FEC data wrangling report

Libby Nachreiner

10/25/2021

Data Wrangling Steps:

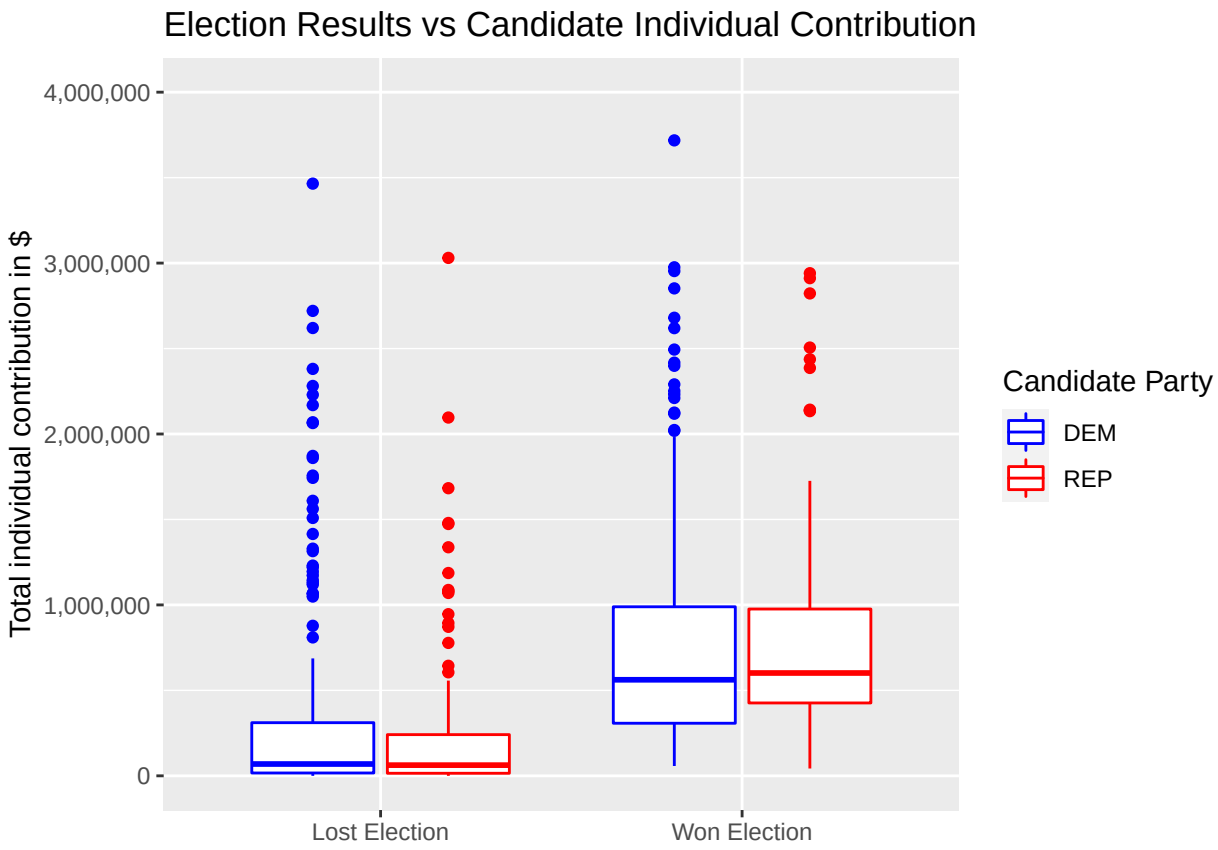
- First, I joined the candidates and campaigns data sets by finding places where candidate id matched in the respective data sets.
- I then joined this newly created data set with the 'results_house' data set using candidate id once again, and set this new data set to 'house_camps'.
- Next, I selected the seven columns from this data set that would be useful to me in my analysis of the results of the elections for house seats.
- Then, I grouped the data by the candidate name, since I was getting multiple rows for each candidate due to the way that the 'house_results' data set was formed. This data set gave the percent of vote received in the general election for each candidate that came from *each* voter party. There were, however, some duplicate rows in this data set. This means that I needed to get distinct voter parties for each candidate, so I used the distinct() function once the data were already grouped by candidate.
- To calculate the total percent of the vote that each candidate received, I then had to use a mutate statement on my data set where the duplicates had been removed. In this mutate statement I summed the percentage of the vote received from each voter party to make a column called overall percent which represents what percent of the vote a given candidate received.
- Next, I filtered my data set to only include democrats and republicans, as the sample size for these two parties was significantly higher than the other parties.
- At this point I still had duplicate candidate names in my data set, and after the wrangling I had done, the 'overall_percent' column for each duplicate row was the same. Therefore all I had to do is filter to get distinct candidate names.
- I now no longer needed the 'voter_party' or 'general_percent' columns so I removed them from the data set.
- Then, I removed all candidates who did not receive any votes, any candidates who started their campaigns in debt, and any candidates who had "NA" as their party.
- Finally, I arranged the rows of my tibble by the candidates who received the highest percent of votes in their respective elections.

