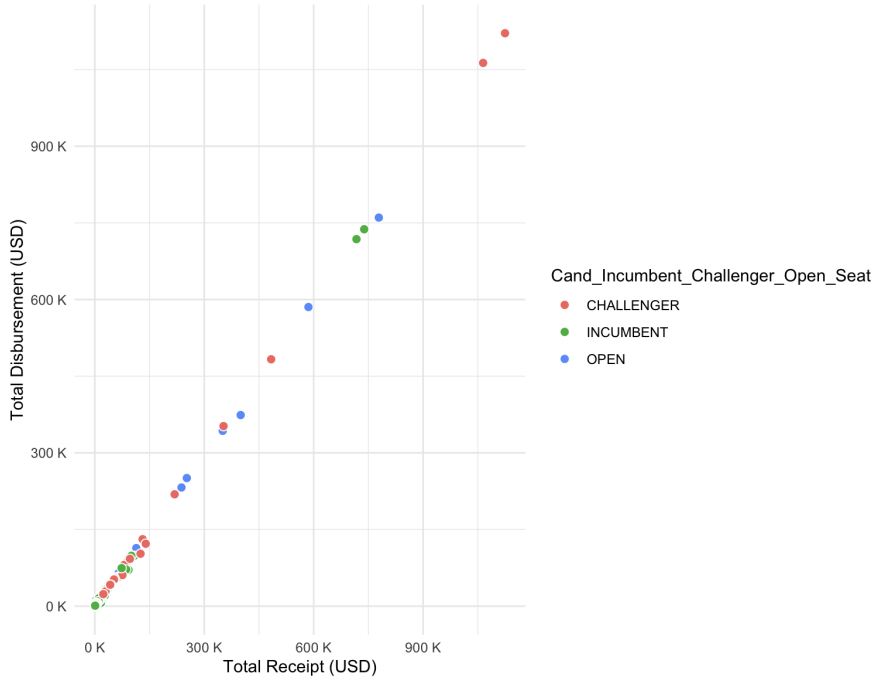
Kentrick Felix Kepawitono (ID: 56462267)

Data Exploration of Candidate Summary File Description

Visualizations

Visualization 1:

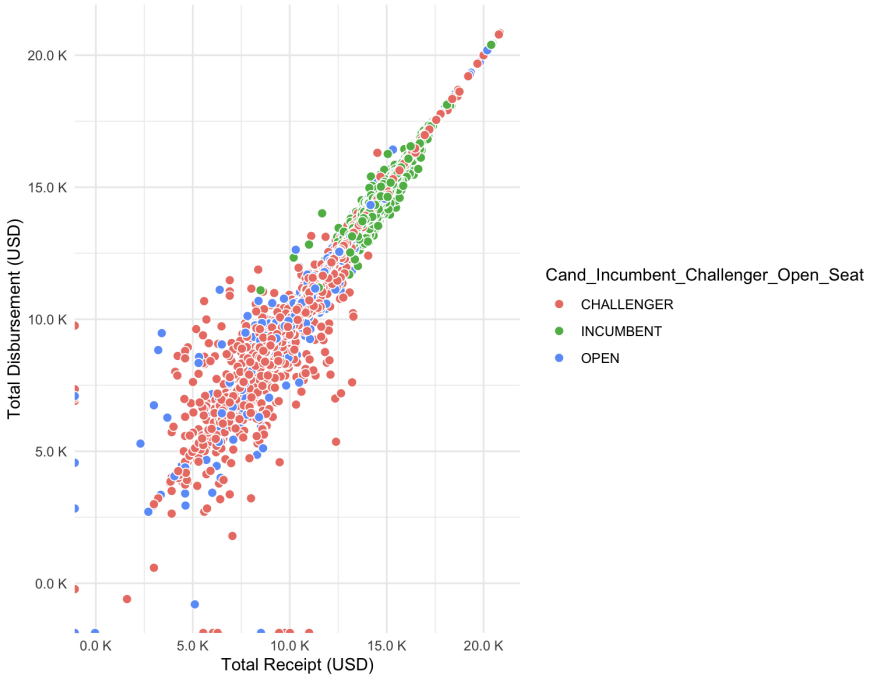



```
#Cite:
https://stackoverflow.com/questions/3445590/subset-filter-rows-in-a-data-frame-base
d-on-a-condition-in-a-column
#Cite: https://www.roelpeters.be/scale-ggplot-y-axis-millions-or-thousands-r/
cha_inc_open <- subset(fec, (Cand_Incumbent_Challenger_Open_Seat == "CHALLENGER" |
                        Cand_Incumbent_Challenger_Open_Seat == "INCUMBENT" |
                        Cand_Incumbent_Challenger_Open_Seat == "OPEN")&
                        Operating_Expenditure > 0 &
                        Total_Contribution > 0)

ggplot(cha_inc_open, aes(Total_Receipt, Total_Disbursement, fill =
Cand_Incumbent_Challenger_Open_Seat)) +
  geom_point(pch = 21, color = "white", size = 2.5, na.rm = FALSE) +
  scale_x_continuous(name = "Total Receipt (USD)", labels = label_number(suffix = "
K", scale = 1e-6)) +
  scale_y_continuous(name = "Total Disbursement (USD)", labels =
label_number(suffix = " K", scale = 1e-6)) +
  theme_minimal()
```

The above visualization is a scatter plot showing that there is a relationship between Total Disbursement and Total Receipt. It clearly shows a trend where the more the candidates receive the money from multiple sources (e.g. party committee contributions, candidate contributions), the more they are likely to spend for their campaign. Besides, the condensed dots between $0 thousand and $150 thousand shows that most candidates receive and spend in the range of $0 thousand and $150 thousand for their campaigns.

