Data Visualization
US Congress Bills Visualization

Process Book

Foivos ANAGNOU MISYRIS - Lucía MONTERO - Keshav SINGH

Overview, Motivation and Target Audience

The purpose of our visualization is to be able to get an insight into the bills that are introduced or introduced and eventually passed in the United States congress.

We can obtain different information when looking at the bills that have been passed and at the total set of bills that were introduced in Congress (regardless of whether they were passed or not), as the former gives insights about the specific congress member who introduced each bill while the latter show the bills that the majority of the congress and the president have agreed with. The reason for this relies behind the process through which a bill becomes a law in the United States.

A simplified intuition on this process starts when a bill is introduced in Congress by a member of either chamber, Senate or House of Representatives. Then members of that chamber vote either *for* or *against* the bill. In the case where the majority of the chamber members accepted the bill, the bill is then passed on to the other chamber, who also get to vote. Lastly, in the event where the majority of both chambers accepted the bill, the president may either veto or sign the bill, and in the latter case the bill would be considered as *passed*.

Given the previous consideration we decided to consider 2 datasets for our visualization, the reason is explained later in About the dataset section, and from here on *number of bills* will refer either of the numbers of bills - number of introduced or number of passed bills.

We can group bills by different characteristics, such as the policy area. In our visualization we wanted to be able to give a holistic overview of the main areas in which bills are introduced, but also see if these main areas have been constant in time or if different political parties have focused on different policy areas. We can extend this idea from different political parties to different states, as each Congress member that may introduce a bill is representing exactly one state.

The motivation of this tool is to give users a tool for carrying out comparisons of different patterns in Congress activity over time, which can be useful in many aspects such as comparing the subjects in which different parties/states are most involved with, or identifying changes in the Congress activity after a specific known event takes place.

The target audience would be people interested in US congress activity, with some background in US and US politics or interested in learning about it. Since it is possible to use this tool to either focus on a specific party only or to compare different parties, the visualization designed is useful for users with either interest.
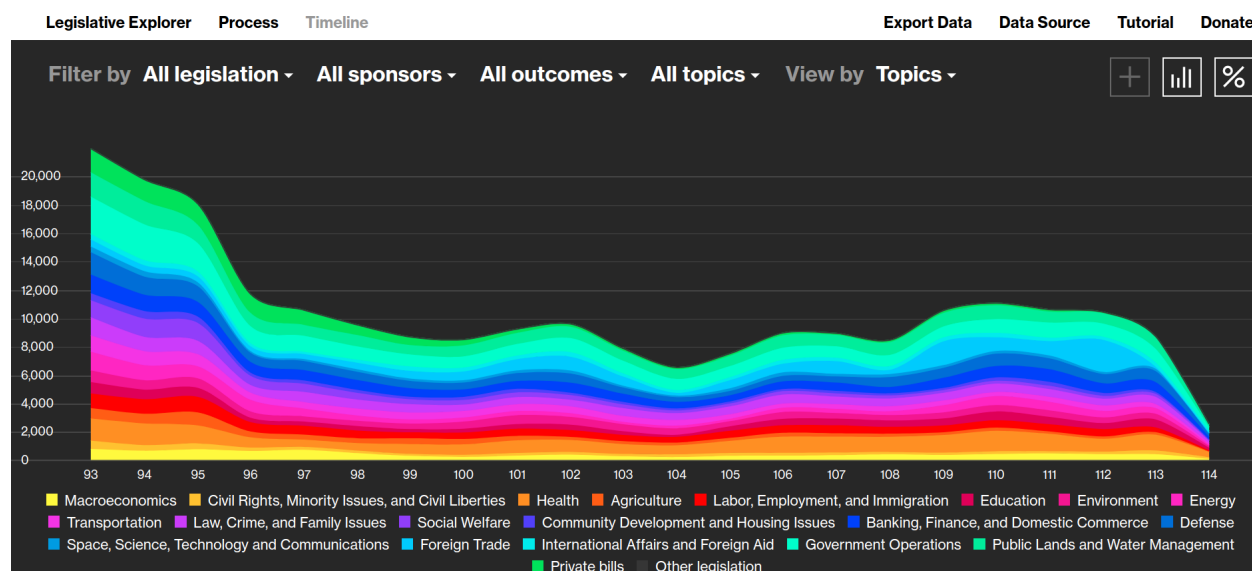
Related work and inspiration

Figure 1: Legislative Explorer, http://legex.org/timeline/index.html

There are several visualizations where the number of bills at different stages (i.e. not only *introduced* and *passed* bills). One of such examples can be found in the Legislative Explorer,

shown in Figure 1.

However, we considered this tool to be quite dense and difficult to use for identifying some patterns, the main reasons being the large number of combinations that can be selected when choosing the data to visualize and the large amount of information being shown at once in a way that does not allow for easy comparisons.

What are we trying to show with our visualization?

In our visualization we are trying to show counts of bills by policy area, year of congress, party and state. We try to do this in a way that allows to compare the bill numbers obtained for the values of each category (giving more of an overview), but also give the possibility to filter the information being shown in order to be able to focus on the category values of interest.