Data Introduction

The data for my project comes from the United States Federal Election Commission (FEC). I focused on the 2015-2016 federal election cycle, specifically the campaigns for the U.S. House of Representatives in this cycle. I was interested to see how various markers of financial status as well as candidate type (challenger, incumbent, or the case where the seat was open) impacted election results overall and within each major political party (Democrat and Republican) for this data set to test my hypotheses that those who have more money (more often Republicans) and those who run as incumbents are more likely to win. To explore the dynamic between finances and election outcome, I specifically analyzed the relationships between election

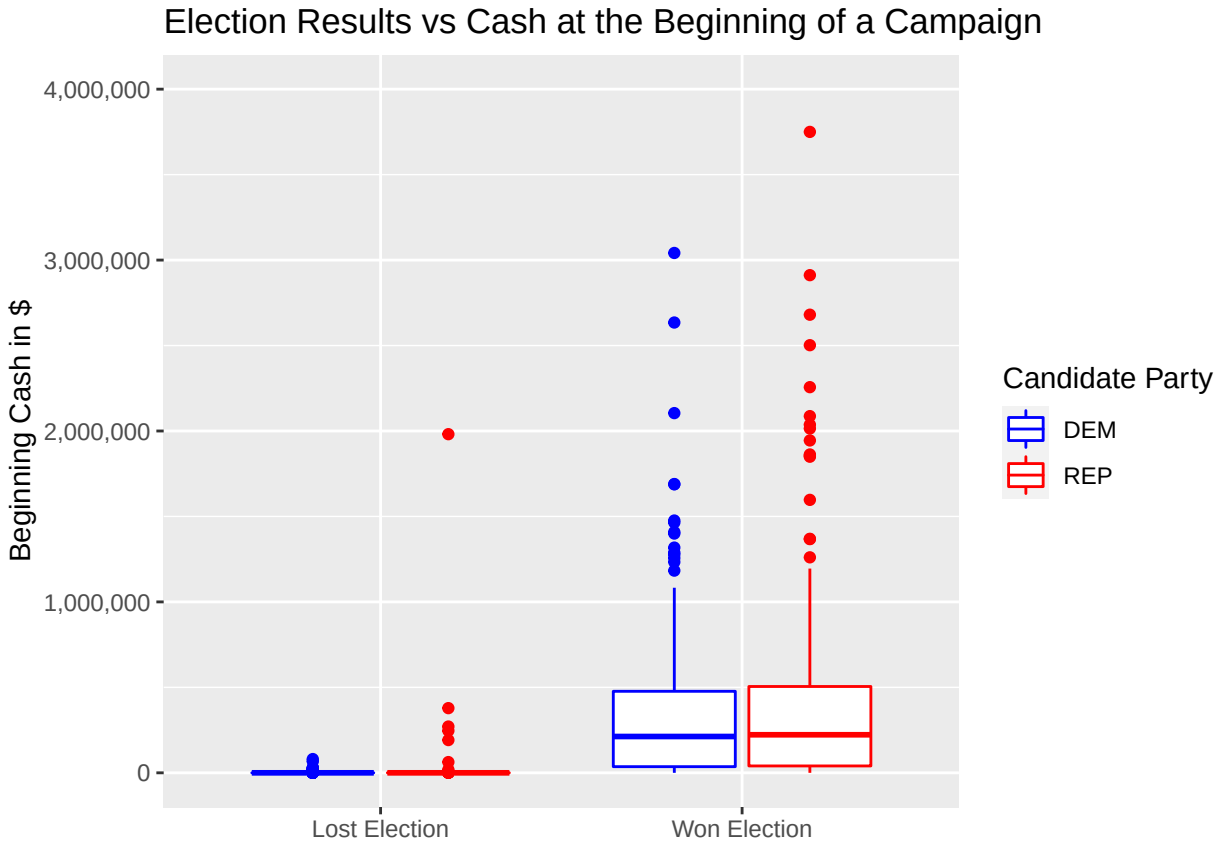
results and a candidate's financial contribution to their own campaign by party, election results and beginning cash by party, and beginning cash and candidate type. Finally, I investigated the relationship between election outcome and candidate type by looking again at the relationship between beginning cash and candidate type as well as exploring the relationship between election outcome and candidate type.

