Where Will Bill End Up?
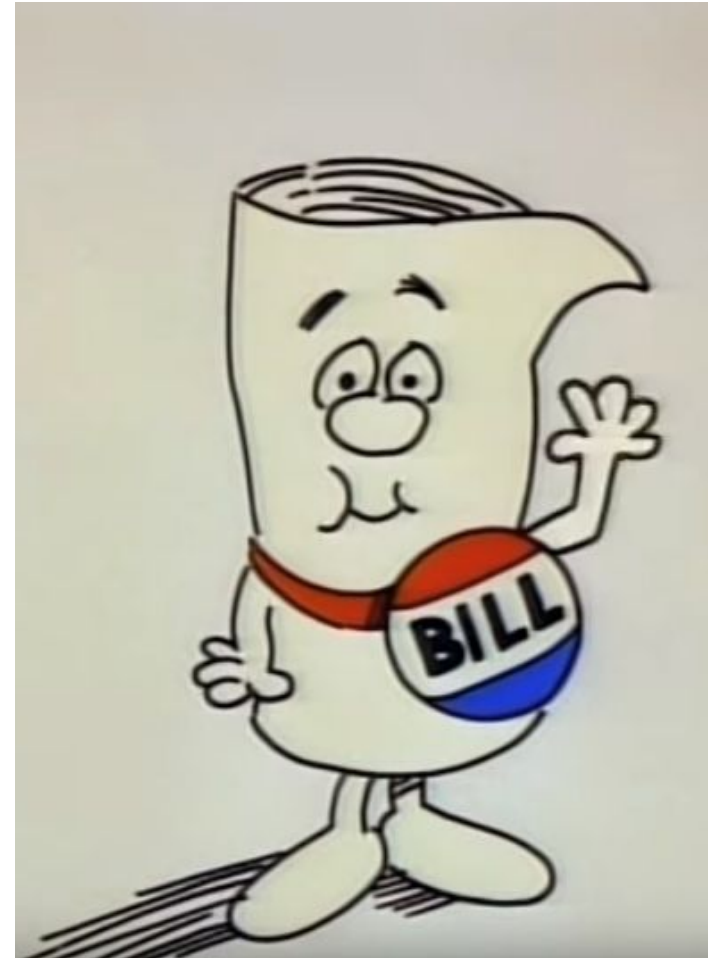
An analysis of 20 years of congressional bills

Overview

- An analysis 20 years of bills introduced across 10 Congresses
  - Congresses 105-115
  - 1997 - 2017
- Predictive model of final status of bill -- will Bill make it out of committee?

Bill's Path

1. Introduced by sponsor in either House or Senate
2. Referred to committee/subcommittee
3. Committee hearings and/or markup
4. Referred to full body
5. Passed by full chamber
6. Referred to other body
7. Passed by other chamber
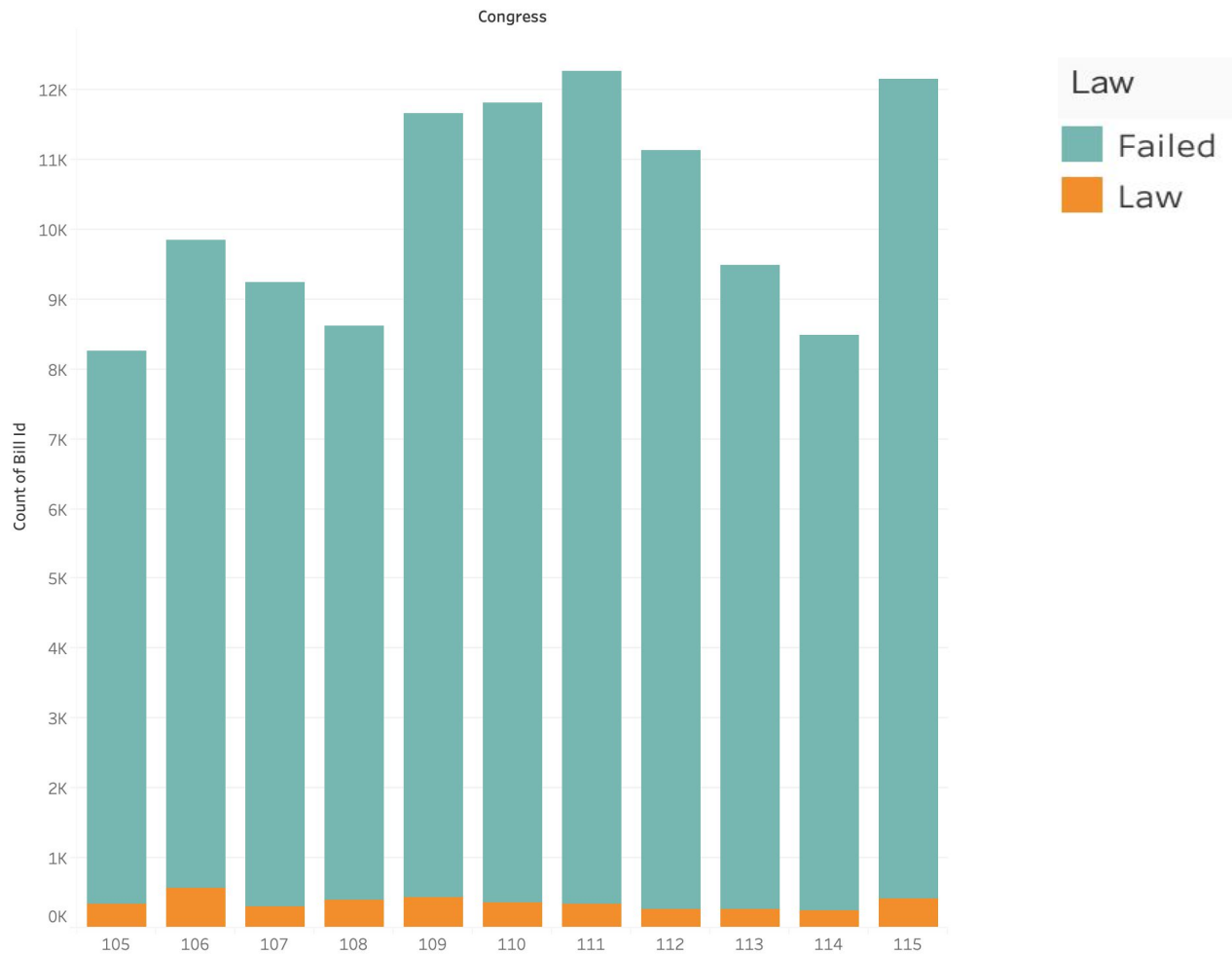8. Sent to president for signature
9. Vetoed or enacted