We want to do this to see specific patterns (e.g. *Has there been a difference in time for the number of bills related to a specific policy area?*), learn about differences between parties/congresses (*Is one party more involved with a policy area than the other?*), exceptions (*Which are usually the main policy areas? Is this not true for some states? Which could lead us to research whether there is a potential meaning hiding behind this exception*)...

Dataset Description

For the purpose of this project we retrieved data from two data sources, the ProPublica Data Store Congess API which was originated at The New York Times in 2009 [1] and the Center for American Politics and Public Policy (CAPPP) from the University of Washington [2].

Initially we chose to make use of ProPublica API where we can retrieve legislative data from the House of Representatives, the Senate and the Library of Congress. A great advantage of this API is that the information regarding the bills, which is the main focus in our project, is regularly updated and thus we can retrieve sufficient and valid information for the current congress too. One limitation is that we each request to the API can download up to 20 bills which make the retrieving process quite long, especially during the exploration phase when we wanted to inspect multiple datasets.

To overcome this issue we chose to use CAPPP, which is the data source used by Legislative Explorer described above, and which includes extensive information about each bill's progress and sponsor. The information is normalized and stored into CSV files rendering the retrieval and exploration of the data an easy task. One limitation of this dataset is that it has not been updated to include information about the current process. To overcome this issue, we had to download the recent bills using the ProPublica API and merge the two datasets. Another problem was caused because of he fact the two data sources use different grouping of Policy Areas for the bills. This problem was solved by doing a manual mapping from the bill

subjects of one dataset to the ones of the other. Although an integrity violation seemed to be quite possible by performing this mapping, we verified that in both projects they retrieve data about the bills from `www.congress.gov` which provides a detailed categorization for bill policy areas, rendering us able to perform an accurate grouping.

After deciding on the content of our visualization we had to process our data in order to keep only the important information and transform them in a form suitable for our implementation. Some of the multiple preprocessing steps we performed was joining tables in order to retrieve bills based on their status (e.g. 'passed' or 'introduced'), to filter out resolutions because they are not laws but rather the statements of intent or declarations that affect the operations of Congress, group data based on the different filtering options we include in our visualization and merge data from different sources.

	congress	party	major	state	idNEW	url	title	status
0	93	D	Government Operations	IL	1	http://beta.congress.gov/bill/93th-congress/ho...	A bill to provide for the establishment of an ...	not_passed

Figure 2: Data row showing with final fields.

congress	party	major	state	status	id_list	url_list	title_list	count
93	D	Agriculture and Food	AL	not_passed		24509	http://beta.congress.gov/bill/93th-congress/se...	A bill to amend the Egg Products Inspection Act.	1
			AR	not_passed	1614;2152;3042;5288;5571;7787;8518;10697;10903...		http://beta.congress.gov/bill/93th-congress/ho...	A bill to amend the emergency loan program und...	17

Figure 3: Data row showing the aggregated information.

Designs

Since our data has multiple levels of details we can aggregate by and let user interact with the visualization to dig deeper, we experimented with sketches of multiple types of plots using which we could meet our visualization objectives.

Discarded Designs

Streamgraph

Initially we considered using something more similar to what we saw in Figure 1, but we ended up rejecting this idea because there are too many policy areas. We also considered using a similar plot to see how many bills of each policy area make it through to each of the next steps, that is, not only considering *introduced* and *passed* but also all the intermediate stages (one chamber approved, two chambers approved and many others). This idea was messy, and we considered that the information on intermediate stages would not be very relevant for the type of questions we wanted to face and information we wanted to show with our visualization. Since the bills flow in monotonically decreasing volumes from being introduced to passed, we thought about visualizing the flow of bills in different policy areas with increasing congress numbers by using Streamgraph. This idea had the problem of selecting the absolute y axis which would enable us to show the distribution between different policy areas. Also, our time axis is discrete due to the Congresses , thus there would be sudden jumps in the plot and would not look aesthetically pleasing. If we made continuum in design by using redundant slope, that would be confusing to the audience and didn't meet our design objectives.

