Kaylan Wallace

POLI 381

Data Project: Correlation

Are elections becoming more competitive or less competitive?

The Independent Variable X (Theoretical Conjecture)

In the US, electoral turnout is often theorized to be positively related to electoral competitiveness for several reasons. In highly competitive elections where the outcome is uncertain, political parties and candidates are motivated to mobilize their supporters and increase voter turnout to secure victory. This can lead to greater efforts in voter registration and canvassing. These activities tend to generate more media attention, public discourse, and political debates, which may increase voter engagement and motivation to participate in the electoral process. In closely contested races especially, individuals may perceive their votes as having a greater impact on the outcome.

However, it's important to consider alternative hypotheses and potential pitfalls. While there may be a correlation between electoral competitiveness and turnout, it's essential to recognize that correlation does not imply causation. Other unobserved factors, such as demographic characteristics, political culture, or historical trends, could drive both variables independently. In such case, these variables are both driven by something in the model that is not accounted for (endogeneity). We can also consider a non-linear relationship between electoral competitiveness and turnout. For example, increased competitiveness may initially lead to higher turnout rates, but beyond a certain threshold, further increases in competitiveness may have diminishing returns on turnout as voters may feel overwhelmed or disengaged by highly contentious campaigns. This can take place within the same electoral season, or may even happen over the course of years, which indicates the potential for both variables to have a temporal lag effect on each other.

Operationalization of X

This project will explore the correlation between electoral competitiveness and turnout. The data for electoral turnout was sourced from University of Florida's Election Lab (McDonald), which defines electoral turnout as the proportion of the voting eligible population (VEP) that actually votes in an election. They further define VEP as an estimate of those eligible to vote in an election, calculated by the voting-age population minus ineligible non-citizens and felons (McDonald). As there is no national central agency that provides electoral statistics, a University publication was one of the next best credible sources to turn to. This dataset contains turnout rates for both presidential and midterm elections from 1789 - 2020. To measure electoral turnout from this data, only data for presidential election turnouts will be used, as the polling average data is also limited to presidential elections. Additionally, only data from 1972-2016 will be used, as this is the time span measured in the polling average data.

One shortfall of this operationalization of electoral turnout is that the data are national aggregates, not data for each state. There was no information found on how each state's turnout contributed to the national turnout average. This may undermine the procedure taken in operationalizing electoral competitiveness to account for the differences in population and assigned electoral votes related to each state. For electoral competitiveness, daily candidate poll ratings per state were aggregated from April 9th to election day, as this was the earliest data collection date all election years had in common. The states were then grouped into 9 geographic divisions based on Census Regions (Bureau), and each region was given a weighting based on the number of electoral votes of all its comprising states (National Archives). National polling ratings per election were calculated as weighted averages of divisional ratings, and the standard deviation of these ratings served as a measure of election competitiveness.

Quality Control

The theoretical range of values for electoral turnout is 0 to 1, as it is a proportion represented in decimals. All the data is within the range 0.51 – 0.62, which is not surprising given that the US is known to have a lower voter turnout than many democracies (FairVote). **Figure 1** is a density plot, which shows the distribution of the data to be bimodal. Though the absolute values of the mean and mode are very close together, in context of the actual range of values for turnout the mean is notably higher. This suggests concentrations of voter turnout around 54.5% and 60%. Seeing as little methodology was given for the sources of data used to calculate turnout, this may also be indicative of measurement errors.

