CPSC 4030/6030: Data Visualization

Visualization of US 2016 Presidential Election Dataset

Group - G03

Harika Naidu Etha, Mugdha Patil

December 8, 2020

0.1 Introduction

Visualization is a key concept to explore and understand complex data. It allows us to get insight into problems from different perspectives. Moreover, it is a quick, easy, and efficient way of understanding the problem. The results of US 2016 presidential election were unexpected as many of the people and businesses were expecting Clinton as their new president. Thus many people are interested to explore and understand 2016 election. In this project, we have developed a website that provides sufficient visualization to get insights from the US 2016 presidential election data. We have used four type of visualizations: Choropleth, Bar graph, Scatterplot and Sunburst chart for effective visualization of the data. Our visualizations on the website clearly demonstrate/depict the data of the election in different perspective.

0.2 Literature Review

Graphical presentation of complex data provides a quick and easy way to understand complex patterns from the data that are not easy to understand using simple observations [5]. There has been a surge in the approaches that follow different visualization approaches to handle various complex problems. Though, there are many visualization methods, barchart, scatterplot and maps are known to be the common visualization methods for understanding complex data. In [7], different scatterplot visualization and design options are studied and guidelines are suggested for scatterplot based on data characteristics, analysis tasks and design choices. Similarly, continuous scatterplots were studied and suggested in [1] to the visualization of spatially continuous input data by a continuous and dense plot. In addition to the visualizations, scatterlplot also helps in identifying outliers and provide a real-world distance measure and-or comparing facility between two points. The authors in [8] provided a comprehensive review of choropleth maps. Choropleth maps are graphical representation highlighting the differences in the geographical distribution of data by spatial unit, often using

administrative boundaries such as countries, states or regions [8]. Choropleth maps have numerous applications in big data, crime analytics, data science, geo sciences and many more. In [2], the effectiveness of barchart is highlighted and several use case has been shown to deploy barchart on different types of data. Among them, monthly rainfall in Toronto and Beijing is shown with barchart. Similarly, several other studies including [3, 4, 9] highlight the applications of bar graphs in different scenarios. Among these visualizations tools, Sunburst is also known to be very useful for visualizing hierarchical data [6]. Each level of the hierarchy is represented by one ring or circle with the innermost circle as the top of the hierarchy. A sunburst chart without any hierarchical data (one level of categories), looks similar to a doughnut chart.

Since there are millions of records in our dataset, the problem of our interest is quite complex. It's extremely hard to make sense of it without a good visualization tool. One example can be the number of votes with respect to states, districts and party. To tackle this problem, we leverage four different visualization methods: Choropleth, scatterplot, barchart, and pie chart. Every chart we used has it's own numerous applications and are well-known because of their diverse usage. The applications include computer science, where machine learning and deep learning models and datasets are largely explored using scatterplots and barcharts. On the other hand, geospatial data such as satellite imagery, maps, and many other have wide applications in geosciences and machine learning. Choropleth is one of the excellent tool to understand the geospatial data. Since our data is spatial and has so many categories, we leverage these tools and designed the website based on these tools to explore this data.

0.3 Dataset and Method

We have chosen the publicly available US presidential 2016 data which is available on the Kaggle[1]. This dataset contains 18 different attributes, among which the most important are: state, county, party, candidate, and vote. We have used

[1] `https://www.kaggle.com/tunguz/us-elections-dataset`

these attributes to produce effective visualizations of different levels that include state level, county level, and candidate level. The attributes are as follows:

- Year
- State
- Office
- District
- Candidate
- Candidate Votes
- Total Votes
- Party
- Special
- Mode
- Version

We hypothesized which candidate, state, and county was significantly involved to change the results of the election. To answer this question, we have developed different Tableau visualizations shown in 1, 2 and 3. In Figure 1, The Choropleth shows the winning party state-wise. In Figure 2, The Scatterplot shows state wise number of votes gained by each candidate. Finally, In Figure 3, The Bar chart shows the percentage of votes gained by each candidate.