Network

We have a networked selection where bills are connected to each other through common variables such as same sponsor, same policy area, same congress number in which they were introduced, so we thought of visualizing the aggregation of bills grouped by 1 variable's domain and connected to another variable's domain which can be accessed when one clicks on it . On iteratively clicking user can see individual bills. This idea met our design objectives partly because one could go from the highest level view to the lowest level view by interaction but it was hard to see the evolution of bills in different variables which is the key requirement of our visualization.
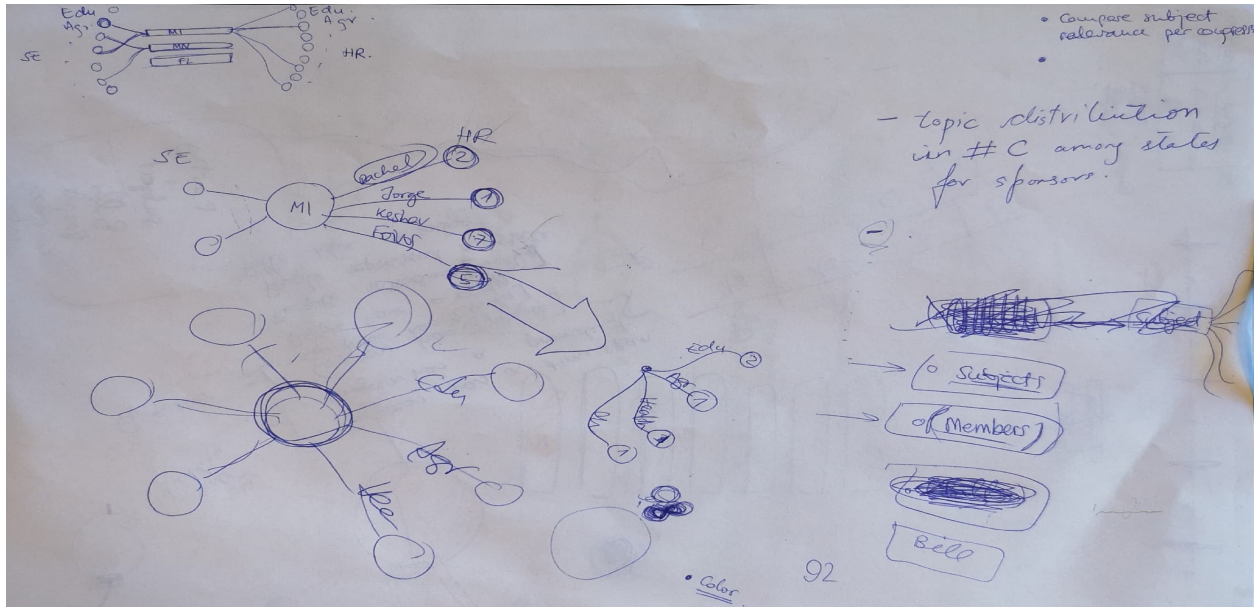
Figure 4: Network layout attempt

Design Implemented

- **Comparison between subject areas** – A bar plot shows the number of bills for the selected congress (after having filtered by party and state, if filtering criteria were chosen), for all different subject areas. Since the policy areas are categorical values without a defined order, they are ordered from largest to smallest bill count values. This sorting allows to easily identify the main policy areas in each scenario.

- **Comparison in time, i.e. comparison between congresses** – A bar plot was chosen in this case as well, to show the number of bills for all congresses (after having filtered by party, policy area and state, if filtering criteria were chosen).

- **Comparison between states** – Map with a logarithmic scale coloring, grouping the values in 6 bins for more clarity and to make it easier to distinguish the states with large or with small numbers of bills.

Bar plots were chosen for comparing the values for different subject areas and for different moments in time, as they are very suitable for comparing different values to each other. design decisions you made using the perceptual and design principles.

We can get an intuition of the fraction of bills that are passed compared to the (considerably larger) number of bills that are introduced, by shifting from *Introduced* to *Passed* bills representations and viceversa.

On the side we wanted to include some information about the chosen congress in order to set the context for the visualization. We also want to see a list of the titles of the bills that matched the selected criteria, and we can click on a specific bill in this side bar to visit the website about the bill and learn more about it.

The colors chosen were very important in our visualization. We chose red and blue for republican and democrat parties respectively. We then chose green to represent independent parties, as this color is not close to either blue or red. Finally we chose purple for the case when several parties are selected at once. Lastly, we use colors in the gray scale in the plots to show that the elements are not selected.
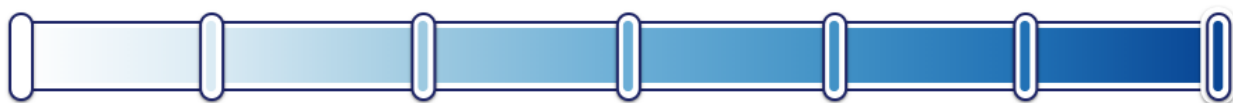

Figure 5: Blue color map.

Figure 6: Green color map.

Figure 7: Red color map.

The buttons in the top bar can be divided in two groups. The ones in the left correspond to the parties and allow to filter by a specific party. Their respective colors are red, blue and green, as described previously for the plots. The buttons in the right group are shown in gray, no color, because they correspond to all parties equally.

A rejected idea was to use the animal symbols that typically represent the parties: an elephant for the republican party and a donkey for the democrat party. However these symbols did not give any more information than the parties' colors that we are using already, and it is not possible to understand their meaning and what they represent unless it was known beforehand.
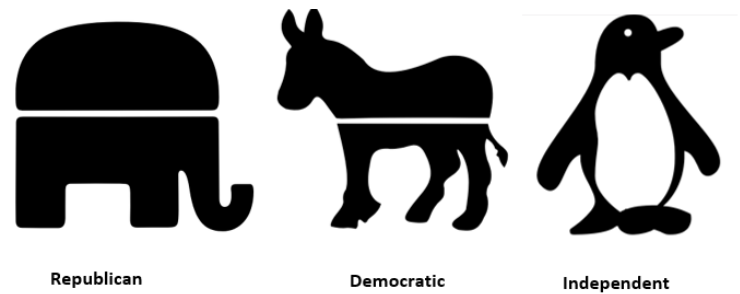

Republican   **Democratic**   **Independent**

Figure 8: Animals representing respective party affiliations .

