Analysis of the 2020 U.S. Democratic Primary

Elise Bergmann Talia Colalancia Savannah Bornstein Joshua Paup Conner Sinjem

ABSTRACT

Our motivating problem that we are trying to solve, or at least attempt to guess, is uncertainty in who will be elected as the primary democratic candidate in 2020. With so many perceptions in the news, on social media, and even in local communities, it can be very difficult to maintain a broad eye's view of what is truly happening politically in other states and especially nationwide. We attempted to solve this problem through three different visualizations. Our first visualization is a map that represents which candidate of the top 4 polling candidates is leading a specific state, and the color of the state changes based on the leading candidate. Our second visualization displays which democratic candidate a specific state is most leaning towards; once hovering over the state, a list of the top 4 polling candidates is revealed. This visualization provides democratic polling data based on a state-to-state basis. Our third visualization is a line graph that represents the top four candidates polling lead averages based on a monthly basis. In summary, we discovered that Joe Biden is leading the Democratic polls, which Elizabeth Warren, Bernie Sanders, and Pete Buttigieg are following.

1 Introduction

Our topic is interesting as our visualizations provide both a state and a national perspective of the 2020 democratic primary. To be able to view both of these perspectives in one location is very helpful to citizens of the United States as it provides more insight into the bigger picture of the whole nation and which states are more highly influencing the national result. The motivating problem for our project is incredibly important as it encourages citizens to be more knowledgeable on an upcoming political election. Since politics can often be overwhelming and something that many individuals avoid to be involved with, our visualizations provide basic information that is easy to understand. The visualizations allow people to be informed of important political information even if they do not want to be extremely involved.

2 Related Work

2.2 Conceptual Citations

This set of citations were useful in providing assistance with conceptual ideas, whether this be for creative types of visualizations, or interesting geographical representations. While these sources did not help with the actual code, they were the first steps in providing overall ideas of what types of visualizations we wanted to create as they provided examples of representations that are often used to showcase political data.

Citation: https://projects.fivethirtyeight.com/2020-primaries/democratic/

This website was used to get ideas of how the dataset we are using has already been used; By referencing these visualizations, we were able to see what types of graphs have already been done so we can create new and innovative representations while avoiding some of the visualizations that may have been executed more poorly/inaccurately.

Citation: https://www.270towin.com/polls/2020-presidential-election/

This source provided visualizations that are based off of the red republican and blue democrat representation of politics. By looking at these visuals, we were able to determine if this was a route we wanted to go in terms of choosing what colors to use to represent our data.

Citation: https://calmatters.org/blogs/california-election-2020/2019/10/california-donor-map-2020-presidential-election/

This reference provided many different types of visualizations, some geographical and some not. By viewing different types of geographical data specifically, this source actually provided us with examples of how not to geographically represent data; Many of the images were overly cluttered and too colorful so this inspired us to have our visualizations still be geographical, but not have too much information represented or too many colors.

Citation: https://www.politico.com/2020-election/democratic-presidential-candidates/polls/

2.3 Technical Citations

This set of citations were useful in providing assistance to the technical aspects of the project such as the actual coding for the project and how to create specific visualizations. These sources allowed us to learn/integrate specific aspects of a certain code that we may have not learned or covered in class.

Citation: https://plot.ly/python/choropleth-maps/

This website was used to see how we can visualize our dataset using Python and the plotly library. We wanted to learn how to do this in order to have more control of how we display the visualizations. As we continued to work further into the project, we realized that we are running into multi-attribute dimension errors with plotting multiple presidential candidates in each state, depending on who was leading in each state.

Citation: https://www.softwareadvice.com/resources/excel-data-cleaning-integration-techniques/

This source allowed us to learn how to data clean in Excel as well as proper techniques of data cleaning such as : correct filtering, sorting, pivot tables, etc. We wanted to clean the data so \want to use for our specific visualizations.

Citation: http://vizdiff.blogspot.com/2016/08/embedding-tableau-public-in-web-page.html and https://towardsdatascience.com/how-to-create-a-plotly-visualization-and-embed-it-on-websites-517c1a78568b

2.3 Political Citations

This set of citations were useful in providing assistance with the actual political data that was needed for the project in order to create the visualizations. In addition, these sources were also useful in understanding our topic as a whole and how democratic political data is gathered.

