Influence of Campaign Financing on Election Outcomes

Leirong (Leia) Feng

Sofiane Nour Hadji

Overview

- Problem
- Data Source
- Exploratory Data Analysis
- Data Preprocessing
- Model Training & Validation
- Results & Extensions

Problem

Objective

We aim to use campaign financing data – specifically, direct contribution amounts from Political Action Committees (PACs) to candidates – to forecast election outcomes for congressional candidates.

Political Action Committee

An organization that pools campaign contributions from members and donates those funds to campaigns for or against candidates, ballot initiatives, or legislation.

Data Source

Center for Responsive Politics

- A non-profit, non-partisan research group based in Washington, D.C.
- Tracks the effects of money and lobbying on elections and public policy
- Maintains a public online database of its information at www.opensecrets.org

Data Tables

- Campaign Finance Data
  - Includes 5 tables: individual contributions, **candidates**, FEC committees, **PAC to candidates**, PAC to PACs
  - 2018, 2016, and 2014 Cycle Tables
- Reference Data
  - PAC industry codes
- Open Data User's Guide
  - Data dictionaries for all tables

Data Cleaning

- Used RecipCode column: two characters
- For Candidates: <Party> + <Status>
  - <Party>: **D** for Democratic, **R** for Republican, **3** for Independent, Libertarian or third party, **U** for unknown
  - <Status>: **W** for Winner, **L** for Loser, **I** for Incumbent, **C** for Challenger, **O** for Open Seat, **N** for Non-Incumbent
  - **N** is reserved for candidates that are neither in office nor running during the cycle in question

Data Cleaning

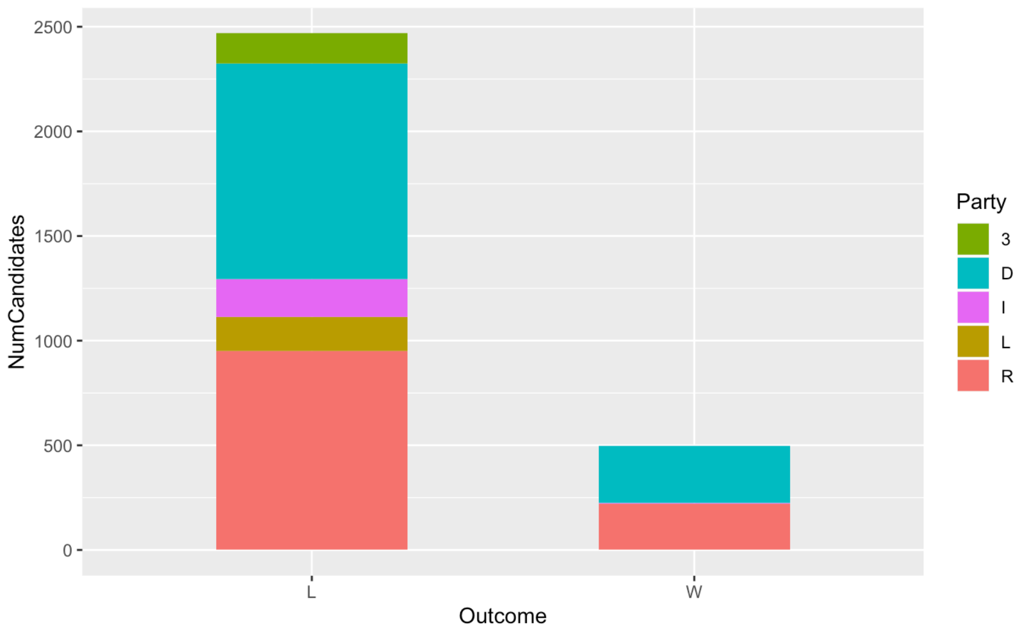
- Filtered out mislabeled data
- Removed rows with RecipCodes ending as **I**, **C**, **O**, **N**
- Mutated new column "Outcome" to indicate the Winner (**1**) or Losers (**0**) in the election
- Removed rows with NA values in contribution amount when calculating total or average contribution
- Kept contributions made before election date (2018-11-6)

Exploratory Data Analysis

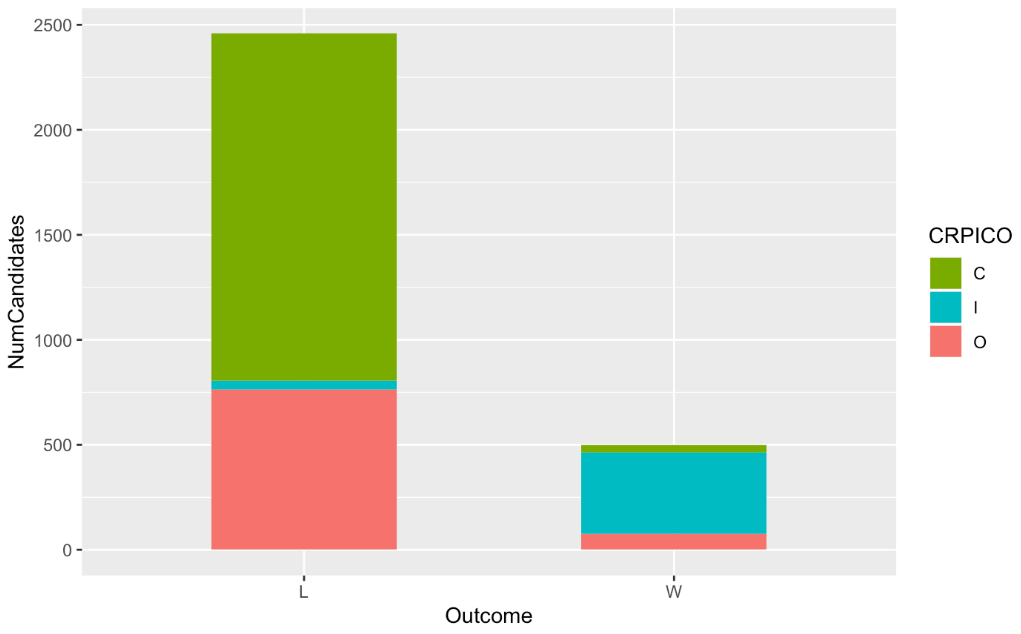
Abbreviations

- Party
  - **D** – Democratic
  - **R** - Republican
  - **I** – Independent
  - **L** – Libertarian
  - **3** – Third Party
- CRPICO
  - **C** – Challenger
  - **I** - Incumbent
  - **O** – Open Seat

