

Scott Renshaw  
Heath Reineke

## Final Project Manifest

We scraped bills from three different sources:

### Using Selenium (a Python tool for web testing/web scraping):

- Source: <https://www.congress.gov/>
  - Originally planned to use this site to scrape all bill statuses, but the site is really unreliable. It also had bot detection, meaning we couldn't just simply use the requests module in Python and instead had to use Selenium, which caused some headaches.
  - Code we started to scrape this site is `ifCongressDotGovWorking()` in the `congress_scraper.py` file.
- Source: <https://www.govtrack.us>
  - Ended up using this site to get all bills that failed because they didn't pass the senate, all bills that failed because they didn't pass the house, all vetoed bills, and all vetoed-then-overridden bills.
  - Still used selenium here because of bot detection. All code used to scrape this site is in `main()` of `congress_scraper.py`

### Using JSON files:

- Source:  
<https://www.propublica.org/datastore/dataset/congressional-data-bulk-legislation-bills>
  - Used this to get all bill data from the 2011-2012 congress session. This included bills in statuses not mentioned as types that we scraped from govtrack.
  - Here we were able to just use the json module that Python offers instead of having to use selenium, requests, etc.