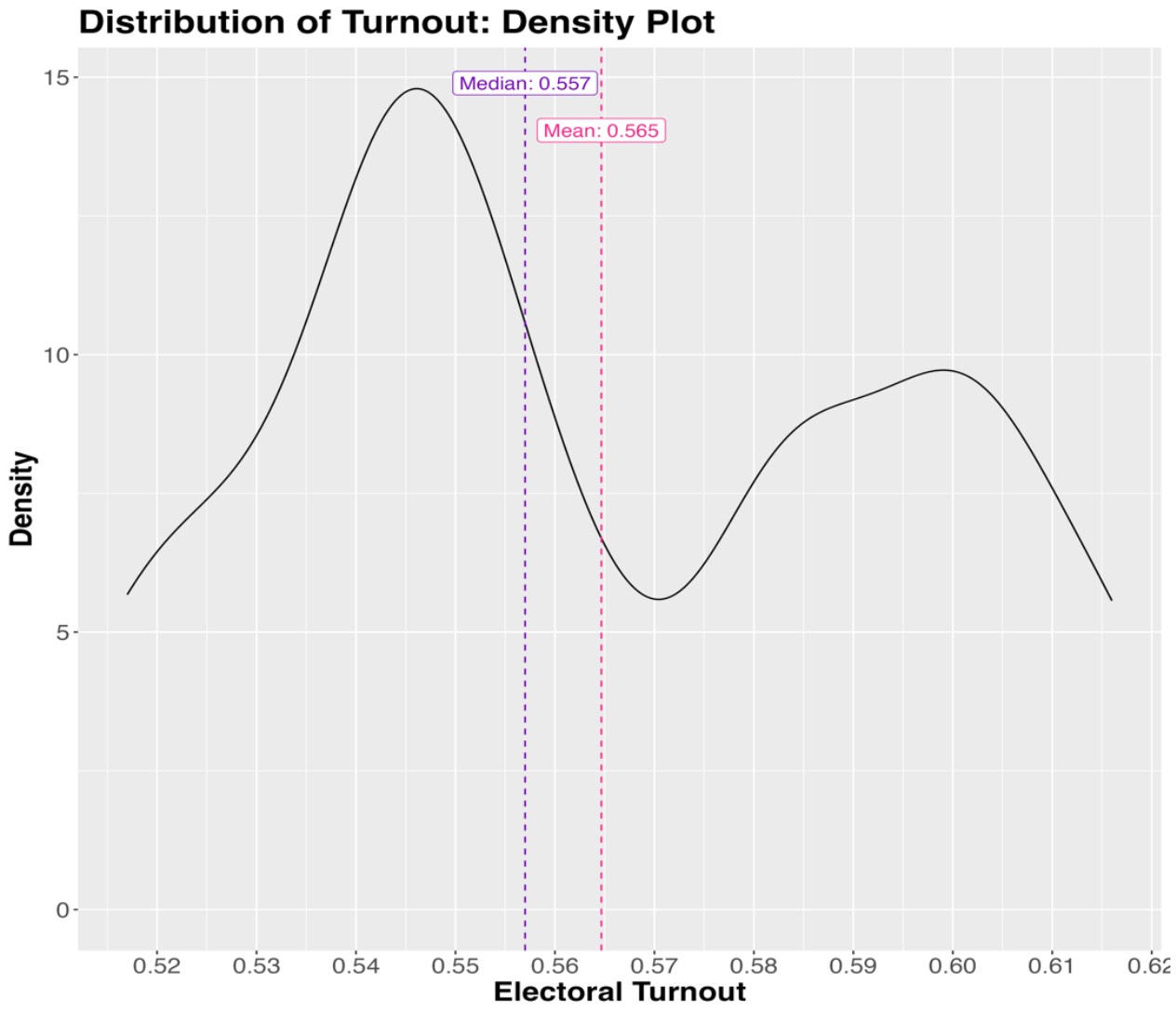

Figure 1: Density Plot for Electoral Turnout

Source: National Voting Eligible Population Turnout Rates, 1789-Present (v1.1).

Figure 2 is a QQ plot, which assesses how well the observed distribution matches the theoretical normal distribution. The points fall close to the pink QQ line that passes through the quantiles of the theoretical normal distribution, suggesting the data is approximately normally distributed. This is a slightly different conclusion than the density plot, which is likely because the QQ plot assumes a unimodal distribution and may have difficulties representing the bimodality observed in the density plot. We should note that the sample size is quite small (12 observations, one for each election year between 1972-2016), which can cause greater variability in both plots. Because of this, even the smallest measurement errors can have a large impact on the conclusions we derive from the data.