Analysis of Election Results and Individual Contribution



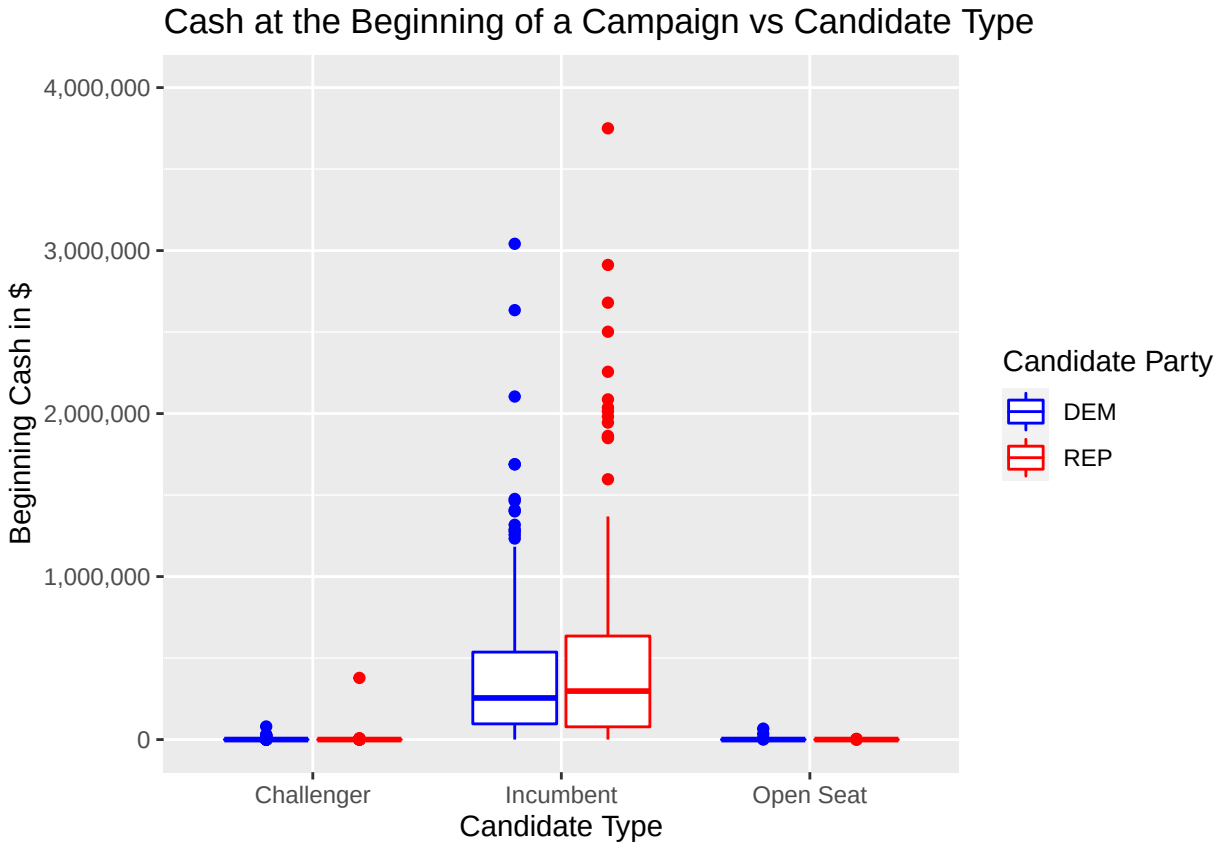
The first aspect of this data I explored was election results and a candidate's individual financial contribution to their campaign. We see that on average, those candidates who lost elections in this data set contributed much less to their campaigns than those who won elections. This shows how having the financial means to contribute to your own campaign is a huge advantage, which is what I expected to find. One thing that surprised me, however, is that within the 'losers' and 'winners', the average individual contribution for the democrats and republicans was extremely similar. This may indicate that candidates who win elections have similar financial statuses regardless of party. Likewise, those who lose elections may have similar financial statuses regardless of party. Independent of party, this data set shows how being able to contribute around $500,000 to one's own campaign is the standard for winning an election.

