



The Bone Scrapers



Sourish Dutta
Prahlad Siwakoti
Teresa Whitesell

NSS Data Science 7
October 24, 2023



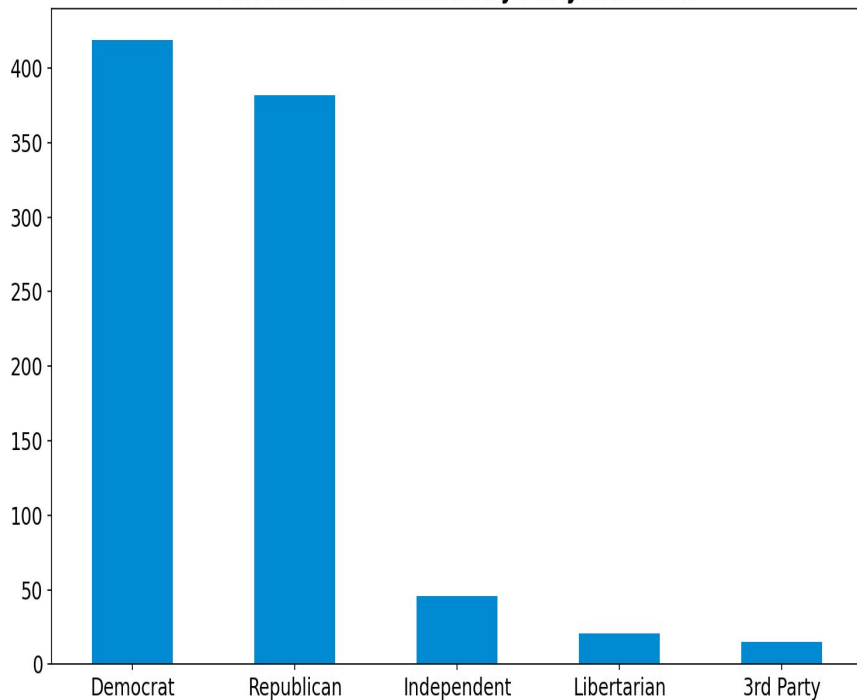
Introduction

- Scraped data for 892 candidates who ran for a House seat in 2020
- 2 districts had no vote or winner information for any candidate and were dropped
- Final analysis included 883 candidates in 433 districts