0.4 Design Solution

This section presents the design solution we have proposed for exploring the the problem which is related to the unexpected results of the 2016 elections in the United States. US presidential elections of 2016 were clearly projected to win by Clinton, however, the results were reversed and Trump won making him

the president of the United States. In order to understand how the places got reversed, we need to identify which candidate, states, and county had played a key role to change the results. For example, can we produce effective visualizations that describe the 2016 US presidential election details? Can we develop a website that will help us to visualize the problem? To answer these questions, we have developed a website that helps to understand the details of the 2016 elections. On this website, we have drawn four different types of graphs which are: map graph (Choropleth), bar graph, scatterplot, and pie chart (Sunburst). Each graph illustrates the details of both parties i.e. Clinton and Trump. These details have been fetched by the dataset of the election 2016, which we have obtained from Kaggle. Figure 1 shows the votes gained by each party state wise, it focuses on showing the winning party state wise, Blue color representing the Democrats and Orange color representing the Republicans. Figure 2 shows state-wise number of votes gained by each candidate. x axis represents states while y axis indicates number of votes gained by each candidate. In this figure the Blue color represents Hillary Clinton and Orange color represents Donald Trump. Similarly, the Barchart in Figure 3 shows percent of votes gained by each candidate. In this figure the Blue color represents Hillary Clinton and Orange color represents Donald Trump.

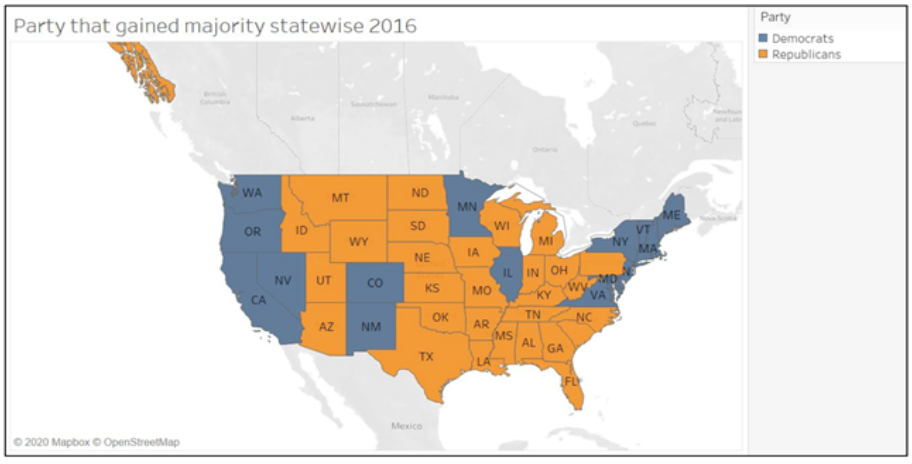

Figure 1: Choropleth graph represents the winning party state-wise.