After we had taken taken the main decisions on what we wanted to see, we still had to reorganize the visualization to make it clearer.
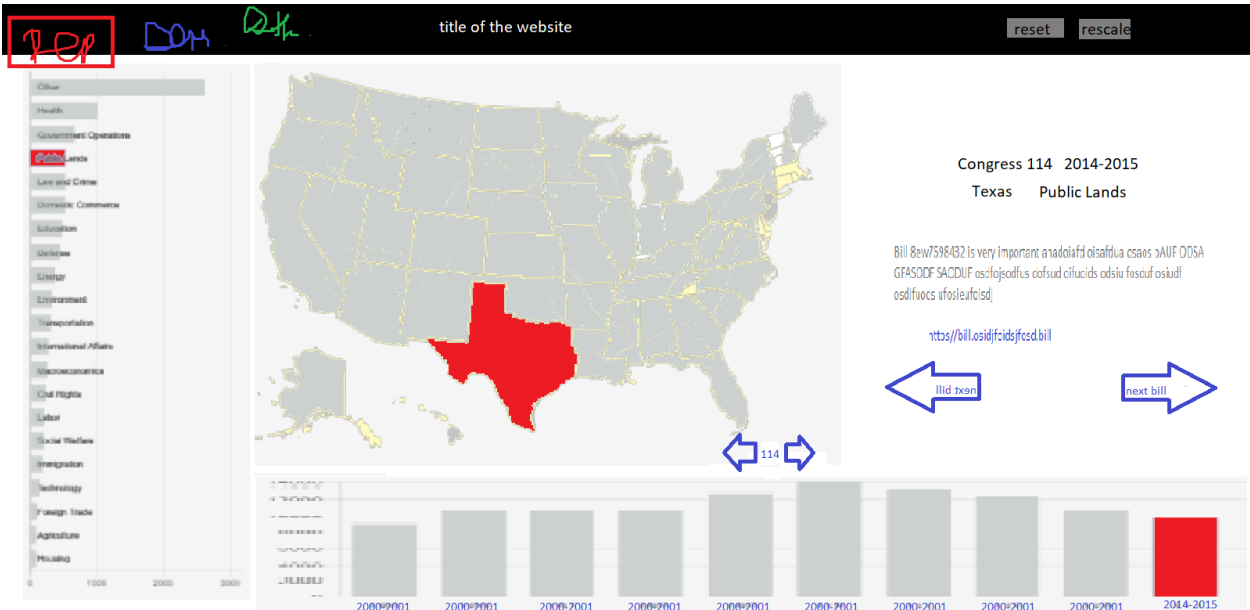
Figure 9: Final sketch.

The intend of our visualization is to provide an interactive and clear way for our target audience to explore information about actions and points of interest of the activities of US congresses and lawmakers. We focused our work on bills because we consider them a straightforward and meaningful way to point out such information. We show the main visualization in Figure 10 . The key components and their philosophies are also delineated below.
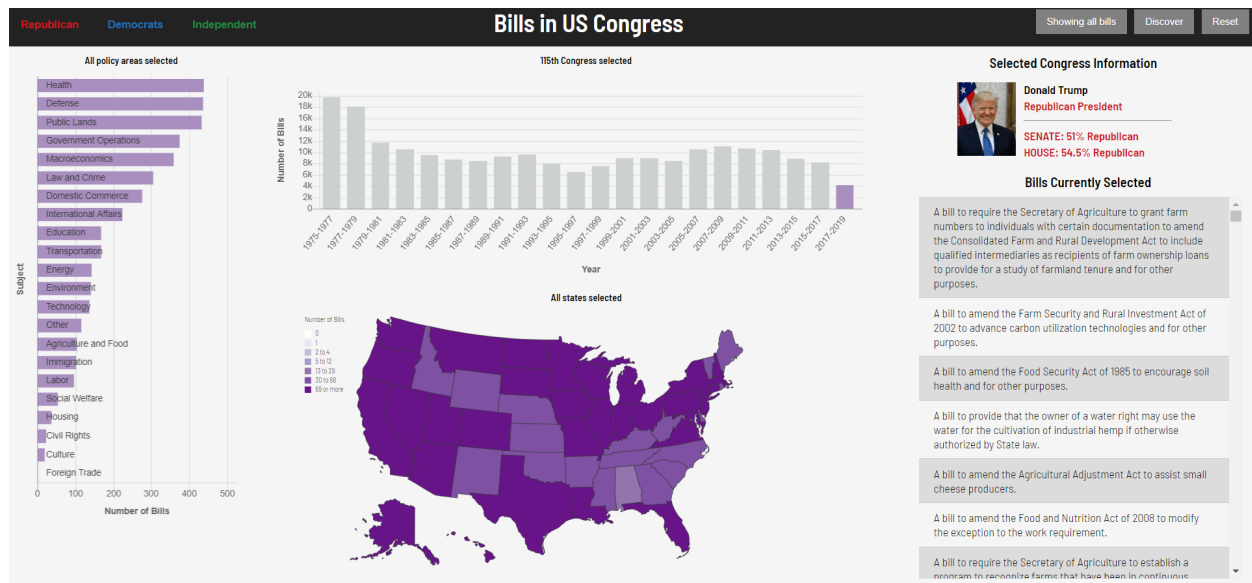