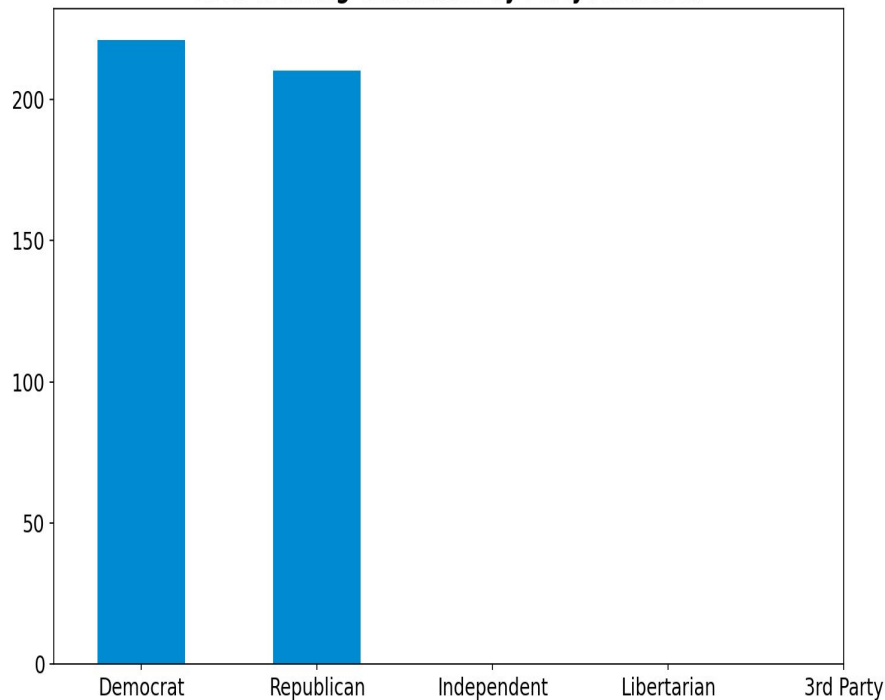
Data Overview



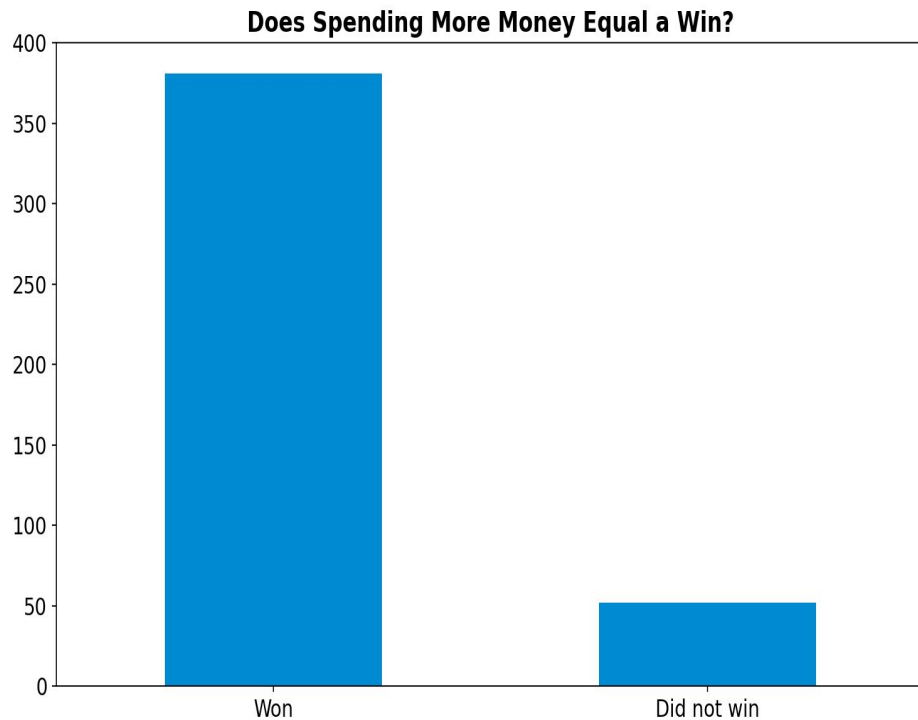
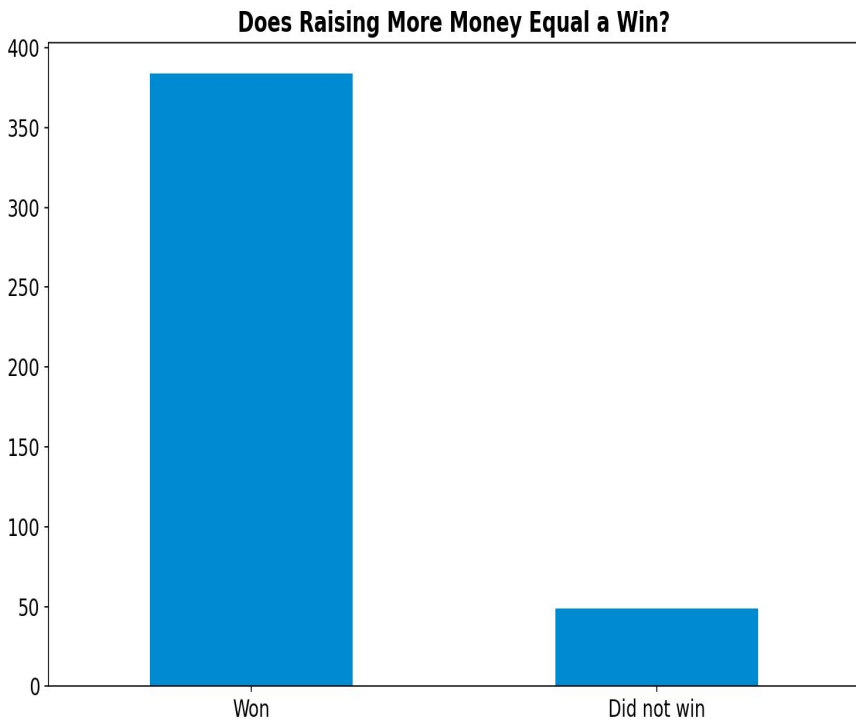
2020 House Candidates by Party Affiliation