Citation: https://github.com/joshuapaup/democratic-primaries-2020-map/blob/master/dem-primaries_v1.0.csv

This was the first CSV data file that we analyzed. This data file had all the raw data needed to analyze the National Democratic Polls. In this file, the state, the pollster id, the sponsors and their ids, sample sizes, population, party, and the leading candidate were included, along with a plethora of other data. We used this as an initial branch for all three individual visualizations. We used the data in this mother-file to create a smaller file, which is also our next source.

Citation: https://github.com/joshuapaup/democratic-primaries-2020-map/blob/master/states_clean.csv

This file is a cleaned up version of the file above. We went through the state and whenever there was a blank entry we placed that into its own national CSV file, and if it had an entry we put that into this state CSV file.

3 Detailed Description

Our project includes the following elements: a website, three visualizations, and this written report. Beginning with the website, we created this using html and CSS through sublime and it consists of the following sections: the three visualizations, three visualizations, and a description of each visualization. In terms of the three visualizations, the first visualization is a geographical map that shows which candidate of the following four candidates: Joe Biden, Elizabeth Warren, Bernie Sanders, and Pete Buttigieg, is leading in the polls for that specific state.
The second visualization is a graph that outlines each state on a United States map. The purpose of this map is to showcase the top four candidates in the democratic primary, again based on the following candidates: Joe Biden, Elizabeth Warren, Bernie Sanders, and Pete Buttigieg. The color scheme consists of the darker the color the stronger the lead a specific candidate is holding in a given state. The third visualization is a line graph that demonstrates the average polling lead based on different months for the four following democratic politicians: Joe Biden, Elizabeth Warren, Bernie Sanders, and Pete Buttigieg, (so the same as the four for the first and second visualization).

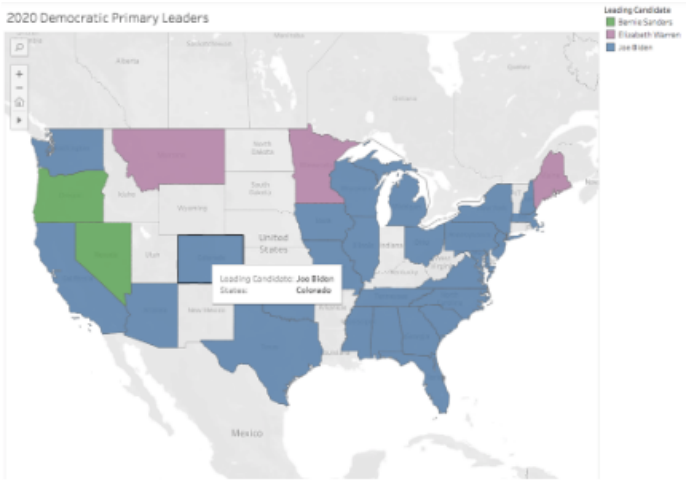
Some of our primary design elements included incorporating geographical representations as well as a visualization that shows time. In our project, we chose to represent the data geographical via a United States map as this is the easiest way to demonstrate our findings to the end user who can hover over a specific state rather than looking at a table/other graph that would be harder to interpret. Since two of our visualizations are geographical, we wanted to display a different type of visualization as well; We chose to use a line graph as in political polling in general, a candidate's ranking can change at any moment. By choosing to create a line graph, this design can represent how a candidate's ranking has changed based on a specific point in time and this reveals to viewers of the graph different trends and levels of consistency. In addition to the types of visualizations we chose to make, an important design element used in our images is the use of color. In one of our graphs, different colors are used to represent different candidates which is very intuitive to the user. The most interesting aspect of the second visualization is that you can specifically see the amount of displacement between the first and second candidate based off of a color distribution, (the darker the color, the stronger the lead and larger disparity). This design was chosen to quickly allow viewers of the visualization to be able to interpret which leads in which states are the strongest.