Figure 10: Main Visualization page

Key Elements of visualization

Party Selection

The first way of interacting with the visualization is to select one button out of 3 that can be seen in Figure 11. The data visualized changes on selection, as can be seen in Figure 12. We decided to have this option since although showing an overview of congress activity can be very useful, some people, insisting on their political beliefs, are interested to be informed only about the party they support or even the members of the congress they have voted for during the elections. For that reason we provide the possibility to filter data by party including both the two dominant parties and the members of the chambers that are politically independent.
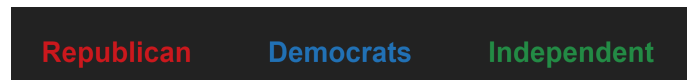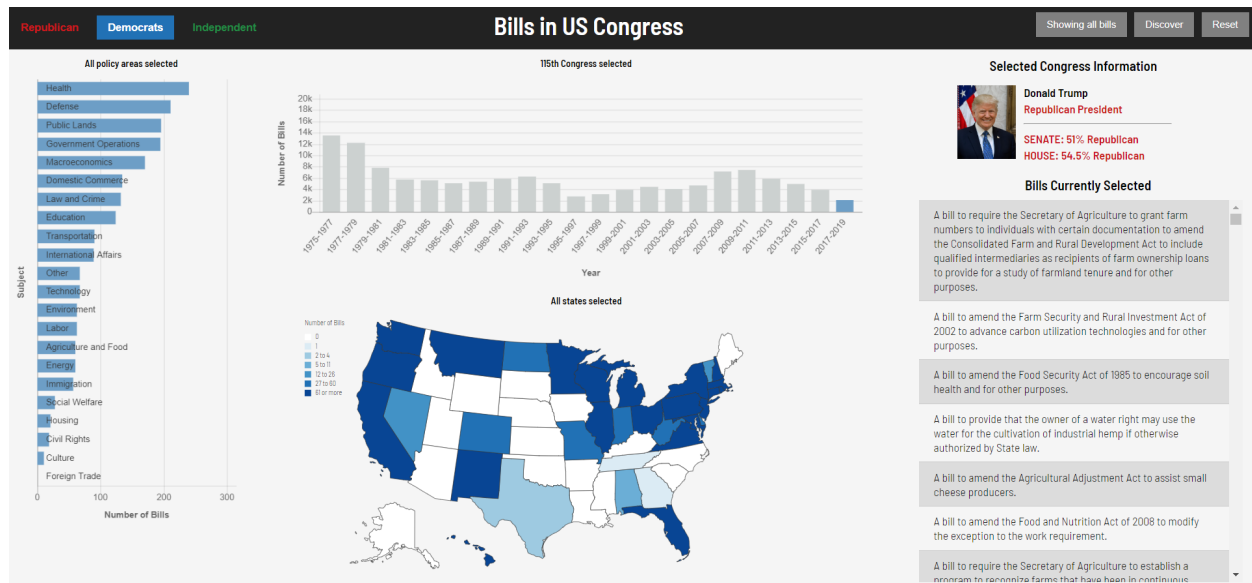

Figure 11: 3 buttons for party selection

Figure 12: Main page On selecting Democrat button

Time Evolution Bar Chart

We visualize the bills introduced or passed (selection can be toggled as explained further) by aggregating the bills by policy areas under selection and states of sponsors under selection

and representing the data by using bars per congress. This is seen in Figure 13. Audience can select different bars to change the main layout depending on which congresses they would like to focus on.
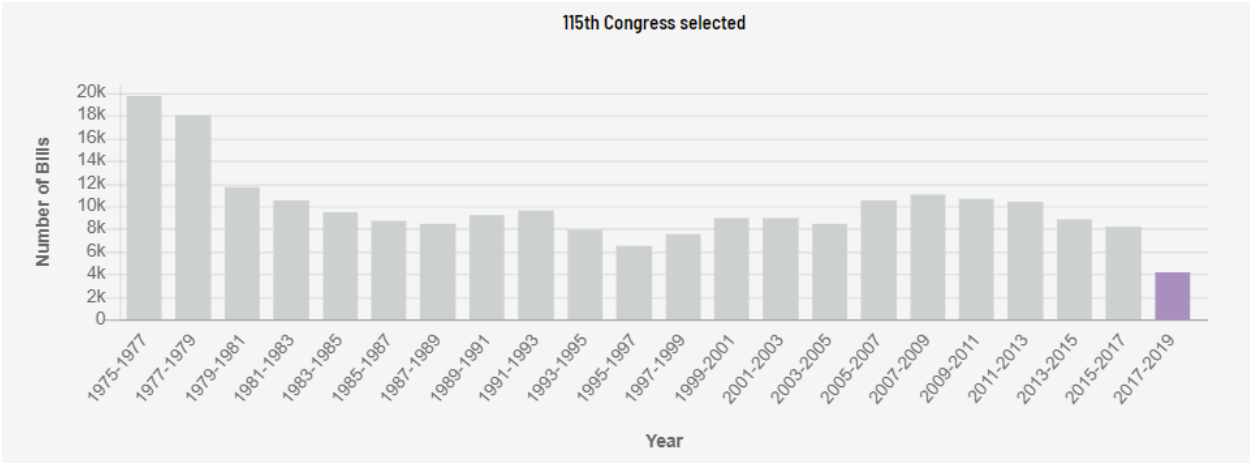

