Politics Are Afoot!

Da Qi Ren

The Setup

There is *a lot* of money that is spent in politics in Presidential election years. So far, estimates have the number at about $11,000,000,000 (11 billion USD). For context, in 2019 Twitter's annual revenue was about $3,500,000,000 (3.5 billion USD).

The work

Install the package, `fec16`.

```
## install.packages('fec16')
```

This package is a compendium of spending and results from the 2016 election cycle. In this dataset are 9 different datasets that cover:

- `candidates`: candidate attributes, like their name, a unique id of the candidate, the election year under consideration, the office they're running for, etc.
- `results_house`: race attributes, like the name of the candidates running in the election, a unique id of the candidate, the number of `general_votes` garnered by each candidate, and other information.
- `campaigns`: financial information for each house & senate campaign. This includes a unique candidate id, the total receipts (how much came in the doors), and total disbursements (the total spent by the campaign), the total contributed by party central committees, and other information.

Your task

Describe the relationship between spending on a candidate's behalf and the votes they receive.

Your work

- We want to keep this work *relatively* constrained, which is why we're providing you with data through the `fec16` package. It is possible to gather all the information from current FEC reports, but it would require you to make a series of API calls that would pull us away from the core modeling tasks that we want you to focus on instead.
- Throughout this assignment, limit yourself to functions that are within the `tidyverse` family of packages: `dplyr`, `ggplot`, `patchwork`, and `magrittr` for wrangling and exploration and `base`, `stats`, `sandwich` and `lmtest` for modeling and testing. You do not *have* to use these packages; but try to limit yourself to using only these.

```
library(tidyverse)
library(magrittr)
library(ggplot2)
library(patchwork)
library(sandwich)
library(lmtest)
library(fec16)
theme_set(theme_minimal())
knitr::opts_chunk$set(dpi = 300)
```

```
candidates    <- fec16::candidates
results_house <- fec16::results_house
campaigns     <- fec16::campaigns
```

