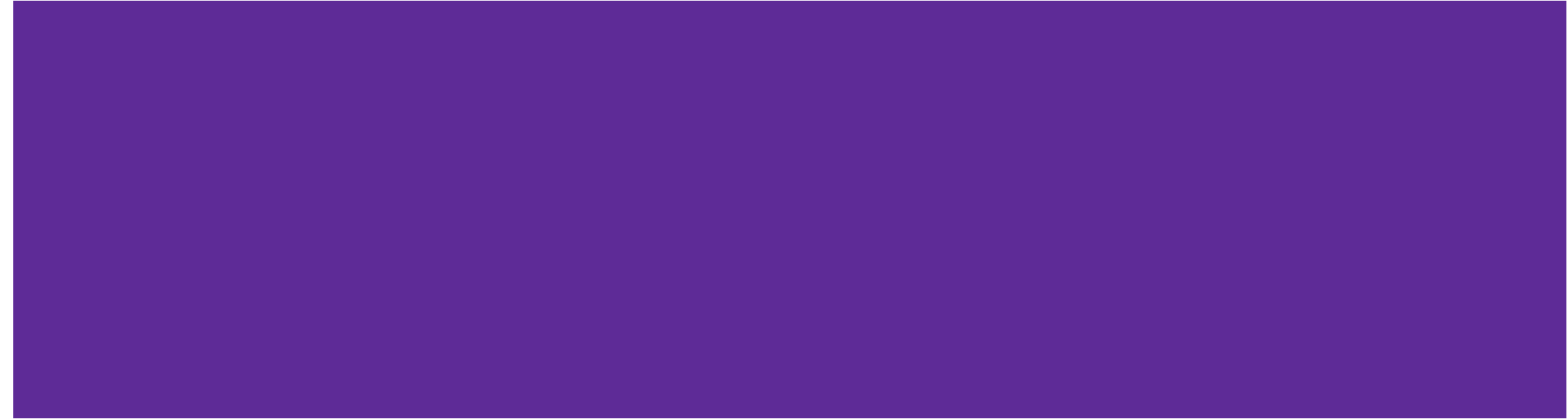


**What can we learn from voting
records of a past congress?**



Dataset

- Data from UC Irvine Machine Learning Repository - 1984 US House Voting Records
- “Key votes” identified by the Congressional Quarterly Almanac
- Categorical Data: Votes are Yes, No, and “?”

Limitations

- Key Vote Selection¹:
 - Controversy or impact
 - Test of presidential power
- 16 votes - limited data
- One year
- “?” votes have multiple categories
 - Voted present - conflict of interest or other reason
 - “Did not vote or otherwise make a position known”

