory, especially for working with larger or more memory-intensive scenes. Nonmobile versions of graphics cards typically have better performance than mobile graphics cards (denoted with m or mobile). The Table 4.1-1 illustrates the requirements for a better performance and the specification of the machine we use in this study [22].

Table 4.1-1 Hardware Requirements for Web3D

	Minimum Requirement	Specs of the host machine
Processor	Intel Core i3 Processor or greater	Intel Core i7
Memory	4GB of RAM or greater	16GB RAM
Video Card	Graphics card that supports WebGL	Intel HD Graphics 520 that supports WebGL.
Operating System	Recent version of Windows, OS X, or Linux	Windows 10

Chrome, Firefox, Internet Explorer, Opera, and Safari are all known to have good WebGL support on both desktop and mobile browsers [21].

4.1.3 Implementation Process

Card et al.'s [1999] Data Reference Model is important for understanding and applying JavaScript properties for visualizing the data. Figure 4.1-1 represents the modified version of the Reference Model for information visualization with JavaScript.

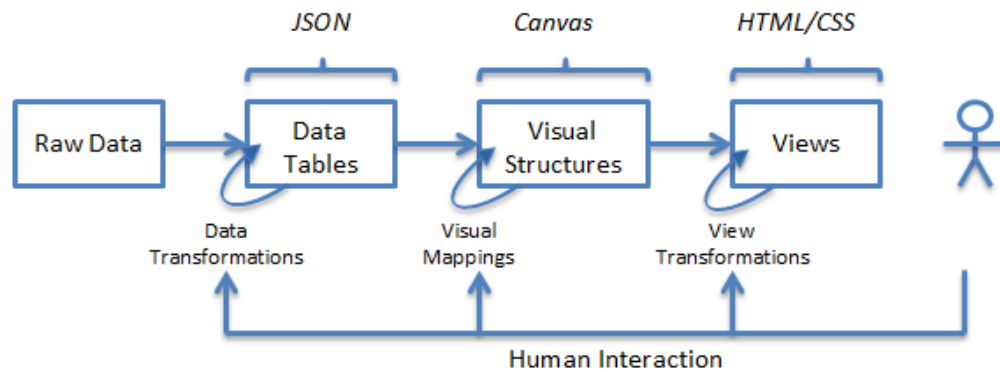


Figure 4.1-1 Reference model adapted for JS web visualization [Card et al., 1999].

JSON corresponds to the data table as it lists the attributes of the metadata. Visual structures, such as a pie chart, are the visualization means that are interpreted directly by the human. Canvas element of HTML5 is being used to draw elements on webpage. View transformations interactively modify visual structures to create interactive visualizations (e.g., zooming and fish eye view). Views are created with HTML elements that are formatted with CSS.

Reference model describes the approximate steps in the information visualization process. Information visualization reference model can be applied easily to the visualizations with JavaScript. JavaScript includes dedicated structures, such as JSON, SVG, Canvas, and HTML, which fit into data tables, visual structures and views of the model.

As mentioned in the design section in order to insert the data into the metaphor it is needed to convert the data into a predefined JSON format as follows,

```
var info = [  
  [  
    'set1', [latitude, longitude, value, latitude, longitude, value, .....]  
  ],  
  
  [  
    'set2', [latitude, longitude, value, latitude, longitude, value, .....]  
  ]  
];
```

Figure 4.1-2 Input JSON format of data

In order to build this JSON format first each of the data sets were arranged in separate csv files containing the longitude, latitude and value. That means each separate 'set1' data set in Figure 4.1-2 will have a separate csv file. Then each of these separated csv file will be put in as an input to a python code snippet (Figure 4.1-3) where it will transform the data set into an array with the required format. Then doing this to each and every csv file, separate arrays were created accordingly and those were arranged in a format as shown in Figure 4.1-2.

```
1 import csv  
2  
3 with open('fileName.csv', 'r') as csvfile:  
4     so = csv.reader(csvfile, delimiter=',', quotechar='')  
5     so_data = []  
6     for row in so:  
7         for i in row:  
8             so_data.append(float(i))  
9  
10    print (so_data)  
11
```

Figure 4.1-3 Python code snippet of csv to JSON conversion

Then the data set is ready to be inserted into the built 3D metaphor and users will be able to interact with each of the data sets that will consist in the JSON file. The 3D metaphor (the globe) is developed using the three.js library which is an open-source JavaScript library which enables high-level programming of browser-based 3D scenes, such as that shown in Figure 4.1-1.

Three.JS features a scene graph, several types of camera and navigation modes, several pre-programmed shaders and materials (and the ability to program custom shaders), Level-of-Detailmesh loading and rendering, and an animation component allowing skeletal and morph-target animation.

