Pratool Gadtaula, Zoher Ghadyali, and Anne LoVerso
Data Science Spring 2016

Introduction

In recent years, the American people seem to care less and less about Congress compared to the offices of the President and the Supreme Court, which have accomplished sweeping changes in recent decades. We want to learn more about Congress, how it functions, how its role has changed over time, and share our results with others in an easy and digestible format.

MVP

For our MVP, we want to create a tool to inform voters about their congresspeople in this current session of Congress. We then want to compare this current session of Congress to previous sessions, using measures for partisanship, diversity, and efficacy. Our final output should consist of beautiful, interactive, and insightful visualizations that are enjoyable to look at, educational to the viewer, and inspire curiosity in our legislative process.

Possible Directions to Take

- Tool to inform voters about their congresspeople, current sessions of Congress
- Diversity in Congress as a function of time, the most recent Congress is the most diverse ever, how does this compare with the American population
- Length of each congressperson's time in office (over time)
    - What if we had congressional term limits? Could we predict how things in Congress might be different?
- History of bills in Congress - what subjects have they focused on over time, how has it changed? How has attitude to certain subjects changed over time through reactions/results of bills?
- Has congressional approval ratings always been so low? Was Congress ever effective? What are the factors that have caused this shift?

Stretch Goals

- Congressional history before 1980s, it would be really cool to get this data but it's not in the format we expect for most recent
- Not only the subjects a bill concerns but being able to understand what the bill is actually doing, is it cutting government funding for planned parenthood or giving government funding to planned parenthood?
- Summarization of bill for laypeople
- Making the tool efficient. This includes minimizing HTTP requests and using good coding style.

- Being able to make our data openly available to anyone who wants to use it, hosting it on something like Amazon EC2, or Olin's supercomputer

Data Sources

- https://www.govtrack.us/congress/votes lets us see how every single member of Congress voted on any given bill in the history of Congress
- http://www.senate.gov/legislative/legislative_home.htm and http://www.house.gov/ and https://www.congress.gov/ provides which includes bills, vote tallies on bills, who sponsored the bill, its current status, among other things.
- https://memory.loc.gov/ammem/amlaw/ for info on historical bills

Skills

- This project relies heavily on scraping data from the federal government's websites. That will most definitely be a skill that we will develop.
- We need to be able to store a large data set and set up some kind of storage system for easy access.
- Visualizations - we view Congress as the most boring and gridlocked part of our government and if we want to inspire excitement and curiosity, we need fascinating visualizations that tell a story, are insightful, enjoyable to look at and easy to understand
- We will likely need to set up a server to host our data and website, which is a new skill to develop and learn how that works.

Timeline

- Week 1: Initial data collection
- Week 2: Data exploration
- Weeks 3-4: Set up web app, scaffold visualizations, add data as needed
- Weeks 5: Getting feedback on visualizations, creating new visualizations as needed
- Week 6: Polish visualization, clean up code, finish documentation
- Throughout: documentation, adding data as needed

Biggest Risk

- Getting bogged down by data collection and overwhelming amounts of data
- Spending a lot of time on scraping and figuring out the best structure for the data, since it is not in an easily available format
- Our timeline is already full so any delays will end up cutting into time we would spend polishing our project, it might be the case we are only just able to get visualizations if we have delays

Individual YOGA Alignment

Anne

I think this project fits well with my YOGA goals. My goal for real world impact is definitely fulfilled - it feels like what we're doing is relevant, and if we succeed, the tool we build could really inform and help voters. I think this project also has a lot of opportunity to work more with data visualization and creating the best figure for the task at hand in a way that will really communicate and persuade well. My final goal was linking machine learning with the visualizations, and I'm not sure how that might incorporate into this project. There is some space for prediction, but I think as a team, we'd prefer to focus on handling large amounts of data and figuring out what aspects we want to focus on presenting. It would be nice to do some machine learning work as a stretch goal, but it's not a priority.

Zoher

This project overall fits very well with my YOGA goals. We want to create interesting and interactive visualizations about Congress and since we are dealing with Congress over time, we will have a lot of data that we will need to figure out the best way to store. I really wanted to spend some time learning D3 and focusing on the visualizations aspect and since our final deliverable will be a web-app, I think this project will satisfy this goal. In figuring out the best way to store our data, I don't necessarily need to be the person to do this because I could always understand my own pain points in getting the data and try and improve that process. Finally, my third goal was to learn how to communicate with my teammates when working on highly related components for this project, which is something I imagine will happen towards the end of this project as we start working on our web app, which should satisfy this goal too.

Pratool

My goals for this project included being able to create clean, persuasive figures and distribute the final product in mass, preferably online. Both of those goals are also goals of the group. The third goal includes doing more statistical analysis and mathematics. I hope to address this goal by somehow incorporating in our political tool a predictive element (e.g. X senator has a Y% of voting down this bill because of it regards the topic of Z) or some compiled statistics about how the elected official generally votes for a certain topic. This statistical element is not yet determined, but there is a plethora of information and data, so I will most definitely find some data to do analysis on.