Figure 13: Bar chart showing bills by every congress

Policy Areas Bar Chart

We visualize the bills introduced or passed (selection can be toggled as explained further) by aggregating the bills for current congress under selection and states of sponsors under selection and representing the data by using bars per policy area. This is seen in Figure 14. Audience can select different bars to change the main layout depending on which policy areas they would like to focus on.
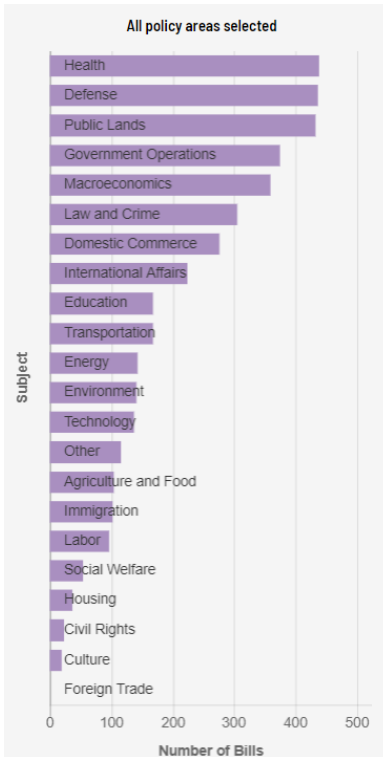
Figure 14: Bar chart showing bills by every policy area

Map Chart

We visualize the bills introduced or passed (selection can be toggled as explained further) by aggregating the bills for current congress under selection and policy areas under selection and group bills by the states from where their sponsors come from. The resulting number of bills/state is shown as a heatmap. This is seen in Figure 15. Audience can select different states to change the main layout depending on which states' sponsors they would like to focus on. Users can also hover over states to see the absolute number of bills from that state.
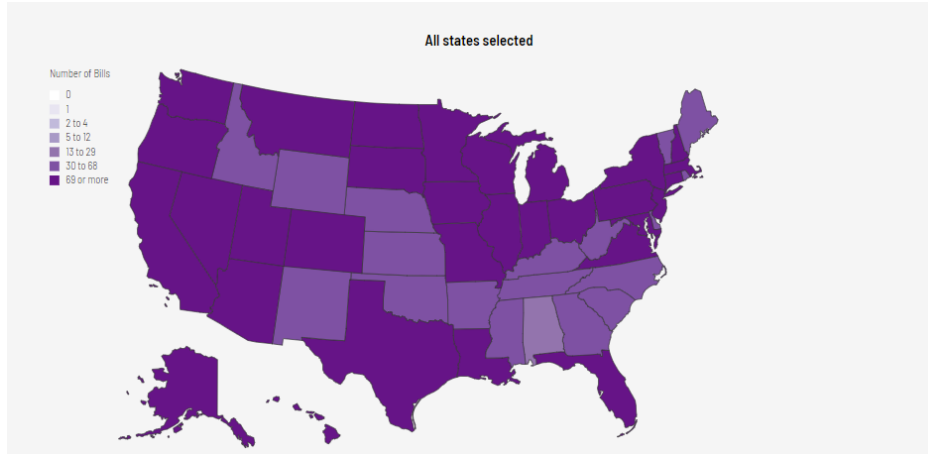
Figure 15: Map chart showing bills by states' sponsorship

List of Bills

Since the plots in Figure 13,Figure 14,Figure 15 focus on the aggregation of blls, we also thought it would be prudent to show what bills are composing this aggregation. Thus we added a scrollable list of bills which make up this aggregation. This can be seen in Figure 16. Every-time the selection is altered by user interaction, this list is updated. User can click on a bill and be directed to the URL explaining that bill in detail.
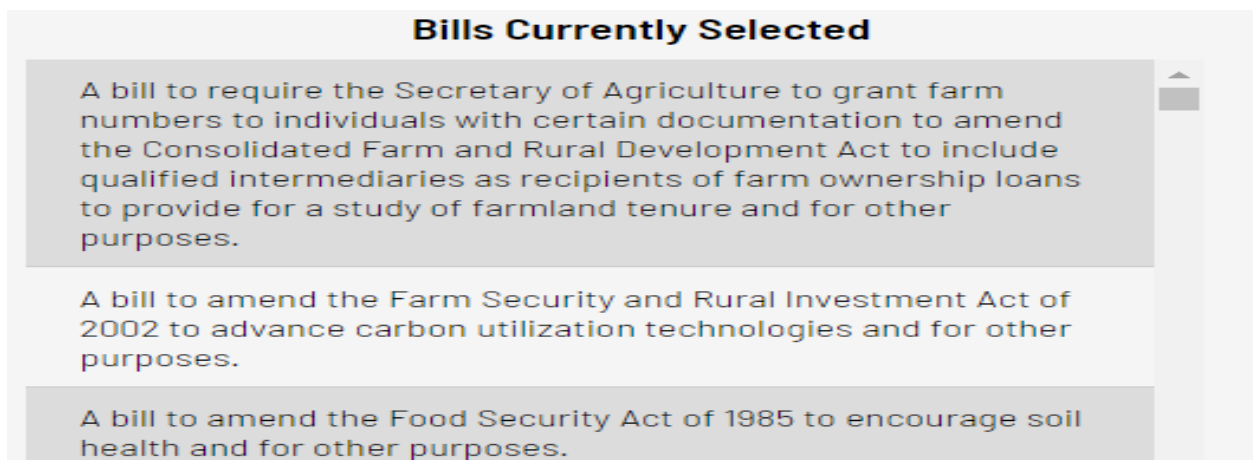