Next task is to import the prepared data into the globe where the user will be able to interact with the metaphor. During the literature review many examples showed that in order to access the data a simple function like in Figure 4.1-4 would be enough. In fact, it won't actually load a JSON document, but it will create a JavaScript object. This technique won't be suitable for true JSON files.

```
1 // 'JSON' data
2 data = '[{"set1" : "testing1", "set2" : "testing2"}]';
3
4 // Function to 'load JSON' data
5 function load() {
6     var someData_notJSON = JSON.parse(data);
7     console.log(someData_notJSON[0].set2); // Will log "testing2"
8 }
```

Figure 4.1-4 JSON parser

If the implementation is not going to be fuzzy using an actual JSON file then the most appropriate decision would be to create a separate.js file. But the correct way to do is mentioned in Figure 4.1-5.

```
2  function loadJSON(callback) {
3
4      var requestObj = new XMLHttpRequest();
5      requestObj.overrideMimeType("application/json");
6      // Replace 'sample' with the path to your file
7      requestObj.open('GET', 'sample.json', true);
8      requestObj.onreadystatechange = function () {
9          if (requestObj.readyState == 4 && requestObj.status == "200") {
10             /* Required use of an anonymous callback as .
11              * open will NOT return a value but simply returns undefined in
12              * asynchronous mode*/
13             callback(requestObj.responseText);
14          }
15      };
16      requestObj.send(null);
17  }
```

Figure 4.1-5 XMLHttpRequest code snippet

XMLHttpRequest has an important role in the Ajax web development technique. It is special JavaScript object that was designed by Microsoft. XMLHttpRequest object call as an asynchronous HTTP request to the Server for transferring data both side. It's used for making requests to the non-Ajax pages.

The clue here is the jQuery method \$.getJSON() which is shorthand for \$.ajax(). The function in Figure 4.1-5 will create a new instance of an XMLHttpRequest and load asynchronously the contents of sample_data.json. By changing the argument to true the reading can be converted into a synchronous load. The main feature is that all the modern browsers support the native JSON.parse method.

But this can be write more elegantly by considering the responseType property in XMLHttpRequest instance. When the property is set to the string 'json' browsers that support this feature automatically handle the JSON.parse() step.

Note when the responseType is set to 'json' , requestObj.response must be used instead of requestObj.responseText (Figure 4.1-6) . When the browser fails to parse the response as JSON, null is returned (instead of throwing an error).

```
2  function loadJSON(callback) {
3
4      var requestObj = new XMLHttpRequest();
5      requestObj.overrideMimeType("application/json");
6      // Replace 'sample' with the path to your file
7      requestObj.open('GET', 'sample.json', true);
8
9      requestObj.responseType = 'json';
10
11     requestObj.onreadystatechange = function () {
12         if (requestObj.readyState == 4 && requestObj.status == "200") {
13             /* Required use of an anonymous callback as .
14              * open will NOT return a value but simply returns undefined in
15              * asynchronous mode*/
16
17             callback(requestObj.response);
18         }
19     };
20     requestObj.send(null);
21 }
22
```

Figure 4.1-6 Handling JSON.parse() automatically

After reading the values the next task is to insert them in to the implemented globe. Implementation is mentioned in Figure 4.1-7. First step is to set the location of the globe and then create the globe on the mentioned location. As the next step a file is requested for the JSON data. Then after importing the data if the data is received successfully parse the JSON and input the data into the created globe. Then the geometry is created and next the animation will be started. As the final step the request will be sent where it will handle the data.

```

191 // Location of the globe
192 var positionHolder = document.getElementById( 'container' );
193
194 // Create the globe
195 var globe = new DAT.Globe(positionHolder);
196
197 // Request a file for the JSON data
198 var requestData = new XMLHttpRequest();
199
200 // import data
201 requestData.open( 'GET', 'infomartion.json', true );
202
203 // next step after importing the data
204 requestData.onreadystatechange = function() {
205
206 // if the data is received successfully
207 if ( requestData.readyState === 4 && requestData.status === 200 ) {
208
209 // Parse the JSON
210 var info = JSON.parse( requestData.responseText );
211
212 // input the data into the created globe
213 for ( var i = 0; i < info.length; i ++ ) {
214 | | globe.addData( info[i][1], {format: 'magnitude', name: info[i][0]} );
215 }
216
217 // Create the geometry
218 globe.createPoints();
219
220 // Start animation
221 globe.animate();
222
223 }
224 };
225
226 // Start request
227 requestData.send( null );
228 }

```

Figure 4.1-7 Main Implementation code snippet

5 Chapter 5

5.1 Results and Evaluation

5.1.1 Experimental Testing

In order to test the functional usage of the proposed 3D metaphor a set of experimental scenarios were carried out comparing its online use to two other data visualization forms (tabular and 2D). The experiment consisted of 15 (n) human participants who ranged in age from 18 to 62 years. Mainly there were 9 (60%) male and 6 (40%) female participants.