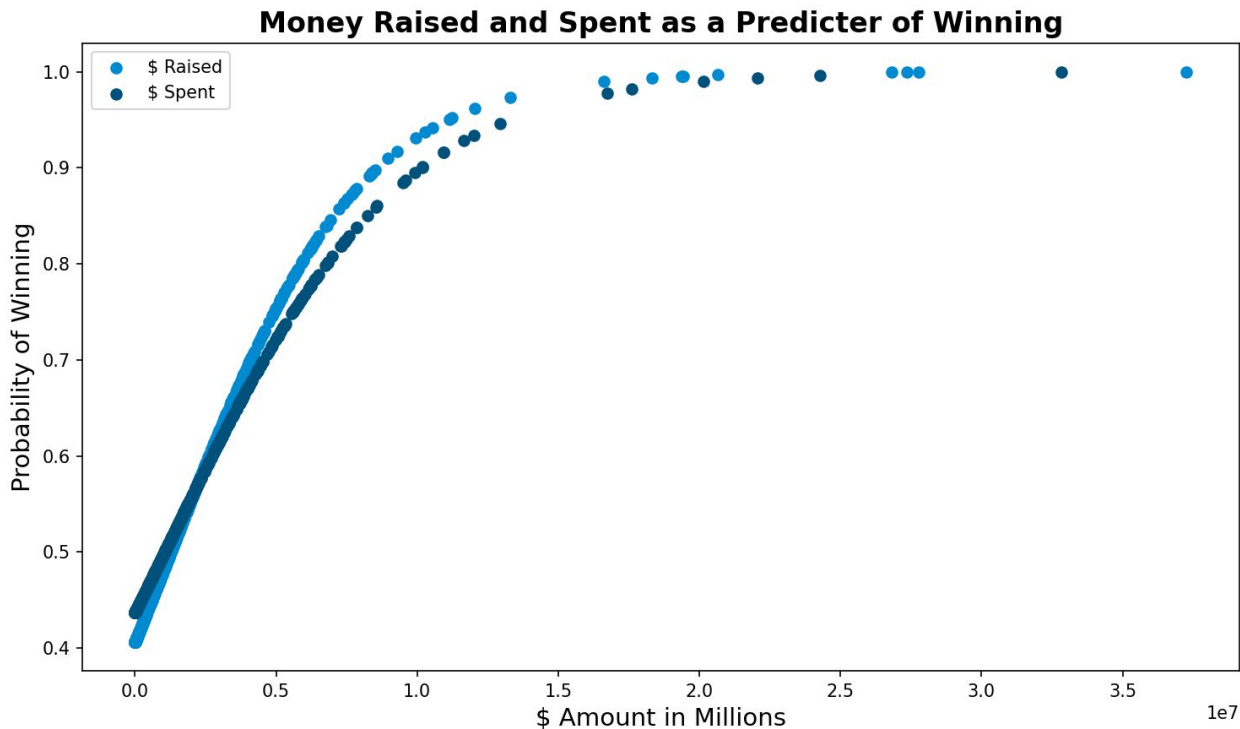
2020 Winning Candidates by Party Affiliation