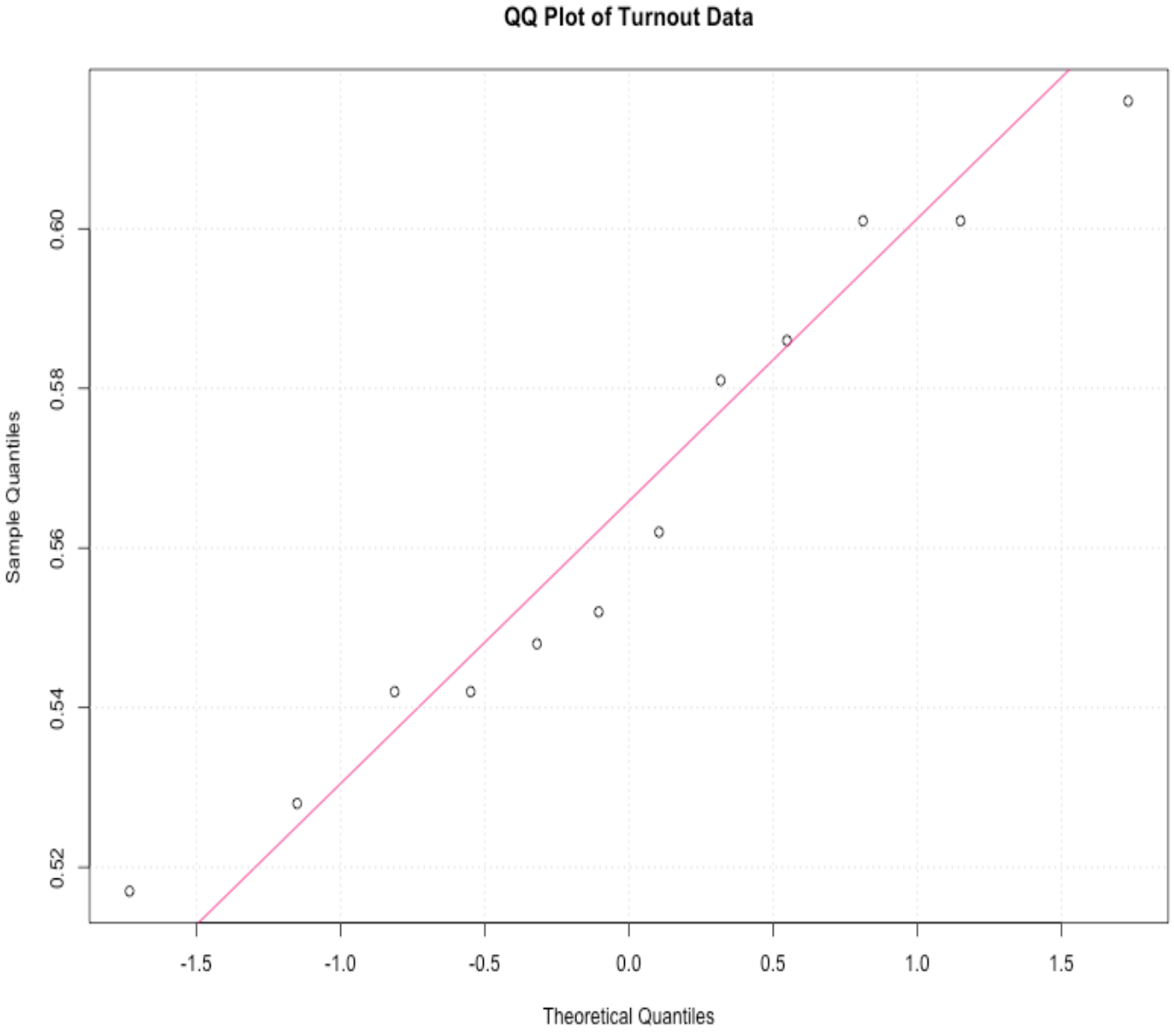

Figure 2: QQ Plot for Electoral Turnout

Source: National Voting Eligible Population Turnout Rates, 1789-Present (v1.1).

The Relationship between Electoral Turnout and Electoral Competitiveness

Figure 3 is time series plot demonstrating the temporal relationship between the variables. Electoral turnout fluctuates much more drastically than competitiveness and takes on a larger range of values as well. Before 1992, both trend lines move in generally the same direction, but after 1992 it appears that they have an inverse relationship. There is a notable spike in turnout and drop in competitiveness between 2000 and 2004, likely due to the aftermath of 9/11. The change in relationship between the two variables following 9/11 may be due to increased cooperation and agreement on policies focused on internal affairs and sense of urgency amongst citizens to become politically involved.

