Project\_eac238

Eric Cartaya

5/1/2020

# A Look At American Politics Over Time

In the modern American political landscape there are two dominant parties, the Democrats and the Republicans, with which you should be familiar with already. There are however many smaller parties and independent candidates that run every election on different platforms, typically to less desirable results. I decided I wanted to look at how these parties have changed over time from two perspectives: how successful they are in presidential elections and how many people identify against the two major parties.

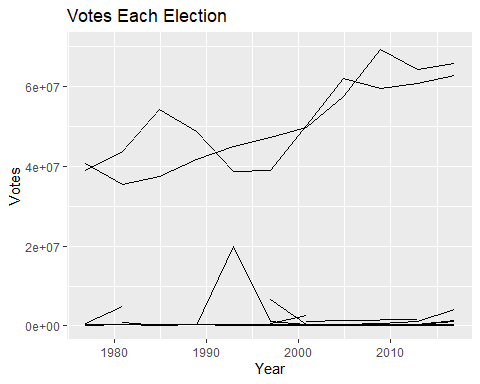
The first dataset I found after only a couple of hours of looking from the MIT Election Lab, which gave me a link to a Harvard database for some reson. This gave me information of the popular vote in presidential elections over time and was tidied and scraped with relative ease. The other dataset was a massive pain and took me nearly a month (I found it on the 1st of May) to find it. I nearly thought about abandonning the project completely on multiple occacions, but I did not. It also took considerably more effore to scrape it, but was eventually molded into a dataset that showed me the results of polls on people’s political identifications over time.

## The Data and The Analysis

First I will go over the data and results, then I will cover the actual analysis and results.

The first dataset (the MIT Election Lab dataset), had to be changed to reflect the national counts for each party above the state counts, but I was able to get a count per election per party. Some parties are divided by states which might skew the numbers, as well as some parties having misspelled names such as the “Communist Party USE”. All the major parties are spelled right in all elections from what I can tell with only the Libertarians changing their name halfway as the only major change so I didn’t bother correcting anything.

Down below we have all of the results per election per party, with the results from all elections being graphed. I couldn’t add color to the graph as the colors take up the whole graph. A full graph will be in the presentation though. The date added is November 16th, as November 15th was the last possible day that could be after the election, but I did my math wrong and thought that this was just as good.



## # A tibble: 377 x 4  
## # Groups: date [11]  
## date party candidatevotes totalvotes  
## <date> <fct> <int> <int>  
## 1 1976-11-16 "" 196372 42583819  
## 2 1976-11-16 "american" 143914 24527755  
## 3 1976-11-16 "american independent party" 131761 27786458  
## 4 1976-11-16 "american party of iowa" 3041 1279303  
## 5 1976-11-16 "communist party use" 44287 33637412  
## 6 1976-11-16 "concerned citizens" 1162 541218  
## 7 1976-11-16 "conservative" 276996 7626107  
## 8 1976-11-16 "constitution party" 26390 6176321  
## 9 1976-11-16 "democrat" 40680446 81601344  
## 10 1976-11-16 "free libertarian" 12197 6668262  
## # ... with 367 more rows

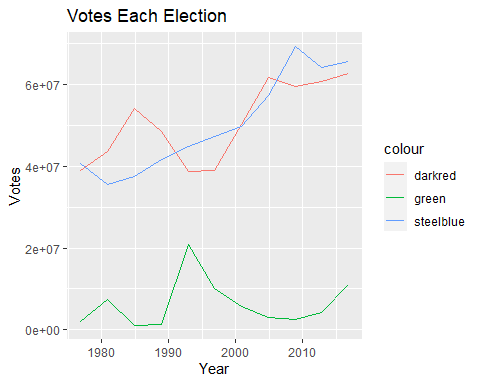
This contained all of the total values for each party and ranks them by how many votes they’ve gotten, which did reveal some issues. Firstly, there is a blank space, which is for blank and spoiled ballots, it is not a mistake and is completely intentional. I believe it was left blank to make sure they wouldn’t get confused for the “Spoiled Ballot Party”. Secondly two of the largest parties are the Independent Party and the Reform party, which will be explained later. In 1992 both hit their record highs and both ran the same candidate, Ross Perot, who was a big part of the reason that I did this topic in the first place. Regardless of Ross Perot, This does skew the vote a little, but not too much more than anything else I’ve mentioned before.