¹ <https://www.jstor.org/stable/439750>

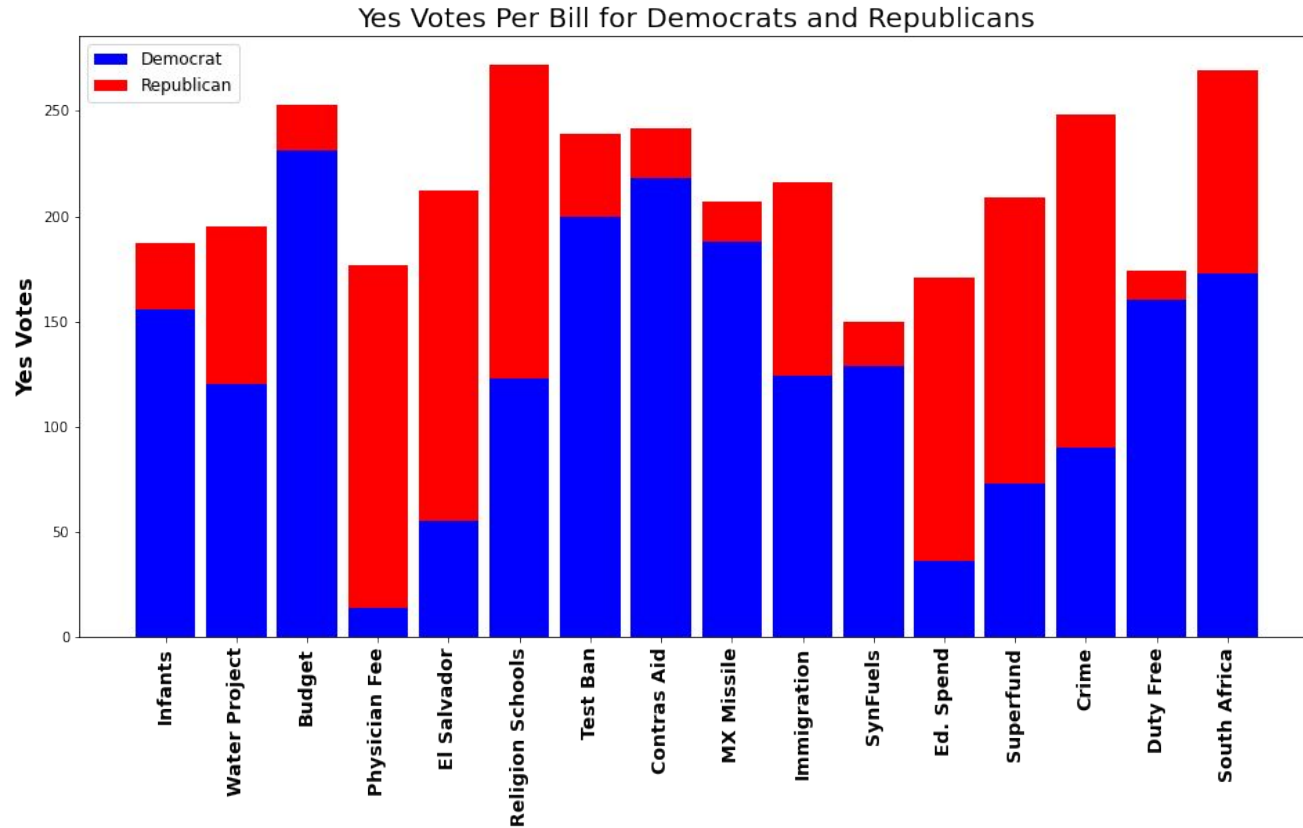
Goals

- EDA to look for anything interesting
- Classification model to predict Republican or Democrat based on voting record
- Cluster analysis to look for more than 2 clusters of voting records
 - Are there important caucuses within the parties, or a 3 way split between moderate, left-wing, and right-wing voting records?

