

**Kaylan Wallace**

**POLI 381**

**Data Project: Description**

**Are elections becoming more competitive or less competitive?**

## Conceptualization

For the first part of the project, I will conceptualize electoral competitiveness as a comparison of election candidates' polling ratings. A competitive election is one where the standard deviation of the mean rating for all candidates is relatively small, given that there are at least two candidates. This will be a continuous variable.

The operationalization of the measure in this project will focus on US Presidential General Elections data, but it can also be generalized to any electoral process for which there is public polling data gathered over a period of time.

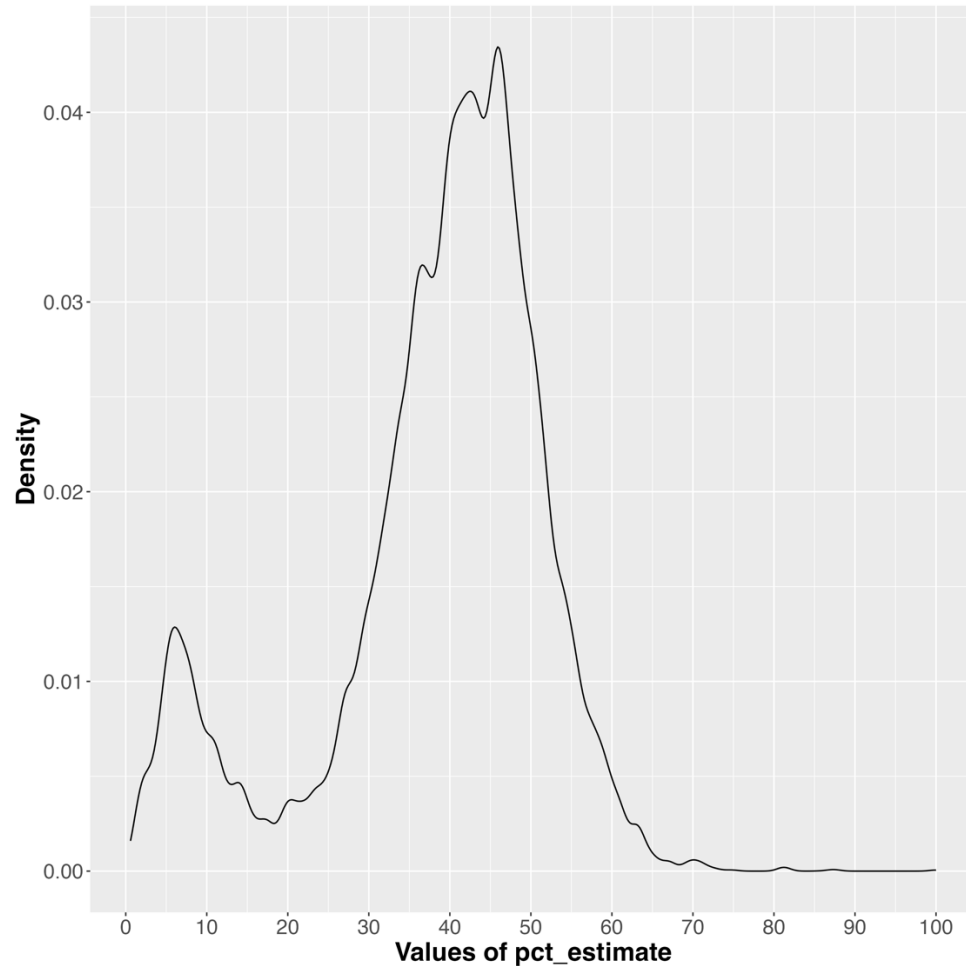
The base data was sourced from 538, a website for opinion poll analysis, politics, economics, and sports in the US. This data set logs polling averages for every Presidential candidate between 1968 and 2016. It shows each candidate's daily polling average in every state leading up to the election date.