Count by Bill Type
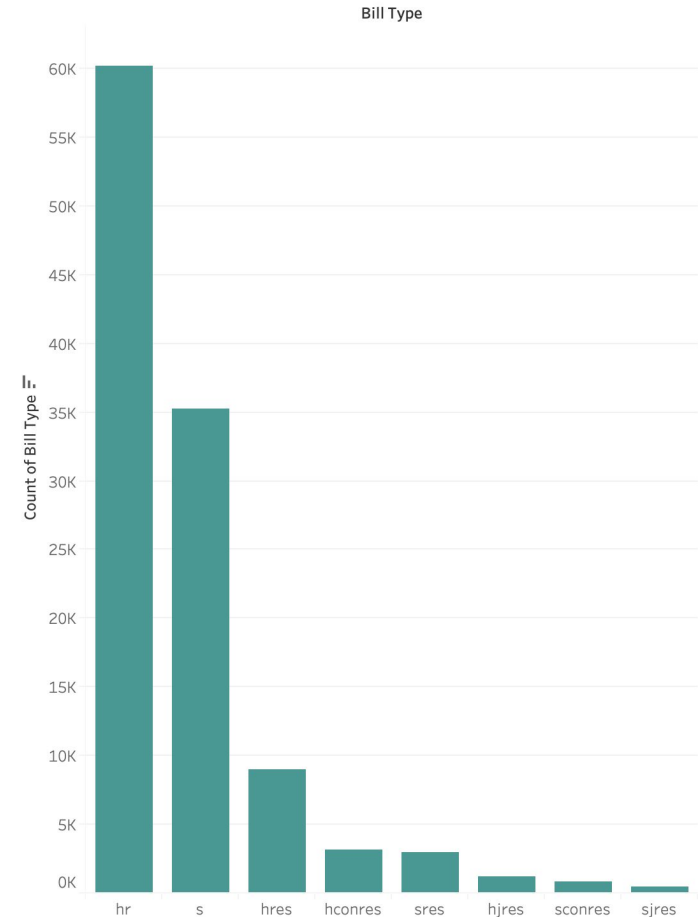
Baseline Statistics

- Average of about 10,000 bills introduced each Congress
- 82% of bills make no progress
  - "Read twice and referred to committee"
- 3.4% of bills introduced over 10 years became law
  - 3-5% per Congress