1. What does the distribution of votes and of spending look like?

1. (3 points) In separate histograms, show both the distribution of votes (measured in `results_house$general_percent` for now) and spending (measured in `ttl_disb`). Use a log transform if appropriate for each visualization. How would you describe what you see in these two plots?

```
summary(candidates)
```

```
##    cand_id           cand_name         cand_pty_affiliation cand_election_yr
##  Length:4699        Length:4699        Length:4699          Min.   :2015
##  Class :character   Class :character   Class :character     1st Qu.:2016
##  Mode  :character   Mode  :character   Mode  :character     Median :2016
##                                                            Mean   :2016
##                                                            3rd Qu.:2016
##                                                            Max.   :2016
##  cand_office_st     cand_office        cand_office_district   cand_ici
##  Length:4699        Length:4699        Length:4699          Length:4699
##  Class :character   Class :character   Class :character     Class :character
##  Mode  :character   Mode  :character   Mode  :character     Mode  :character
##
##
##
##  cand_status          cand_pcc           cand_st1           cand_st2
##  Length:4699        Length:4699        Length:4699        Length:4699
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##   cand_city           cand_st            cand_zip
##  Length:4699        Length:4699        Length:4699
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
summary(results_house)
```

```
##     state           district_id          cand_id          incumbent
##  Length:2110        Length:2110        Length:2110        Mode :logical
##  Class :character   Class :character   Class :character   FALSE:1660
##  Mode  :character   Mode  :character   Mode  :character   TRUE :450
##
##
##
##
##     party           primary_votes    primary_percent   runoff_votes
##  Length:2110        Min.   :     1   Min.   :0.0000   Min.   : 1096
##  Class :character   1st Qu.:  4776   1st Qu.:0.1155   1st Qu.: 1464
##  Mode  :character   Median : 13994   Median :0.3206   Median : 8206
##                     Mean   : 25469   Mean   :0.4300   Mean   :11274
##                     3rd Qu.: 35234   3rd Qu.:0.7420   3rd Qu.:20082
##                     Max.   :326988   Max.   :1.0000   Max.   :25322
##                     NA's   :618      NA's   :619      NA's   :2098
##  runoff_percent   general_votes    general_percent     won
##  Min.   :0.3427   Min.   :     1   Min.   :0.0000   Mode :logical
##  1st Qu.:0.4624   1st Qu.: 12164   1st Qu.:0.0407   FALSE:1615
##  Median :0.5000   Median :102234   Median :0.3385   TRUE :495
##  Mean   :0.5000   Mean   :101632   Mean   :0.3417
##  3rd Qu.:0.5376   3rd Qu.:172444   3rd Qu.:0.5856
##  Max.   :0.6573   Max.   :718591   Max.   :1.0000
##  NA's   :2098     NA's   :819      NA's   :820
##   footnotes
##  Length:2110
##  Class :character
##  Mode  :character
##
##
##
##
```

```
summary(campaigns)
```

```
##    cand_id           cand_name           cand_ici             pty_cd
##  Length:1898        Length:1898        Length:1898        Min.   :1.00
##  Class :character   Class :character   Class :character   1st Qu.:1.00
##  Mode  :character   Mode  :character   Mode  :character   Median :2.00
##                                                           Mean   :1.73
##                                                           3rd Qu.:2.00
##                                                           Max.   :3.00
##  cand_pty_affiliation  ttl_receipts       trans_from_auth
##  Length:1898          Min.   :        0   Min.   :        0
##  Class :character     1st Qu.:    16249   1st Qu.:        0
##  Mode  :character     Median :   141860   Median :        0
##                       Mean   :  1713080   Mean   :   197650
##                       3rd Qu.:  1087465   3rd Qu.:        0
##                       Max.   :585669599   Max.   :160760000
##     ttl_disb         trans_to_auth        coh_bop            coh_cop
```

```
##  Min.   :    -1458   Min.   :      0   Min.   :  -18681   Min.   :  -37238
##  1st Qu.:    15984   1st Qu.:      0   1st Qu.:       0   1st Qu.:       0
##  Median :   142803   Median :      0   Median :       0   Median :    1032
##  Mean   :  1689547   Mean   :  11988   Mean   :  168253   Mean   :  187980
##  3rd Qu.:   977213   3rd Qu.:      0   3rd Qu.:   20778   3rd Qu.:   63073
##  Max.   :585346281   Max.   :3642535   Max.   :18055430   Max.   :10676815
##   cand_contrib        cand_loans        other_loans     cand_loan_repay
##  Min.   :       0   Min.   :       0   Min.   :     0   Min.   :      0
##  1st Qu.:       0   1st Qu.:       0   1st Qu.:     0   1st Qu.:      0
##  Median :       0   Median :       0   Median :     0   Median :      0
##  Mean   :   29765   Mean   :   88093   Mean   :  1530   Mean   :  12695
##  3rd Qu.:    1288   3rd Qu.:    7959   3rd Qu.:     0   3rd Qu.:      0
##  Max.   :18633209   Max.   :47508505   Max.   :500000   Max.   :1655854
##  other_loan_repay   debts_owed_by     ttl_indiv_contrib   cand_office_st
##  Min.   :     0.0   Min.   :  -1786   Min.   :        0   Length:1898
##  1st Qu.:     0.0   1st Qu.:      0   1st Qu.:     5587   Class :character
##  Median :     0.0   Median :      0   Median :    73973   Mode  :character
##  Mean   :   755.7   Mean   :  59302   Mean   :  1114190
##  3rd Qu.:     0.0   3rd Qu.:   9966   3rd Qu.:   500779
##  Max.   :500000.0   Max.   :8158515   Max.   :405664230
##  cand_office_district other_pol_cmte_contrib pol_pty_contrib
##  Length:1898          Min.   :      0        Min.   :     0
##  Class :character     1st Qu.:      0        1st Qu.:     0
##  Mode  :character     Median :   2000        Median :     0
##                       Mean   : 245702        Mean   :  1702
##                       3rd Qu.: 322028        3rd Qu.:     0
##                       Max.   :4134430        Max.   :449200
##    cvg_end_dt         indiv_refunds      cmte_refunds
##  Min.   :2015-02-01   Min.   :  -1150   Min.   : -1000
##  1st Qu.:2016-09-30   1st Qu.:      0   1st Qu.:     0
##  Median :2016-12-31   Median :      0   Median :     0
##  Mean   :2016-10-24   Mean   :  24834   Mean   :  2198
##  3rd Qu.:2016-12-31   3rd Qu.:   5397   3rd Qu.:     0
##  Max.   :2017-01-31   Max.   :5994031   Max.   :310035
```