To create a new index from this dataset to measure electoral competitiveness:

1. Only data from April 9th onward will be considered, so that each election's data covers an approximately equivalent time frame. This is to ensure a fair comparison.
2. The states will be grouped into 9 divisions, based on classifications from the Census Regions and Divisions of the United States (Bureau). This is to simplify analysis.
3. Each candidate's polling average from April 9th to election day per division will be calculated.
4. Each division will be assigned a weight by summing the assigned number of electoral votes for all states within that division (National Archives).
5. Each candidate's national polling rating will be computed as a weighted average of their division rating. The influence of each division's rating on the national rating is determined by the previously assigned weights.
6. Summary statistics for each election's candidate ratings will be calculated.
7. For a given election, a comparison of the standard deviation of candidates' national polling rating will provide information about that election's competitiveness.

## Quality Control

The theoretical range of values for the Quantity of Interest (QOI) is 0 to 100, seeing as it is a percentage value. Before any processing to create the index, the distribution of the data is bimodal, as seen in **Figure 1**. The local maxima occur around approximately 5% and 45%. The distribution is left, or positively, skewed, with very few values after 70 on the right tail. This is affirmed with a median (41) that is higher than the mean (38). As such, the median might be a better measure of central tendency than the mean for this data.



**Figure 1:** Distribution of the variable of interest, pct\_estimate.

## Description

While the average weighted rating fluctuated between 40% in 1972 and 30.5% in 2016, a 23% decrease, the most significant changes occurred between 2004 and 2012, corresponding to the controversial Bush and Obama administrations. This period saw heightened political engagement, potentially reflected in the increased voting turnout and wider public discourse.

Election Year	Mean	Standard Deviation
1972	40.08	18.75
1976	34.14	1.79
1980	29.38	14.38
1984	44.99	12.69
1988	43.98	0.16
1992	31.65	7.31
1996	30.42	19.37
2000	29.65	22.4
2004	45.34	3.05
2008	45.31	0.27
2012	45.62	0.08
2016	30.52	19.99

**Table 1:** Summary statistics for candidates' weighted polling rankings per election.

**Table 1** suggests a non-competitive 2000 election due to a high standard deviation but analyzing the actual data (see Appendix A) reveals a close race between the two main candidates. A third-party candidate's low rating heavily influenced the standard deviation. Conversely, 2012's minimal standard deviation reflects the tight race between the two frontrunners, highlighting its competitiveness. These comparisons underscore the importance of considering both the standard deviations and the actual data to avoid misinterpretations based solely on **Table 1** figures.

Like the raw data, this distribution exhibits left skewness (median higher than mean), indicating more presidents with lower ratings. Additionally, the kurtosis of 3.52 suggests a "peaked" distribution with more data points clustered around the mean compared to a normal bell curve. This, in combination with a high variance indicates potential limitations with using traditional statistical methods (designed for normally distributed data) to estimate population parameters for the average rating.

## References

Bureau, U. C. (2021, October 8). *Geographic Levels*. Census.gov. <https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html>.

National Archives and Records Administration. (n.d.). *Distribution of electoral votes*. National Archives and Records Administration. <https://www.archives.gov/electoral-college/allocation>

## Appendix A

Table showing candidate's weighted polling rating for every US Presidential General Election.

Candidate Name	Election Year	Weighted Polling Score
George S. McGovern	1972	26.83
Richard M. Nixon	1972	53.34
Gerald R. Ford	1976	32.88
Jimmy Carter	1976	35.41
Jimmy Carter	1980	35.7
John B. Anderson	1980	12.93
Ronald Reagan	1980	39.52
Ronald Reagan	1984	53.96
Walter F. Mondale	1984	36.02
George Bush	1988	44.09
Michael S. Dukakis	1988	43.87
Bill Clinton	1992	34.15
George Bush	1992	37.38
H. Ross Perot	1992	23.42
Bill Clinton	1996	45.56
Bob Dole	1996	37.1
H. Ross Perot	1996	8.59
Al Gore	2000	39.79
George W. Bush	2000	45.18
Ralph Nader	2000	3.97
George W. Bush	2004	47.5
John Kerry	2004	43.18
Barack Obama	2008	45.12
John McCain	2008	45.5
Barack Obama	2012	45.67
Mitt Romney	2012	45.56
Donald Trump	2016	41.5
Gary Johnson	2016	7.45
Hillary Rodham Clinton	2016	42.61