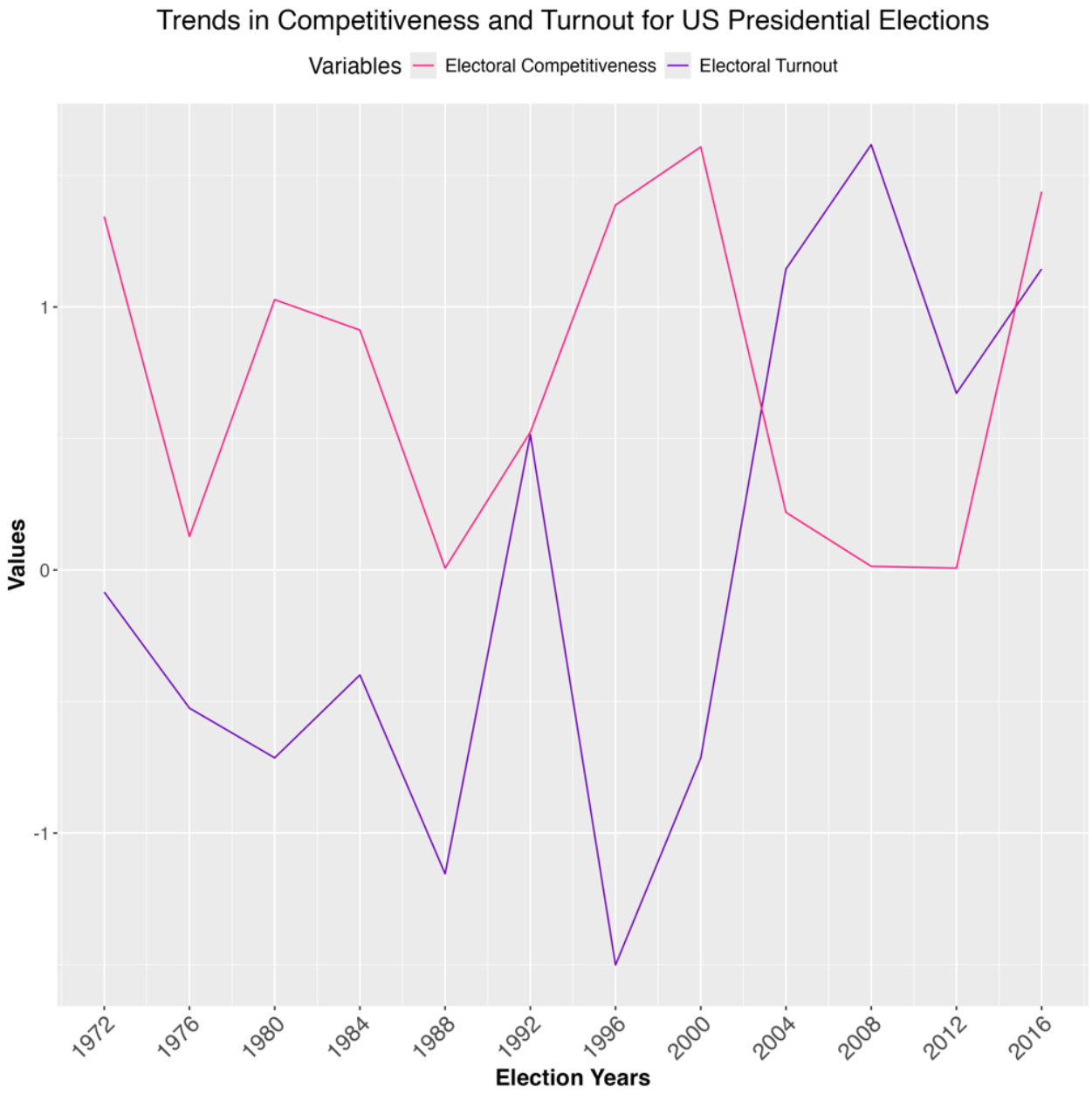

Figure 3: Time Series Plot for Electoral Competitiveness and Turnout

Source: National Voting Eligible Population Turnout Rates, 1789-Present (v1.1).

Figure 4 shows a scatterplot of electoral competitiveness and electoral turnout, with a LOESS curve fitted. A span of 0.75 was chosen because of the limited size of the dataset. Since fewer data points typically runs the risk of greater variability, a higher span ensured that there was a more stable estimate of the underlying trend. The graph displays a negative relationship between the two variables, affirming the conclusion derived from **Figure 3**.