Election Outcomes

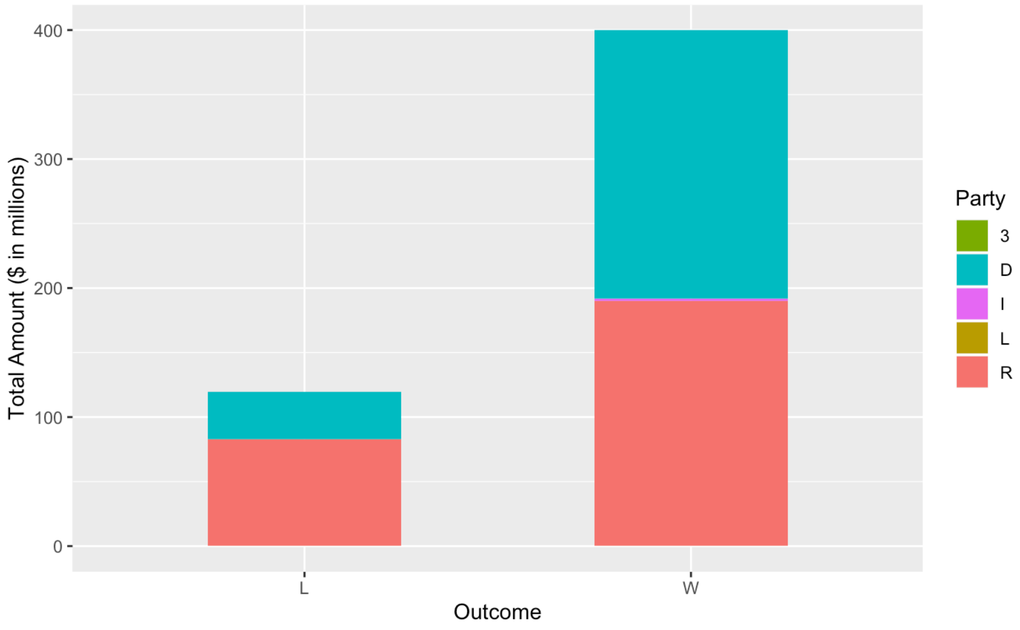


Party

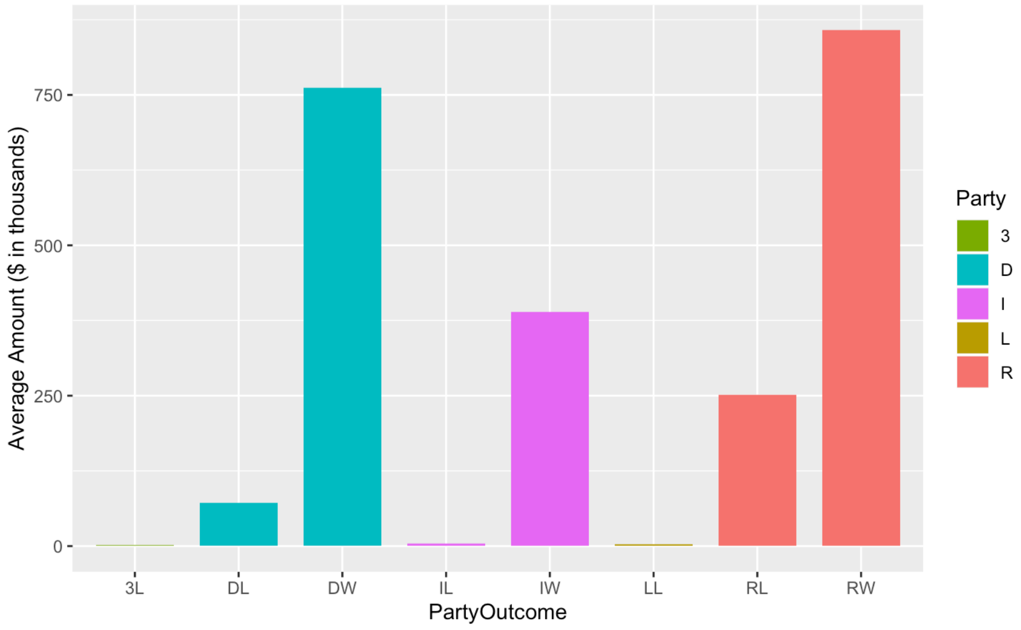


CRPICO

PAC Contribution Amounts by Party

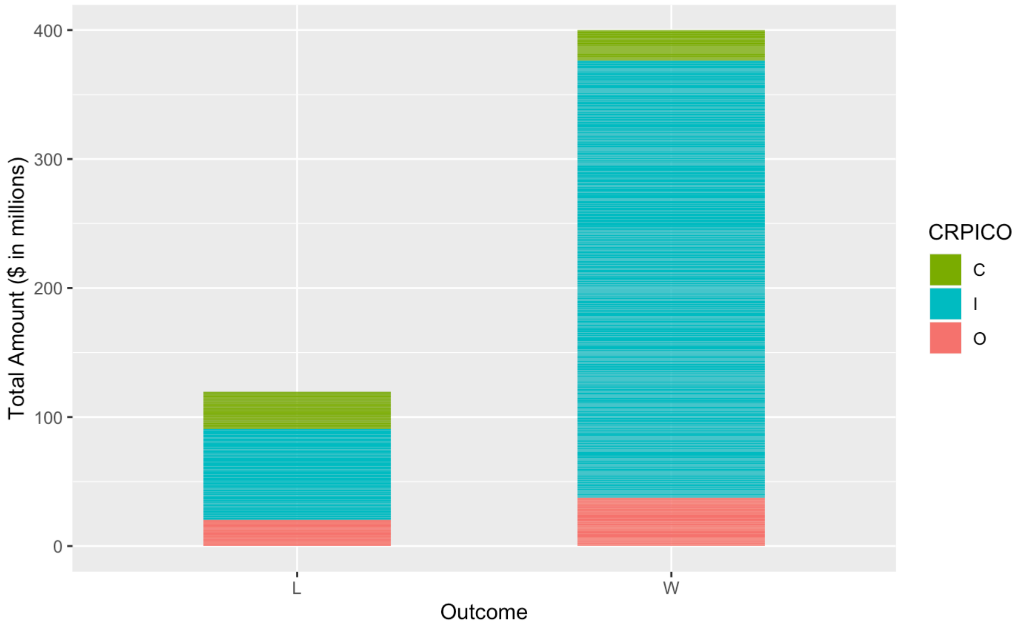


Total Amount

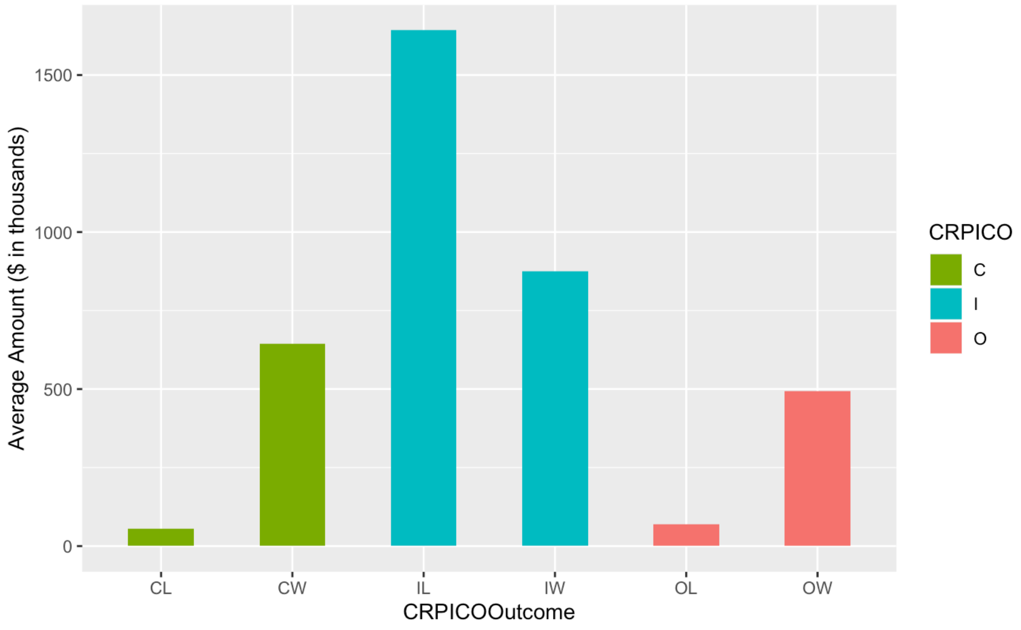


Average Amount

PAC Contribution Amounts by CRPICO

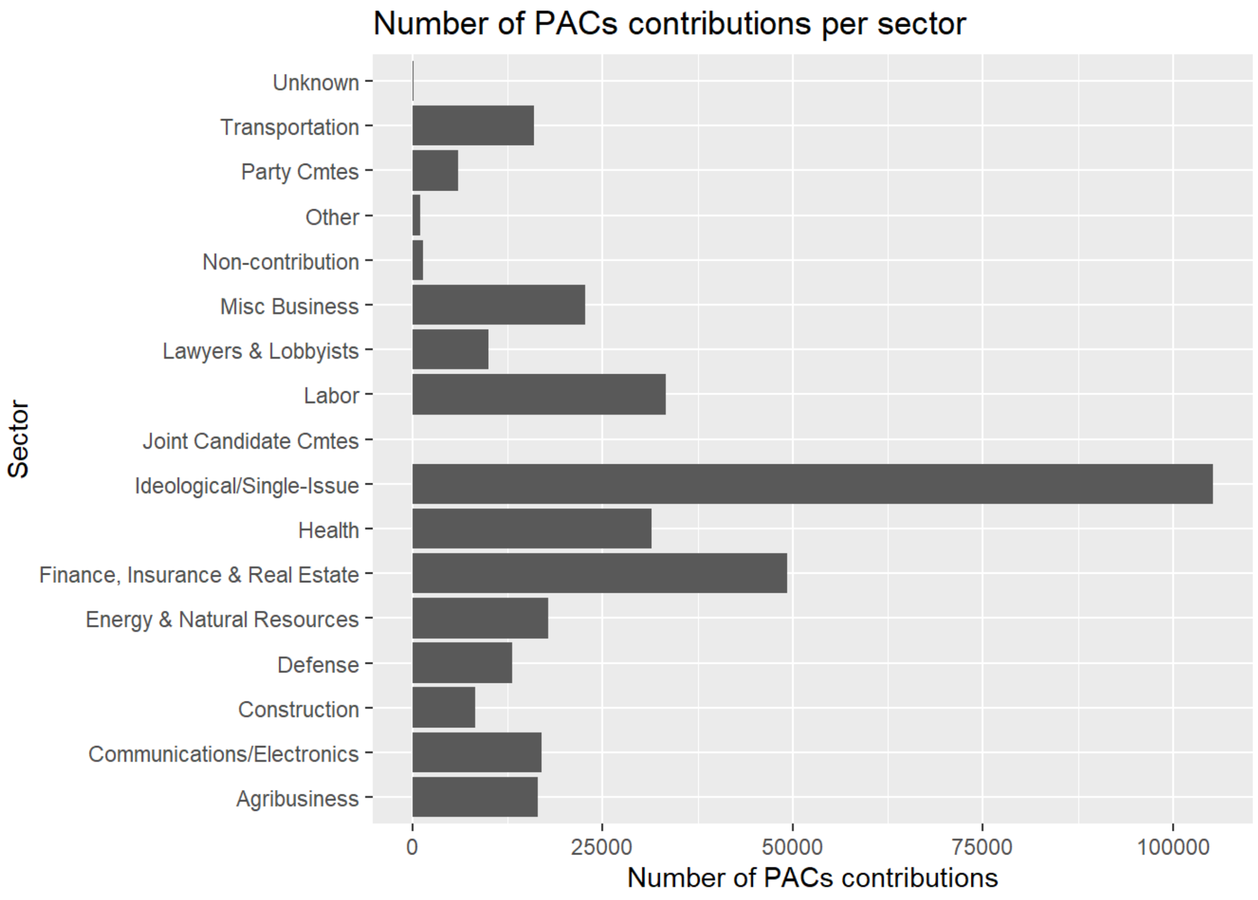


Total Amount

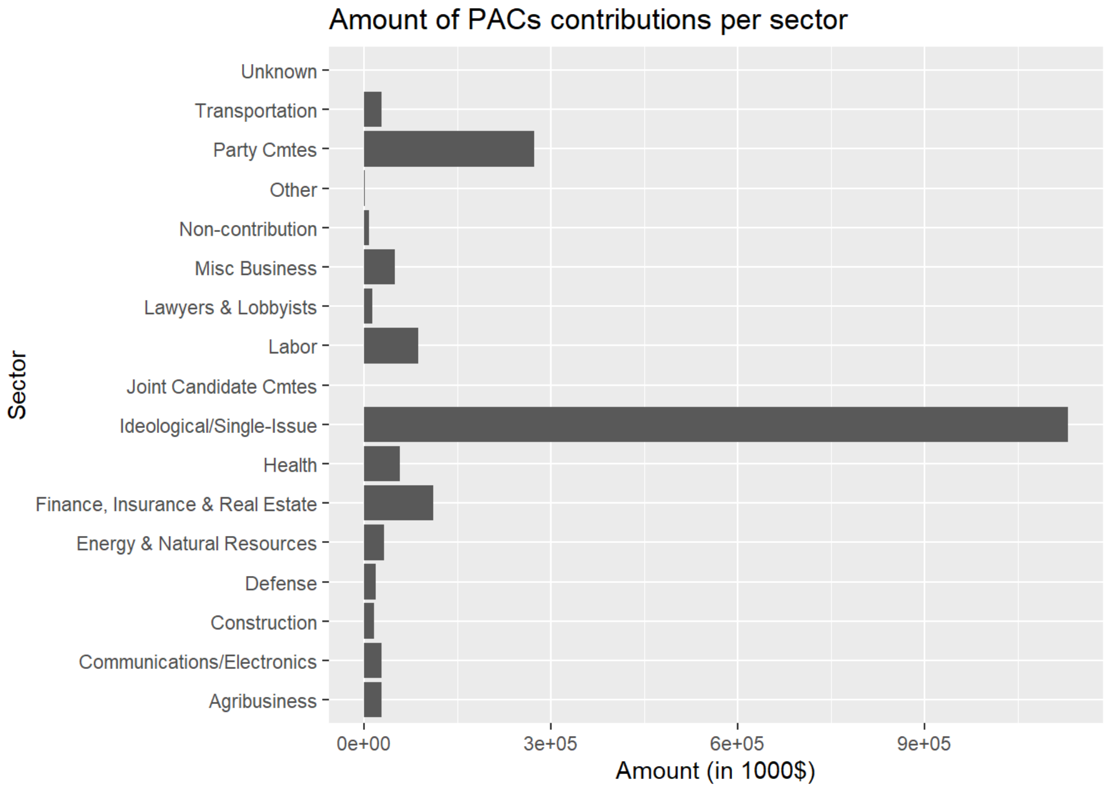


Average Amount

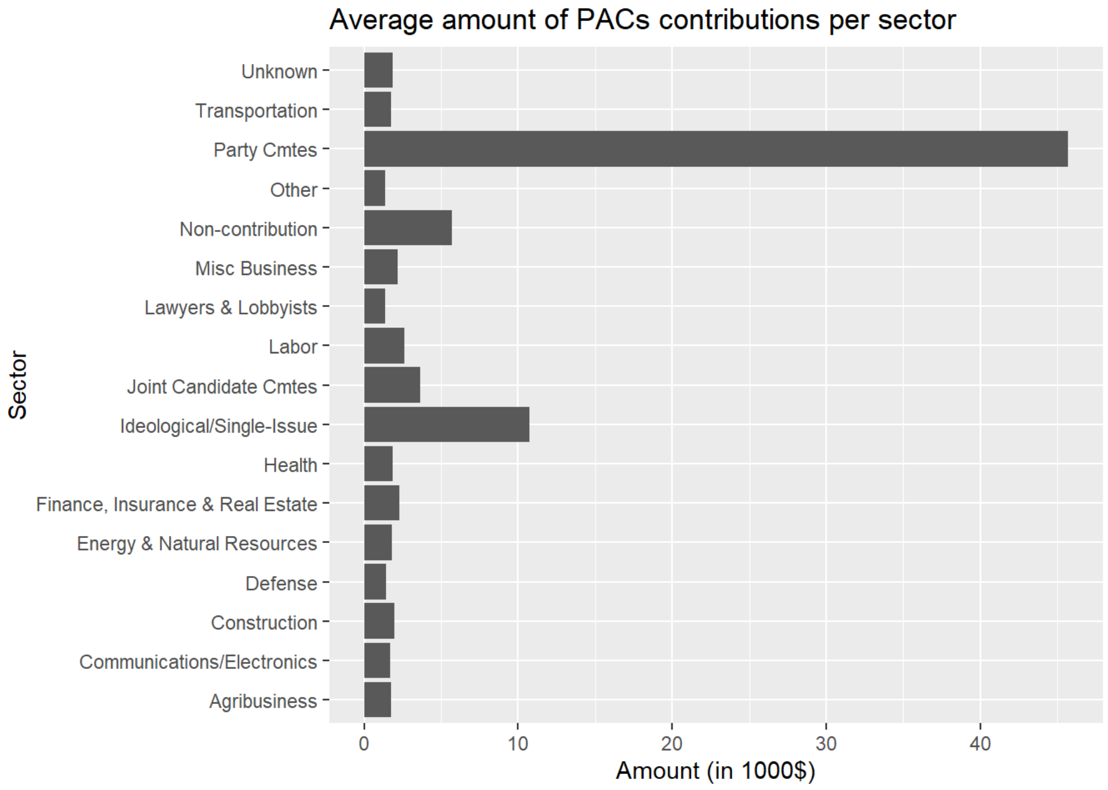
Number of PAC Contributions per Sector



PAC Contribution Amounts by Sector

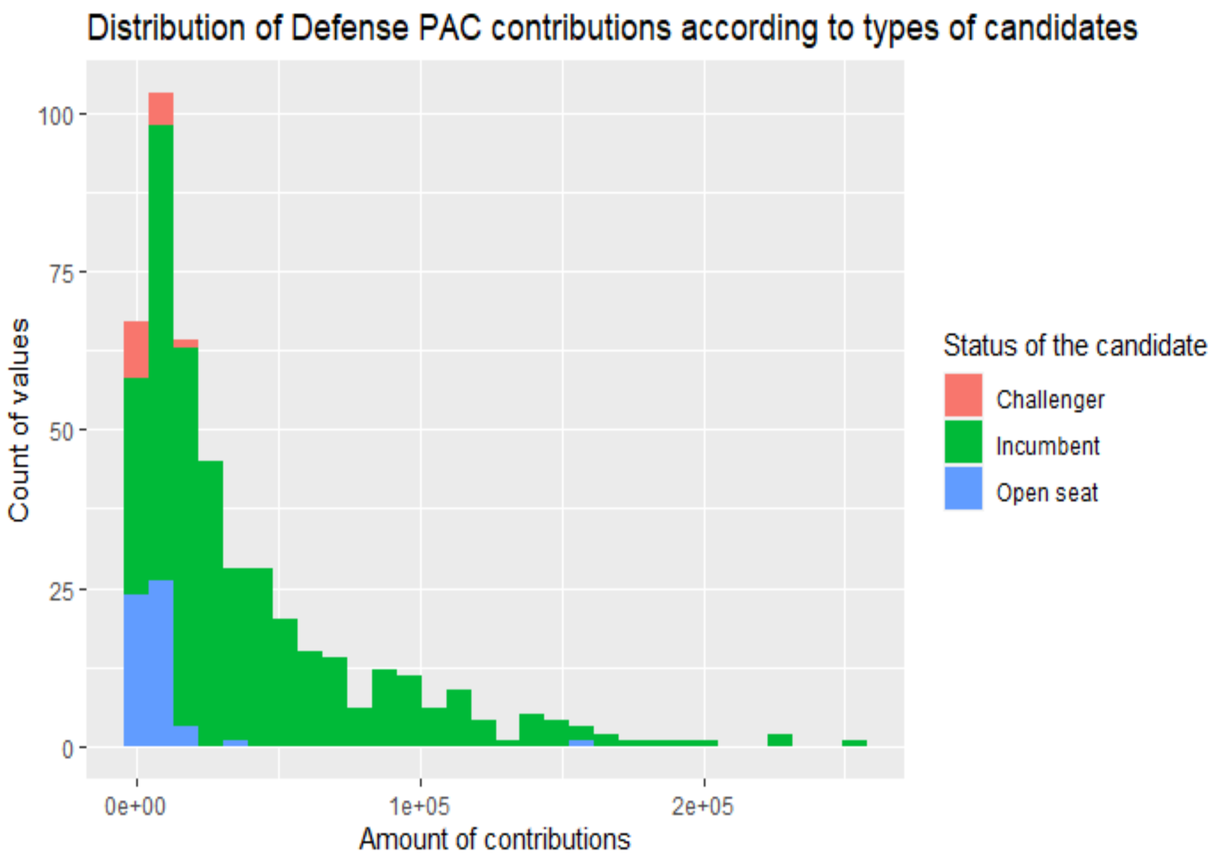


Total Amount

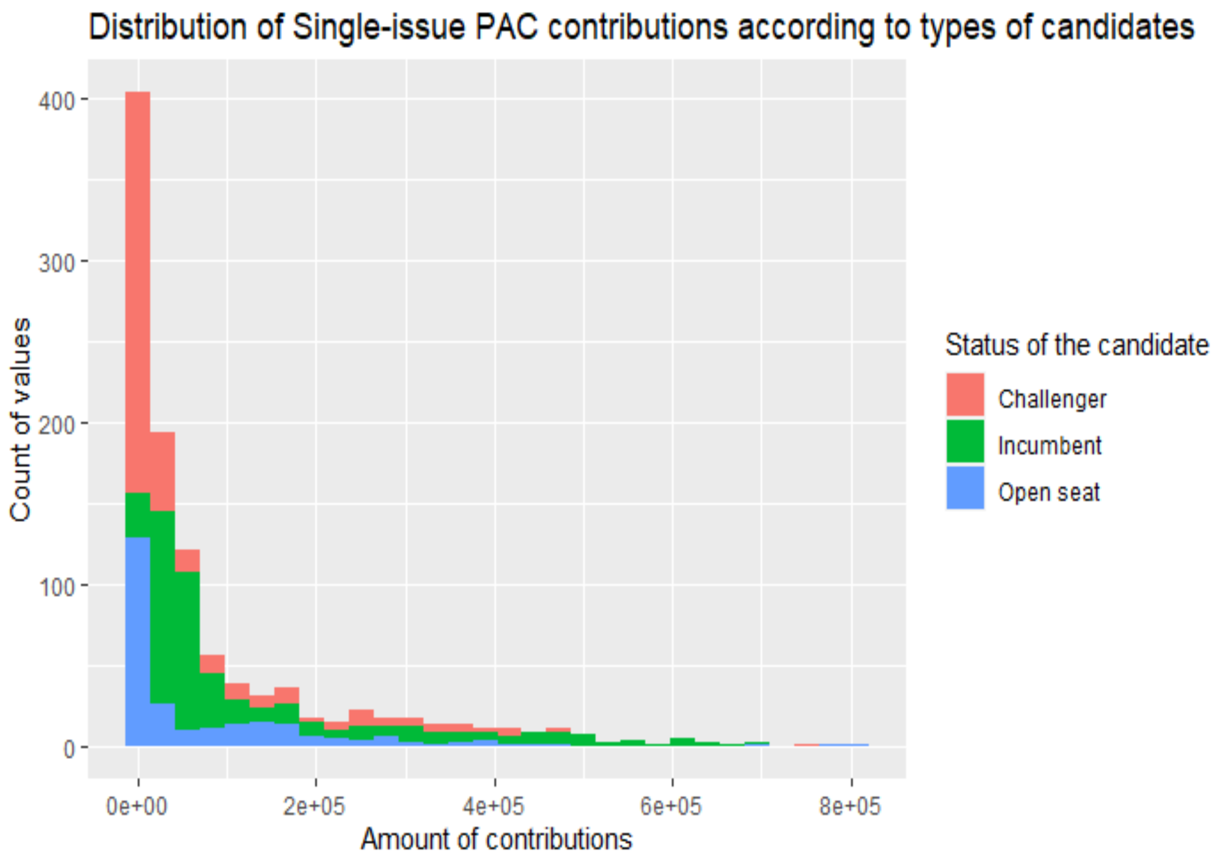


Average Amount

Distribution of PAC Contributions per Sector