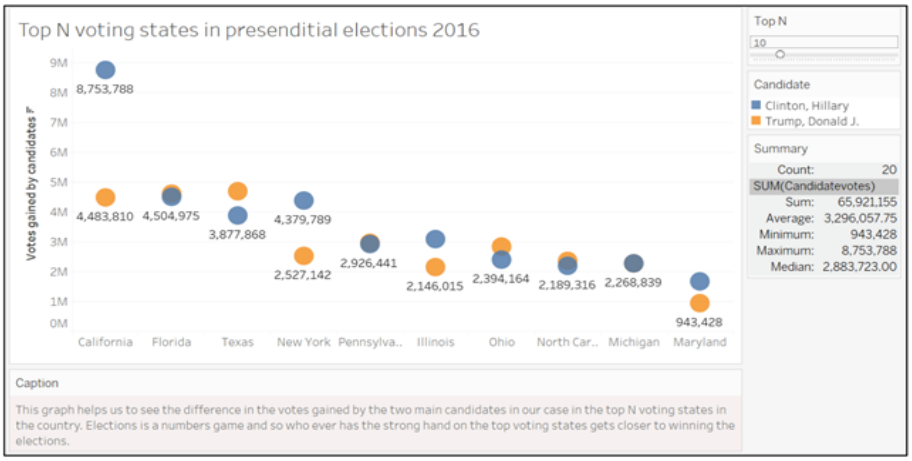

Figure 2: Scatterplot of vote gained by both parties

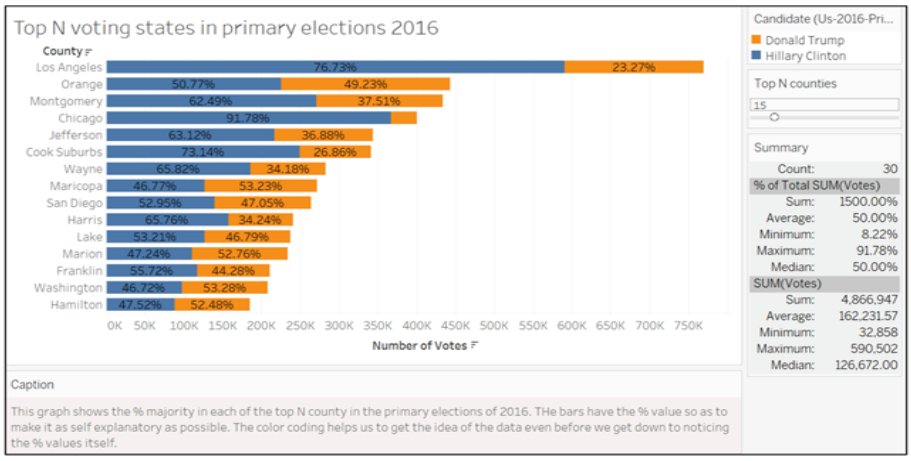

Figure 3: Scatterplot of vote gained by both parties

The website allowed us to apply different filters to know about the more specific and detailed results state-wise and county-wise. The dashboard of the website we have developed can be seen in Figure 4 below.

The design of this website is in such a way, that tells the whole story itself. For example, after the basic description of the project, we have drawn four graphs. The working of each graph is illustrated below:

- **Chloropeth Map**:

  The Choropleth map graph illustrates the party that gained the most votes state-wise. The states with Blue color are democrats, whereas the states with Orange color are republicans. We have drawn this graph because it is a good way to present the state-wise results.

  We have chosen the Chloropeth map to get an idea about the state and the number active voters help us understand the importance of a particular state in the elections. In a state with a fair amount of active voters, the parties competing will try to have a more substantial hand than their opponents. Such top voting states can drive the elections in favor of the party who knows how to plan their campaign and promotion in such state.

  The attributes used here are state name/ state code which is a categorical attribute and total votes which is a quantitative attribute. Visual channels such as color(hue) and shape are used in this particular graph.

  For example, in this particular graph we have selected the Texas state. This selection will change the rest of the graphs which are barchart, scatterplot and sunburst chart.

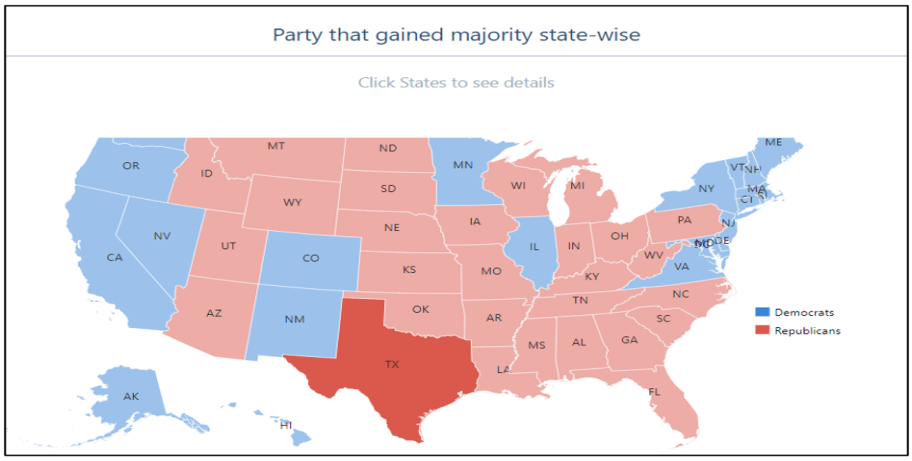

Figure 4: Choropleth graph represents the winning party state-wise.

