**Bill Sorting Methodology**

**Overview of the Process**

In our last semester, we began sorting bills in a few of our topics of interest, with the goal of sorting through every bill in the last two presidential administrations (2017 – 2024) and identifying which bills should be used to measure each legislator’s stance and behavior towards the topics.

To do so, we used a three part process:

1. Create a *.csv* file containing important information on each bill in the time period.
2. Do a preliminary sort of the data for bills that may be related to the topic (via code).
3. Manually sort the resulting data to confirm whether each bill is both relevant and useful in determining legislators’ stances.

***Topics of Interest:*** *NASA, Domestic Oil and Gas, U.S. Veterans, Wildfire Prevention and Recovery, Flooding and Tropical Storm Prevention and Recovery, Domestic Agriculture*

**Retrieving Bill Data**

There are two Python notebooks in this repository. We used *pull\_all\_bills.ipynb* to return a *.csv* file containing data related to each bill in the time period. The program will save that data as *all\_bills.csv*.

Data attributes:

* ***Unique ID*** – The bill’s ID, which is strictly unique. It contains the bill type (HR, HRES, etc.), number, and year in lowercase.
* ***Bill ID*** – The bill’s type and number, in lowercase.
* ***Year*** – The year the bill was introduced.
* ***Congress*** – The House congressional number in which the bill was introduced. Each House congressional number spans a two-year period containing the same legislators (with some exceptions!).
* ***Link*** – A link to the bill’s *congress.gov* website. More information on the bill (such as the bill’s text) can be found here.
* ***Title*** – The bill’s title.
* ***Subjects*** – A list of relevant topics that the bill is associated with. These topics are assigned by *Congress.gov*, a government source.
* ***Committees*** – A list of House Committees that the bill was assigned to.

**Sorting Bills**

The second Python notebook, *bill\_sorter.ipynb* does a fast initial sort of *all\_bills.csv* and returns a shortened *.csv* file with the same attributes, titled *sorted\_bills.csv*. To make this initial sort, edit the following variables:

* ***keywords*** – A list of words or word parts that will be searched for in each bill’s subjects and title. All words and word parts should be entered in lowercase. Note that searching part of a larger word (such as *health*) will return all bills whose subjects include the whole word *health* as well as those that include *health* as part of a larger word (such as *healthcare*).
* ***search\_comms*** – A list of words or word parts that will be searched for in the bill’s committees. The same principles as above apply here.
* ***remove*** – A list of words or word parts that will be searched for in the bill’s subjects and title. Any bill containing these words or word parts will be removed from the output. This can be used to easily sort out large funding bills (that cover numerous subjects, and are thus too broad to measure specific congressional support for a topic, via *“appropriation”*) and bills to rename institutions (that are too frivolous to be used to measure a legislator’s behavior towards a topic, via *“rename”* and *“renaming”*).

**Manual Sort**

Once a shortened list of bills have been returned, we did a manual search of the returned bills to determine whether or not each bill is both relevant and useful. This part of the process is more open ended, but follows a couple guiding questions:

* *Is the bill related to this topic, and this topic alone? In other words, could a legislator’s behavior regarding this bill be attributed to some other subject?*
* *Is the bill substantive enough to reflect the legislator’s position or behavior regarding the topic? In other words, would the legislators’ positions be communicated through this bill?*
* *Could it be argued that both the YEAs and the NAYs support the topic, just from different ideological perspectives – or is one group of voters in support and one against? If the latter, the bill is more useful in determining stance.*