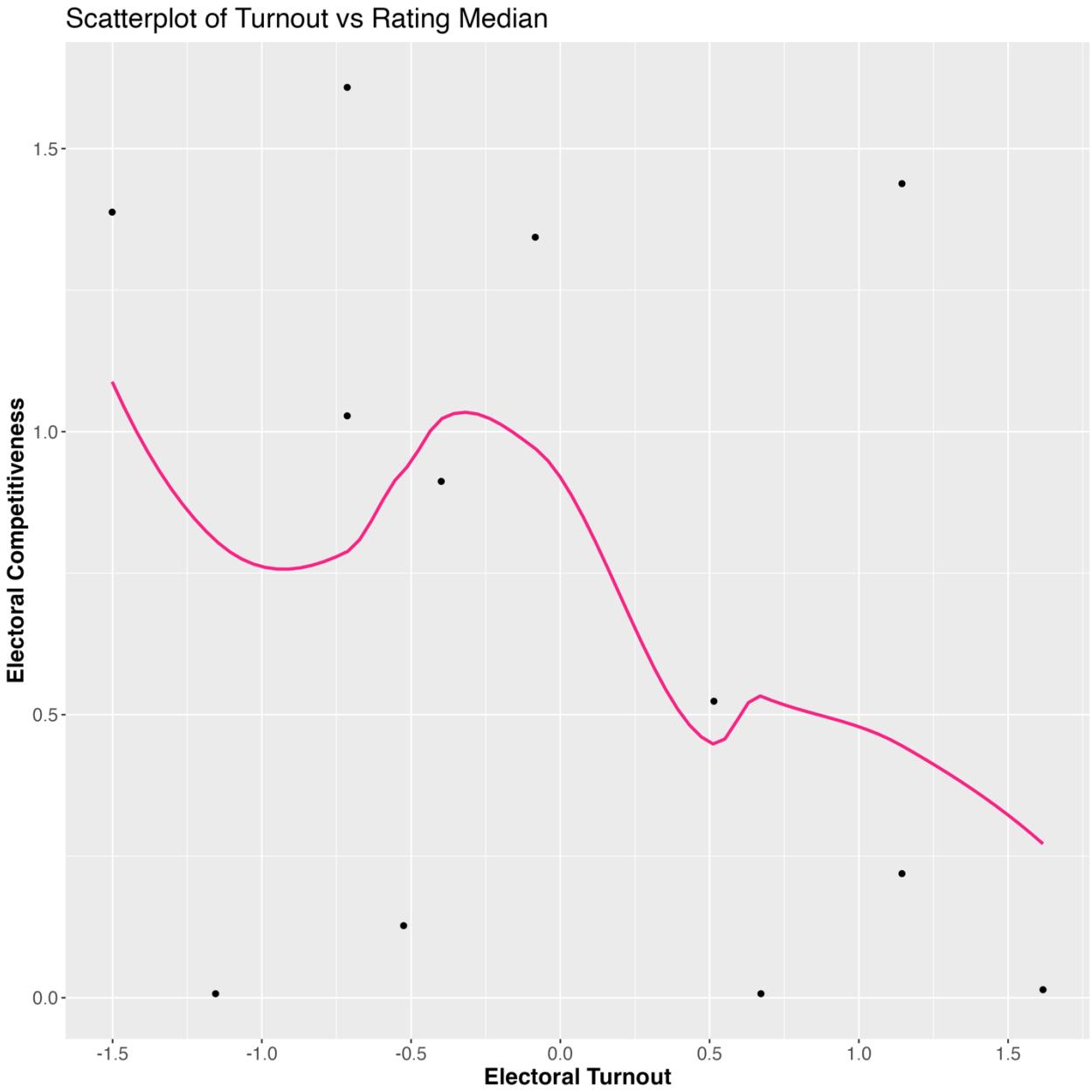

Figure 4: Scatterplot with fitted loess curve.

Span = 0.75

Source: National Voting Eligible Population Turnout Rates, 1789-Present (v1.1).

References

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

Michael McDonald. 2023. "National Voting Eligible Population Turnout Rates, 1789-Present (v1.1)." https://election.lab.ufl.edu/dataset/national-vep-turnout-rates-1789-present-v1-1/

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

Voter turnout. FairVote. (2023, June 16). https://fairvote.org/resources/voter-turnout/#voter-turnout-by-state