FinalReport

Owen Turnbull

Preface

This report is meant exclusively for analysis; while it can be rendered, it will not create the Shiny app on its own. This is because the process of scraping the data and creating the subsequent Shiny app takes around 1.5 hours to run in full. The submission instructions say that this report should be render-able in full; however, given the 1.5 hour run time, I felt that this may not be practical. Instructions for running this app can instead be found in the README.md document. That document will provide instructions for running the scraping code and creating the Shiny app. The process is simple, but too time-intensive to include here.

This report will instead provide context and analysis for this Shiny app. To see this app without having to render it locally, visit the shiny URL: https://owenturnbull.shinyapps.io/state-delegation-scorecards/

Overview and Goal

The purpose of this project is to gauge the relative effectiveness of each state's delegation to the House of Representatives. The driving question for this project was whether certain states had outsized influence in the House. Specifically, I wanted to see whether certain states introduced disproportionately more bills than other states. In doing so, we can see which states tend to have more "effective" House delegations.

This project was designed for two purposes. First, it was created as a way to answer the above question. Second, it was created to assist constituents in analyzing their own state's delegation. As such, a Shiny app was created that allows users to interact with each state to get said state's "scorecard". Users are therefore able to interact with the app and become more informed on how state delegations impact the House.

Data Sources

The data for this app comes from a few key sources. They are as follows:

- www.congress.gov -> This site was scraped using Selenium in order to get the number for each bill, as well as each bill's description.
- www.clerk.house.gov -> This site was also scraped using Selenium in order to get the biographies of each member, as well as how each member voted for each roll call.
- congress.gov API -> The congress.gov API was employed to pull in the sponsorship status of each member for each bill. Note that the congress.gov API was *not* used for bill numbers and descriptions due to pull limits. We scraped using Selenium (see bullet 1) since there were ~10,000 bills, and the API has a 5,000 pulls/hour limit.

Data Scope

The data used in this project is specifically related to the 117th congress. All bills and members analyzed were therefore from the 117th congress. The 117th was chosen as it was the most recent congressional session with a complete data set.

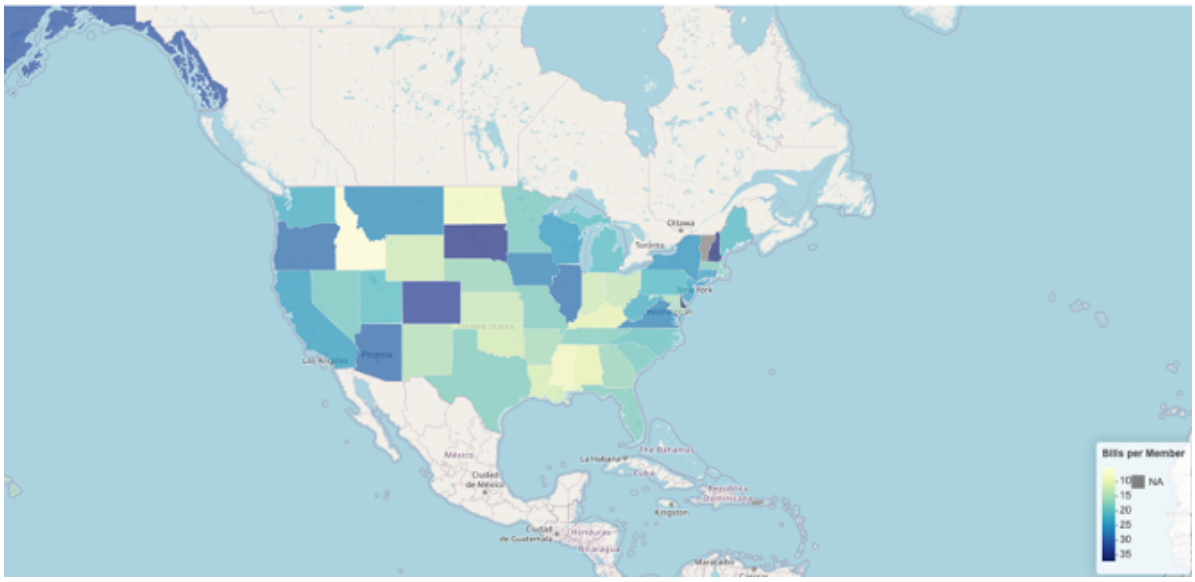
This analysis also only included data for the House of Representatives. It contains no information on Senators or senatorial bills. This is an integral part of this analysis; by looking only at the House, we can get a sense of disproportionate influence. House representation is clearly directly correlated with a state's population. Including Senators would only obfuscate the "proportionate influence".

Finally, not all U.S. states and territories were included in this analysis. First, Vermont was excluded due to issues with scraping. This was an unfortunate bug that could not be resolved in the project's time frame. Luckily, from a user perspective, Vermont is a relatively small state and shouldn't significantly impact analysis. Second, no U.S. territories were included in this analysis. While territories can have non-voting members, the issue of a "voting" vs. "non-voting" member has significantly more impact on "influence" than the number of bills introduced. Put more bluntly, it doesn't matter, for example, how many bills Norton-Holmes may introduce — D.C. still has less influence than any state. Territories were therefore left out of this analysis.

