

Isaac Getto and Zoher Ghadyali

Data Science 16

CTW Project

Project Reflection

Assessment evidence and interpretation

Our core deliverable is contained in [/deliverable/deliverable.ipynb](#) which is an iPython notebook in our Github repository. We began this project wanting to tell a cohesive story about Congress and explore how it functions and why it has had an approval rating under 10% for several consecutive years. Unfortunately, due to the amount of time we spent creating our own dataset, we were not able to move much beyond our data exploration phase or explore beyond the US Senate during the 1st session of the 114th Congress. We have put those results together in as cohesive a way as we can to look at partisanship in the Senate.

In beginning this project, we also wanted to impact the political apathy, especially about Congress, that we see in our generation and in our notebook, we have a section where the viewer can select individual senators that they care about and explore the top contributors to an individual senator's campaign, how that senator votes on a given topic, how often legislation sponsored by that senator passes, and how in-line with their own party that senator votes.

Beyond our deliverable, we have pushed all of our data from the FEC on individual senator's campaigns to Github in a usable SQL database and our .csv files generated from our scraping of senate.gov and congress.gov. It would be absolutely awesome if someone was looking for data on an individual senator and was able to use our work. We are documenting how to use our data in the README.md in our repository. Also, the work in our explorations, some of which is in our deliverable and some of which isn't, is contained in our [/explorations](#) folder in 4 iPython notebooks: [ZG_exploration.ipynb](#), [contributions.ipynb](#), [ma_senators.ipynb](#), and [party.ipynb](#). In assessing this project, all that should really be taken into consideration is the amount of time we spent getting our data into a format that we could work with after scraping it.

Changing the world

Our project does have the potential to change the world in the sense that we can actually do something about our generation's political apathy about Congress. In working on our project, we both learned a lot about how the Senate functions, from what Motion to Table Agreed to means or Motion for Cloture Rejected. We were able to show partisanship in our data and we would have been really interested in getting data for previous congresses and seeing if the effect of partisanship was as significant. Working with this data, especially seeing all of the 662 topics that legislation concerned during this one session in the

Senate, was both motivating and educational and we think if other people played with this data, or looked at our results, they'd find interesting things they wanted to know more about in the Senate.

Learning goals

Zoher:

- creating interactive and easy to understand visualizations

I didn't really get to create interactive visualizations beyond specifying topic and senator for some of the visualizations in our deliverable notebook but I did spend time when creating our visualizations thinking about what the best way to display this data would be. I read several articles vilifying the pie chart but since we were reporting such binary data, I thought a pie chart was fine. There is a lot more work visualizing this data beyond what we did, like showing legislation over time which we did not do, or incorporating previous sessions of congress and looking at how votes change on a given topic with time, but I think what we created was easy to understand and well-designed.

- working with multiple uncleaned data sources and formatting them in such a way that they can be merged and easily used

I got a lot of experience on this learning goal. Before this project, I had no experience scraping data and I learned a lot from having to do it to create our rollCallVotes csv files. I don't think I followed a lot of best practices and would love to spend more time making our scraping easier to apply to the House and to multiple sessions of Congress.

Isaac:

- Clustering, unsupervised learning

For this project, I did not actually do any clustering or unsupervised learning. I initially thought that this would be an interesting technique to visualize our data, but I found that there were a lot more visualizations that could be made without clustering. I think that it would be possible to apply some of these techniques to the data we found in the future. I hope that I get a chance to learn about unsupervised learning in the final project.

- Data visualization techniques

I enjoyed creating data visualizations during the project. I was able to make some time based plots, bar charts, and pie charts. Doing this gave me more practice in creating visualizations. These visualizations were made more interesting because they could be generated dynamically for each senator. My favorite part about the deliverable we produced was that people could interact with the data to investigate senator's and issues that they are interested in.