Distribution of PAC Contributions per Sector



Data Preprocessing

Joining Data Tables

- PAC Tables (2014, 2016, 2018)
  - Selected direct contributions
  - Grouped by candidate ID and PAC industry
  - Summed total contributions to each candidate (2014/2016)
  - Summed total contributions per industry to each candidate (2018)
  - Grouped by timeframe before 2018 election (3 months, 6 months,12 months)
- Candidate Table (2018)
  - Left joined with PAC contributions by candidate ID
  - Replaced all NA contributions with 0
  - Kept distinct candidate IDs, removed duplicate rows
  - Added number of candidates per district (competitiveness of race)

Model Training & Validation

Model Variables

- Party – categorical (D, R, I, L, 3, U)
  - The party of the candidate
- CurrCand – binary
  - Indicates whether the candidate is currently running for federal office
  - If a candidate loses a primary or drops out of the race, this field becomes blank
- CycleCand – binary
  - Indicates whether the candidate ever ran for federal office during the cycle in question
  - Any candidate who filed to run for office or otherwise formally declared intention to run
  - This does NOT change if the candidate drops out or loses a primary

Model Variables

- CRPICO – categorical (I, C, O)
  - Identifies the type of candidate
- NoPacs – binary
  - Identifies whether candidate has publicly committed to forego contributions from PACs
- Amount14, Amount16, Amount18 – numeric
  - Total amount of contributions from PACs toward each candidate for 2014, 2016, and 2018 cycle, respectively

Model Variables

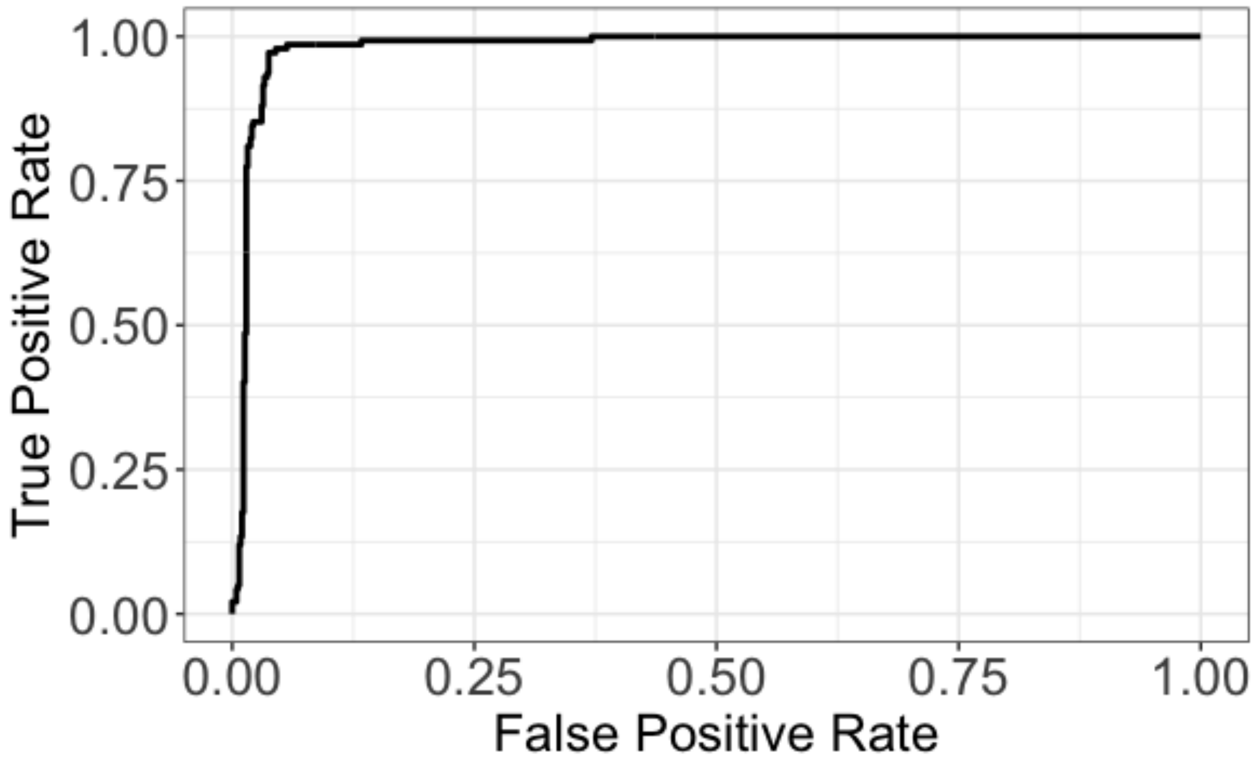
- Amount_Sector_2018 – numeric
  - Total amount of contributions from PACs of a given sector toward each candidate for 2018 cycle
- Three_M_Cont_2018, Six_M_Cont_2018, Twelve_M_Cont_2018 – numeric
  - Total amount of contributions from PACs toward each candidate for 2018 cycle, during the time period (months) preceding the election
- N_candidates_final – numeric
  - Number of candidates on the ballot on election day (2018 cycle)

