

"2020 Election EDA Report"  
Juan Alfaro  
11/19/2020  
MSDS 430 Python for Data Science

## Introduction

A good quote to mention in regard to election polling data used this cycle was once famously said in “Duck Soup,” the 1933 Marx Brothers comedy, “Who ya gonna believe, me or your own eyes” (1). Election cycle data has been scrutinized as misleading as far back as 2016 despite the confidence of pollsters this year. In 2020, this election cycle has provided interesting data to project the outcomes for the U.S. Congress, Senate, and the Presidency elections. This was an area of interest was chosen as it would give more context to how election data is used, provide a better sense of what are the key indicators to winning an election, and how to interpret a dataset such as the one used by a respected data site.

The decision to use a data set that contains the final state-level topline of the 2020 election was easily made as this election became a touchstone for many people throughout the country affected by a pandemic and economic recession. With the benefit of hindsight as the election draws to a close, the outcome of this EDA report will help to achieve several goals such as: informing how swing states effected the outcome of the election, which states delivered the decisive vote in the Electoral College based on state turnouts, voter power index and tipping points, and what effect the challenger and incumbent chances of winning certain states had on the election results.

## Data Preparation and Analysis

This year as much as 239.2 million Americans were eligible to vote in 2020 with a projected turnout rate of 65.1%, the highest voter turnout since 1900 (2). The data set used for this EDA project was from FiveThirtyEight, a polling aggregation website that focuses on opinion polls, politics, economics and sports. The data was collected through aggregating polling data conducted by firms and organizations that includes the names of the pollster, survey dates, sample sizes, and details about the population sampled. The population being represented by the data is all registered voters 18 years-of-age or older. This dataset is one of several used in the current model by FiveThirtyEight. The model is simulated 40,000 times and a sample size of 100 outcomes is used in their 2020 election projections according to their website.

Pollsters used in the data included their methodology, including but not limited to the medium through which their polling was conducted (e.g., landline calls, text, etc.), the source of their voting files and their weighting criteria. In regard to weights, weights are based on sample size and pollster rating. The pollster ratings, in turn, reflect a combination of the pollster's past performance and if it meets the needs of the current industry best practices. For instance, since 57% of American adults actually vote in presidential elections, polls of adults have their weights multiplied by 0.57 (3).

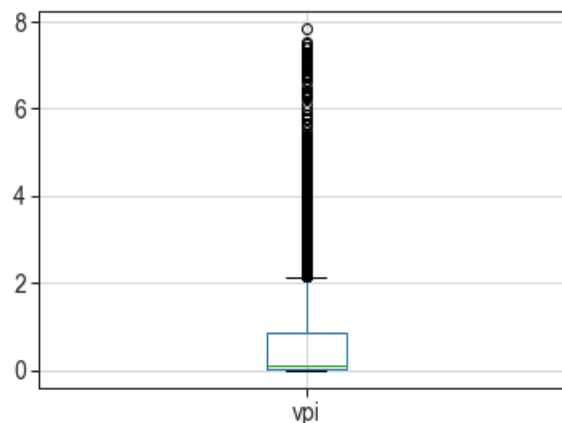
The data set used in this EDA report initially had missing values there were excluded in this report pertaining to third-party candidates as there were none influential to the overall election result to report. Another detail excluded were data values relating to dates and times of FiveThirtyEight's model simulations and election cycle information which was redundant.

## Data Dictionary Table

| Field Name    | Description                                  | Sample Data                        |
|---------------|--|------------------------------------|
| state         | States that hold electoral votes             | “Georgia”, “Texas”, “Pennsylvania” |
| tipping       | Chances a state has to deliver decisive vote | 0.153613, 0.001388, 0.0551134      |
| vpi           | Voter power index                            | 7.166093, 2.796491, 0.333934       |
| winstate_inc  | Chances the incumbent has to win             | 0.137910, 0.746798, 0.418699       |
| winstate_chal | Chances the challenger has to win            | 0.983607, 0.859989, 0.703103       |
| margin        | Forecasted margin for the incumbent          | -15.024, 43.66, 1.607              |
| state_turnout | Forecasted state-level voter turnout         | 209006.5, 2658756.0, 582574.0      |

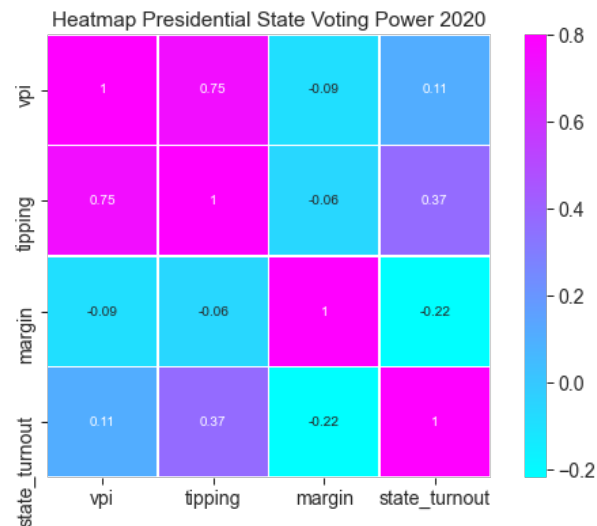
To begin with the most relevant data contained in the heatmap below, the voter power index (VPI) is the relative likelihood that an individual voter in the state will determine the Electoral College winner. For instance, if a state has a VPI of 3.5, this means a vote in that state is 3.5 times more powerful than the national average.