Analysis of Election Results and Beginning Cash



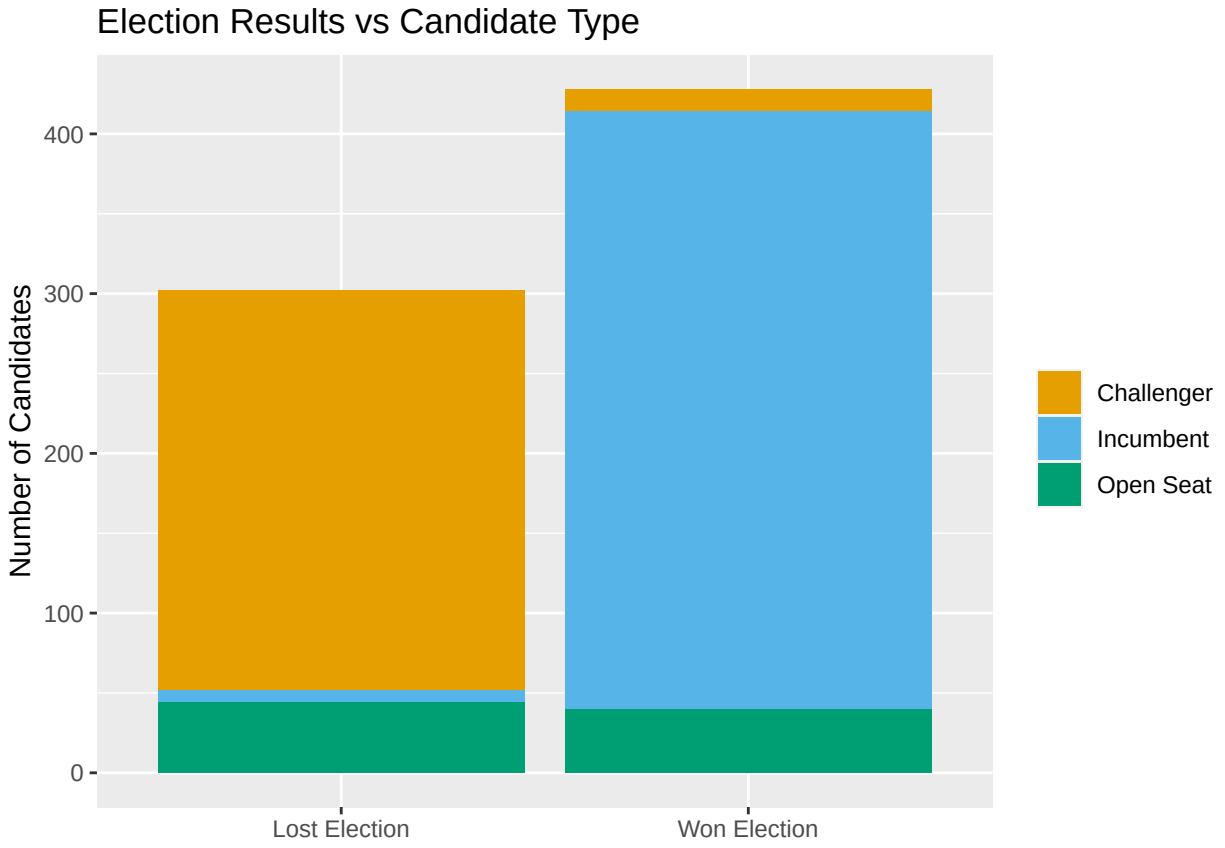
Next, I analyzed the relationship between election results and a candidate's beginning cash. In the graph above we can see that on average, those candidates who lost elections in this data set had less cash at the beginning of their campaigns than those who won elections. This again shows how having more money is an advantage, though not as much of an advantage as being able to contribute to one's own campaign given the smaller difference between the winner and loser average in this plot compared to my first one (which looked at individual contribution). Similar to the previous plot, the average beginning cash level for the democrats and republicans was almost identical within each winning status, with the exception of a Republican who had $2,000,000 and still lost their election. Independent of party, then, this data set shows how having more money is helpful in winning an election.

Analysis of Beginning Cash and Candidate Type



Another aspect of this data that I investigated further was beginning cash and candidate type (challenger, incumbent, or the case where the seat was open). In the graph above, it is evident that in general, incumbents start with a much higher amount of money than challengers do. Additionally, I saw that there were some incumbents who started their campaigns with well upwards of $1,000,000, while no challengers started with above $500,0000. Something that surprised me, however, is that within each candidate type, the beginning cash for the democrats and republicans was extremely similar. This was interesting to us because this seems to go against the message pushed by many democrats that they value distributing money more equally among the population. Finally, this shows that once one is in power, it is much easier to raise money and thus be better equipped to run a good campaign and stay in power.

Analysis of Election Results and Candidate Type



The fact that starting with money and power makes it significantly easier to keep said money and power is illustrated in the graph above. We see that of those who lost elections in this data set, the vast majority are challengers. Furthermore, challengers make up just a tiny fraction of election winners. Incumbents, on the other hand, make up the vast majority of election winners and just a small fraction of election losers. This furthers the idea that being an incumbent and, as I showed earlier, having money leads to more successful elections. As expected, open seats have a much more balanced distribution, as no one has the edge being an incumbent with money and power.