Method: Penalized Logistic Regression

- Performed stratified split of candidates into training (70%) and testing (30%) set based on outcome
- Used cv.glmnet for cross-validation to select best lambda
  - nfolds = 10, type.measure = “auc”, lambda.min
- Obtained estimates of each coefficient
- Determined in-sample threshold, accuracy, prediction and recall
- Used threshold for out-of-sample prediction
- Generated out-of-sample test results

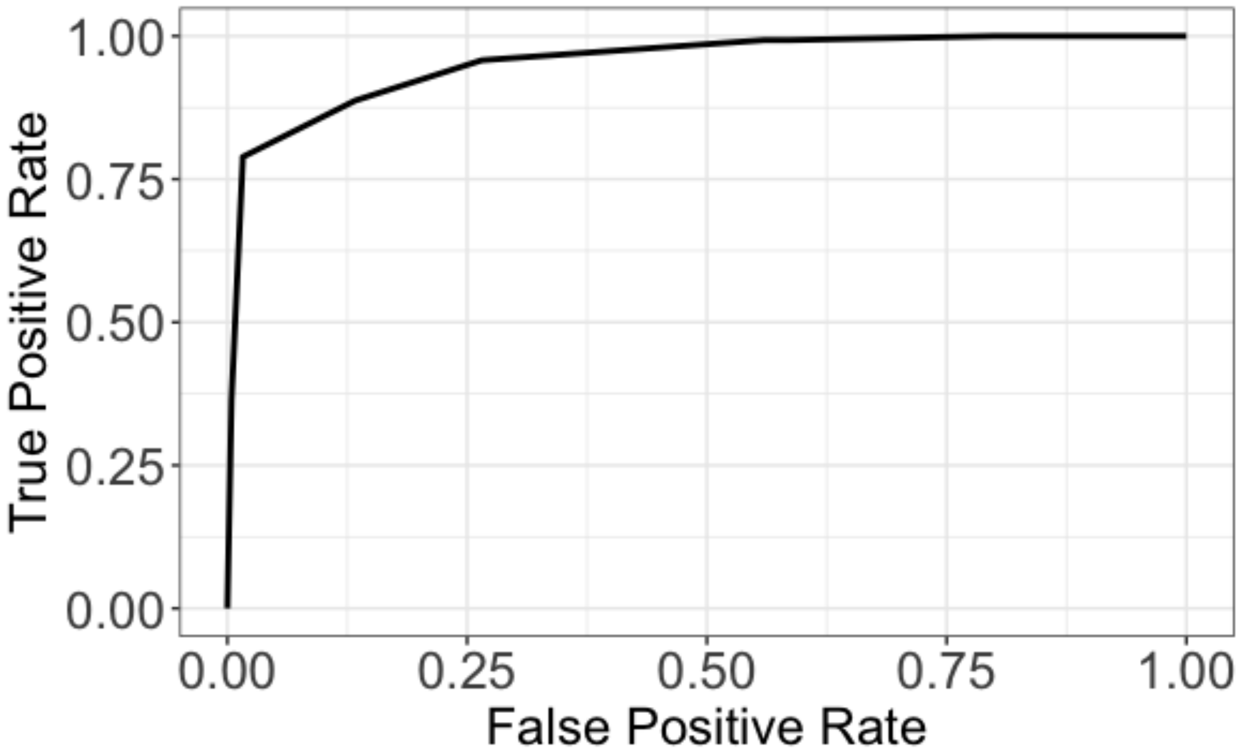
Model 1

- Variables
  - Party, CRPICO, Amount14, Amount16, Amount18
- In-Sample
  - Threshold: 0.0622
  - Accuracy: 0.9624
  - Precision: 0.8346
  - Recall: 0.9729
- Out-of-Sample
  - AUC: 0.9812

