

Debugging the 2020 Election Projections

Problem:

We will examine the disparities between polling and election outcomes in the 2020 general election.

Dataset:

We are in the process of reviewing and collecting a variety of datasets, which are being collected in a large list [here](#).

Datasets are subject to change as the project progresses. Some websites depict data, but that data may not be easily exported as a CSV file, for example. As a result, we may need to compile that data into a datasheet ourselves.

Proposed Analysis and Real World Application :

Polling in both the 2016 and 2020 elections underestimated the proportion of votes Donald Trump ended up receiving. These disparities varied by region and were particularly pronounced in the upper midwest, but he also outperformed expectations among certain demographic groups. We intend to compare the shift between exit polls for the 2020 general election and a variety of pre-election polls, described above. Our analysis will focus on finding correlations between these shifts and the various regions and demographics specified in the exit polling data.

Identifying the subpopulations which are being missed or mischaracterized by existing polling methods is exceedingly important as political parties design their messaging and campaign efforts based on the results of this polling. It follows that the results of our analysis can hopefully be applied to more appropriately modeling the voting population going forward, which would allow for more accurate polling and more well calibrated, well informed political campaigns.

Project steps

Step	Estimated completion time	Person(s) in charge (among the group of 3)
1. Dataset gathering and research	One week	Jacob
2. Coherent data schema design, data cleaning and reorganization.	Two weeks	Fahim
3. Exploratory analytics, and model selection	One week	Greg
4. Run more detailed analysis, carry out any laborious computations,	One week	Srinivas
5. Finalize conclusions and prepare presentation	One week	Zhehui