4 Discussion

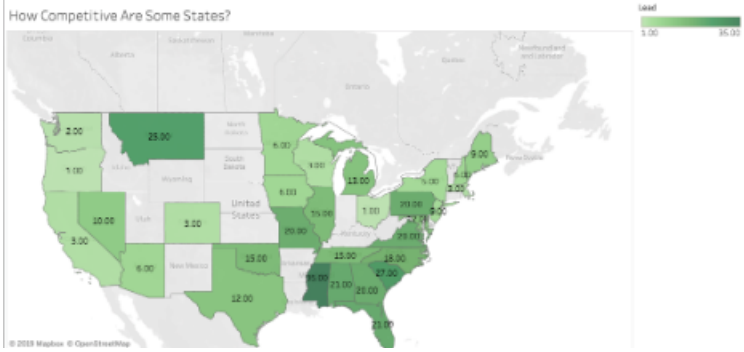
Upon analyzing these three visualizations, on the top four Democratic Candidates, we were able to deduce a few things. First and foremost, we noticed that Joe Biden is generally winning most states in the US, in particular he is dominating the South. Meanwhile, we noticed that Elizabeth Warren is winning a few states in the North and Bernie Sanders is winning a few states in the West. We then noticed that, although Joe Biden has very strong leads in the South, his leads in states, such as Iowa and New Hampshire, were not as strong in these early voting states, meaning that these states could be up for grabs among any of these Democratic Candidates.

5 Figures

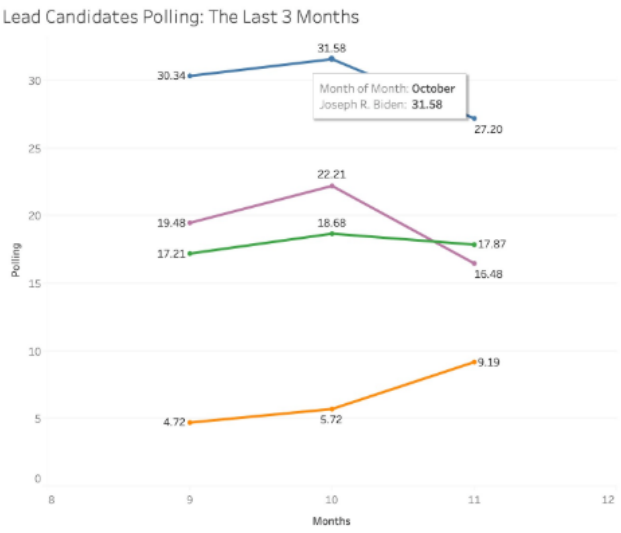
Visualization 1



Visualization 2



Visualization 3



References

Bycoffe, Aaron, et al. "The 2020 Democratic Presidential Primary." *FiveThirtyEight*, 13 Dec. 2019, https://projects.fivethirtyeight.com/2020-primaries/democratic/.

"2020 Presidential Election Polls." *270toWin.Com*, https://www.270towin.com/polls/2020-presidential-election/.

Christopher, Ben. "7 Data Visualizations: Which Presidential Candidates Are Californians Funding - Month by Month, Zip by Zip?" *CalMatters*, 19 Oct. 2019, https://calmatters.org/blogs/california-election-2020/2019/10/california-donor-map-2020-presidential-election/.

"Democratic Primary Polls: Who's Ahead in the 2020 Race?" *POLITICO*, https://www.politico.com/2020-election/democratic-presidential-candidates/polls/."Democratic Primary Polls: Who's Ahead in the 2020 Race?" *POLITICO*, https://www.politico.com/2020-election/democratic-presidential-candidates/polls/.

https://plot.ly/python/choropleth-maps/

"Techniques for Data Cleaning and Integration in Excel." *Software Advice*, 30 Mar. 2017, https://www.softwareadvice.com/resources/excel-data-cleaning-integration-techniques/.

"Embedding Tableau Public in a Web Page." *Vizible Difference*, http://vizdiff.blogspot.com/2016/08/embedding-tableau-public-in-web-page.html.

https://github.com/joshuapaup/democratic-primaries-2020-map/blob/master/dem-primaries_v1.0.csv

https://github.com/joshuapaup/democratic-primaries-2020-map/blob/master/states_clean.csv