Count of Bills/Laws by Congress

Data

- ProPublica Data
  - ID number
  - Sponsor name, state, party
  - Type
  - Title
  - Number and party of the cosponsors
  - Committees to which the bill was assigned
  - Primary subject of the bill
  - Summary
  - Latest major action on the bill
- MIT Data
  - Member seniority
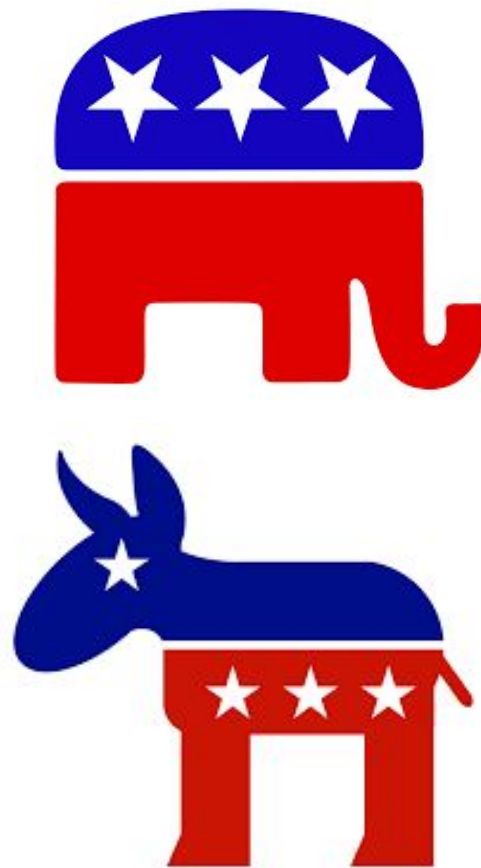 - Member committee assignments

