Visualizing Voter Demographics in the 2016 United States Presidential Election

Avinash Sharma (sharmaav@usc.edu)[1], Joshua Haug (jhaug@usc.edu)[1], Henry Jiao (henrygji@usc.edu)[1], and Griffin Weinhold (gweinhol@usc.edu)[1]

University of Southern California, Los Angeles CA 90007, USA

Abstract. Over the past few decades, the United States has become a divided nation. The 2016 United States Presidential Election demonstrated the partisan polarization of U.S. voters. Regional factors like employment level, education level, racial demographics, and income level play an important role in a determining a region's voting behavior. For our project, we will be visualizing these factors in relation to voting behavior in the 2016 election.

Keywords: Election · Visualization · D3 · Voting · Demographics

1 Introduction

The socioeconomic, cultural and geographic factors that dictate an American voter's affiliation has long been a topic of interest. How someone votes is not a random act at all, but rather a result of confounding factors such as wealth, education and location. Therefore, we will look deeper in an attempt to visualize and convey this paradigm further. With the 2020 election finishing up, having a good understanding of how Donald Trump beat Hillary Clinton in the 2016 election could provide useful insight to how the 2020 election, as well as future elections, could turn out. Our visualizations are addressed to anyone who is interested in learning about how regional demographics are correlated with voting behavior.

Many others have created visualizations on the same topic. However, we made our work original by showing specific correlations and adding interactivity not seen in other charts. Our visualizations focus on the demographics of unemployment, education level, median household income, and geographic region in relation to voting behavior in the 2016 election. We have attempted to show these correlations using geographical, area, and linear encodings we have not seen before as well as improve upon common electoral visualizations by increasing the functionality, lightness, and originality as seen on Alberto Cairo's visualization wheel.

As an outline, in section 2, we present related visualizations on the same topic. In section 3, we present the datasets we used for our visualizations. In section 4, we explain how we designed the dashboard and the process we used to build the system. Finally, in section 5, we showcase the main aspects of the website we have built.

2 Related Work

Many researchers and organizations have already created visualizations on this topic. The New York Times created an interactive map that shows voting results by region, size of lead, and change from 2012. [2] They created stacked bar charts to show the differences in voting trends of these demographics and scatter plots over time of the share of Republican votes and income per person.

In addition, Nate Silver of FiveThirtyEight created predictions of the 2016 election based on a machine learning model trained on almost all state and national polls. [9] He created several interactive visualizations to show the expected chance of winning each state in a map, the expected margin of victory in each state in a stacked bar chart, and a novel chart of the winding path to 270 electoral votes necessary to win the election. [1] [3]

Our work takes prior visualizations as inspiration and improves upon them by combining hierarchical datasets into a single visualization as well as making them not only more detailed, but also easier to understand.

3 Data

The main dataset we used was the U.S. Elections dataset from Kaggle, which includes the voting results by county, party, and candidate choice with descriptive statistics of many socioeconomic factors, including unemployment, education level, and median household income. [11] We also used county-level voting statistics from the MIT Election Data and Science Lab. [6] In order to retrieve more accurate data for median household income in 2016, we parsed each State government's website. We also obtained state-level results from Townhall to use in our zoomable circle packing layout. [4]

In terms of pre-processing, we used Python 3 and the Pandas package in Jupyter notebooks to organize the data, aggregate county-level voting, and demographic statistics into state-level, and create csv files for d3. We created Altair graphs for brainstorming as well as quickly displaying correlations of interest.

4 Approach

Our design process began with storyboarding our web application in Figma before coding it in React to ensure the user experience was optimal. We decided to keep the layout simple as we wanted to tell our story in a sequential order with our visualizations and accompanying text. The rest of our process was decidedly data-driven; we determined correlations between demographic variables and voting behavior before designing our visualizations to avoid cluttering the graphics with irrelevant factors.

In our research of other works done on this topic, we found several correlations that were visualized, such as party preference versus education level. According to Pew Research, "people with a four-year college degree or more education made up 37% of all validated voters". Among these voters, 57% said they voted for

Clinton. While among those who had not completed college, Trump shifted the distribution in his favor (50% to 43%). [5] We decided to demonstrate this in our d3 choropleth map to show how it directly relates to the election results by county.

Moreover, The New York Times created a scatterplot showing how "Republicans became the affiliated party of the poorer areas over the past few decades." [8] However, this scatterplot only showed the trend lines of the data and failed to explain if their bubbles represented states or counties. So, we represented the bubbles in our scatterplot by state and county, added a color scale as well as a tooltip, along with clicking interactivity to create more functionality and originality on a state and county level. Figure 1, shown below, shows our scatterplot with incomes broken down by state as well as counties of the chosen state.

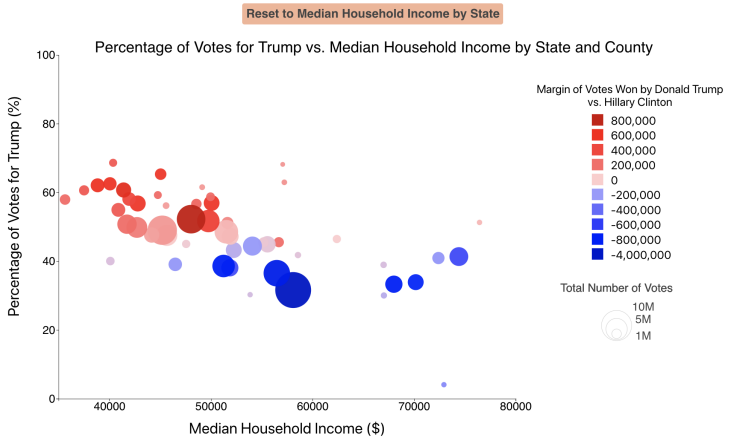

Fig. 1. Our scatterplot compares vote outcomes per state with the median household income of each state. Bubbles can be clicked to reveal all counties within that state as separate bubbles.