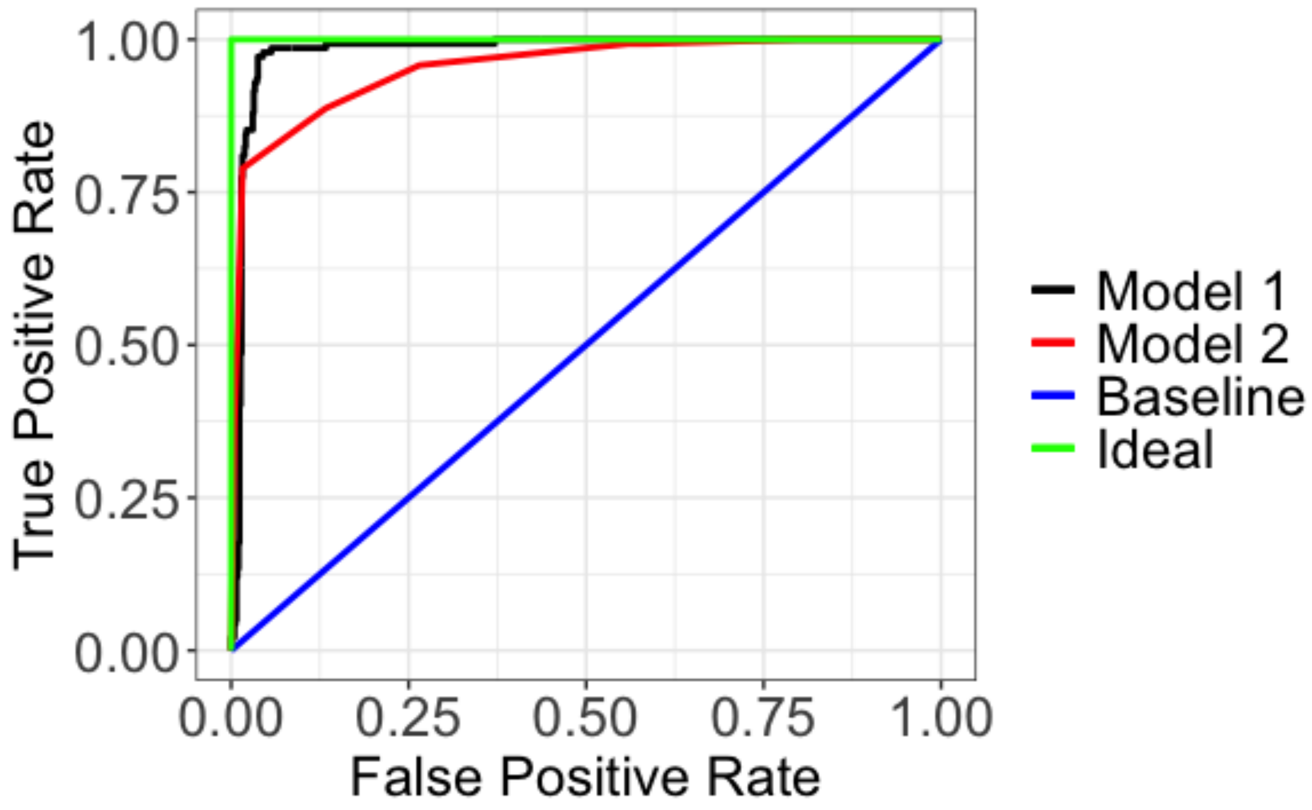


Model 2

- Variables
  - Party, CRPICO
- In-Sample
  - Threshold: 0.4008
  - Accuracy: 0.9485
  - Precision: 0.9000
  - Recall: 0.7861
- Out-of-Sample
  - AUC: 0.9541



Comparison



1.00
0.75
0.50
0.25
0.00
True Positive Rate
0.00
0.25
0.50
0.75
1.00
False Positive Rate
Model 1
Model 2
Baseline
Ideal

Model 3

- Variables
  - Amount14, Amount16, Amount18 (3m,6m,12m), Amount_Sector_18 (18 sectors)
- In-Sample
  - Threshold: 0.0985
  - Accuracy: 0.9398
  - Precision: 0.7535
  - Recall: 0.9668
- Out-of-Sample
  - AUC: 0.9775

	variable	estimate
1	(Intercept)	-2.626132e+00
3	Amount_2014	7.267247e-07
4	Amount_2016	5.325595e-07
5	Three_M_Cont_2018	2.217340e-06
6	Six_M_Cont_2018	2.248738e-06
8	Amount_Agribusiness_2018	1.041766e-06
12	Amount_Defense_2018	1.490617e-05
18	Amount_Labor_2018	1.033565e-05

Results & Extensions

Key Findings

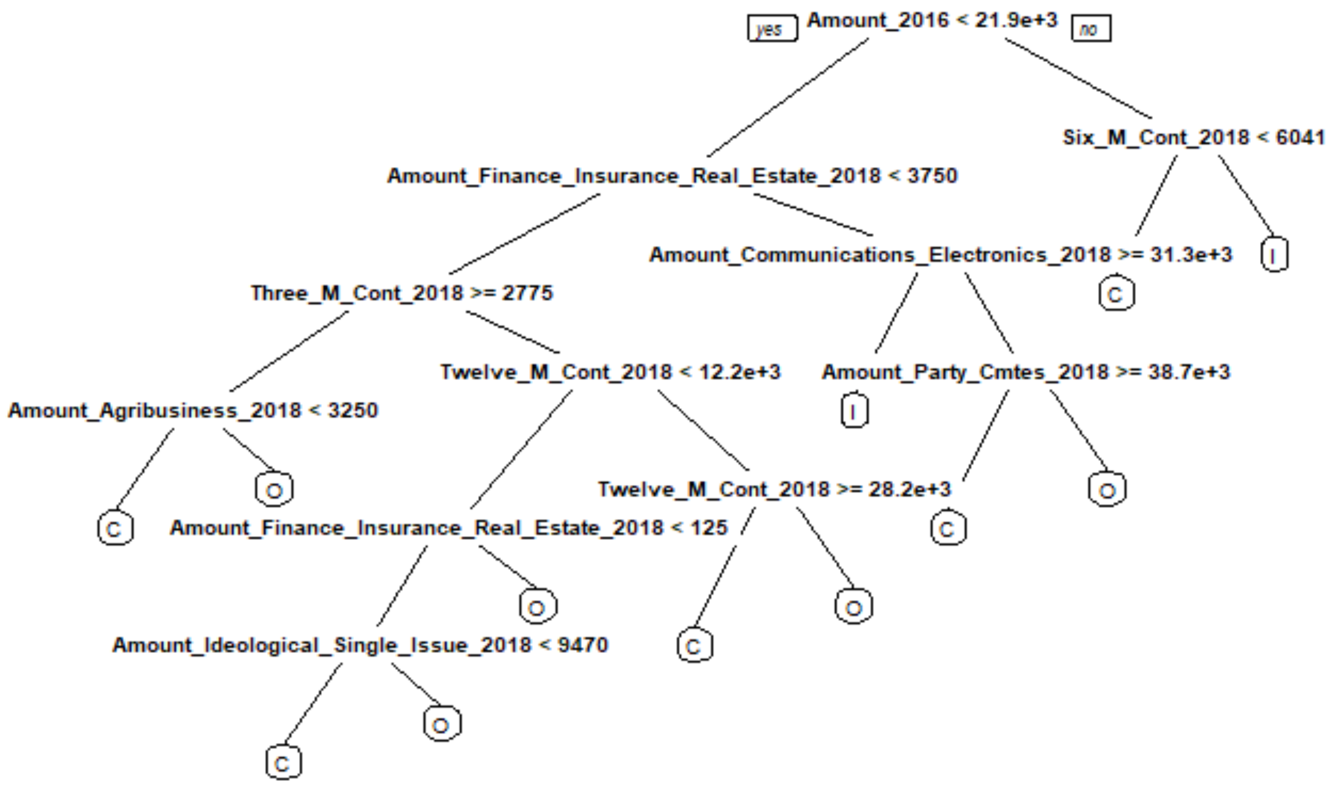
- PAC contribution is a good indicator of electoral success
- Ex : A contribution increase of $45000 from Defense PAC contributions doubles the odds of winning
- However, we also found that incumbents have a very high probability of reelection, and they also tend to get more contributions from PACs
- The causal relationship is unclear – we do not know whether the best candidates raise more money, or if raising more money contributed to their electoral victory

Predicting Incumbency from Contributions

- Predicting CRPICO variable (Challenger, Incumbent, Open Seat) from amount of contributions
- Used classification trees to provide an interpretable decision path
- Determined in-sample and out-of-sample confusion matrixes
- Evaluated in-sample and out-of-sample accuracy

Predicting Incumbency from Contributions

- Variables
  - Amount14, Amount16, Amount18 (3m,6m,12m), Amount_Sector_18 (18 sectors)
- In-Sample
  - Accuracy: 0.7563
- Out-of-Sample
  - Accuracy: 0.744