- **Bar Chart**:

  The Bar graph illustrates the top 10 voting counties, after selecting the state. The x axis represents the states and the y axis represents the number of votes. The Blue color represents the Clinton votes percentage and the Orange is represents Trump. We have used this graph because it can represent both parties' results in the same bar. This means it provides an easy form to understand the election results.

  The attributes used here are county being a categorical attribute, total votes being a quantitative attribute and candidates being categorical attribute. Visual channels such as color, position and size(length) are used in this particular graph.

  For example, As we have selected the state of Texas in the Choloropeth map, the barchart in this particular graph displays the top 10 counties in that state with the number of votes won by Clinton and Trump respectively. The county with the highest votes is displayed at the top and the one with the least votes is displayed at the bottom.

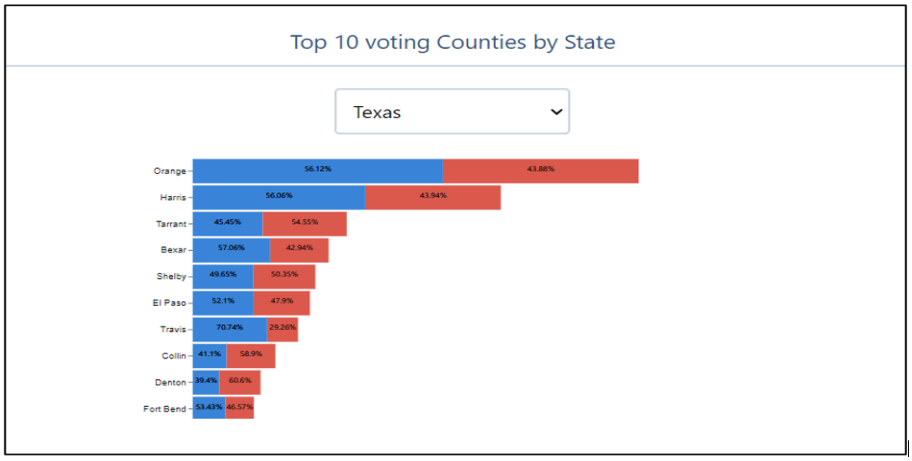

Figure 5: Barchart of top 10 voting counties by state.

- **Scatterplot**:

  The Scatterplot represents the exact results gained by Clinton and Trump in different states. The Blue color represents the number of votes gained by Clinton and the Orange color represents the votes gained by Trump. We used this graph because it provides an easy way to identify the difference of votes between Clinton and Trump.

  The attributes used here are country being a categorical attribute, total votes being a quantitative attribute and the candidates being a categorical attribute. The visual channels used here are color and position.

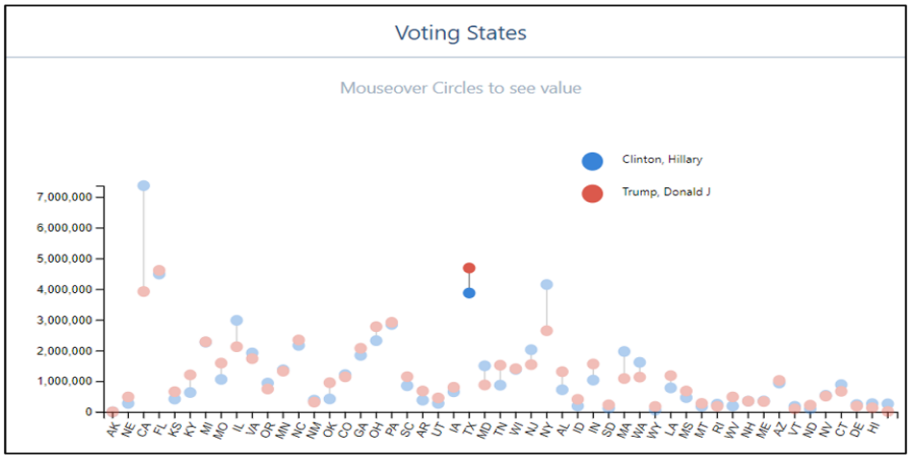

Figure 6: Scatterplot of total votes gained by each candidate state-wise.