Analysis

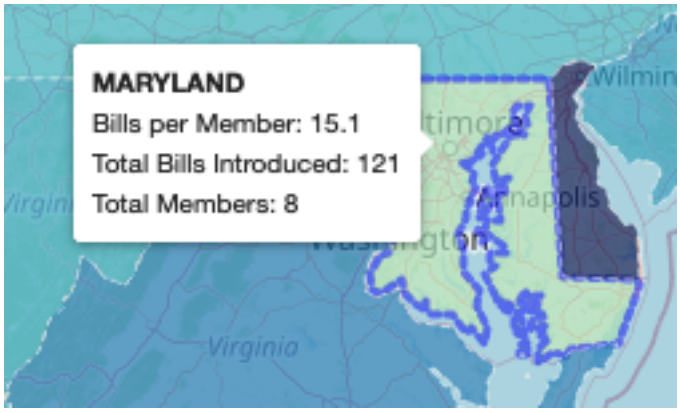
In accordance with our driving question and data sources, the Shiny app specifically shows a single unit of comparison: the number of bills introduced per member per state. At first glance, there are massive disparities in this weighted count of bills. For example, it's clear that Idaho and Mississippi produce very low rates of bills per member. On the other hand,

states like New Hampshire and Colorado produce very high rates of bills per member. We of course have to take the number of representatives into account; for example, Idaho only has 1 House member, so a slight variation in bills introduced by that member will have huge sway over Idaho's rate. The overall map can be seen here:

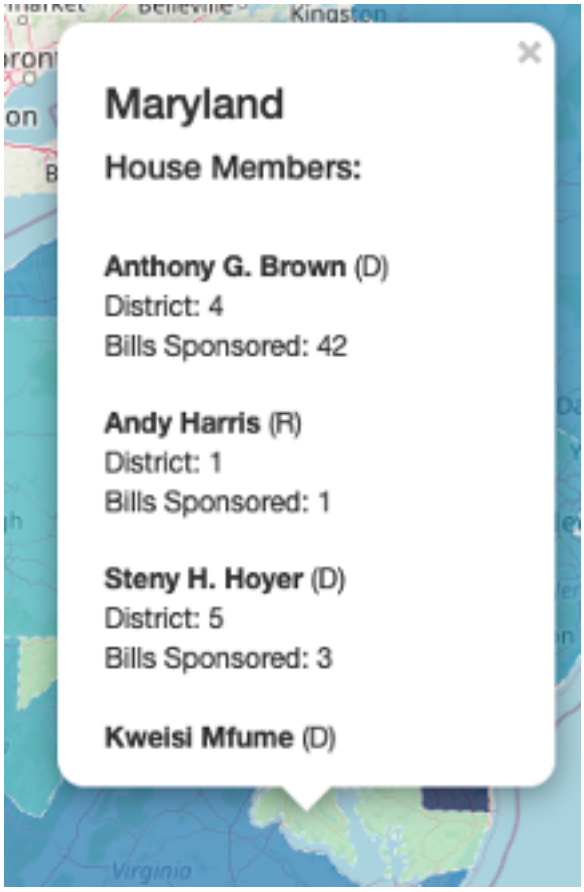


Interestingly, there are no clear trends among states with high (or low) bills per member. For example, if we looked at the app from a partisan perspective (such as who each state voted for in the 2024 election), there would be no clear through-line. This is true from poverty, religious, etc. perspectives. From an analytic perspective, there are no true reasons why one state may introduce more or less bills per member. This is an expected outcome; the decision to introduce bills falls solely on the shoulders of the legislature. Maybe there are slight difference in a state's culture that may lead to more effective House members, but this isn't quantifiable or obvious.

After analyzing this app in relation to our driving question, we can also analyze it from the user's perspective; after all, the second goal of this app is to help users in understanding their own state's House delegation. The app meets this goal exceptionally well. As you can see in using the app, each user can get an overall understanding of the bills per member rate. This is the first thing that's displayed when opening the app. Users can then get a specific state's "scorecard" by hovering over the state. This will show raw facts, such as the number of members in a state, the number of bills introduced by the state, and more.



Finally, users can get in-depth information for a state by clicking on the state itself. This will open a pop-up box that shows the name, party, district, and count of bills introduced for each member from that state.



Future Improvements

Overall, the app serves it's initial purposes. It both answers out driving question and gives users an informative experience. Nevertheless, there are still some key improvements that could be made.

- First, there are additional statistics that could be added to the "scorecard" box that displays when hovering over a state. Specifically, there could be data relating to the "unity" of a state's members. This would essentially be the number of bills introduced by a member of a state that were co-sponsored by at least one other member from the same state. Adding this indicator would be helpful for users and really flesh out the idea of a state's "scorecard". However, this was not included in the app as it stands now for a few issues. Mainly, there were issues with bot detection when trying to scrape the co-sponsor names for each bill. Doing so caused a captcha box to appear. While I am confident that I could bypass this in the future, it wasn't possible given the timeline of this project.
- Second, I would also want to add the historical "success" of a state's delegation. This was mentioned previously in the report. Basically, the state's scorecard would show the percent of bills from members in that state that pass the House. However, this would require historic data; using only data for the 117th congress would cause issues since confounding variables (mainly which party controlled the House) would greatly affect this rate. This historical data could not be scraped, given how long it takes to scrape even a single session's data. This will therefore be added as a feature in months to come, once said scraping is able to be implemented.
- Third, the scraping issue for Vermont would be resolved so that the app shows data for all states.
- Fourth, I would increase the amount of biographic information available to the user for each House member. The app currently shows the name, party, district, and number of bills introduced by each member. However, enhanced features such as the member's picture could be added. In fact, the data for this has already been pulled. I similarly scraped the URL for each member's website. These two points would greatly enhance the user experience. However, they have not yet been added due to time constraints; adding these in as an embedded JPEG or click-able hyperlink were too time intensive.

These improvements would together greatly enhance the user experience with this app. Nevertheless, they are critical components of this project. The app is still successful in answering our main question and giving valuable insight to users. These improvements would simply increase these insights.