**CHAPTER 1**

**INTRODUCTION**

This chapter explains introduces about the project and explains about its background study, the problem statement that led to the development of this project. Furthermore, it follows by highlighting of the aim of the study and objectives. To complete this chapter, the scope of the study is outlined, together with the significance.

**1.1 Background Study**

In general, data visualization is a process of representing data in visual form to help people understand the data (Ephrati, 2017). Based on Ottinger (2017), there are two types of data visualization namely the ‘explanation’ and ‘exploration’. Explanation data visualization concerns on explaining about the data while exploration data visualization concerns on presenting the patterns of the data to the audience and both types are required to draw the full potential of the data itself.

There are various usages of data visualization which is to simplify the data and display the data pattern. In addition, data visualization can be used to represent data in various domains including emotions (Montanez, 2016), social network (Desale, 2015), election data (Krum, 2017), etc.

These visualizations involve different types of visualization techniques, such as Hierarchical techniques, Network techniques, Graph techniques, Temporal techniques, Map Visualization techniques, etc. Most of the data are visualized using these interactive visualization techniques, especially in US, UK, Australia, Russia, and India. This kind of visualization can ease the process of learning the context of a data (Gillet, 2014), and human brains can make use of the visual processing much better (Balm, 2014).

In Malaysia, most of the data are represented using tabular format (Pepinsky, 2009), simple bar charts (Nehru, 2013), and infographic (Zairi, 2011). These data include election data, population statistics, budget data, economic, financial statistics, gross domestic product (GDP), etc. The viewer must read and interpret the related articles one-by-one to understand the information and pattern of the data.

**1.2 Problem Statement**

Election data consist of multiple information including state, district, name of candidate, political parties, population, demographic data, etc. These data are hard to be understood by the viewers without prior knowledge to interpret the data. The viewers need to analyze each data in detail before they can understand the data and the stories behind it (Steele, 2012).

Therefore, most of the countries, like US (Lilley, 2016), UK (Hanretty, 2017), Russia (Andrew, 2017), and Australia (Lach, 2010), have presented their election data using interactive data visualization techniques. Figure 1.1 shows the example of visualization used in US Presidency Election 2016.

|  |
| --- |
| **FiveThirtyEight’s 2016 Election ForecastFiveThirtyEight’s 2016 Election Forecast5-30-8_box-plot**  **Figure 1.1** Example of Data Visualization  (*Source*: Lilley, 2016) |

Based on Figure 1.1, there are three types of visualization techniques, which are bar chart, tabular bar chart, and winding chart. These visualizations are interactive, and the user can click and view the details of the data. However, most of previously Malaysia election data are represented in tables and some of them are represented in simple but non-interactive graphs. It is very problematic and time consuming to analyze the election data one-by-one. The previously common Malaysia election data visualization are shown in Figure 1.2.

|  |
| --- |
| http://4.bp.blogspot.com/-4siL9dGW2OY/UaRLpr0MhyI/AAAAAAAAAEI/YHllXC1U-mg/s640/GE13+DUN+RESULTS.png**http://4.bp.blogspot.com/-vU3xnuHzMU4/Tkjwc1f7LqI/AAAAAAAAAHs/V9FAsFRD1QM/s400/2.notakananpartimenangpp.jpg**  **Figure 1.2** Example of data visualization for Malaysia election data  (*Source*: Zairi, 2011) |

With the help of data visualization, it could help the people to analyze the election data much faster and easier. With this, viewer could understand the election data without prior knowledge of data analytics.

The motivation for this project is to solve the problem of interpreting the tabular data and improve the data visualization of Malaysia election data to a better representation by conveying more information in a single visual. This project would also implement a dynamic data visualization of Malaysia election data by changing the output of the visual when the input changes.

**1.3 Objective**

This project aims to visualize Malaysia election data using dynamic and interactive representation. With this aim, the following objectives are outlined:

1. To identify the suitable data visualization technique for Malaysia Election Data.
2. To develop a system based on the chosen data visualization technique.
3. To validate the accuracy of visual with the actual election data.

**1.4 Scope**

This project only focuses on the data visualization for Malaysia election data based on Parliaments and not Dewan Undangan Negeri (DUN). Based on Malaysia General Election 2013, there are a total of 222 parliament seats. Each of the seats represents an area in each states of Malaysia.

The chosen technique to visualize Malaysia election data is the map data visualization. This is because most of the election data visualization in Malaysia is based on simple charts and most of the visualization that includes map are represented in the form of infographic (static) and most of the visualization are not interactive, cannot dynamically change the visualization based on inputted data and most of it are based on explanatory data visualization which means that they do not allow the user to explore the data.

This project also focusses on applying a similar visualization technique that had been applied by other selected countries to enhance this project’s data visualization. Some of the similar systems that this project is focusing on is the data visualization from United States (US), United Kingdom (UK), Australia and India.

**1.5 Significance**

This project is intended for users who needs to visualize and interpret election data dynamically and interactive while it is also beneficial to political figures, public audience and data scientist.

From the visualized data, political figures could understand what their current conditions are and how they can make use of the data to improve themselves for the next election. For example, they could determine which party has won the most seats and why by looking at the demographic information provided. After that, they can try to come out with a solution to their weaknesses.

Public audience also can understand the political condition of the country easier and faster. For example, people are always overwhelmed with a tabular format of election data. This tabular format will not make them understand the election data easier and faster because it is quite hard to read the table row-by-row.

From the visualization, data scientist also could learn the relationship between the data much faster. For example, they could learn the relationship between the number of votes and geographical information that could cause a certain behavior of the election such as the cause towards the winning party.

**1.6 Summary**

This chapter presents the background study, problem statement, objectives, scope and the significance of this project. The background study shows on the origin of the terms of ‘Data Visualization’, its usages and types. It focuses on the current representation of Malaysia election data. With the comparison with other countries’ election data visualization, the weaknesses of Malaysia election data representation are outlined as problem statement. Based on this problem statement, this project sets its aim and objectives to solve this problem. The scope specifies the type of data, techniques, and similar systems to set the boundary of the project. Lastly, the significance of this project is highlighted as their potential usage.