EDA Mini Project 2

Logistics and timeline:

- Final write-up due Friday, April 1, 5pm.

- No need to submit a draft this time, but if you would like feedback feel free to talk with me or Haoran.

- Work in groups of up to 3, and make one submission for each group.

Assignment:

Many people studying politics in the United States believe that the country is becoming more ideologically polarized: either liberals are becoming more liberal, conservatives are becoming more conservative, or both. In either case, they believe that the middle of the political spectrum is being hollowed out.

One way to investigate whether this is true is to look at the congressional voting record. I've given you a partially cleaned version of roll call votes taken by the Senate, and they are available at https://jfukuyama.github.io/teaching/stat670/assignments/congress.zip.

When you unzip the folder, you should get one subdirectory for each year between 1989 and 2014. Each folder contains two csvs: one giving each senator's vote on each of the bills for which there was a roll call vote in that year, and one describing the senators (name, id, state, party). If you want more data or more information, I got the vote and member information using the tools at https://github.com/unitedstates/congress.

We'll investigate polarization in a couple of ways.

- *Polarization in two years*: Take a look at the voting records for one of the earlier sessions in the dataset and one of the later sessions (e.g., 1989 and 2014). Using either PCA on the senators by bills matrix or multi-dimensional scaling on a measure of distance between senators that you construct, make a plot describing the relationships between the senators. Does it look like they fall on a one-dimensional liberal/conservative axis? Are there outliers? Are the patterns similar in the two years that you chose? What are the differences in the PCA/MDS plots in the two years that you chose?

- *Polarization over time*: Next, we want to look at this more systematically over time. Repeat the same analysis you used in the first part for every year in our dataset. Construct a measure describing polarization, and plot it over time (think about variance explained by the first principal component, average distance between the scores for Democrats and Republicans, or similar measures). If you are able to make a clean visualization showing the PCAs or MDSs for each session of congress and you think it helps to support your claims, include that as well.

  Do you think that polarization has increased over time? Is something more complicated happening? Back up your claim with specific aspects of your analyses and graphs, and describe why they support your claim.

– *Ideological position of one senator*: The analyses in the previous parts told us about the relative ideological positions of and the amount of separation between the parties, but not about absolute changes from year to year. We might want to know about absolute changes in the ideological positions of the parties so that we can assign blame to one or both parties for polarization, but it's a tricky question to get at because both the senators voting on the bills and the bills being voted on change from year to year.

One way of starting to answer the question is to take one senator who served each year and compare the parties to that senator. If the senator's politics stayed the same over time, we can measure the two parties' positions relative to that senator.

On the other hand, if you think that there hasn't been an increase in polarization and that the ideological distributions of the parties has stayed basically the same over time, we can investigate whether individual senators have changed their politics over time in the same way, by measuring the position of a given senator relative to the positions of the two major parties.

Choose one senator who served in all (or at least many) of the years we have represented in our dataset, and make a plot describing his or her position relative to the two major parties over time. Describe the implications of your plot either in terms of which party is more to blame for polarization or relating to the question of whether that senator's politics has changed over time. You should also describe the limitations of your analysis, and whether there are other possible explanations for the phenomena you see.

Note: An example of a senator whose ideology we might expect to have been constant over time is Mitch McConnell (currently the minority leader and a senator from Kentucky). A senator whose ideology we might suspect has changed over time is John McCain, the late senator from Arizona.

Grading: Submit a report as a PDF, the body of which should be no more than ten pages, including graphs (if you don't feel like you need ten pages, submit fewer). Additional technical graphs can be included in an appendix, which will not count toward the ten page limit and which we might not bother to read. Submit your code as a separate file. Also upload any additional sources required to reproduce your work.

Point distribution:

– 10 points for the first set of questions.

– 10 points for the second set of questions.

– 10 points for the third set of questions.

– 10 points for presentation.