**Mid-Term Project Proposal**

Bruce Mallory

*“Bruce, life’s not fair – get over it.” -Bruce’s dad*

1. Personal Statement

My job goal is to work for social change using my limited data analysis skills. A dream job for me would be to work for the Brennan Center for Justice, which “focuses on the fundamental issues of democracy and justice.”

1. My Question

Chart, scatter chart

Description automatically generatedIF the election of state representatives to the US Congress was “fair,” it would be proportional. In North Carolina, in 2018, the Democratic candidates (as a group) received 48.3% of the votes cast. That year, only 3 of the 13 (23%) representatives sent to congress from North Carolina were Democratic.

The graph to the right shows the Democrats’ vote-share vs. seat-share for the 50 states in the 2018 election.

After the 2010 Census, all states redistricted to account for the reapportionment of the 435 House Representatives. Across the country, there were six types of redistricting authorities as identified in a 2012 report done by the Brennan Center for Justice, including: Democratic controlled, Republican controlled, independent, and court controlled.

For this project, I propose to examine whether the fairness of representation is related to who did the 2010 redistricting. I will use data from 2012, 2014, 2016, 2018 elections (and 2020 if I can get it in time). And I will use two measures of “fairness.” First is the vote-share vs. seat-share comparison shown above. The second measure, referred to as the “efficiency gap”, was devised by N. Stephanopoulos and E. McGhee from the University of Chicago. It’s a metric that quantifies the percentage of a given party’s votes that are “wasted” (e.g. if a congressional district is “packed” so that most voters are of one party, then the votes above the 50% threshold are “wasted.”)

1. My Data

Through the “MIT Election Data + Science Lab” I was able to download congressional election results for the years 1976 to 2018. This .csv is in my GitHub.

1. My Timeline

I appreciate the outlining of the steps of this project (EDA -> Data Processing -> Modeling & Validation -> WriteUp). And I will strive to complete each of the four steps in a manner proportional to the 3.5 weeks between project approval and submission.

Write-up:

Introduction

Every ten years, after the census is completed, the 435 House Representatives are reapportioned among the states based on the current populations. Then each state redraws its electoral districts, so the districts within a state have approximately equal populations. Aside from mandating that the districts have equal populations, the Constitution is silent about the process and leaves the redrawing of districts to the states. Redistricting laws in the different states vary as does the responsibility for redistricting. In some states redistricting is done by the courts, in other states there are independent commissions. And in the majority of states, redistricting is done by the state legislator and the governor. Of these states, in 2011, twenty-three had a state legislature and a governor who belonged to the same political party.

In assessing the fairness of the redrawn electoral districts, there are four overlapping and often conflicting, perspectives.

First is the hope that districts will be politically proportional, yet that doesn’t always happen. In North Carolina, in 2018, the Democratic candidates (as a group) received 48.3% of the votes cast. That year, only 3 of the 13 representatives (23%) were Democratic. And in Massachusetts in 2018, Republican candidates received 36% of the votes cast, yet none of the 9 congressional representatives were Republican.

A second concern in redistricting is the competitiveness of the districts. Clearly there is a relationship between the number of competitive districts and proportional-political-representation. In the appendix I discuss my observations of the number of uncontested and competitive districts that I observed in the years that I’ve analyzed 2002 to 2018).

The third perspective about fairly drawn congressional districts is stipulated by the Voting Rights Act of 1965 and focuses on a minority groups’ ability to elect representatives of their choice. The law addresses instances where this ability is diminished during redistricting and focuses on assuring that there are “majority-minority districts.”

And the fourth perspective on drawing “fair” electoral districts is most clearly articulated in California’s 2010 voter initiative that established a redistricting commission tasked with creating “communities of interest.” The goal was to create districts that allow like-minded communities to have representation in congress.

The analysis that I am doing in this paper will only focus on proportional-political-representation and will not address minority representation or community representation. All the while understanding that the Voting Rights Act impacts the drawing of districts in a manner that can skew proportional-political-representation (ironically, assuring majority-minority districts is a form of packing and can dilute Democratic representation). And the goal of keeping communities of interest together can also impact proportional-political-representation.

In my current analysis, I will focus on the proportion of seats to votes. Specifically, the (% of seats) to the (% of votes). From a proportional-political-representation perspective, if an election is “fair,” then the proportion of seat to votes should be 1. And if you look at the seats and votes from one party’s perspective, if the proportion is greater than 1 then that party was advantaged. As an example, in Massachusetts the representation proportion in 2018 was (100% seats) / (66% votes) = 1.52. (In the appendix, I introduce another measure of political fairness called the “efficiency gap.” As I continue to work on this paper, I will look at how the “efficiency gap” compares with the proportional-political-representation outcome that I’m currently using.)

In my analysis, I am looking for differences in redistricting between those states that were controlled by Republicans, by Democrats, and by Others.

Who draws the districts can and often does have an impact on the political fairness of the districts. There is no need for statistical analysis to convince you of this statement, nor do I need to argue the fact that gerrymandering is an aspect of politics. But there are other factors aside from who drew the lines, that can impact the “fairness” of the resulting districts.

Gerrymandering writ-large, is accomplished by either “packing” (drawing district lines to pack one group into a small number of districts, which gives the other group an advantage in the remaining districts), or by “cracking” (drawing district lines to dilute one group’s votes by distributing those voters among numerous districts).

Though we know that those who are drawing political boundaries can pack a district, voters also pack districts when the choose where they will live. And in a political environment where Democratic correlates with urban and Republican correlates with rural, this natural clumping can have an impact on proportional-political-representation. In the appendix I discuss some measures of clumping that I analyzed, and others that I will look at further. For this version of my analysis though, I have not included clumping as a predictor in my model fitting, and fully realize that any model that doesn’t include clumping is incomplete. I will include clumping as I continue to refine my model.

The other predictor variable that I did include in my model fitting, aside from who did the redistricting, is the number of districts in the state. Clearly if there are only 2 congressional seats, proportional-political-representation is unlikely. And at the other end of the spectrum as an example are the 61 seats the Democratically controlled legislature in California had to work with when they drew district boundaries in 2001 – a number that gives numerous opportunities for cracking and packing.

Methods

In building the data frames for my analysis I’ve worked from: 1) election results from the Harvard Dataverse website, 2) An analysis done by the Brennan Center for Justice of who was responsible for redistricting in each state, and 3) population and land area data for 2000 and 2010, downloaded from Census.gov.

Aside from some data cleaning and data frame joining, the major data wrangling issue that I dealt with was imputing vote numbers for candidates who ran unopposed. The vote tallies for unopposed candidates are not reflective of the number of Democratic and Republican voters in a district. In it’s extreme, there are Florida’s vote records where the winning candidate is noted as having 1 vote. Since my outcome measure was comparing % of seats (an accurate number) I needed a method get a more accurate total of votes within a state.

For each unopposed election, I calculated the maximum number of votes received by candidates of that party in the other districts in that state and then assigned that maximum vote total to the unopposed candidate. And for the party that didn’t field a candidate, I created a candidate named “Imputed” and assigned them the minimum vote total from the other candidates of that party. (When I do my analysis using the “efficiency gap” and wasted votes, I will revisit this method of imputing uncontested elections.)

Finally, for each county within each state I calculated the population density for each year. Those values were used in my attempt to build a “clumping” metric (see the appendix).

When done, I had a data.frame for both Democratic results and Republican results. And I did my data analysis from the Democra’s perspective and using the Democratic Results data frame.