Participants were in different education levels and had different ratings for their computer literacy.

The experiment involved participants initially being asked to use each of the three forms of data presentation interfaces (presented in random order). In order to maintain the consistency of the data, it has been kept consistent in each and every presentation, showing the International Diabetes Federation figures on people with diabetes around the world. And also in order to make it more convenient the environment also being kept consistent for all the participants where each of them used the same hardware devices. The only changing variable was the presentation interface in each scenario.

Each of the used example presentations are shown in Figure 5.1-1, Figure 5.1-2 & Figure 5.1-3. First presentation included data presented in a tabular format which can be scrolled, and it consisted of the longitude, latitude and the amount of people with diabetes (Figure 5.1-1).

The second presentation was involved the data being presented in a 2D map where data is presented in colored areas (Figure 5.1-2). As the final presentation the users were presented the implemented 3D metaphor where the users will get the chance to interact with the globe by zooming in/out and investigate the tube based data points (Figure 5.1-3).

Initially the users were not given any training time since all of the presentations were simple to use without any prior knowledge. They were given the chance to interact with each of the presentation and gain an understanding of the data is presented. Then the users were given a set of tasks to be completed.

Task 1: Locate low diabetes areas.

Task 2: Locate high diabetes areas.

Task 3: Locate areas where the low diabetes areas are bordered by the high diabetes areas.

Estimates by Country/Territory				
Country/Territory	People with diabetes	Lower limit of uncertainty interval	Upper limit of uncertainty interval	
Afghanistan	935.8	741.3	1,295.8	
Albania	243.6	213.2	275.6	
Algeria	1,679.5	1,157.6	2,359.5	
Andorra	6.2	5.4	8.1	
Angola	345.3	216.2	566.1	
Anguilla	1.3	1.0	1.6	
Antigua and Barbuda	8.3	7.3	9.7	
Argentina	1,723.9	1,292.0	2,502.1	
Armenia	158.0	106.2	294.8	
Aruba	10.7	8.9	13.3	
Australia	1,079.6	840.7	1,301.4	
Austria	611.4	518.9	697.9	
Azerbaijan	428.6	291.8	801.8	
Bahamas	36.0	30.7	43.0	
Bahrain	154.3	141.6	170.4	
Bangladesh	7,138.9	5,319.3	12,009.6	