Exploratory Data Analysis



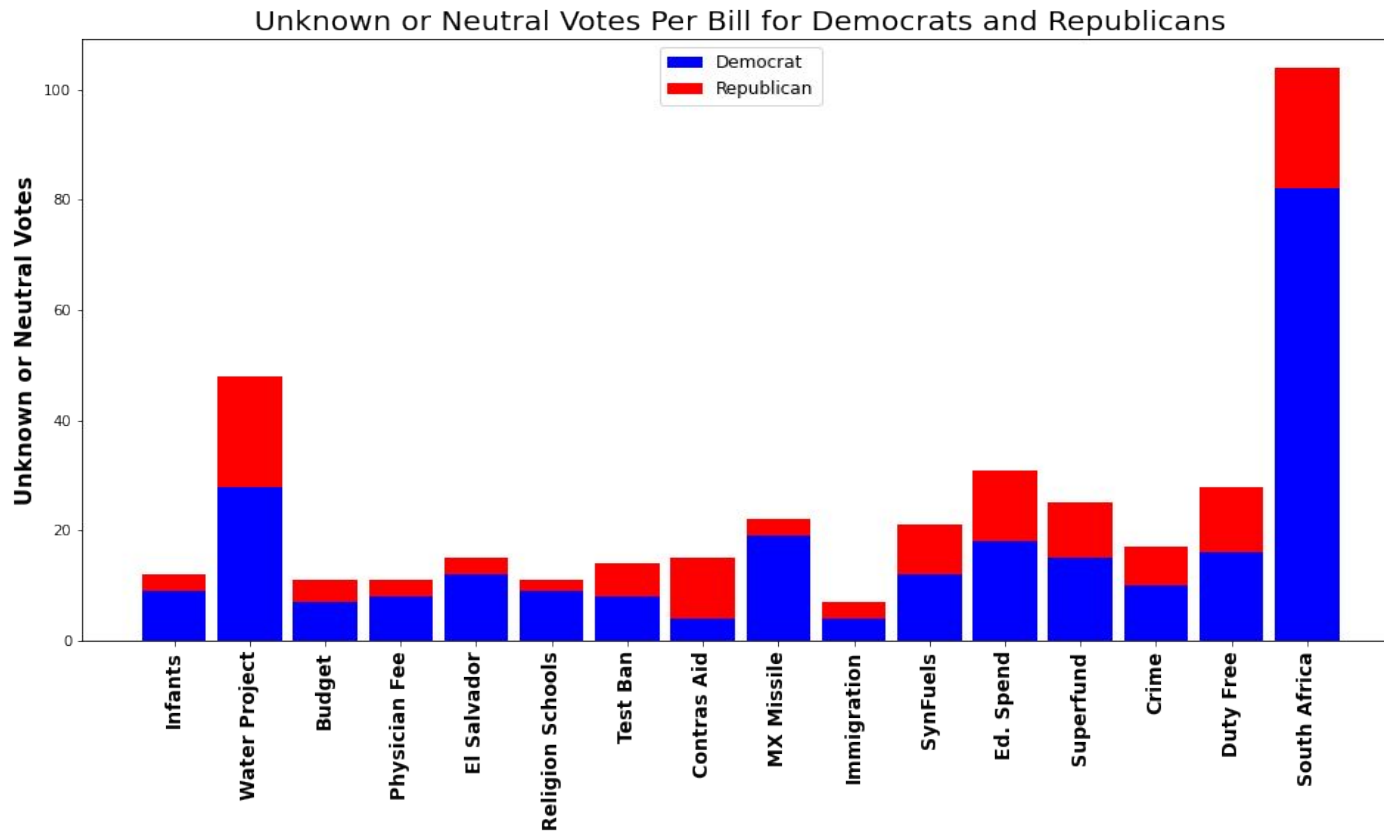
Yes votes



Important features of the data

- 267 democrats, 168 republicans
- Votes are more bipartisan than would be expected for the current congress
- Bills cover range of issues
- Some bills passed, some did not

'?' votes per bill



Neutral Vote Analysis

- “Water Project” - environmental policy bill with environmentalist groups advocating for a ‘No’ vote. Bill passed.¹
- “South Africa” - Sanctions on apartheid South Africa. Bill passed²

¹<https://scorecard.lcv.org/roll-call-vote/1984-540-water-project-cost-sharing-maintenance-and-repairs>

²<https://www.congress.gov/bill/98th-congress/house-bill/4230/summary/40?r=92>

Classification Model



Prepare and Find Model

- Goal: predict Republican or Democrat based on voting record
- Outliers:
 - Look for Republicans and Democrats with very unusual voting pattern for the party. Boxplot of # yes votes.
 - No important outliers
- Tested several models with F1 score for precision-accuracy balance: logistic regression, k-nearest neighbors, bagging classifier, Adaboost, random forest

Final Model

- Random Forest: train F1 = 1.0, test F1 = 0.98
- No need to change parameters or features
- Evaluate (positive and negative class both matter):
 - Sensitivity (Democrat) - 0.98
 - Specificity (Republican) - 0.98