Data Wrangling

- Matching names to join data
  - Same lawmaker has slightly different names in different datasets
    - James → Jim, Elizabeth → Liz, Charles → Chuck
  - Fuzzywuzzy string matching
    - Levenshtein Distance between two sequences
    - Matched everything with score over 85
- Consolidating final status categories
  - 7000 unique final actions
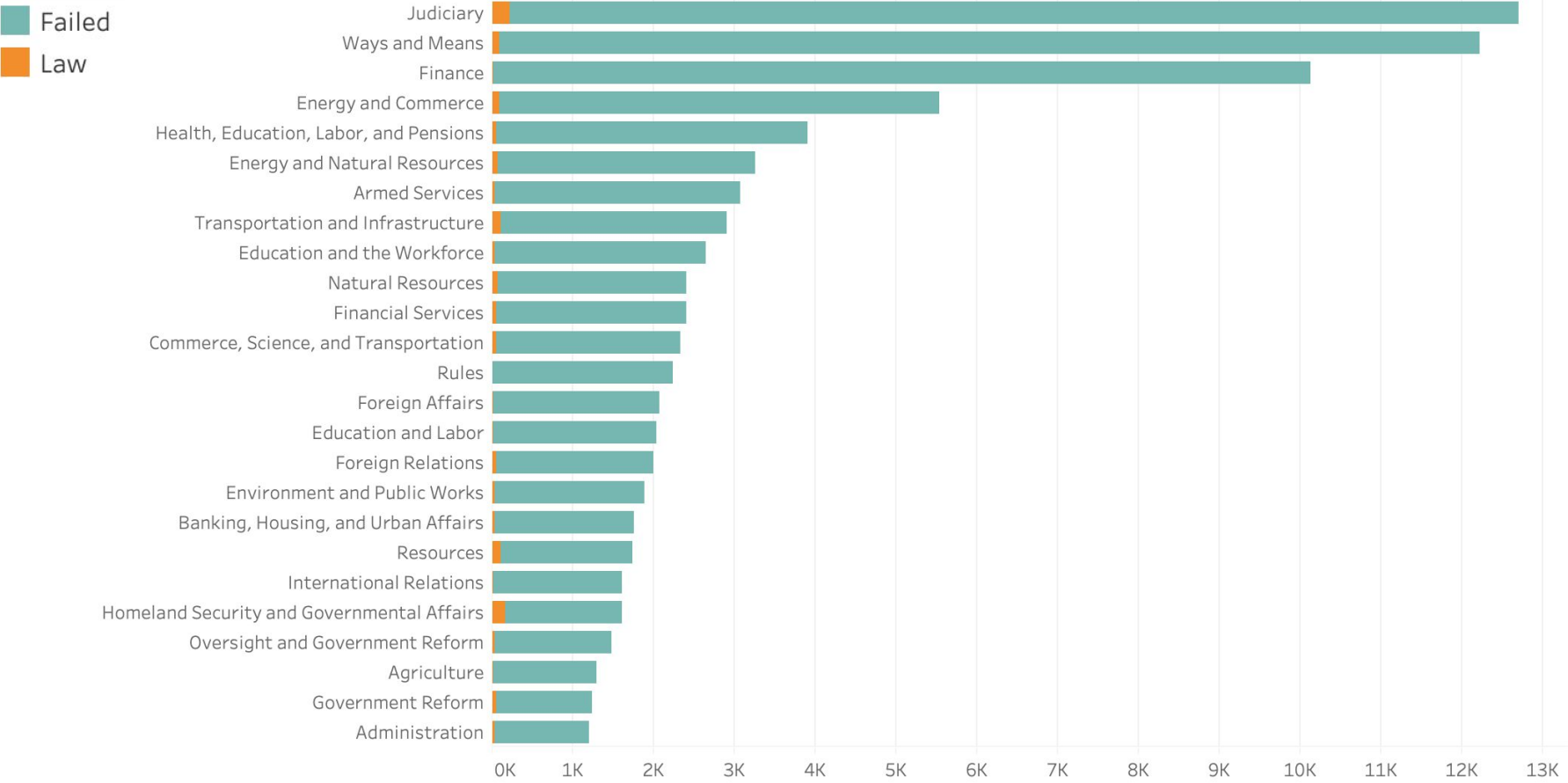 - Elaborate else if function used to consolidate into five categories

```python
def get_final_status(action):
    action = action.lower()
    if ("Hearings held").lower() in action:
        return "Hearings held"

    elif ("Read twice and referred to Committe").lower() in action:
        return "Referred to committee"

    elif ("filed written report").lower() in action:
        return "Written report filed by congressperson"

    elif("Became public law").lower() in action:
        return "Became public law!"

    elif ("Held at the desk").lower() in action:
        return "Held at the desk"

    elif("Submitted in the Senate, considered, and agreed to without ar
        return "Senate agreed to unanimously"

    elif ("Placed on the Union Calendar").lower() in action:
        return "Placed on the Union Calendar"

    elif ("Motion to reconsider laid on the table Agreed to without obj
        return "Dead on the floor"

    elif ("Received in the Senate").lower() in action:
        return "Received in the Senate"

    elif ("Senate Legislative Calendar").lower() in action:
        return "Placed on Senate Calendar"

    elif ("Placed on the House Calendar").lower() in action:
        return "Placed on House Calendar"
```
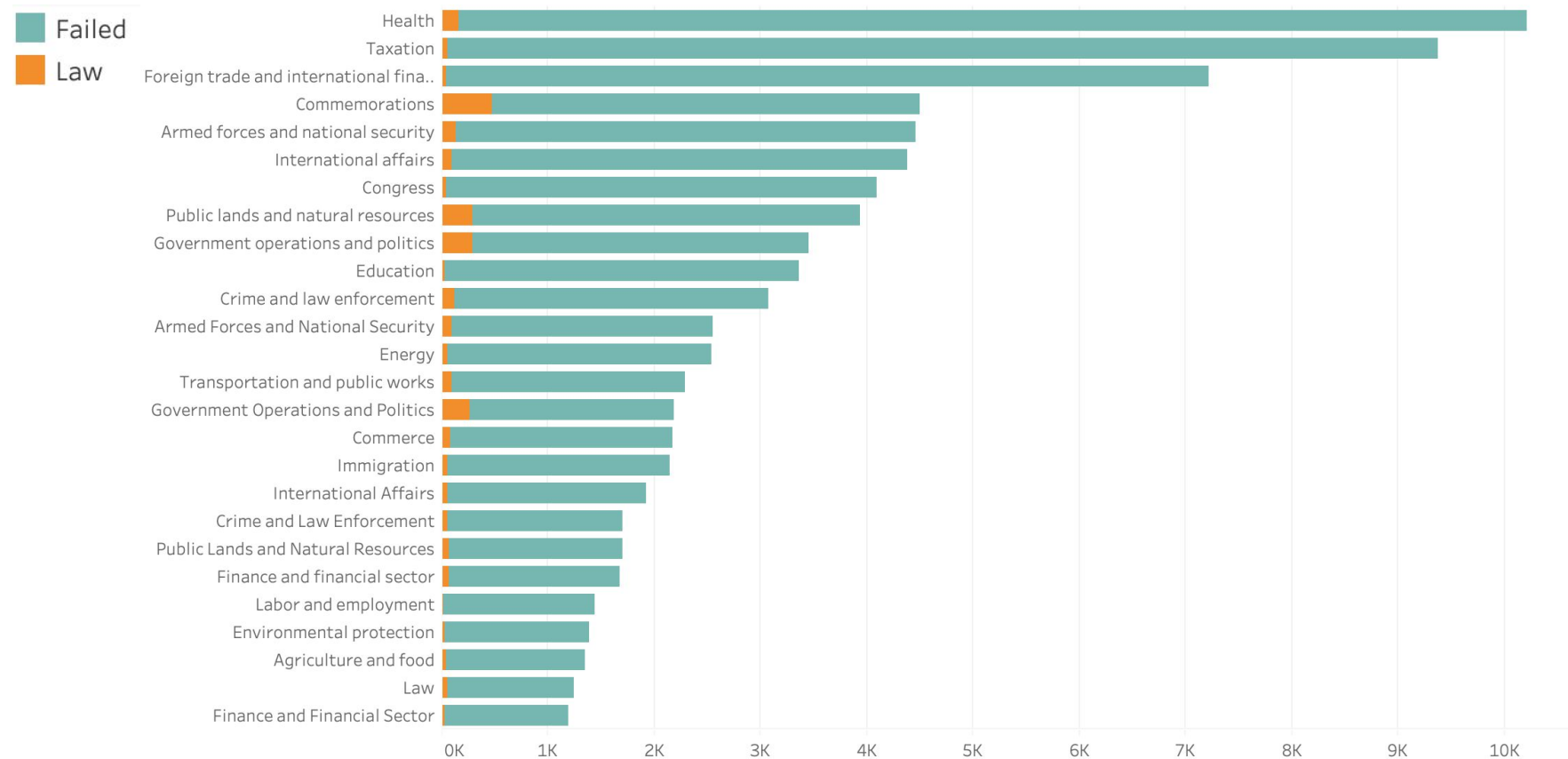
Feature Engineering