Figure 5.1-1 Tabular presentation of IDF data

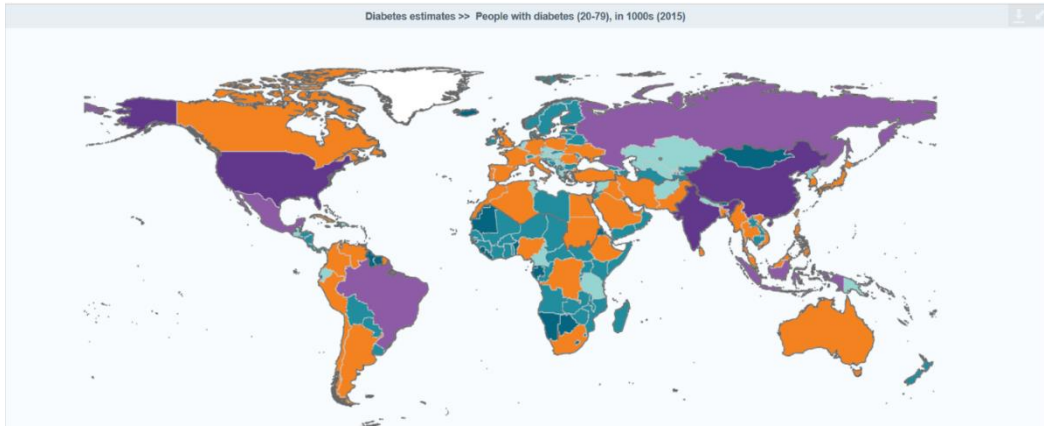


Figure 5.1-2 2D presentation of IDF data

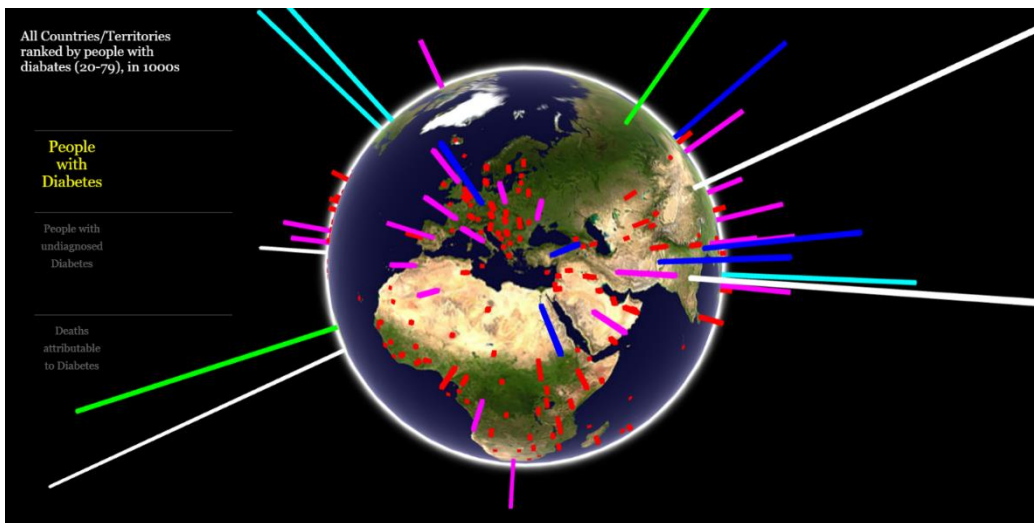


Figure 5.1-3 3D metaphor

In the experiment data was collected based on three main phases.

1. Assigning the participants to several set of tasks.
2. Questionnaire.
3. Visual Observation.

5.2 Results

Mainly the analysis was carried under both quantitative and qualitative perspectives. Then those findings was thoroughly analyzed in order to make conclusions.

5.2.1 Quantitative Results

In this approach the data was gathered by assigning each of the participants to tasks mentioned above. All the participants were given five attempts for each of the presentations. After analyzing the results it can be clearly identified that each of the presentations have different results when it comes to speed and accuracy.

Task 1 - Locate low diabetes areas

Analyzing the data in Table B 1, Table B 2 & Table B 3 in Appendix B, the average and the accuracies are listed for each of the presentations in Table 5.2-1.

Table 5.2-1 Average speed/accuracy of the task1

Presentation	Average Response Speed(s)	Average Accuracy (%)
Tabular	99.93	88.00
2D	88.27	89.33
3D	95.60	81.73

Task 2 - Locate high diabetes areas

Analyzing the data in Table B 4, Table B 5 & Table B 6 in Appendix B, the average and the accuracies are listed for each of the presentations in Table 5.2-2.

Table 5.2-2 Average speed/accuracy of the task2

Presentation	Average Response Speed(s)	Average Accuracy (%)
Tabular	81.73	90.67
2D	79.00	93.33
3D	66.13	96.00

Task 3 - Locate areas where the low diabetes areas are bordered by the high diabetes areas.

Analyzing the data in Table B 7, Table B 8 & Table B 9 in Appendix B, the average and the accuracies are listed for each of the presentations in Table 5.2-3.

Table 5.2-3 Average speed/accuracy of the task3

Presentation	Average Response Speed(s)	Average Accuracy (%)
Tabular	299.67	44.44
2D	121.40	72.00
3D	110.40	78.67

When considering the first task 2D format tend to perform well than both the tabular and the 3D metaphor with an average response time of 88.27s and accuracy of 89.33%. But when it comes to the second task 3D metaphor was more effective by having average response time of 66.13s and accuracy of 96%. Final task was much difficult and most of the users were failed to complete the task using the tabular format. Only 9 people were able to complete the task out of 15 but having a less accuracy of 44.44% and an average response time > 250s. All the users were able to complete the final task using the 2D and 3D presentations. Out of those techniques people were able to answer with a 78.67% when they were involved with the 3D metaphor.

5.2.2 Qualitative Results

Mainly the qualitative research was carried out based on the questionnaire. Set of information was gathered from the users based on the survey. The questions asked from the users as follows. And for each presentation user ratings were gathered accordingly.

Q1: How informative was each of the presentations?

Q2: How suitable the presentation for the given task?

Q3: User Interaction

Q4: Overall rating

The main idea of conducting a survey is that the user experience is very much important when it comes to data visualization. People get more familiar with the data they will be able to make very valuable decisions. So the visualization tool should be crafted in a way that it would not lead to any conflicts and it would give the user the utmost experience that he desire by using the visualization tool. The questionnaire was mainly intended to find out how people react to the newly implemented metaphor which is a totally new experience to them and compare it with other existing techniques and find out whether it enhances their knowledge finding capacity.