As Edward Tufte said, "Confusion and clutter are failures of design, not attributes of information". [10] We utilized Tufte's Design Principle to create elegance in our visuals by matching the complexity of the 2016 U.S. Election data to the simplicity of our design. In all of our visualizations, we focused on making sure we showed the data first and foremost, while maximizing the data-ink ratio. This fits into our use of Alberto Cairo's visualization wheel in Figure 2 by minimizing our use of unnecessary decoration. We used these design principles along with Alberto Cairo's recommendations to maintain the audience's understanding of the data along with a Boom effect "with the use of more original, yet familiar forms to attract the reader". [7] We also made sure to use colorblind-safe colors in our charts.

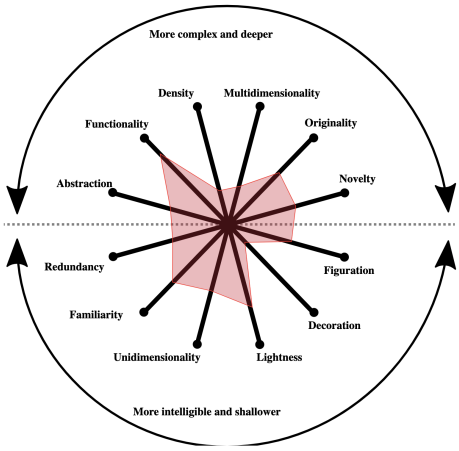

Fig. 2. As we brainstormed our charts, we created a visualization wheel for each of them. This is the wheel we used for not only our scatterplot chart, but also our circle packing layout chart. We focused on increasing lightness, functionality, and originality.

5 System

Our system consists of several visualizations, generated using D3 and Mapbox, encapsulated in a React app using the Create React App framework.

Our web application includes five visualizations to tell our story. Our first visualization shows the magnitude of the polarizing effects of the 2016 election in a line chart, with voter turnout in 2020 increasing by 7.2% from 2016, making it the largest voter turnout in history. In our second visualization, we show the impact economic status had on current voting trends. According to the New York Times, "By 2016, the Republican Party won almost twice the share of votes in the nation's poorest counties than it won in the richest." [8] Our scatterplot corroborates this correlation and shows the evenly divided voting distribution between states and counties in a preattentive manner. Next, we wanted to showcase some more relationships between 2016 voting trends and other variables in a dynamic d3 choropleth map, which shows both voting outcomes as well as county-level demographics. Furthermore, we wanted to analyze the impact geographical regions and subregions have on voting trends in a hierarchical circle packing layout by region of the U.S., state, and county results. Clicking on each state shows a comparison of the number of counties won by Trump versus Clinton. Naturally, there are more red counties than blue counties for most states as most Democrat-leaning cities have high population densities. Finally, we wanted to visualize the counties that flipped from Democratic to Republican in 2016 compared to 2012 to highlight how Trump took away the presidency by focusing on certain areas, such as the Rust Belt. Light salmon counties are ones that flipped from blue to red while blue and red counties are counties that stayed Democrat and Republican, respectively, or flipped red to blue. Our visu-

alizations tell the story from beginning to end of the 2016 election, the effects it has had on the 2020 election, as well as on different regions and socioeconomic groups in the US.

6 Conclusion

In our research we confirmed our hypothesis and assumptions of unemployment, education level, median household income, and region in relation to voting behavior in the 2016 election, while also discovering surprising trends within the data. Furthermore, by effectively visualizing the data we are able to see the troubling paradigm we as Americans face today in our political climate. Someone's wealth, education, and location not only impact their day to day life, but also strongly dictate what party they support.

In terms of future work, we considered breaking down voter demographics in Los Angeles by neighborhood to find interesting local voting trends. We also considered analyzing how the 2020 presidential election results would have changed if a split electoral college system was used, but the Georgia recount had not finished by the time we finished planning. Regarding the website, we could improve the overall design by connecting our d3 map to our other visualizations by state or county level depending on the user's selection of a particular state versus a particular county.

References

1. 270 to Win - Election Map. https://www.270towin.com/maps/2016-actual-electoral-map (2016)
2. New York Times - Election Results Map. https://www.nytimes.com/elections/2016/results/president (2016)
3. Politico - 2016 Election Results. https://www.politico.com/2016-election/results/map/president (2016)
4. Townhall - 2016 Election Results and Analysis. https://townhall.com/election/2016/president (2016)
5. Pew Research Center - An Examination of the 2016 Electorate Based on Validated Voters. https://www.pewresearch.org/politics/2018/08/09/an-examination-of-the-2016-electorate-based-on-validated-voters/ (2018)
6. MIT Election Lab Dataset. https://electionlab.mit.edu/data (2020)
7. Cairo, A.: The Functional Art: An Introduction to Information Graphics and Visualization. Voices That Matter, Pearson Education (2012), https://books.google.com/books?id=xwjhh6Wu-VUC, p. 86
8. Porter, E.: How the G.O.P. Became the Party of the Left Behind. https://www.nytimes.com/interactive/2020/01/27/business/economy/republican-party-voters-income.html (2020)
9. Silver, N.: FiveThirtyEight - 2016 Election Forecast. https://projects.fivethirtyeight.com/2016-election-forecast (2016)
10. Tufte, E.R.: Envisioning Information. Graphics Press (1990), p. 52
11. Wilson, J.: Kaggle 2016 Presidential Election Dataset. https://www.kaggle.com/joelwilson/2012-2016-presidential-elections (2018)