The range of the VPI within the data set used for this EDA report were limited between the values of 0 and 8 as one of the main goals within this report is to determine the influence contested states have on the election. The VPI mean is 0.78 with a minimum of 1 and a maximum of 7.84.



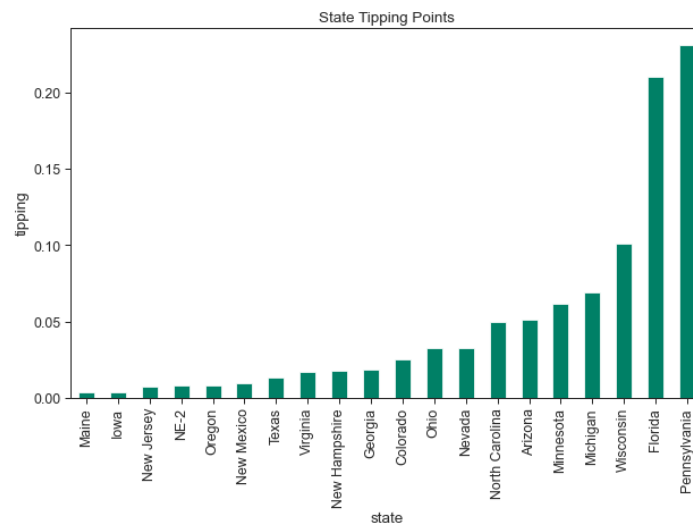
Tipping (tipping-point chance) is the chance the state will deliver the decisive vote in the Electoral College. Also, state turnout values are the forecasted voter turnout based on past

turnout, estimates of population growth, polls about whether voters are more or less enthusiastic about the election than usual and other factors in each state. The margin is the forecasted margin for the incumbent for each state or district.

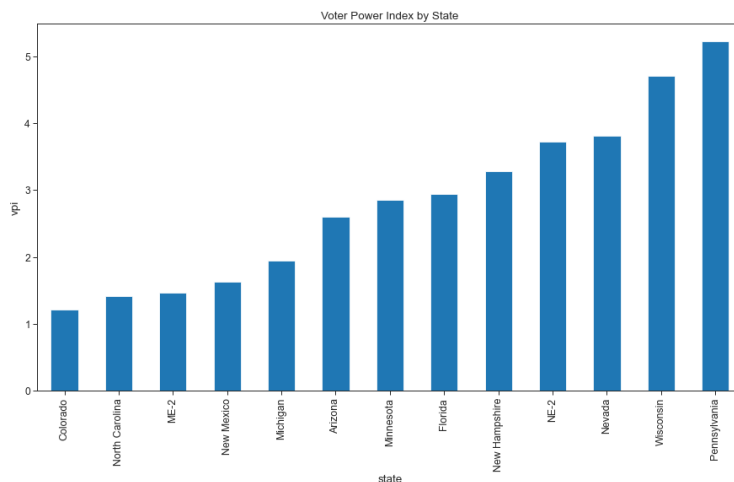


In the heatmap above, the values show the VPI and the tipping point chance playing the a more vital role in the relevance to state voting power when the margin and state turnout are low to moderate. In contrast, state turnout has a real effect in regard to the state voting power when coupled with a higher tipping point chance. However, the margin of victory as an incumbent only matters if turnout, tipping, and VPI are low according to the data.

Further analysis of tipping points per state indicates which states were considered high importance in the 2020 election. According to the data, Pennsylvania had the highest chances of delivering the decisive vote in the Electoral College. Without the benefit of hindsight, the presidential candidates should be persuaded to pursue all avenues to win the state of Pennsylvania as it may be the most crucial state besides Florida, Wisconsin, and Michigan to winning the election based off this alone. However, there are other factors to consider in this report.