## Adding missing grouping variables: `date`

## # A tibble: 148 x 2  
## party partyvotes  
## <fct> <int>  
## 1 "republican" 558286316  
## 2 "democrat" 553813829  
## 3 "independent" 28193567  
## 4 "libertarian" 8947877  
## 5 "reform party" 6915968  
## 6 "green" 5046219  
## 7 "democratic-farmer-labor" 4159447  
## 8 "" 2745756  
## 9 "conservative" 2467312  
## 10 "liberal party" 1107541  
## # ... with 138 more rows

Here we look at the dataset I ended up sticking with as my “tidied” dataset. In reality I did some more stuff after, but it was more for calculations than anything else. The thirdparty votes were calculated by seperating out the Democrats and the Republicans and subtracting them from the total, allowing for there to be an entire column for the third party and independent candidates. This is the dataset I am submitting by CSV. The colors for the graph correspond to their typical party with green for the Independents to represent the fact that many third parties revolve around legalizing marijuana or protecting the enviroment.

## # A tibble: 11 x 5  
## # Groups: date [11]  
## date demvotes repvotes totalvotes thirdpartyvotes  
## <date> <int> <int> <int> <int>  
## 1 1976-11-16 40680446 38870893 81601344 2050005  
## 2 1980-11-16 35480948 43642639 86496851 7373264  
## 3 1984-11-16 37449813 54166829 92654861 1038219  
## 4 1988-11-16 41716679 48642640 91586825 1227506  
## 5 1992-11-16 44856747 38798913 104599780 20944120  
## 6 1996-11-16 47295351 39003697 96389818 10090770  
## 7 2000-11-16 49662314 50311372 105593982 5620296  
## 8 2004-11-16 57449547 61872711 122349450 3027192  
## 9 2008-11-16 69338846 59613835 131419253 2466572  
## 10 2012-11-16 64205850 60670117 129139997 4264030  
## 11 2016-11-16 65677288 62692670 139568633 11198675



The data had to be gathered to get a decent model of the data, giving a model where the total votes was not included. This gave the conclusion that third parties are in general becoming less popular over time and are being outpaced by both partied particularly the Democrats. It can also be seen the since the 90’s the Democrats have tended to do better than the Republicans, even in elections the Republicans win.

##   
## Call:  
## lm(formula = votes ~ ., data = consolidatedPresidentTable2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11749604 -4755962 18911 3790711 16581096   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.733e+07 3.412e+06 10.941 8.25e-12 \*\*\*  
## date 1.326e+03 2.694e+02 4.921 3.16e-05 \*\*\*  
## partyrepvotes 4.066e+05 3.049e+06 0.133 0.895   
## partythirdpartyvotes -4.405e+07 3.049e+06 -14.446 8.82e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7150000 on 29 degrees of freedom  
## Multiple R-squared: 0.9132, Adjusted R-squared: 0.9042   
## F-statistic: 101.7 on 3 and 29 DF, p-value: 1.717e-15

This second model that I created was the same but I included the total votes. There is not much of not here and this is an overall terrible model that I included specifically becuase there is a P-value of 1 in it. I wasn’t even sure that was possible, but here it is. The report is still 5 pages long without this interlude.

##   
## Call:  
## lm(formula = votes ~ ., data = consolidatedPresidentTable3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11839881 -4556534 160101 4349522 15577492   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.455e+07 1.648e+07 0.882 0.385   
## date 8.422e-13 9.758e+02 0.000 1.000   
## totalvotes 3.333e-01 2.361e-01 1.412 0.169   
## partyrepvotes 4.066e+05 2.998e+06 0.136 0.893   
## partythirdpartyvotes -4.405e+07 2.998e+06 -14.692 1.1e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7031000 on 28 degrees of freedom  
## Multiple R-squared: 0.919, Adjusted R-squared: 0.9074   
## F-statistic: 79.38 on 4 and 28 DF, p-value: 7.304e-15