Figure 16: List of bills composing aggregation

Congress related information display

In order to provide more depth into the Congress selected, we decided to show the image of the president and the majority composition of the 2 houses of congress. This can be seen in Figure 17. Once the selection of congress changes, this area is automatically updated.
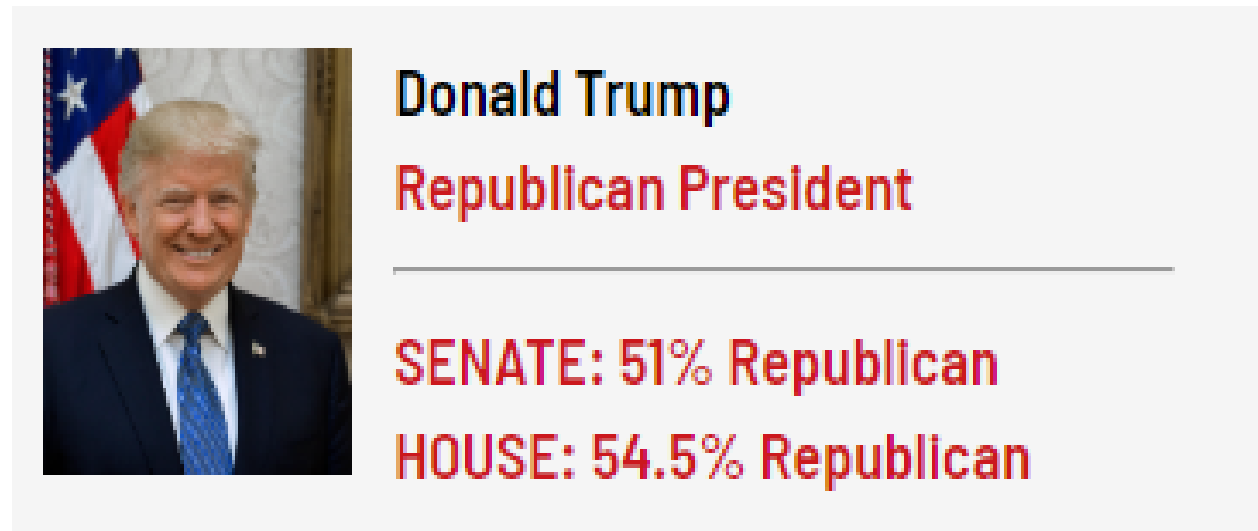

Figure 17: Congress Information

Discover Mode

We have a very interesting feature called the discover mode which the users can click on and obtain more insights about the data. We have custom curated some trends and related information which would enable the user to understand more effective utilization of the tools we have made. The Discover button can be seen in Figure 18 and can be clicked. On clicking, the main visualization changes as can be seen in Figure 19.
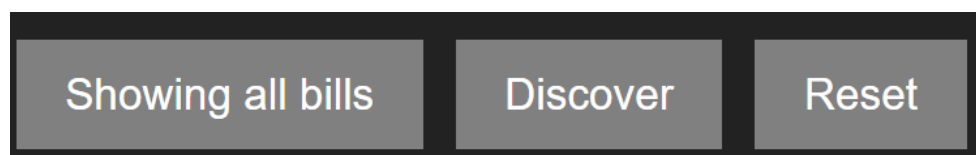

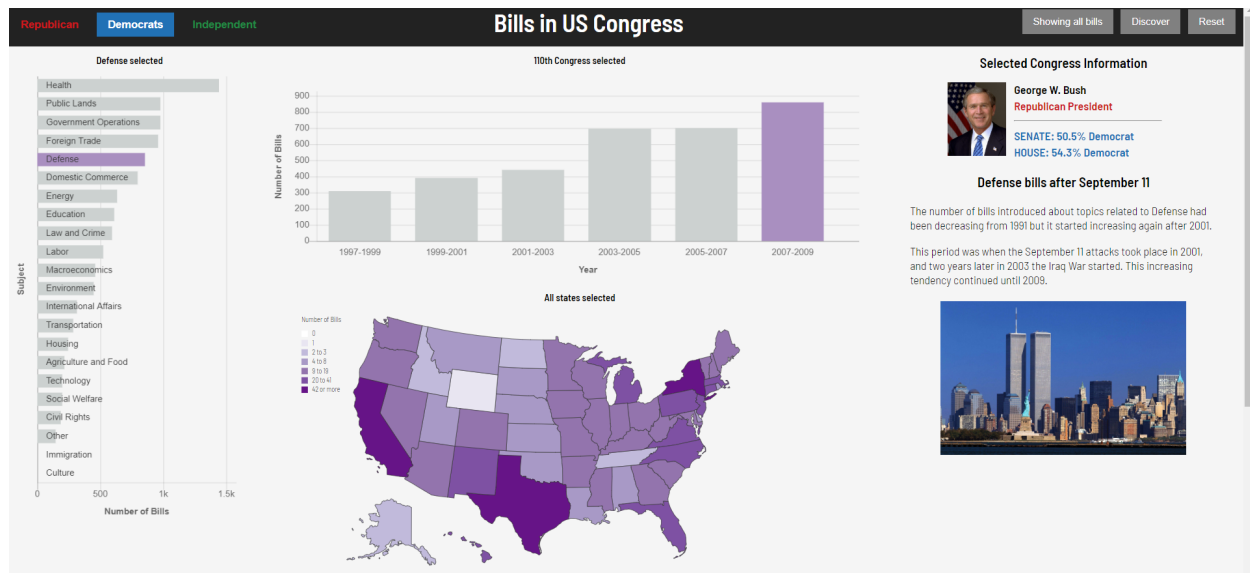Figure 18: Discover , Reset and Show which bills buttons

