


Gov 50 Final Project

Lauren Cruickshank

Contents

	https://media-cldnry.s-nbcnews.com/image/upload/rockcms/2023-11/231106-Election-Preview-Joanna-Nebor	1
Code Chart		1
Introduction		2
Background		2
Hypotheses		2
Data Interest		2
Data Links:		3
Results		4
Linear Regression Between the Share of Votes and Overall Funding Broken Down by Loss and Win		4
Linear Regression between the Share of Votes and Overall Funding		5
Conclusion		7

	https://media-cldnry.s-nbcnews.com/image/upload/rockcms/2023-11/231106-Election-Preview-Joanna-Nebor
---	---

Code Chart

The data file for this study is `fpd.csv` and contains the following variables:

Name	Description
------	-------------

chamber	whether the candidate is in the House or Senate
candidate	the names of candidate, party, and state
selffunding	how much money the candidate funded themselves
collectedfunding	how much money the candidate funded from outside sources
overallfunding	the overall money the candidate received in funding
Percent Self-funded	percent of funding the candidate funded from themselves
result	whether or not the candidate won or loss the election
pervote	share percentage of votes received by the candidate

Introduction

Background

The 2024 presidential election cycle is currently approaching. As it is known, there is a host of factors that can effect ones campaign run and thus the results of the election. The U.S states government has formed very specific rules around the use of campaign funds and the way that they are used. Nonetheless, this study aims to find if there is a correlation between the amount of campaign funds a candidate has and their election results. The data set represents the campaign funding and election results of House and Senate candidates from 2016 to 2020. The explanatory variable in this study is is the amount of funding that each candidate receives. The response variable or outcome is whether the candidate saw success in the election.

Hypotheses

The Hypotheses of this Study are as follows:

- 1) As the **amount of available campaign funding increases** so will the **wins of the candidates increase**
- 2) As the **amount of available campaign funding increases** so will the **share percentage of votes received by candidates in the elections increase**

Data Interest

```
library(tidyverse)
library(dplyr)
library(readr)
library(gapminder)
library(infer)
library(moderndiver)

cand_sp <- read_csv("fpd.csv")

cand_sp
```

The sample used is comprised of data compiled by the OpenSecrets the “nation’s premier research group tracking money in U.S politics and its effect on elections and public policy. In this study, OpenSecret’s data on candidate spending and results from 2016 to 2020 were used. In order to include the percentage share by which the candidate won or loss, it had to be individually searched in the database of OpenSecret.

Data Links:

OpenSecrets

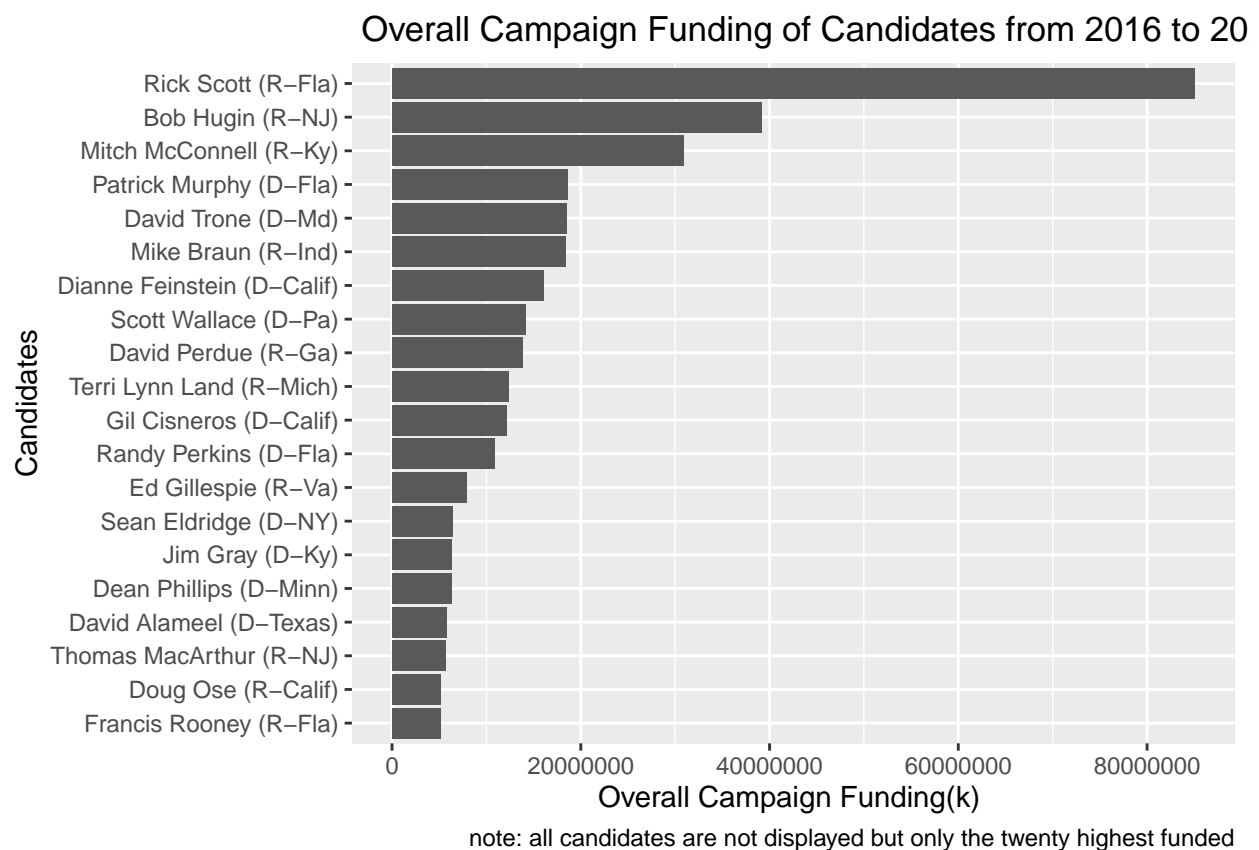
The **independent variable** in this study is the *amount of money that each candidate has for campaign funding*. The **dependent variable** is the *share percentage of votes the candidates receive and the results that each candidate yielded from the election*.