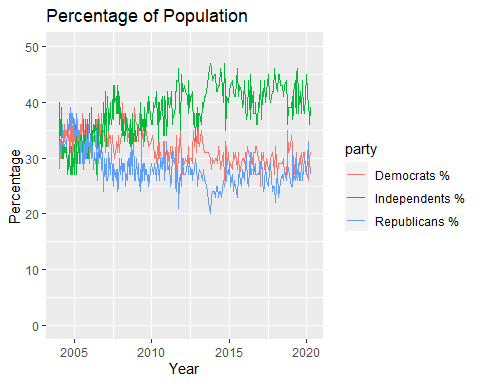
This dataset took considerably longer for me to find, and was much more difficult to tidy, but this was exactly what I wanted out of a dataset. This dataset came from Gallup Polls, asking the question of political party affiliation. The data was asked much more frequently than every election, but did not go back as far as the other dataset. I also do not know how many people were asked in this endeavor, but this was the only thing I could find. As this one was simpler there were fewer problems, even if it was a pain.

First up I wanted to include a view of what the dataset initially looked like after all the garbage was removed, if only to prove that it was an uphill battle to get anything done in it.

## x  
## 1 2020 Apr 1-14  
## 2 27  
## 3 39  
## 4 31  
## 5 2020 Mar 13-22  
## 6 30  
## 7 36  
## 8 30  
## 9 2020 Mar 2-13  
## 10 30

After an ungodly ammount of trying to get the data to make sense I was able to not only fix the dates, but also get all the percents in one column with a party value, I did not make that mistake twice. From here I displayed the information on a graph with the colors messed up, but the actual lables were better for this one so you can see each of the different parties labled. Here we could see the truth, people don’t want to identify as Democrats or Republicans anymore, even if they keep voting for them. This was also the dataset that I decided to submit as the tidied one.

## Date party percent  
## 1 2020-04-01 Republicans % 27  
## 2 2020-03-13 Republicans % 30  
## 3 2020-03-02 Republicans % 30  
## 4 2020-02-17 Republicans % 30  
## 5 2020-02-03 Republicans % 33  
## 6 2020-01-16 Republicans % 30  
## 7 2020-01-02 Republicans % 27  
## 8 2019-12-02 Republicans % 28  
## 9 2019-11-01 Republicans % 30  
## 10 2019-10-14 Republicans % 28



For the analysis I tried to make an identifier for party that wasn’t a string, but got the same exact result so it didn’t matter. I kept this version becuase I didn’t wasnt the effort to go to waste, and came to the conslusion that people don’t want to be Democrats or Republicans anymore and that either people in thrid partied aren’t really voting or are voting for the Democrats and Republicans.

##   
## Call:  
## lm(formula = percent ~ Date + party, data = gallup)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.4236 -2.2962 -0.2027 2.7348 9.9226   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.330e+01 9.628e-01 34.586 <2e-16 \*\*\*  
## Date -8.081e-05 6.345e-05 -1.273 0.203   
## partyIndependents % 5.143e+00 2.754e-01 18.678 <2e-16 \*\*\*  
## partyRepublicans % -3.197e+00 2.754e-01 -11.609 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.674 on 1064 degrees of freedom  
## Multiple R-squared: 0.4679, Adjusted R-squared: 0.4664   
## F-statistic: 311.9 on 3 and 1064 DF, p-value: < 2.2e-16

## The Conclusion

The final conclusion that I have come to and the main reason that I chose to do this project is that I believe that the Democrats and Republicans aren’t the only options. Ross Perot on a ballot of social liberalism, taxation of the rich, and assistance for businesses small and large alike got nearly a quarter of the votes in 1992. After that no other candidate has done as well but with their wanning popularity I think the two main parties have only some much time. I was originally going to touch on what the larger political partied beleive but I don’t think that is as necesary, even if the United States keeps the Democrats and Republicans, I don’t think they could stay the same and would see very different political ideas rising up to take their place.

The final and truest reason that I wanted to do this project is that I believe in the importance of plurality in a political system. The idea that everyone can vote for a candidate that they believe in an actually have a chance to win. That is what I see as fair and honest and something that is lacking from our current political system. I wanted to do this project to see if there was hope. I live in a bubble like everyone else and an army of anarchists, libertarians, and “gamers” tell me that they outnumber the Dems and GOP. I had no way of truly knowing without looking it up and I am glad to see that I was right, that we were right. There is hope. I know that this report is a bit more strangely written than what you are used to but this project was a passion of mine and that certainly came out in the report. I am sorry that the presentation is basically the report in slides. Thank you for humoring me and I am sorry this took so god damn long.