Table 5.2-4 Average user ratings for each of the presentations

Presentation	Q1	Q2	Q3	Q4
Tabular	1.67	1.27	1.20	1.87
2D	3.33	3.33	3.20	3.27
3D	4.07	4.33	4.20	4.13

TAB indicates the average user ratings for each of the questions asked from them. Differences between them can be clearly identified from the tabular formats (Figure 5.2-1, Figure 5.2-2, Figure 5.2-3 & Figure 5.2-4).

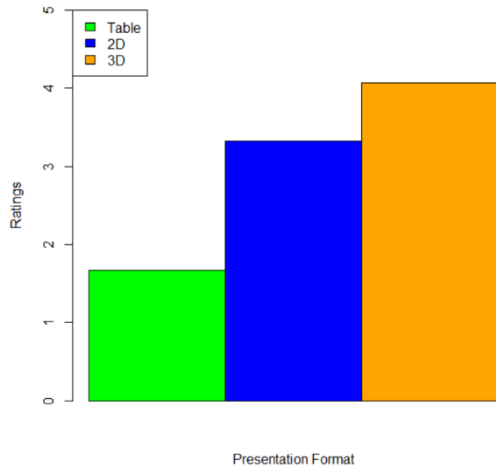


Figure 5.2-2 Chart of user response to "How informative was each of the presentation?"

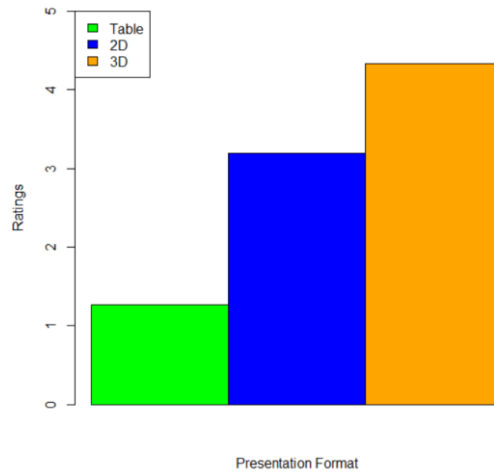


Figure 5.2-3 Chart of user response to "How much suitable, the presentation for the task?"

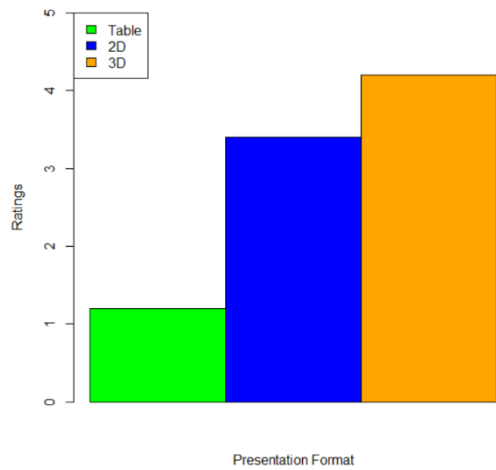


Figure 5.2-1 Chart of user response to "User Interaction"

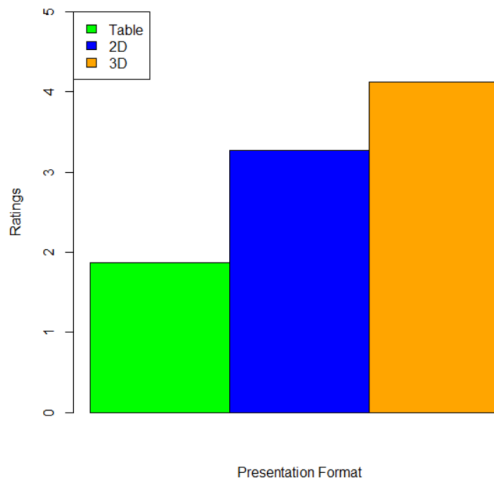


Figure 5.2-4 Chart of user response to "Overall rating"

Based on the facts gathered from the questionnaire in Appendix A we can clearly see that the users were more likely to use the 3D metaphor than any other presentations. Even though the sample size is limited based on these graphs it clearly indicates that there is a significance difference in each of these presentations.

5.2.3 Observational Study

Table B 10 in Appendix B shows the time taken by each user at the start of the experiment before they were given the tasks.

According to the stats people spent a long time playing with the 3D metaphor than any of the other presentations averaging 106.68s comparatively higher than other two presentations (tabular 21.47s & 2D 27.80s).

5.3 Analysis

Considering the results obtained from the experiments it is clearly identified that using this 3D metaphor can be effective in presenting International Diabetes Federation data. And another finding is that the results indicates that in most case the tabular format tend to produce less effectiveness when it comes to speed and accuracy. It may be the fact that it is less interactive than other presentation methods as found out in the qualitative results section.