In order to identify a relationship between overall campaign funding, share percentage of votes from election, and candidates results from the election, this analysis is utilizing a correlation research design.

The graph below represents the top twenty candidates with the highest campaign funding from 2016 to 2020.

```
options(scipen = 999)

cand_sp|>
  slice_max(overallfunding, n = 20)|>
  ggplot(
    mapping = aes(x = overallfunding, y = fct_reorder(Candidate, overallfunding))) +
  geom_col()+
  labs(x = "Overall Campaign Funding(k) ",
       y = "Candidates",
       title = " Overall Campaign Funding of Candidates from 2016 to 2020",
       caption = "note: all candidates are not displayed but only the twenty highest funded")
```



Results

Linear Regression Between the Share of Votes and Overall Funding Broken Down by Loss and Win

The following multivariate regression model demonstrates the relationship between the amount of money received for campaign funds, the share percentage of votes the candidate received from the election, and whether the candidate won or loss the overall election.

```
regr_1 <- lm(overallfunding ~ pervote + Result, data = cand_sp)

r1 <- get_regression_table(regr_1)

knitr::kable(head(r1), digits = 2)
```

term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
intercept	2509749.46	5177210	0.48	0.63	-7899729.1	12919228.0
pervote	87475.74	125817	0.70	0.49	-165496.2	340447.7
Result: Won	6217351.12	4271917	1.46	0.15	-2371913.3	14806615.5