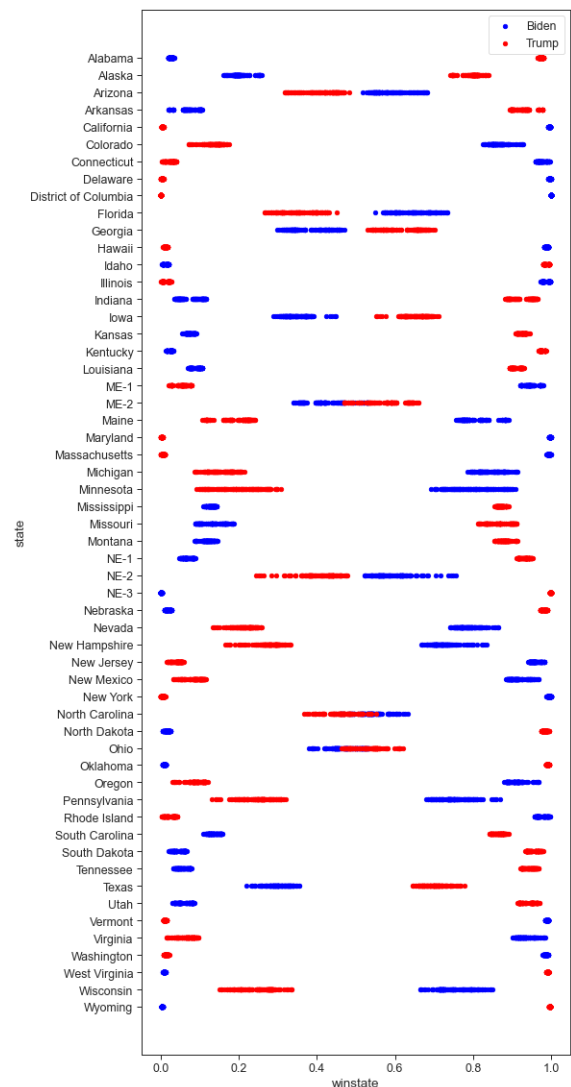


Another factor to observe is how profound the VPI is to the election. If we held a national popular vote for president, everyone's votes would be counted equally. However, this is not the case as some states such as Texas. For instance, Texas is divided into six City Council districts, all equal in terms of population. One of the districts includes a significant immigration population, however, rendering half of its district ineligible to vote. Therefore, the eligible voters left remaining have more political power than those in the other districts (4). Or in the case of Maine and Nebraska, where electoral votes can be split among the candidates depending on the votes in their respective congressional districts.



In the graph above the data shows the VPI of 1x or higher for a state or district. Out of 50 states and 6 congressional districts there were at 15 that had at more than 1x the voter power or more attributed to its voters. The graphs show Pennsylvania, Wisconsin, Nevada, Arizona and Nebraska's 2nd Congressional district all have a high VPI. All these states played a pivotal role in the presidential challenger winning the election despite losing Florida, ME-2, and North Carolina.

The `winstat_inc` and `winstat_chal` refers to the chances either the incumbent or challenger could win a state. Looking at the scatter plot on the right, it seems the challenger had a higher chance of winning key swing states such as Pennsylvania, Michigan, Wisconsin, Florida, but had a lower chance of winning others such as Georgia. As we now know, the incumbent has won Florida despite having a lesser chance of winning the state and in contrast the incumbent lost Georgia and North Carolina despite have a greater chance of winning. This was the most interesting part of the data as it shows that despite polls collected the wins and losses of any candidate is not guaranteed.



## Conclusion

Findings from the data used in this EDA report show several instances in which the tipping, VPI, and state turnout proved to be major indicators in the incumbent and challenger either benefiting or losing electoral votes. In the data, states with larger VPI's showed to be more beneficial for the candidate with a higher chance of winning the state. Swing states were won by their prospective candidate with higher tipping points coupled with high VPI's in most cases with a few exceptions. States that delivered the decisive votes to the Electoral college were ones that the challenger had a higher chance of winning the state such as Pennsylvania. The incumbent managed to win Florida despite the chances of winning being less than the challenger. Nevertheless, with challenger had 29 instances where the chances of winning electoral votes were greater than the incumbent's 27 chances. The challenger managed to flip 5 states from 2016 (Arizona, Wisconsin, Michigan, Pennsylvania, and Georgia).

Going forward, a further detailed analysis would benefit this EDA report as exit data collected after the election could offer more information in regard to mail-in voting vs. ballot casting the day of the election, the amount of money that was spent on advertising in a state with higher tipping points and voter power index. Lastly, the demographics of voters (age, sex, ethnicity, education, location, religion) could offer a more detailed concatenation of the election.

## References

1. <https://www.youtube.com/watch?v=cHxGUe1cjzM>
2. Schaul, Kevin et al. (Nov. 5, 2020). 2020 Turnout is the Highest in Over a Century. *The Washington Post*. Retrieved from:  
<https://www.washingtonpost.com/graphics/2020/elections/voter-turnout>
3. <https://projects.fivethirtyeight.com/2020-election-forecast/>
4. <https://www.wnycstudios.org/podcasts/takeaway/segments/276850-does-one-person-one-vote-mean-one-voter-one-vote>
5. <https://www.politico.com/2020-election/results/president/>