In the observational study it is found out that users spent a long time playing with the system when it comes to the 3D metaphor. This may be the fact that the 3D presentation enhanced the interaction with the data and thus presented its information effectively and in a manner that enhances its audience's interest.

6 Chapter 6

6.1 Conclusions and Future Works

This chapter includes a review of the research objectives and results, limitations of the current work and the directions for future researches.

6.1.1 Conclusion

At the beginning of the research the thesis tried to address the fact how a 3D interactive metaphor can offer the potential to enhance the user's interaction with data in finding hidden patterns inside the data by data visualization.

This study introduced an interactive 3D metaphor using diabetes data and produce three-dimensional data plots which provide interpretability that makes the data analysis more transparent for a person who is interested in the assessment of individual patients and populations.

In the literature review it is clearly identified the potential to utilize the 2D animation to significantly enhance the effectiveness of visual tools for statistical global data visualization [26, 27, 28, 29]. This research mainly focused on extending the usage of animations as a mechanism to use other entertainment based forms to present IDF data. The proposed 3D metaphor was built as a web based 3D globe where the people got the chance to interact with the IDF data in a more innovative manner. Even though using a small sample ($n=15$) the proposed 3D metaphor tend to provide more significant findings than the other conventional techniques.

Apart from the main design, two other architectures shown in the design section illustrate the idea that this concept can be expanded in a broader way where it can be more complex and can be used to identify different sets of clusters and hidden patterns. But these kinds of visualization tools will have to be used by experienced people and also a lot of preprocessing has to be done before showing the data in a metaphor.

Analyzing all the findings and the facts gathered from the literature review, the proposed approach can conclude the fact that by enhancing the interactions of the visualization tools, it provides a knowledge extraction mechanism to the user in a different perspective when considering the speed and accuracy.

6.1.2 Future Works

The results may be sometimes biased based on two main factors.

1. Experiences users have with the existing systems. (e.g. people have experienced with tables and charts more than they have with 3D metaphors).
2. Novel nature of the proposed 3D metaphor.

The 3D metaphor provided good results may be because of its new nature where the people may have had a more interest in the new approach rather than the conventional techniques. As a future work it is essential to carry out long-term comparative studies between the conventional techniques and the 3D metaphor to see the case.

The metaphor should be tested out with real-time data which is larger in size and can be compared as Big Data. Interactive visualization in the field of Big Data will open up many findings in the data mining field using visualization tools.

There is also the question of why the 3D technique has not come further in software modeling although the 3D technique has been around for decades. When it comes to hardware modeling, 3D tools have been standard for many years. Although the technical issues is one limiting factor behind this difference in technique adaptation, other reasons such as socio-technical and economic factors may be behind that the software industry have not put this technique into practice. This should also be analyzed in a more practical way to come out with good results.

References

- [1] Sun GD, Wu YC, Liang RH et al. A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology* 28(5) Sept. 2013.
- [2] Shixia Liu, Weiwei Cui, YingcaiWu, Mengchen Liu. "A survey on information visualization: recent advances and challenges". *The Visual Computer: International Journal of Computer Graphics*, Volume 30 Issue 12, December 2014 Pages 1373-1393.
- [3] Jeffrey Heer and Sean Kandel. "Interactive analysis of big data". *Magazine XRDS: crossroads, The ACM Magazine for students – Big Data*, Volume 19 Issue 1, Fall 2012 Pages 50-54.
- [4] Jinson Zhang, Mao Lin Huang."2D approach measuring multidimensional data pattern in big data visualization", *Big Data Analysis (ICBDA) 2016 IEEE International Conference on*, pp. 1-6, 2016.
- [5] Daniel Cheng, Peter Schretlen, Nathan Kronenfeld, Neil Bozowsky, William Wright. "Tile Based Visual Analytics for Twitter Big Data Exploratory Analysis" *IEEE International Conference on Big Data* 2013.
- [6] Naveed Ejaz, Irfan Mehmood, Jeong Joong Lee, Su Mi Ji, Myung Ho Lee, Sung Mmahn Anh, Sung Wook Baik. "Interactive 3D visualization of social network data using cloud computing". *International Conference on Cloud Computing and Social Networking (ICCSN) 2012*.
- [7] Praveen Murthy, Anurag Bharadwaj, P. A. Subrahmanyam, Arnab Roy, Sree Rajan. *BIG DATA WORKING GROUP Big Data Taxonomy*, September 2014.
- [8] 3D Visualization World interviewed Dr. Helwig Hauser: *Advanced Visualization Research Explores New Paths*.
- [9] Shneiderman, B., 2002, "Inventing Discovery Tools: Combining Information Visualization with Data Mining," *Information Visualization*. 1(1), 5-12.
- [10] Card, S.K., Mackinlay, J.D., and Shneiderman, B., 1999, *Information Visualization: Using Visualization to Think*, Morgan Kaufmann, San Francisco.
- [11] J. Heer and B. Shneiderman, "Interactive Dynamics for Visual Analysis," *Communications of the ACM*, vol. 55, no. 4, pp. 45–54, Apr. 2012.