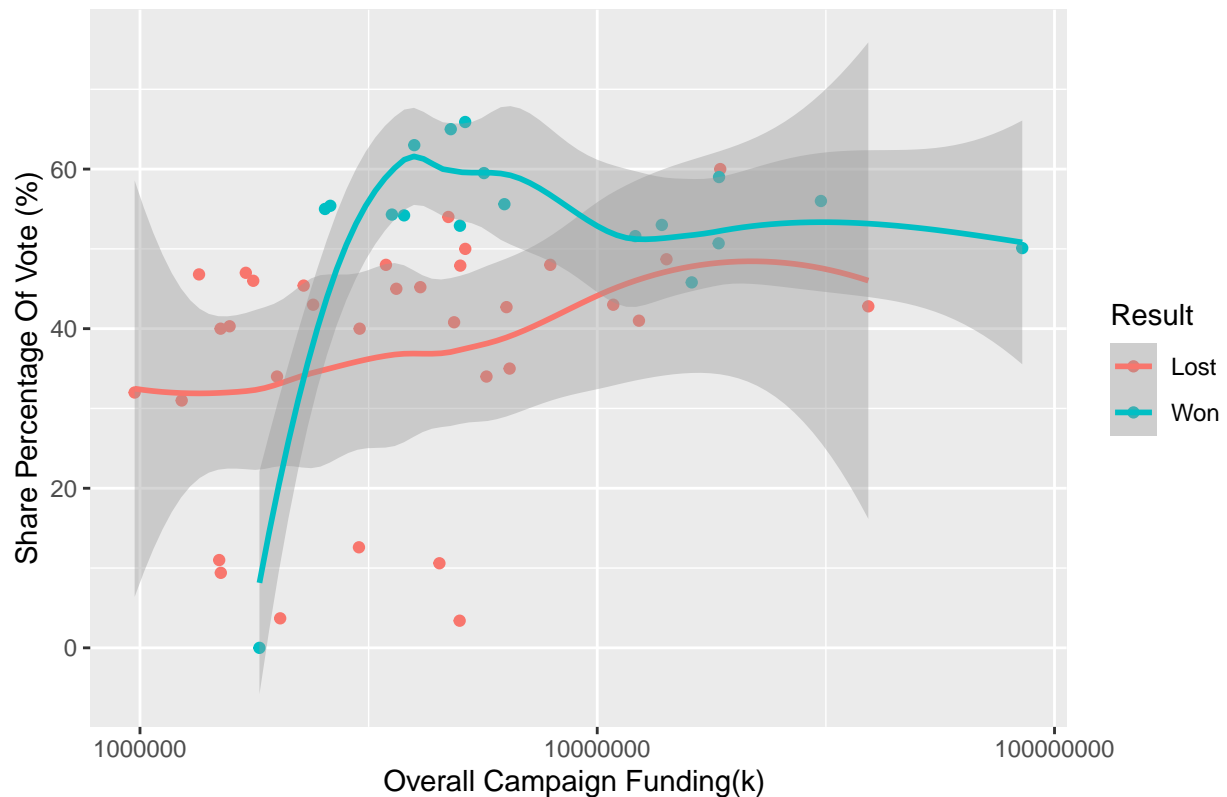
Interpretation

The intercept of 2,509,749.46 , is the estimated campaign funding when the share percentage of votes received by the candidate is zero and the candidate did not win the election. There is an expected increase of 87,475.74 in campaign funding for each unit increase in the share percentage of votes received by the candidate , also assuming that the candidate winning status stays constant.

The coefficient for 'Result: Won' of 6,217,351.12, shows that when a candidate wins the election, there is an expected increase of 6,217,351.12 in campaign funds in comparison ton when the candidate does not win, also assuming that the share percentage of votes stays constant.

```
cand_sp|>
  ggplot(
    mapping = aes(
      x = overallfunding,
      y = pervote,
      color = Result)) +
  geom_point() +
  geom_smooth() +
  scale_x_log10() +
  labs(x = "Overall Campaign Funding(k) ",
       y = "Share Percentage Of Vote (%)",
       title = " Share Percentage of Vote vs. Campaign Funding (with Result)")
```

Share Percentage of Vote vs. Campaign Funding (with Result)



The graph above suggest that with both losses and wins there is a slight relationship between an increase of money and an increase of that result. Within the graph it is also evident that candidates who lost their elections experienced more variability in campaign funding than those who won.