Predicting Incumbency from Contributions

- Confusion matrix on train
  - Good recall on Challenger and Incumbent
  - But 77% of Open seats candidates are predicted Challengers
- Confusion matrix on test
  - Same issue

pred_train

		C	I	O
	0	0	0	0
C	0	1057	7	22
I	0	2	284	5
O	0	434	3	127

pred_test

		C	I	O
	0	0	0	0
C	0	467	3	21
I	0	3	116	3
O	0	178	5	36

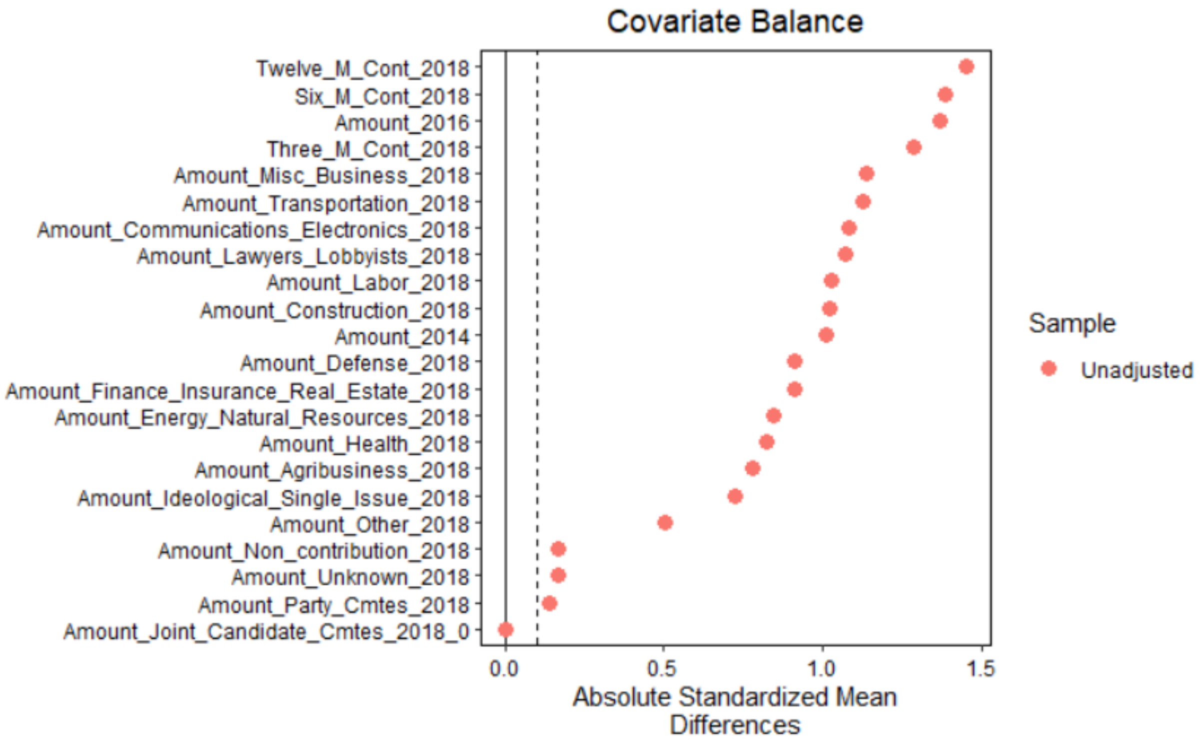
Effect of Incumbency on Election Results

- Many covariates used are highly correlated
  - Incumbents receive more money
  - Similar industries give to same candidates
- Matching: Dividing candidates into incumbents (I) and non incumbents
- 413 I and 1564 non-I: Pair each incumbent with a similar non incumbent
- Measure the effect of incumbency on election outcome if the two groups are similar enough

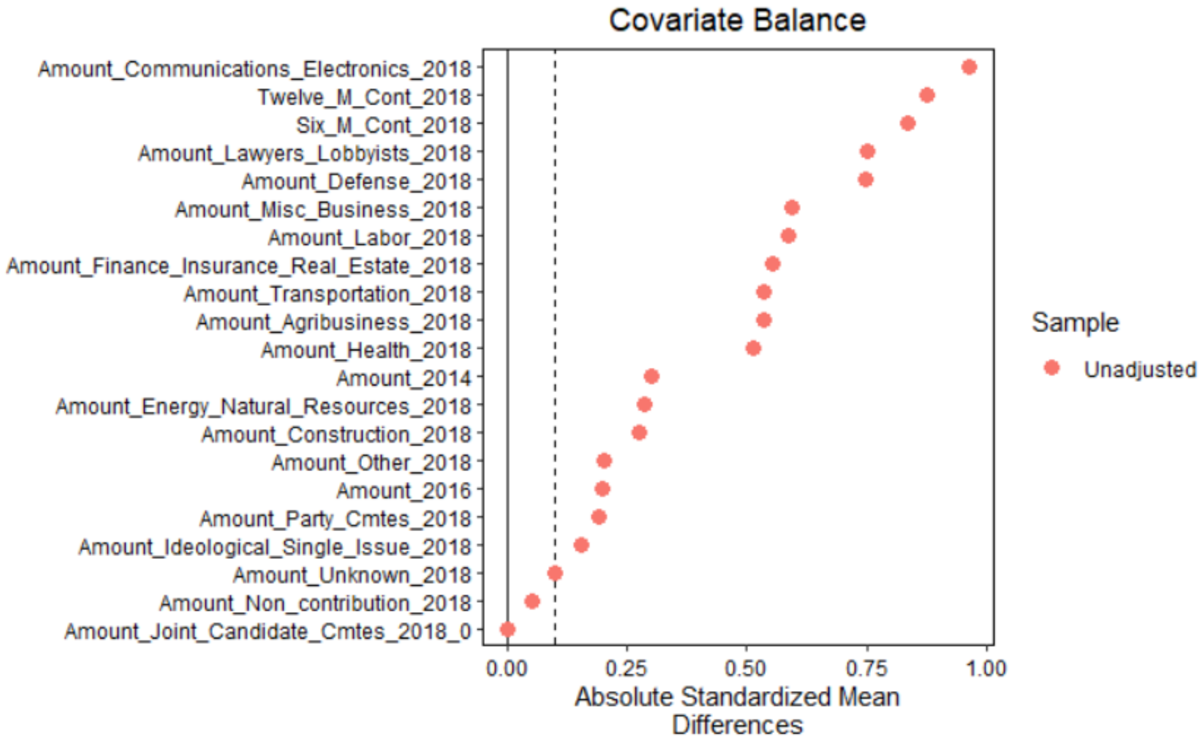
Matching

- Standardized mean difference defined for each covariate, measure of similarity between the two groups
- A = Amount of Contributions
- $SMD(A) = \frac{Mean(A\ for\ incumbents) - Mean(A\ for\ non\ incumbents)}{Standard\ deviation\ of\ A\ between\ the\ two\ groups}$

Matching



Before Matching



After Matching

Results

- Average Treatment Effect = 0.88 (1 is being elected)
- After controlling for the covariates (contributions), high effect of incumbency over election results

Future Steps

- Can examine more data from previous cycles
- Since congressional elections take place every 2 years, and all 435 Representatives serve 2-year terms, data is more frequent and comprehensive
- On the other hand, only 35 seats in the Senate were up for election in 2018, and Senators serve 6-year terms, so our model was unable to capture all Senators
- Use more efficient causal inference techniques to understand the role of incumbency / monetary contributions

Thank You!