- [12] C. Ware, *Visual Thinking: for Design*, 1st ed. Morgan Kaufmann, 2008.
- [13] J. Thomas and K.A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, 2005.
- [14] C. A. Blok, U. D. T. Turdukulov, R. Zurita-Milla, V. Retsios, M. Schouwenburg, and M. Metaferia, "Development of an Open-Source Toolbox for the Analysis and Visualization of Remotely Sensed Time Series," *Cartographica*, vol. 46, no. 4, pp. 227–238, 2011.
- [15] J. Grady, "Metaphor," in *The Oxford handbook of cognitive linguistics*, D. Geeraerts and H. Cuyckens, Eds. Oxford; New York: Oxford University Press, 2007.
- [16] Christof Nuber, Ralph W. Bruckschen, Bernd Hamann and Kenneth I. Joy, "Interactive Visualization of Very Large Medical Datasets using Point-based Rendering". *Proceedings of the SPIE*, Volume 5029, p. 27-36 (2003).
- [17] Beyer, Johanna, Markus Hadwiger, and Hanspeter Pfister. 2014. "A Survey of GPU-Based Large-Scale Volume Visualization". In the *Proceedings of The Eurographics Conference on Visualization (Eurovis 2014)*, Swansea, Wales, June 9-13, 2014.
- [18] Eleftherios E. Koutsofios, Stephen C. North, Russell Truscott, Daniel A. Keim, "Visualizing Large-Scale Telecommunication Networks and Services". *VIS '99 Proceedings of the conference on Visualization '99: celebrating ten years*, pages 457-461.
- [19] Wim de Leeuw and Robert van Lier, "Methods for Interactive Visualization of Large Flow Data Sets".
- [20] Nathaniel Rossol, Irene Cheng, John Berezowski, and Iqbal Jamal, "An Extensible Interactive 3D Visualization Framework for N-Dimensional Datasets Used in Heterogeneous Software Display Environments". *ISVC'11 Proceedings of the 7th International conference on Advances in visual computing Volume Part 1*, pages 508-517.
- [21]"Getting a WebGL Implementation", [Get.webgl.org](https://get.webgl.org), 2017. [Online]. Available: <https://get.webgl.org/get-a-webgl-implementation/>. [Accessed: 18- Oct- 2017].
- [22] Alun Evans, MarcoRomeo, ArashBahrehmand, JaviAgenjo and JosepBlat, "3D graphics on the web: A survey", *Interactive Technologies Group, Universitat Pompeu Fabra, Barcelona, Spain*.
- [23]H. Piringner, "Large Data Scalability in Interactive Visual Analysis ausgef", *Dipl.-Ing., Vienna University of Technology*, 2017.
- [24] Nuzzo, R. 2014. Statistical errors. *Nature*, 506(7487), 150-152.

- [25] G. Ellis and A. Dix. Enabling Automatic Clutter Reduction in Parallel Coordinate Plots. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):717 – 724, 2006.
- [26] Purchase, H. C. 2014. Twelve years of diagrams research. *Journal of Visual Languages & Computing*, 25(2), 57-75.
- [27] Rosling, H. 2006. Debunking third world myths with the best stats you've ever seen. online]. *Ted Talks-Ideas Worth Spreading*. Retrieved July, 25, 2011.
- [28] Rosling, H. 2010. What showbiz has to do with it? In *Data and context in statistics education: Towards an evidence based society*. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8).
- [29] Shankar, P. R. 2015. Gapminder: Understanding Health through Interactive Statistics. *Education in Medicine Journal*, 7(2).
- [30] J. Bottger, M. Balzer, and O. Deussen. Complex Logarithmic Views for Small Details in Large Contexts. *IEEE Transactions on Visualization and Computer Graphics*, 12:845– 852, 2006.
- [31] D. A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, editors. *Mastering The Information Age – Solving Problems with Visual Analytics*. Eurographics, 2010.
- [32] R. Becker and W. Cleveland. Brushing Scatterplots. *Technometrics*, 29(2):127–142, 1987.
- [33] A. Cedilnik, B. Geveci, K. Moreland, J. Ahrens, and J. Favre. Remote Large Data Visualization in the ParaView Framework. In *6th Eurographics Symp. On Parallel Graphics and Visualization*, pages 163–170, 2006.
- [34] "International Diabetes Federation - Home", *Idf.org*, 2017. [Online]. Available: <https://www.idf.org/>. [Accessed: 18- Oct- 2017].