Figure 19: Effect of clicking Discover button

Reset Button

Audience can click on Reset button to revert back to initial visualization we had when website loaded. The reset button can be seen in Figure 18.

Select which bills to visualize.

Audience can select if they want to focus on visualizing bills introduced in congress or bills passed by congress. The Showing all bills toggle to Showing Passed bills button can be seen in Figure 18.

Our opinion on implementation

Bringing all this options together we have a quite effective tool that enables users to delve deeply into the data and retrieve the information they want. For example, a farmer form Texas that has voted for Republicans the past years, can can easily navigate to bills about Agriculture and Food introduced from congress members of his state and see if they are in favor of him, or even make a comparison with the respective information for Democrats.

In our visualization we provide the option to navigate both through introduced and passed bills. The latter show a good indication about the quality of coordination within the members that form a congress and and even show the effectiveness of different congresses throughout the years. Meanwhile, introduced bills might have not been passed because of disagreements among the members but can denote the overall interest of the parties to specific policy areas, displayed as aggregated information for the whole U.S. or even investigated by state.

The whole process for a bill from being introduced to be passed is quite complicated. It includes multiple steps including advocacy with the intention of influencing decisions (Lobbying), debates and marks up on proposed bills by committees and so on. Showing all these information in an interactive manner could be quite difficult and would probably lead to a confusing visualization. For that reason, we decided to attribute a bill to the members that sponsor the bills and more specifically to the main sponsor. Although by applying this approach we abstract information from the actual process, we still give a good indication about the area of interests of the parties and the congress members. This information is being preserved because these days the two main parties are fairly polarized, which means that rarely a bill has cosponsors that do not belong to the same party.

What did you learn about the data by using your visualizations? How did you answer your questions?

In general we could also see that when the majority of both Senate and House of Representatives belong to the same party, the number of bills that are introduced by members of that party are larger than the ones introduced by members of the other party, which makes sense with what we were expecting to find in the data. Afterwards we used two different approaches:

The first one was to brainstorm different events and patterns that we knew could possibly be present in our data, and then to verify if we could indeed identify them. This was the case for the increase in bills introduced in the Defense policy area after September 11.

The second approach consisted on looking at the common patterns that we were seeing and filtering by different criteria in order to try and find interesting insights. The insights obtained were later verified by doing some research on the topic. An example for this approach was how we could see from the main topics that are present in bills by Alaska representatives include Public Lands, which led us to guess that Alaska is one the states with the largest number of Native American populations in the US.

In both approaches it was useful to look at some of the specific bills that are included in the subset selected – especially in the case of policy areas, as we do not know for some topics in which policy areas they would be considered to be included.

18

How well does your visualization work, and how could you further improve it?

The use of colors and the types of plots chosen work nicely in this visualization.

Possible improvements that could be done in the visualization are to find a way to highlight in the map the state that has been selected without *hiding* that state's original color by using a solid, constant color. On the one hand, the representation that we chose makes it very clear that a specific state has been selected. On the other hand, we can no longer compare the value for the chosen state with the values for all the un-chosen states. It is possible that we could find a way to face this issue and reach a solution that meets both needs.

The visualization would also be improved if we could somehow sort the bills displayed on the side by some set of criteria, so that the bills that have been more relevant or that have been considered the most interesting by some users are shown first (for example, based on the number on times each bill has been visited). To do this we would need to modify the current implementation, as it would be considerably more computationally expensive.

It would also be a very interesting improvement to use a data source that updates with a high frequency, so that we can keep our visualization up-to-date.

Lastly, as the policy areas have slightly changed over time (some new ones being defined, split, merged...), some preprocessing work could be done to obtain more consistent plots.

- Preparation – were they prepared during team meetings?
  Yes (for all)

- Contribution – did they contribute productively to the team discussion and work?
  Yes (for all)

- Respect for others' ideas – did they encourage others to contribute their ideas?
  Yes (for all)

- Flexibility – were they flexible when disagreements occurred?
  Yes (for all)

[1] ProPublica Congress API.
https://projects.propublica.org/api-docs/congress-api/

[2] Center for American Politics and Public Policy, University of Washington. 2014. "Legislative Explorer Dataset."
http://cappp.org/index.php/data/.