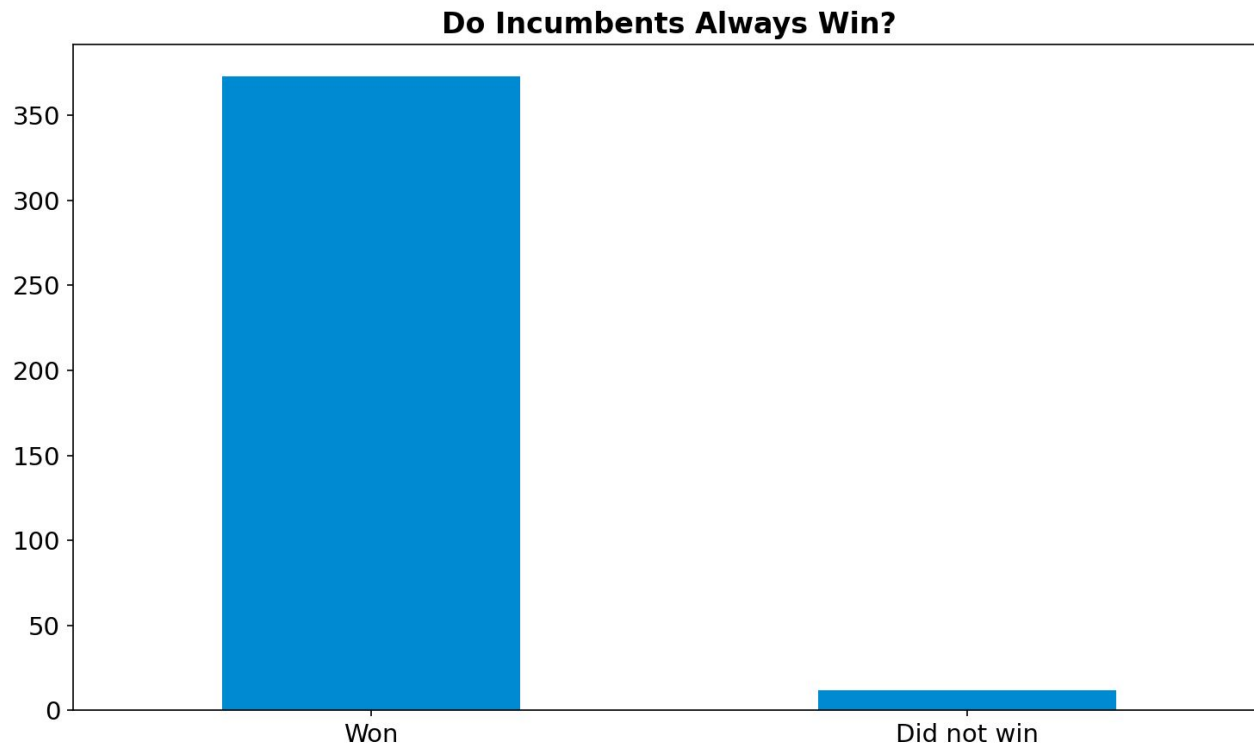
More Money, More Winning?