Feature Importance

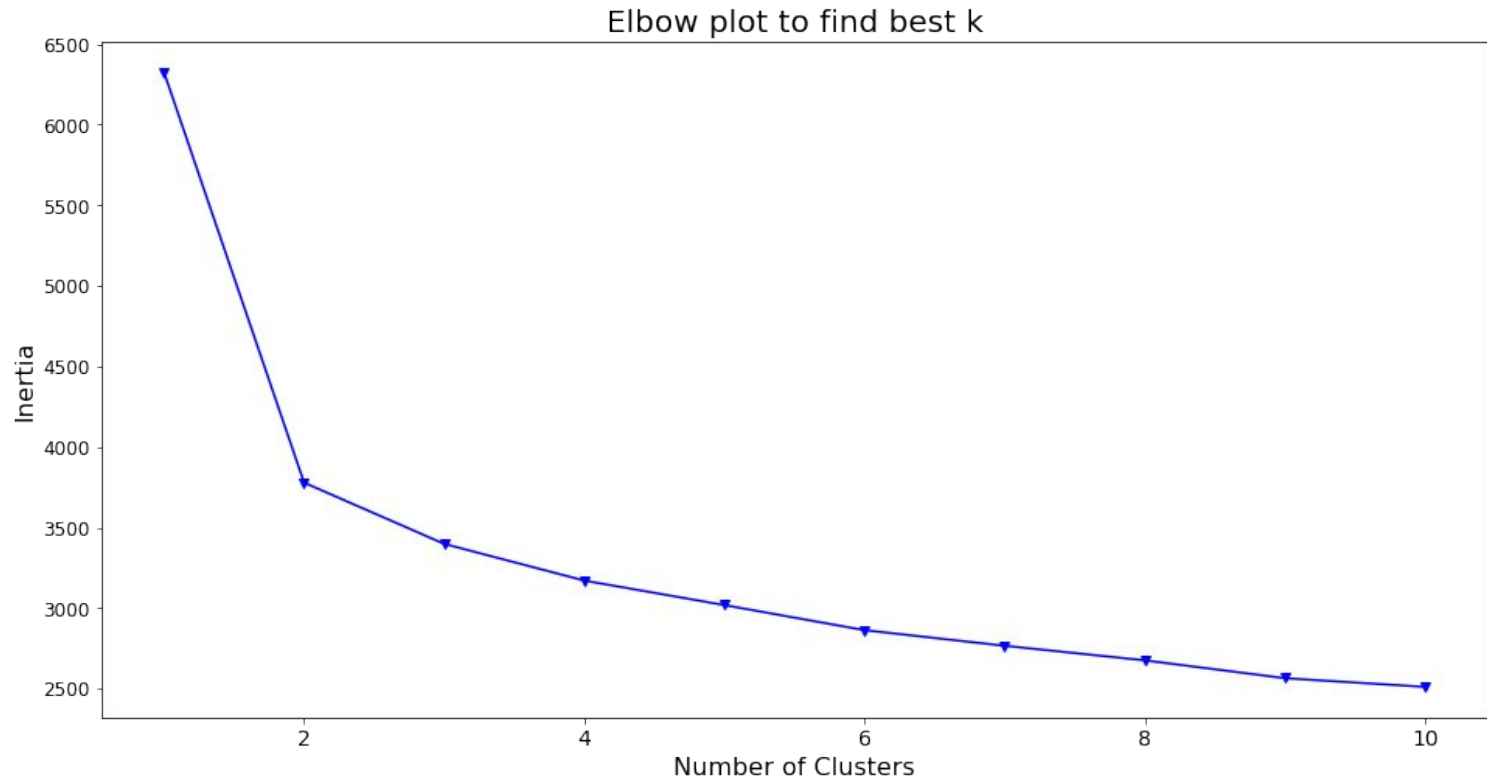
- Model predicting Republican vs. Democrat largely on a health care policy vote
- Vote on budget is expected to be partisan

Feature	Feat. Importance
Yes Physician Fee	0.21
No Physician Fee	0.19
Yes Budget	0.09
No Budget	0.08
No Ed. Spend	0.05
Yes El Salvador	0.04
No El Salvador	0.04