- **Sunburst graph**:

  The Sunburst graph represents the total electoral votes between Clinton and Trump. As the electoral votes play a vital role in US elections, they have been represented using a sunburst chart. The blue color is represents the Clinton electoral votes and the orange is representing Trump. We used this graph because it provides a faster and clearer output than other graphs.

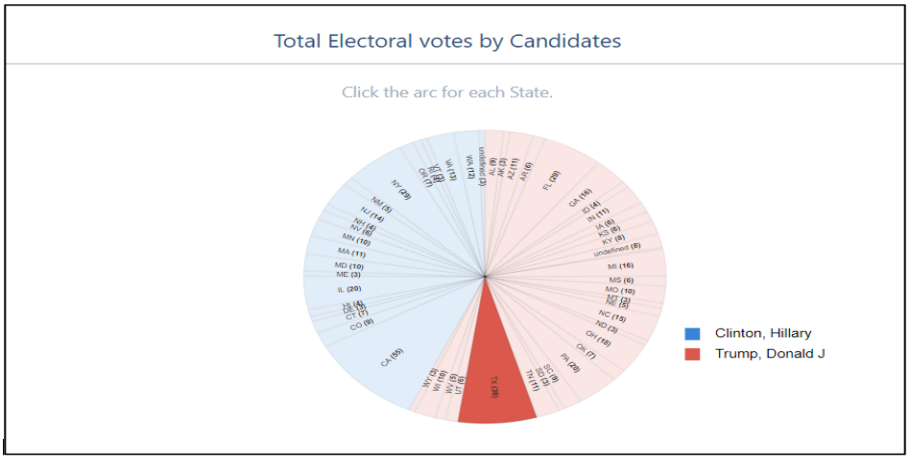

Figure 7: Sunburst chart of total electoral votes gained by each candiate.

Each graph allows the user to apply the filter, The selection of any state in the Choloropeth map updates the rest of the graphs i.e. the barchart, scatterplot and the sunburst chart automatically display the data accordingly. Moreover, we have mentioned links to data sources and reference codebases at the bottom of our website.