Raising and Spending Money



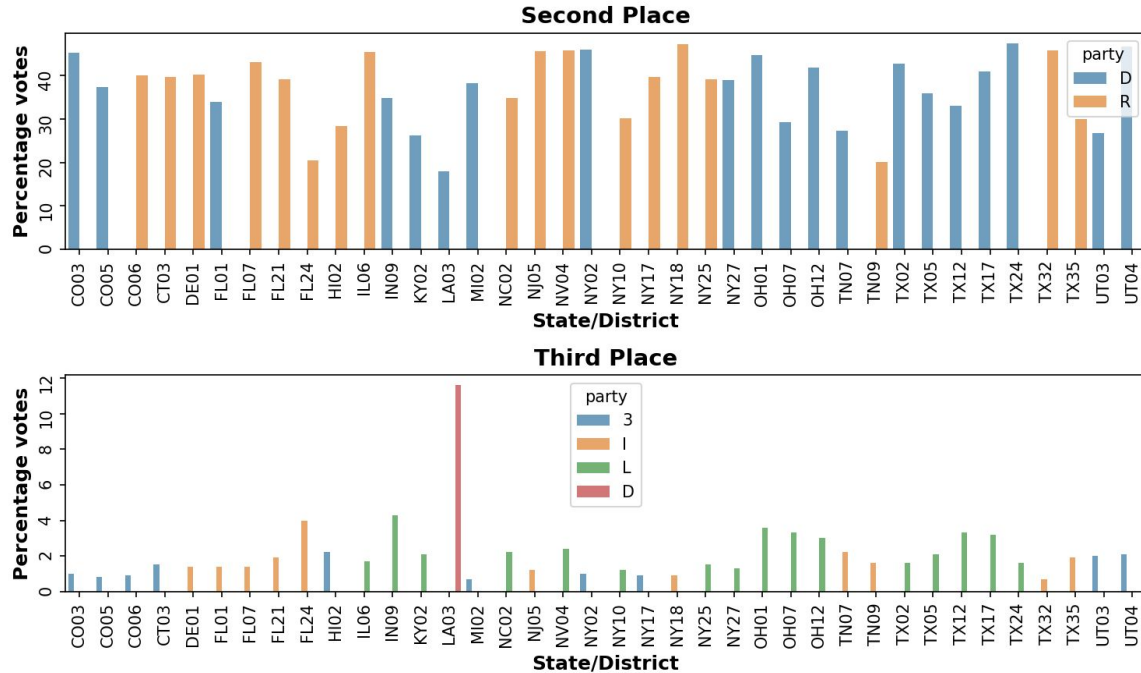
The Effect of Incumbency



The Effect of Incumbency



What percentage votes do third place candidates receive?



Logistic Regression

We built a logistic regression model to predict the probability (p) of a candidate winning using the following predictors:

- The candidate's incumbent status (i)
- The candidate's spending as a percentage of total amount spent in their race (s)

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_i i + \beta_s s$$

$$p = \frac{\exp(\beta_0 + \beta_i i + \beta_s s)}{1 + \exp(\beta_0 + \beta_i i + \beta_s s)}$$

Logistic Regression

- To see if our model has predictive power, we randomly split up our dataset into a training set (containing 90% of the data) and a test set containing the rest
- We trained the model using our training set

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-4.3969	0.363	-12.129	0.000	-5.107	-3.686
perc_money_spent	0.0611	0.006	9.515	0.000	0.048	0.074
incumbent	3.6326	0.396	9.166	0.000	2.856	4.409

Interpretation of coefficients

- The intercept term of -4.39 implies that a candidate who is not an incumbent and has not spent any money has a 1.21% probability of winning
- The coefficient corresponding to i (incumbent status) of +3.63 implies that an incumbent candidate who has spent
 - no money would have win probability of 31.5%
 - 90% of the total amount spent in their race would have a win probability of 97.9%
- The coefficient corresponding to s (spending percentage in race) of +0.0611 implies that increasing the spending by 1% (of total spending in race) raises the probability of winning by at most 1.53%.

Model Predictability

- We used our test set to check the predictions of our model
- Defined accuracy as $(\text{Number of Correct Predictions})/(\text{Number of Predictions})$
- Ran the analysis 1000 times and obtained an average accuracy of 93.5%!
(85.2% - 100%)
- Surprisingly high predictive power.
- Would be interesting to see how this model does on a different election (e.g the 2022 elections)

Conclusion

- Money matters
- Incumbency matters
- Party affiliation matters

Data Sources

- Election data
 - <https://www.opensecrets.org/>
- List of House districts
 - https://en.wikipedia.org/wiki/2020_United_States_House_of_Representatives_elections