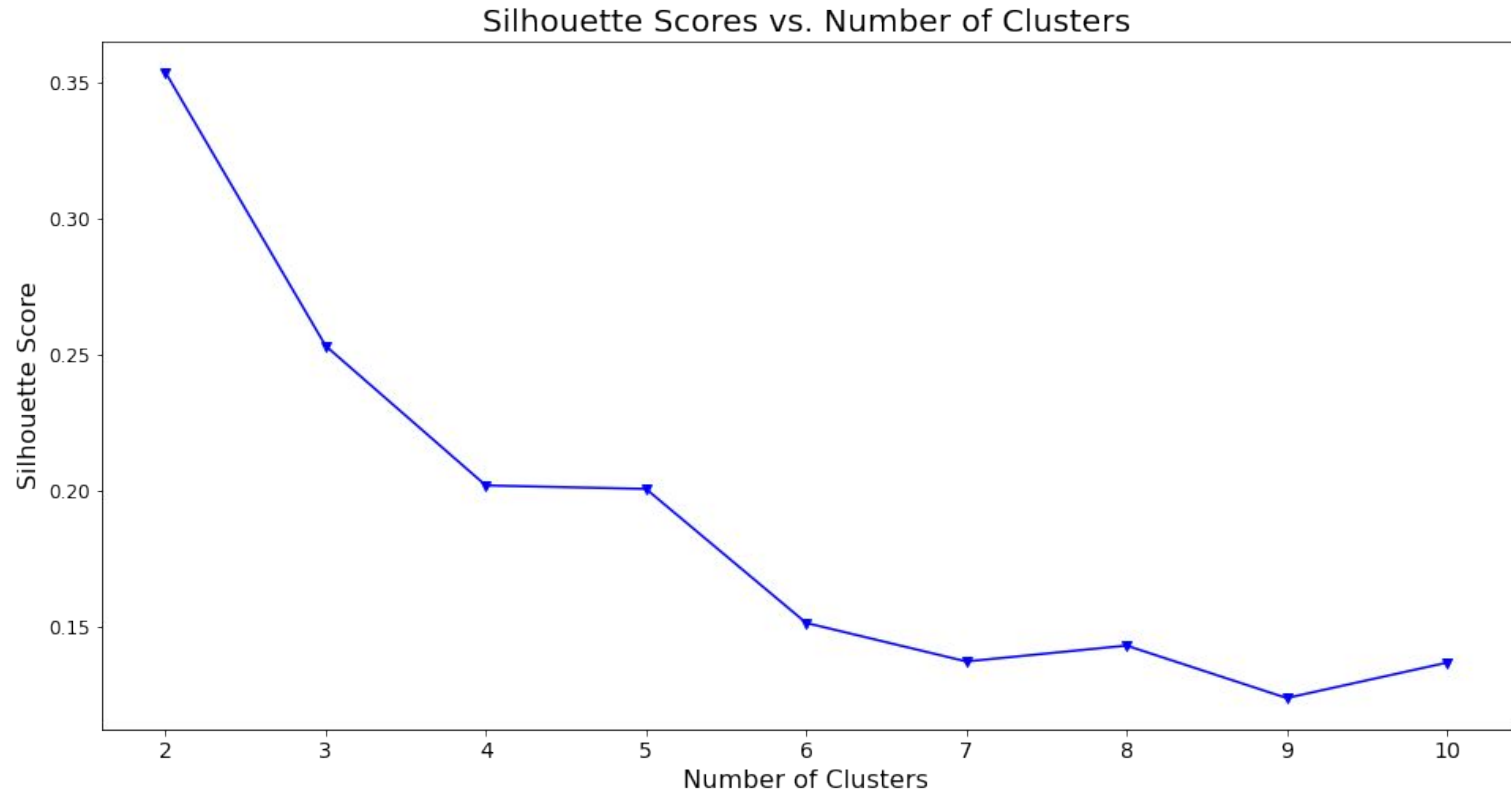
Cluster Model



Elbow Plot



Silhouette Plot



Conclusions and Next Steps

- Conclusions:
 - Caucuses other than parties are not important enough to show up in modeling
 - Health care policy may be the area that is most likely to divide Republicans and Democrats
- Possible Extension
 - Get data from all votes and model with lasso regularization and logistic regression to see if the conclusions from 'key votes' match conclusions from all votes