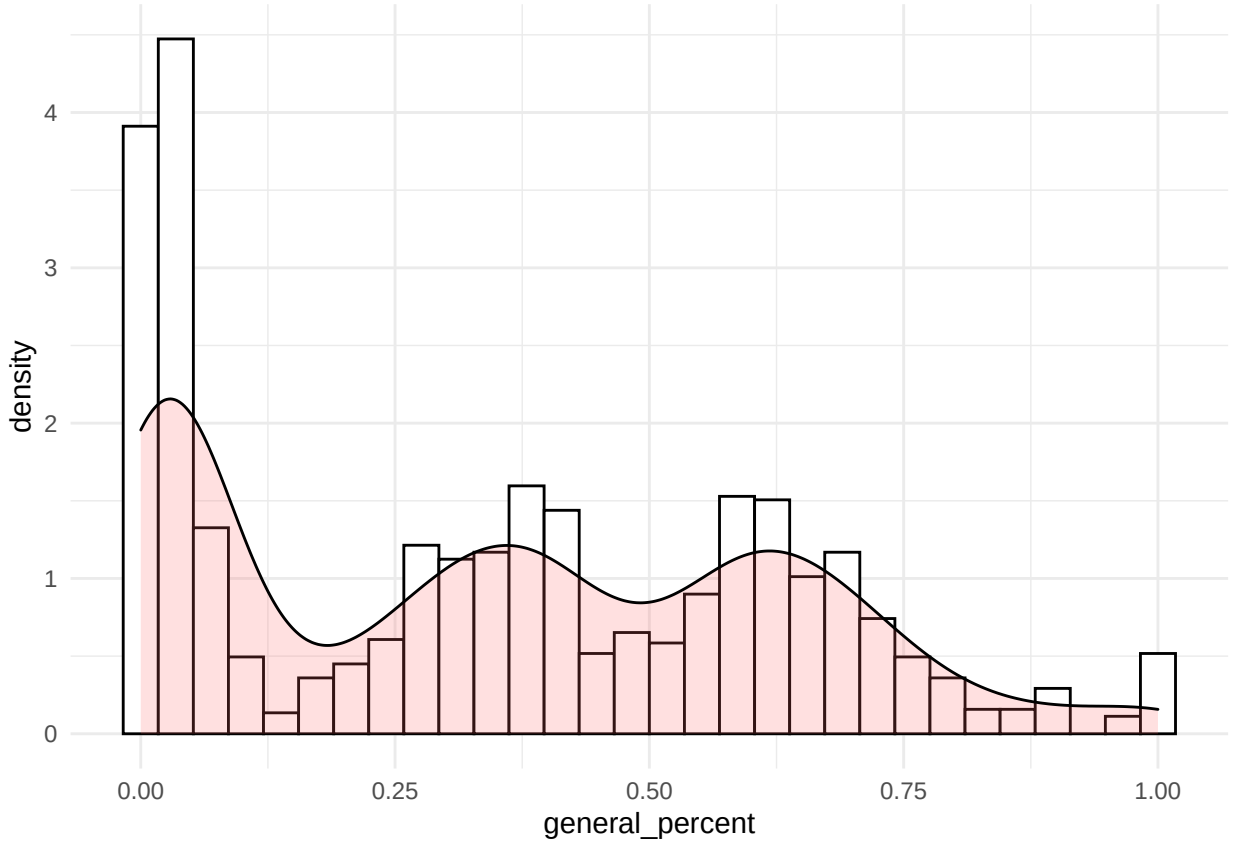
```
#ggplot(data= results_house, aes(y=general_percent))

#discrete.histogram(results_house$general_percent )

library(ggplot2)
# # Basic histogram
# ggplot(results_house, aes(x = general_percent )) + geom_histogram()
# # Change the width of bins
# ggplot(results_house, aes(x =  general_percent )) +
#   geom_histogram(binwidth=1)
# # Change colors

p<-ggplot(results_house, aes(x= general_percent)) +
   geom_histogram(color="black", fill="white")
#
# # Add mean line
# p+ geom_vline(aes(xintercept=mean( general_percent)),
#             color="blue", linetype="dashed", size=1)
# Histogram with density plot
```

```
ggplot( results_house  , aes(x= general_percent)) +
 geom_histogram(aes(y=..density..), colour="black", fill="white")+
 geom_density(alpha=.2, fill="#FF6666")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

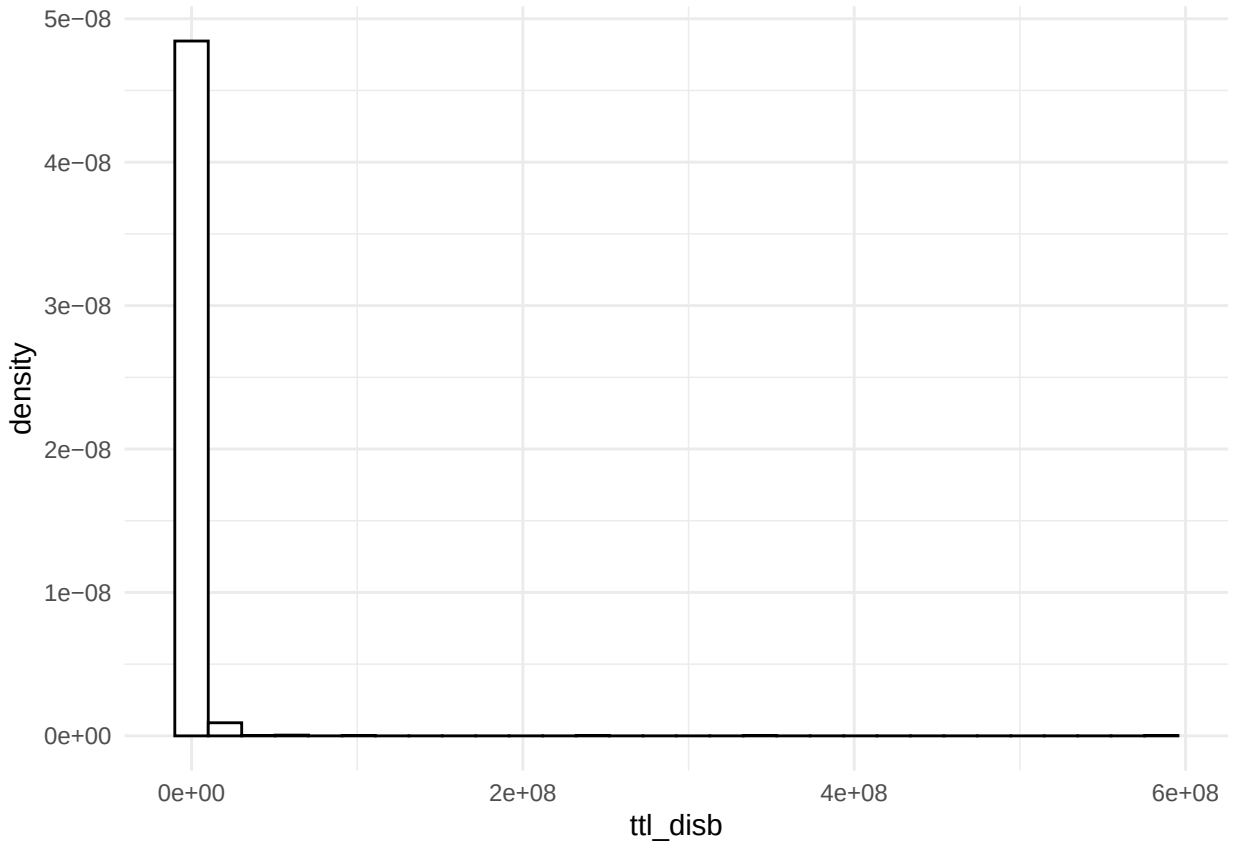
```
## Warning: Removed 820 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 820 rows containing non-finite values (stat_density).
```