Appendix A: Questionnaire

Rate the following from 1-5

- 1: very bad
- 2: bad
- 3: average
- 4: good
- 5: very good

Question	Presentations		
	Tabular	2D	3D
Computer Literacy			
How informative was each of the presentations?			
How suitable the presentation for the given task?			
User Interaction			
Overall rating			

Appendix B: Diagrams

	Average Response Time (s)	Average Successful Attempts
User1	110	0.8
User2	90	1
User3	80	1
User4	88	1
User5	106	1
User6	112	0.6
User7	82	1
User8	101	0.8
User9	105	0.6
User10	90	0.8
User11	101	0.6
User12	89	0.6
User13	157	0.6
User14	90	1
User15	98	1

Table B 1 Tabular representation stats for finding low diabetes areas

	Average Response Time (s)	Average Successful Attempts
User1	80	1
User2	70	1
User3	80	1
User4	83	1
User5	89	1
User6	90	0.8
User7	70	1
User8	89	0.4
User9	96	0.6
User10	88	1
User11	100	1
User12	94	1
User13	134	0.6
User14	72	1
User15	89	0.8

Table B 2 2D representation stats for finding low diabetes areas

	Average Response Time (s)	Successful Attempts
User1	94	1
User2	98	1
User3	78	1
User4	90	1
User5	103	1
User6	101	0.6
User7	79	1
User8	95	0.8
User9	102	0.4
User10	94	1
User11	94	1
User12	90	1
User13	132	0.6
User14	89	1
User15	95	1

Table B 3 3D representation stats for finding low diabetes areas

	Average Response Time (s)	Average Successful Attempts
User1	70	1
User2	64	1
User3	65	1
User4	70	1
User5	90	1
User6	100	0.6
User7	58	1
User8	73	1
User9	75	0.8
User10	85	1
User11	79	0.8
User12	80	1
User13	160	0.6
User14	82	1
User15	75	0.8

Table B 4 Tabular representation stats for finding high diabetes areas

	Average Response Time (s)	Average Successful Attempts
User1	78	1
User2	69	1
User3	76	1
User4	70	1
User5	75	1
User6	120	1
User7	76	1
User8	76	0.8
User9	75	0.8
User10	73	0.8
User11	78	1
User12	69	0.8
User13	107	1
User14	73	0.8
User15	70	1

Table B 5 2D representation stats for finding high diabetes areas

	Average Response Time (s)	Average Successful Attempts
User1	62	1
User2	54	1
User3	69	1
User4	60	1
User5	68	1
User6	78	0.8
User7	63	1
User8	69	0.8
User9	64	1
User10	61	1
User11	68	1
User12	75	1
User13	81	0.8
User14	63	1
User15	57	1

Table B 6 3D representation stats for finding high diabetes areas

	Average Response Time (s)	Average Successful Attempts
User1	320	0.4
User2	350	0.4
User3	290	0.4
User4	278	0.6
User5	-	-
User6	-	-
User7	290	0.6
User8	450	0.4
User9	-	-
User10	-	-
User11	-	-
User12	420	0.2
User13	-	-
User14	475	0.4
User15	540	0.6

Table B 7 Tabular representation stats for finding low diabetes areas bordered by high diabetes areas

	Average Response Time (s)	Average Successful Attempts
User1	120	1
User2	110	1
User3	127	0.8
User4	122	0.6
User5	115	0.8
User6	145	0.6
User7	123	1
User8	110	0.6
User9	120	0.8
User10	105	0.6
User11	143	1
User12	110	0.4
User13	138	0.4
User14	110	0.4
User15	123	0.8

Table B 8 2D representation stats for finding low diabetes areas bordered by high diabetes areas

	Average Response Time (s)	Average Successful Attempts
User1	100	0.8
User2	120	1
User3	102	1
User4	108	0.4
User5	110	0.8
User6	124	1
User7	105	0.8
User8	103	0.6
User9	110	1
User10	98	1
User11	132	1
User12	101	0.6
User13	140	0.4
User14	99	0.4
User15	104	1

Table B 9 3D representation stats for finding low diabetes areas bordered by high diabetes areas

	Time (s)		
	Tabular	2D	3D
User1	23	28	90
User2	24	29	132
User3	23	24	124
User4	19	27	117
User5	19	29	136
User6	24	20	99
User7	19	30	101
User8	24	34	104
User9	12	34	89
User10	24	32	105
User11	23	33	103
User12	23	27	109
User13	19	20	78
User14	24	30	93
User15	22	20	119

Table B 10 Average playing time with presentations initially