Linear Regression between the Share of Votes and Overall Funding

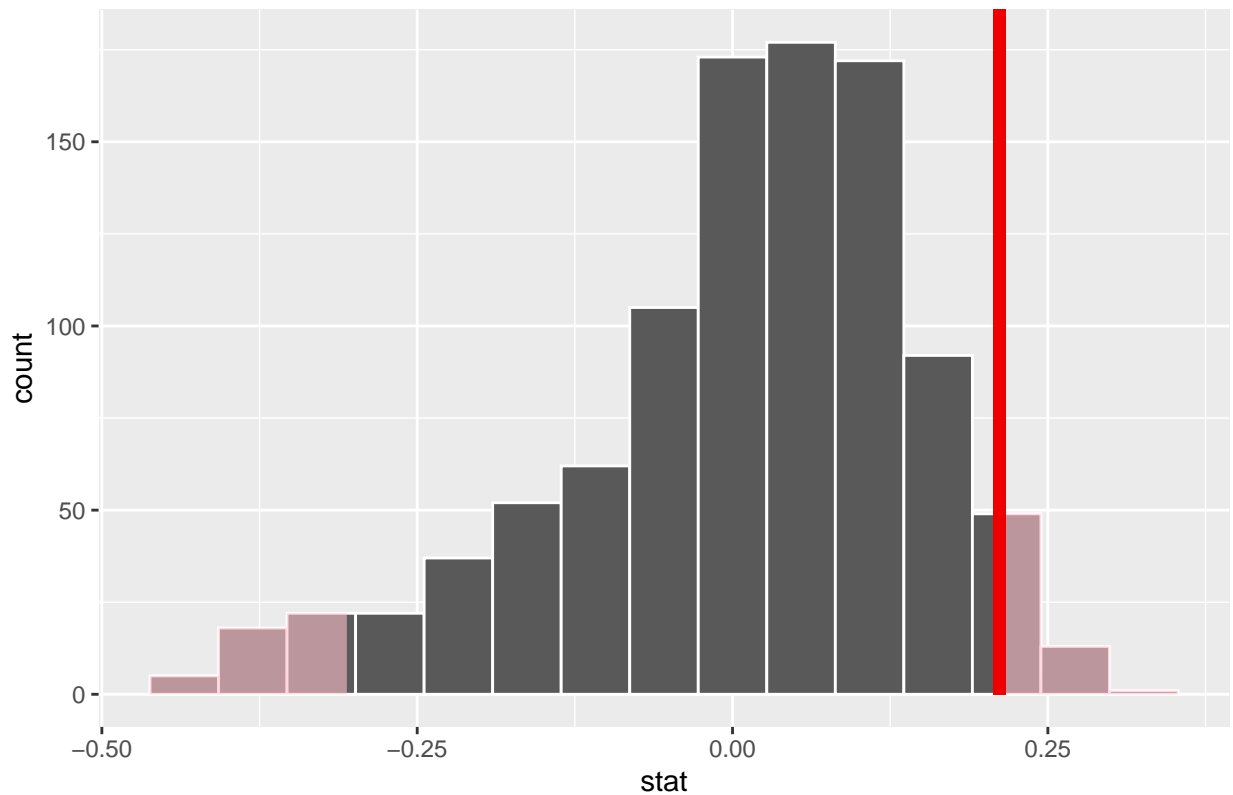
```

null_dist <- cand_sp |>
  specify(overallfunding ~ pervote) |>
  hypothesize(null = "independence") |>
  generate(reps = 1000, type = "permute") |>
  calculate(stat = "correlation")

correlation <- cand_sp |>
  specify(overallfunding ~ pervote) |>
  calculate(stat = "correlation")

null_dist |>
  visualize() +
  shade_p_value(correlation, direction = "both")
  
```

Simulation-Based Null Distribution



```
get_p_value(null_dist, correlation, direction = "both")
```

In this study, the null hypothesis suggests that there is no relationship between the amount of campaign funding a candidate received and the share percentage of votes they received in the election. With a significance level of $\alpha = 0.05$, we would fail to reject the null hypothesis with a p-value of 0.072. Therefore, we conclude that there is not a statistically significant relationship between campaign funds and the share percentage of votes candidates receive in elections.

The following regression displays the relationship between the amount of funding a candidate received and the share percentage of votes they received during the election process.

```
regr_2 <- lm(overallfunding ~ pervote, data = cand_sp)
```

```
r2 <- get_regression_table(regr_2)
```

```
knitr::kable(head(r2), digits = 2)
```