- Committee/sponsor match
- Bipartisan
- Summary length
- Sponsor in majority/minority
- Sponsor/president party match
- Dummying categorical variables:
    - Type
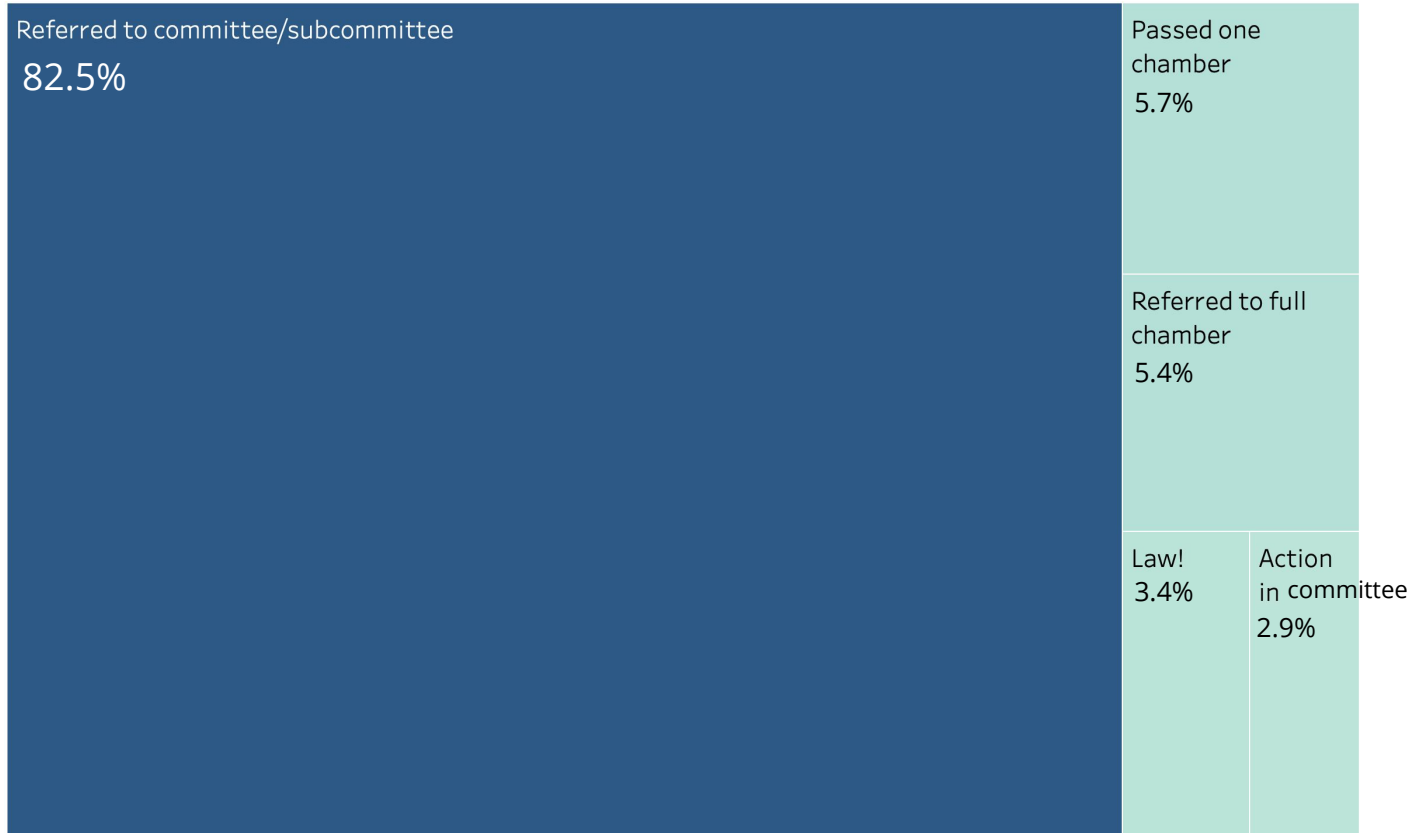   - Subject
    - Committees
    - Sponsor region