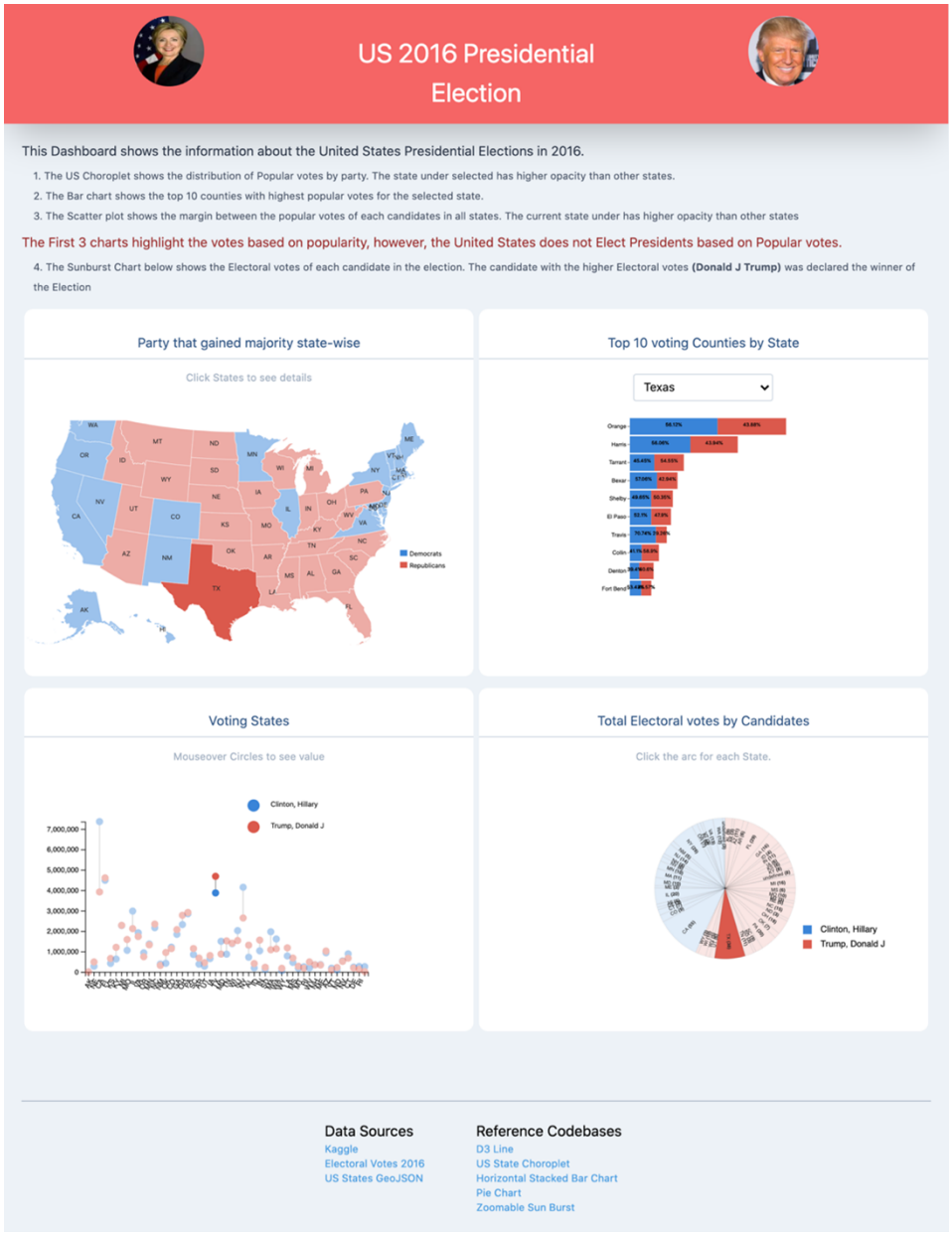

Figure 8: Dashboard of our developed website

Bibliography

[1] Sven Bachthaler and Daniel Weiskopf. Continuous scatterplots. *IEEE transactions on visualization and computer graphics*, 14(6):1428–1435, 2008.

[2] Linda L Cooper and Felice S Shore. The effects of data and graph type on concepts and visualizations of variability. *Journal of Statistics Education*, 18(2), 2010.

[3] Thomas Deane, Kathy Nomme, Erica Jeffery, Carol Pollock, and Gülnur Birol. Development of the statistical reasoning in biology concept inventory (srbci). *CBE—Life Sciences Education*, 15(1):ar5, 2016.

[4] Patricia B Humphrey, Sharon Taylor, and Kathleen Cage Mittag. Developing consistency in the terminology and display of bar graphs and histograms. *Teaching Statistics: An International Journal for Teachers*, 36(3):70–75, 2014.

[5] Stephen R Midway. Principles of effective data visualization. *Patterns*, page 100141, 2020.

[6] Bahador Saket, Alex Endert, and Çağatay Demiralp. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics*, 25(7):2505–2512, 2018.

[7] Alper Sarikaya and Michael Gleicher. Scatterplots: Tasks, data, and designs. *IEEE transactions on visualization and computer graphics*, 24(1):402–412, 2017.

[8] Alsino Skowronnek. Beyond choropleth maps: A review of techniques to visualize quantitative areal geodata. 2016.

[9] Douglas Whitaker and Tim Jacobbe. Students' understanding of bar graphs and histograms: Results from the locus assessments. *Journal of Statistics Education*, 25(2):90–102, 2017.