```
# p+ geom_vline(aes(xintercept=mean(ttl_disb)),
#             color="blue", linetype="dashed", size=1)
# Histogram with density plot


p<-ggplot(campaigns, aes(x= ttl_disb)) +
   geom_histogram(color="black", fill="white")

ggplot(campaigns  , aes(x= ttl_disb)) +
 geom_histogram(aes(y=..density..), colour="black", fill="white")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#+
# geom_density(alpha=.2, fill="#FF6666")
```

2. Exploring the relationship between spending and votes.

2. (3 points) Create a new dataframe by joining `results_house` and `campaigns` using the `inner_join` function from `dplyr`. (We use the format `package::function` – so `dplyr::inner_join`.)

3. (3 points) Produce a scatter plot of `general_votes` on the y-axis and `ttl_disb` on the x-axis. What do you observe about the shape of the joint distribution?

4. (3 points) Create a new variable to indicate whether each individual is a "Democrat", "Republican" or "Other Party".

- Here's an example of how you might use `mutate` and `case_when` together to create a variable.

```
starwars %>%
  select(name:mass, gender, species) %>%
  mutate(
  type = case_when(
    height > 200 | mass > 200 ~ "large",
    species == "Droid"        ~ "robot",
    TRUE                      ~ "other"
    )
  )
```

Once you've produced the new variable, plot your scatter plot again, but this time adding an argument into the `aes()` function that colors the points by party membership. What do you observe about the distribution of all three variables?

Produce a Descriptive Model

5. (5 Points) Given your observations, produce a linear model that you think does a good job at describing the relationship between candidate spending and votes they receive. You should decide what transformation to apply to spending (if any), what transformation to apply to votes (if any) and also how to include the party affiliation.

6. (3 points) Interpret the model coefficients you estimate.

- Tasks to keep in mind as you're writing about your model:
  - At the time that you're writing and interpreting your regression coefficients you'll be *deep* in the analysis. Nobody will know more about the data than you do, at that point. *So, although it will feel tedious, be descriptive and thorough in describing your observations.*
  - It can be hard to strike the balance between: on the one hand, writing enough of the technical underpinnings to know that your model meets the assumptions that it must; and, on the other hand, writing little enough about the model assumptions that the implications of the model can still be clear. We're starting this practice now, so that by the end of Lab 2 you will have had several chances to strike this balance.