Committee Count

Primary Subject Count

Legend: Failed, Law

Subject	Value
Health	
Taxation	
Foreign trade and international fina..	
Commemorations	
Armed forces and national security	
International affairs	
Congress	
Public lands and natural resources	
Government operations and politics	
Education	
Crime and law enforcement	
Armed Forces and National Security	
Energy	
Transportation and public works	
Government Operations and Politics	
Commerce	
Immigration	
International Affairs	
Crime and Law Enforcement	
Public Lands and Natural Resources	
Finance and financial sector	
Labor and employment	
Environmental protection	
Agriculture and food	
Law	
Finance and Financial Sector	

Final Status of Bill: Five Categories

Prediction

- Unbalanced classes
  - Bootstrapping with sklearn's resample
- Don't trust accuracy
  - Recall: the percentage of total relevant results correctly classified by your algorithm
  - Precision: the percentage of your results which are relevant

Important Features

1. Number of cosponsors
2. Sponsor party rank
3. Sponsor committee match
4. Bipartisan
5. Length of summary
6. Sponsor in the majority
7. Multiple committees
8. Committee
   a. Energy and Natural Resources (positive)
   b. Finance, Ways and Means (negative)
9. Subject
   a. Public lands and Natural Resources (positive)
   b. Taxation (negative)
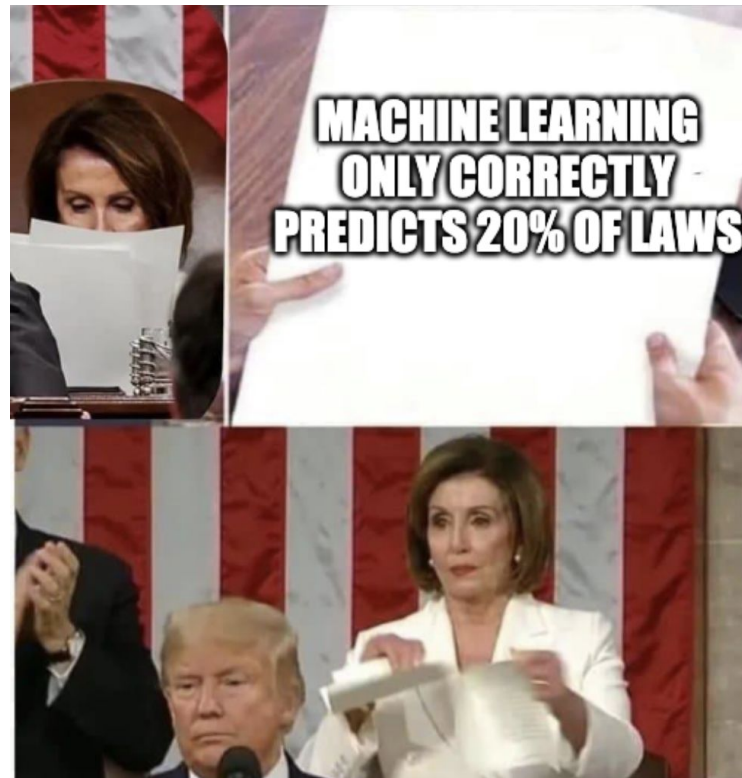
Law Model

Will Bill become a law?

Random Forest Classifier:

- Accuracy: 96%
- Recall: 20%

The model is basically classifying every bill as failed - but getting high accuracy because the vast majority do fail.
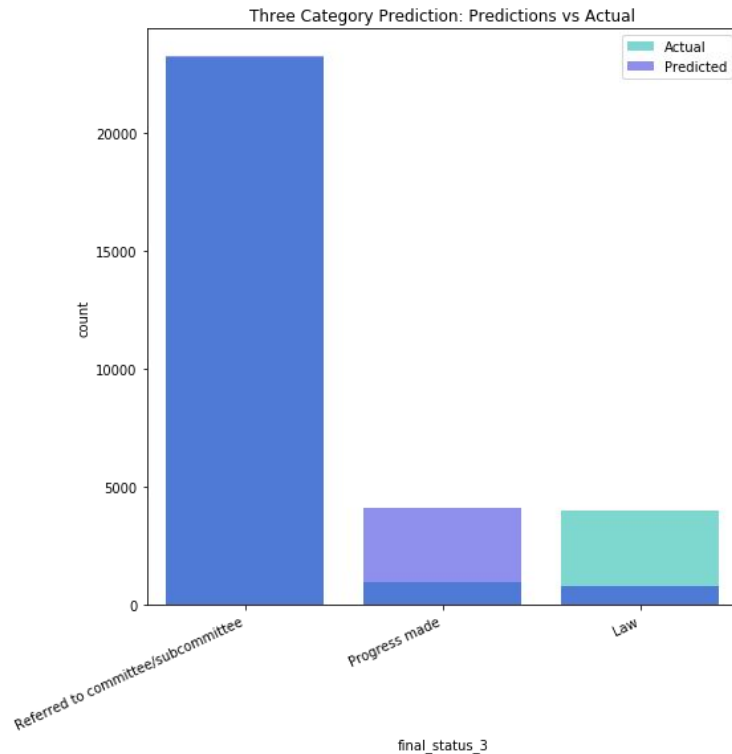
Three Category Model

Logistic Regression

- Accuracy: 71%
- Recall:
  - Law: 67%
  - Progress Made: 51%
  - Referred to committee: 75%

Somehow, once the model is given more choices, it's better at classifying the laws. It is now over predicting law vs progress made.

Three Category Prediction: Predictions vs Actual
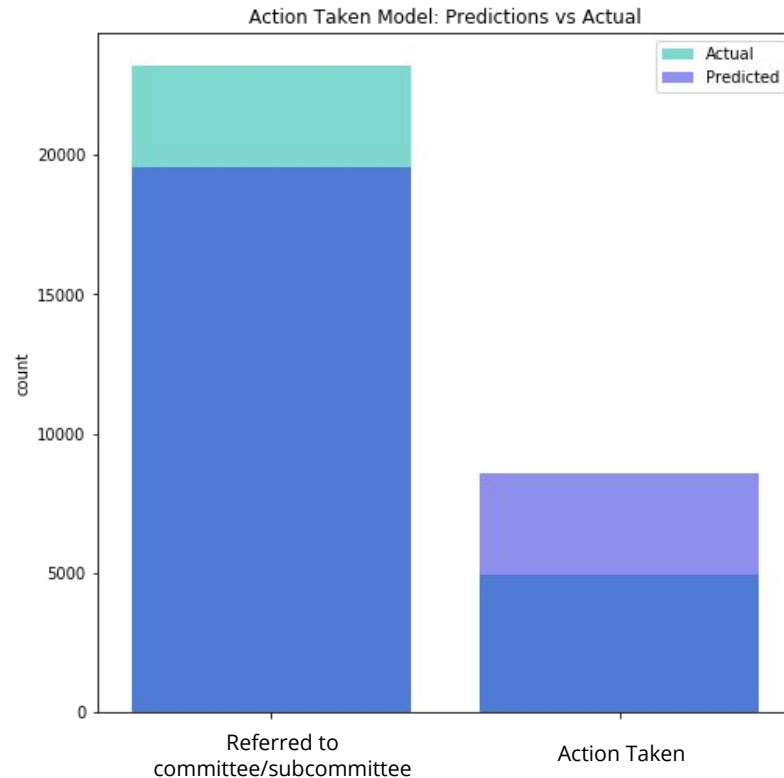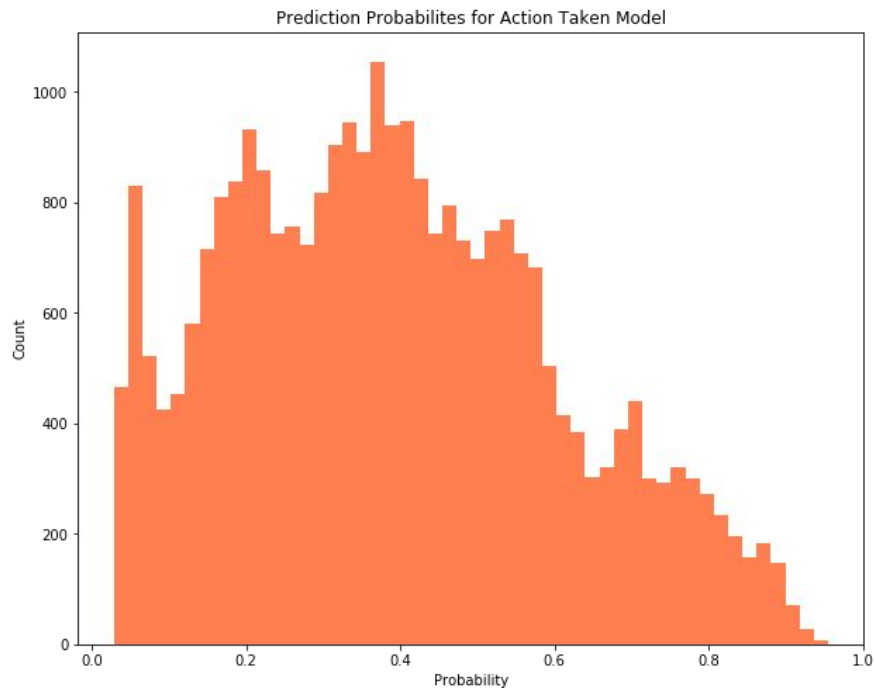
"Action Taken" Model