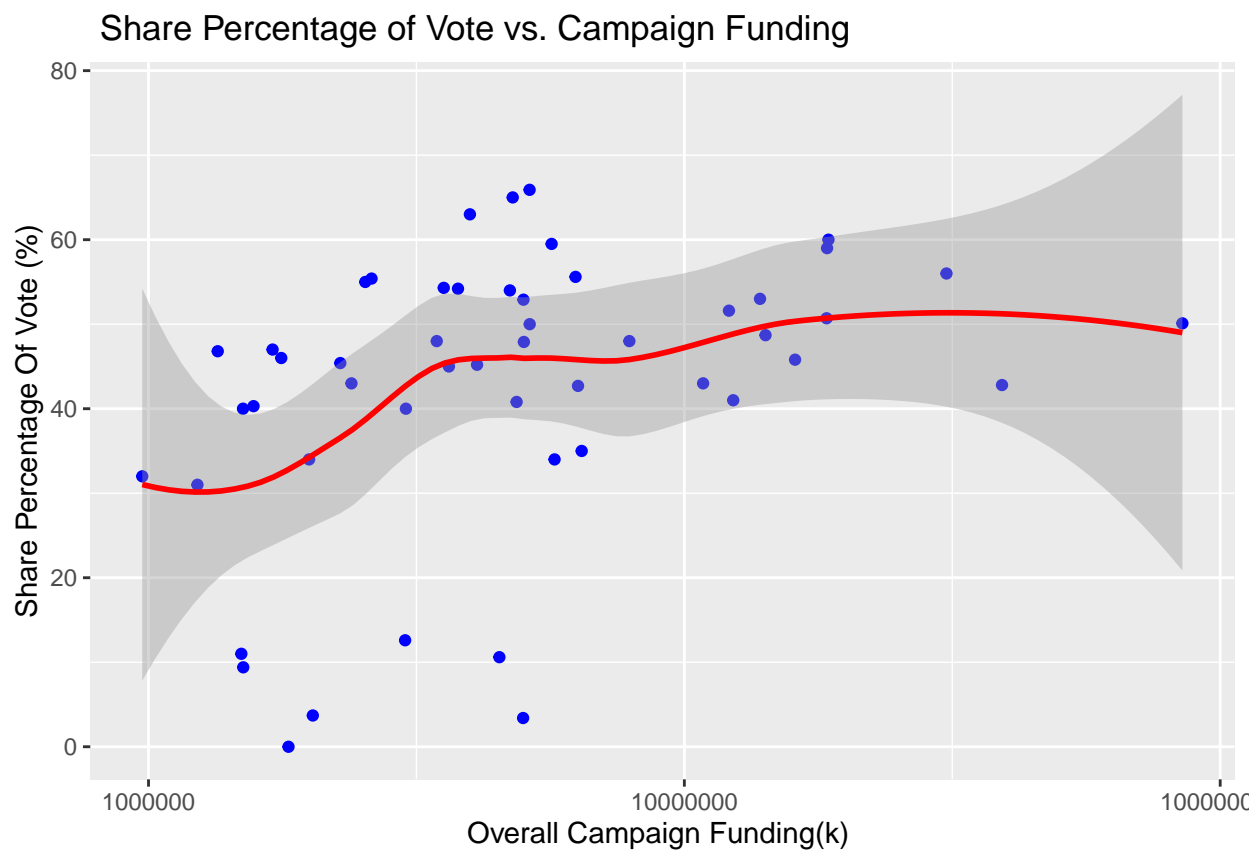
term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
intercept	1132003.8	5147677.4	0.22	0.83	-9212641.27	11476648.8
pervote	171455.5	113073.6	1.52	0.14	-55774.31	398685.3

Interpretation

The intercept of 1,132,003.8, is the estimated campaign funding when the share percentage of votes received by the candidate is zero. There is an expected increase of 171,455.5 in campaign funding for each unit

increase in the share percentage of votes received by the candidate.

```
cand_sp|>
  ggplot(
    mapping = aes(
      x = overallfunding,
      y = pervote)) +
  geom_point(color = "blue") +
  geom_smooth(color = "red") +
  scale_x_log10() +
  labs(x = "Overall Campaign Funding(k) ",
       y = "Share Percentage Of Vote (%)",
       title = "Share Percentage of Vote vs. Campaign Funding")
```



The graph above denotes the relationship between the amount of campaign funding a candidate has and the share percentage of votes they received in the election. In this case, without evaluating which candidates won or loss, there seems to be a *slight positive correlation* between an increase in funding and an increase in the share percentage of votes the candidate received.

Conclusion

From the findings above it is clear that there is not a statistically significant relationship between campaign funds and the share percentage of votes candidates receive in elections. This can be seen in the p-value of `pervote` at .14 which we would fail to reject the null at the significance level of 0.05.

There is also not a statistically significant relationship between campaign funds, the results of the campaign, and the share percentage of votes candidates receive in elections. This can be seen by both the p-value of 0.072 at a significance level of 0.05 and with the scatterplot of both loss and wins as the both have slight positive relationships with the increase of campaign funding.

(talk about my hypothesis.

Limitations

It should be noted that this study only considers monetary funding as the main independent variable and thus the variable that is causing the outcomes. However, in governmental elections there are a host of factors that actually effect why a candidate may or may not be able to secure the seat. There were also candidates that had to be removed from the original data set because they were included in the main data frame but may have not made it to the main election.

Future Improvements

In future studies more factors can be evaluated that are related to the outcome of elections. They could possibly introduce how many commercials a candidate used or polls that evaluate public opinion around how candidates do in debates. With the introduction of more factors, the causal reasoning of the study will be much stronger because there will be less confounding variables.