Will Bill get any action?

Random Forest Classifier

- Accuracy: 79%
- Recall: 77%

The model overpredicts bills becoming laws, but does correctly catch 77% of laws.

A Closer Look at the Results

Prediction Probabilites for Action Taken Model
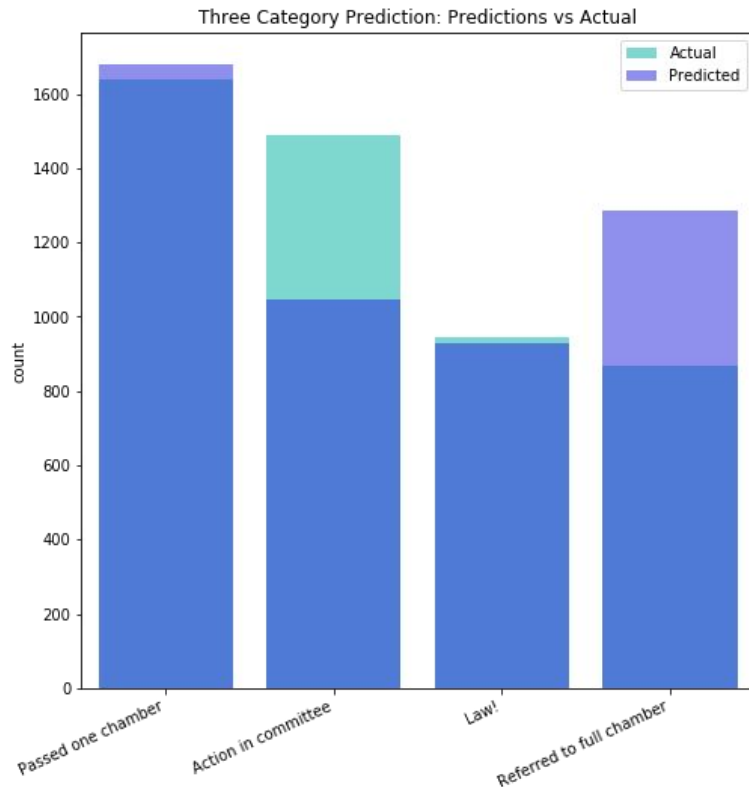

Action Taken Model: Predictions vs Actual

Two Stage Model: Second Stage

Random Forest Classifier using only data for bills that got some action (excluding all that died at referred to committee).
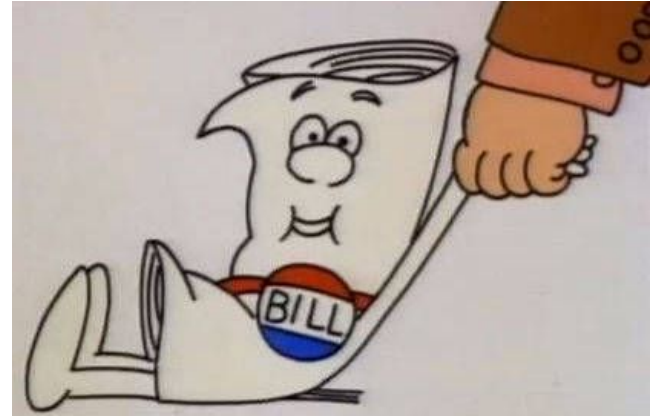
Accuracy: 62%

- Recall:
  - Law: 57%
  - Passed one chamber: 49%
  - Referred to full chamber: 67%
  - Action in committee: 66%

Conclusions

- With the current data, predicting which bills will become law is not productive
- It is possible to predict with fairly high recall which bills will make it out of committee -- and this is a valuable that allows us to focus on bills that actually matter

Further Analysis

- NLP predictive model based on bill text
- Statistical analysis to look for relationships in data
  - Are members from a certain region more likely to propose certain types of bills?
  - Is there a statistical difference between male and female lawmakers?
    - Do they propose different types of bills?
    - Do they vote differently?
- Bayesian statistics
  - Given x characteristic, what are a bill's chances of becoming law, making it out of committee, etc.