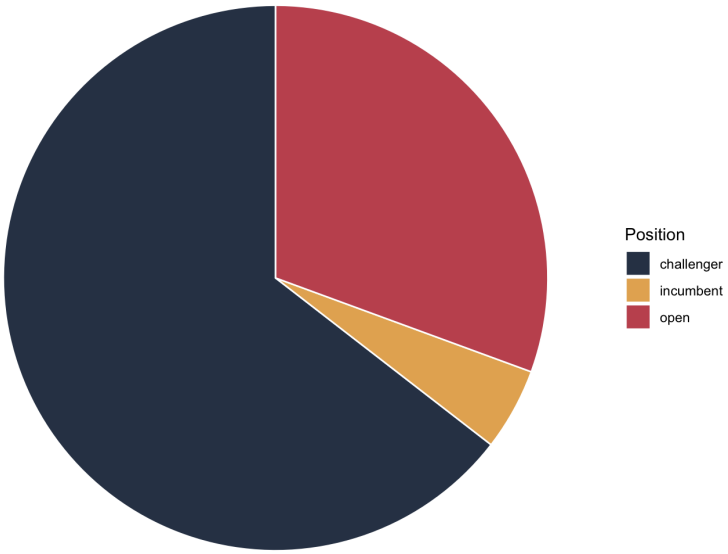
Visualization 2:




```
ggplot(cha_inc_open, aes(log(Total_Receipt), log(Total_Disbursement), fill =
Cand_Incumbent_Challenger_Open_Seat)) +
  geom_point(pch = 21, color = "white", size = 2.5, na.rm = FALSE) +
  scale_x_continuous(name = "Total Receipt (USD)", labels = label_number(suffix = "
K")) +
  scale_y_continuous(name = "Total Disbursement (USD)", labels =
label_number(suffix = " K")) +
  theme_minimal()
```

The above visualization is the same scatter plot as the first visualization, but with the application of logarithmic scale. The logarithmic scale helps us visualize the part of the plot that has a lot of points, in this case from $0K to $20K. From the visualization above, we can see that the majority of incumbents tend to receive and spend money more than challengers and opens. However, a few challengers also seem to receive and spend more than most incumbents.

Visualization 3:



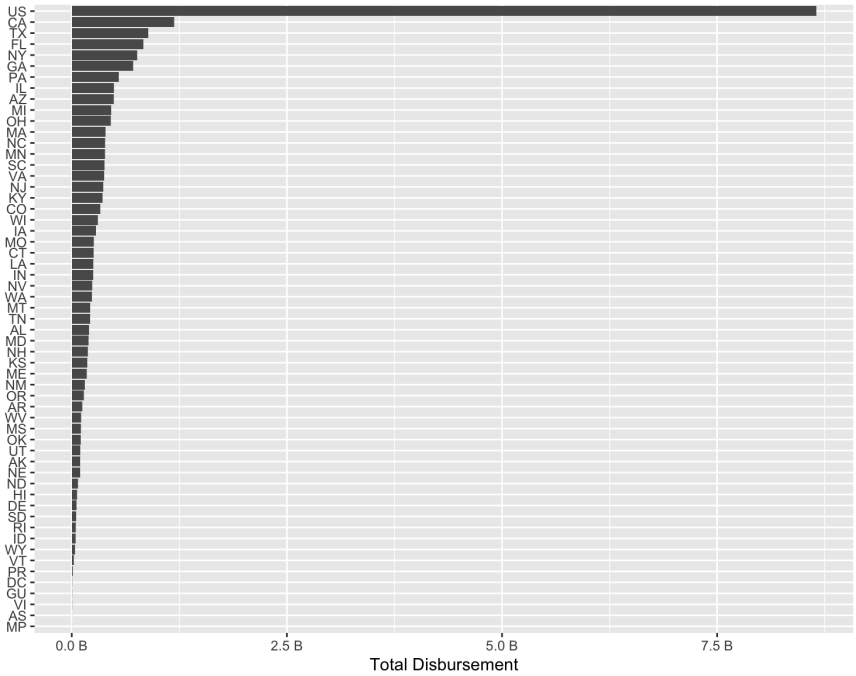

```
incumbent <-
data.frame(incumbent=unique(fec$Cand_Id[which(fec$Cand_Incumbent_Challenger_Open_Se
at == "INCUMBENT")]))
challenger <-
data.frame(challenger=unique(fec$Cand_Id[which(fec$Cand_Incumbent_Challenger_Open_S
eat == "CHALLENGER")]))
open <-
data.frame(open=unique(fec$Cand_Id[which(fec$Cand_Incumbent_Challenger_Open_Seat ==
"OPEN")]))

inc_cha_open <- data.frame(
  Position=c(colnames(challenger), colnames(incumbent), colnames(open)),
  value=c(nrow(challenger), nrow(incumbent), nrow(open))
)

ggplot(inc_cha_open, aes(x="", y=value, fill=Position)) +
  geom_bar(stat="identity", width=1, color="white") +
  coord_polar("y", start=0) +
  scale_fill_manual(values = c("#2e4057", "#edae49", "#d1495b")) +
  theme_void()
```

The above visualization is a pie chart. It clearly shows that there are more candidates who applied as challengers in the campaign. Challenger takes more than half of the pie chart. The open position is the second most popular position applied by the candidates. Then the incumbent position comes up as the least popular position.

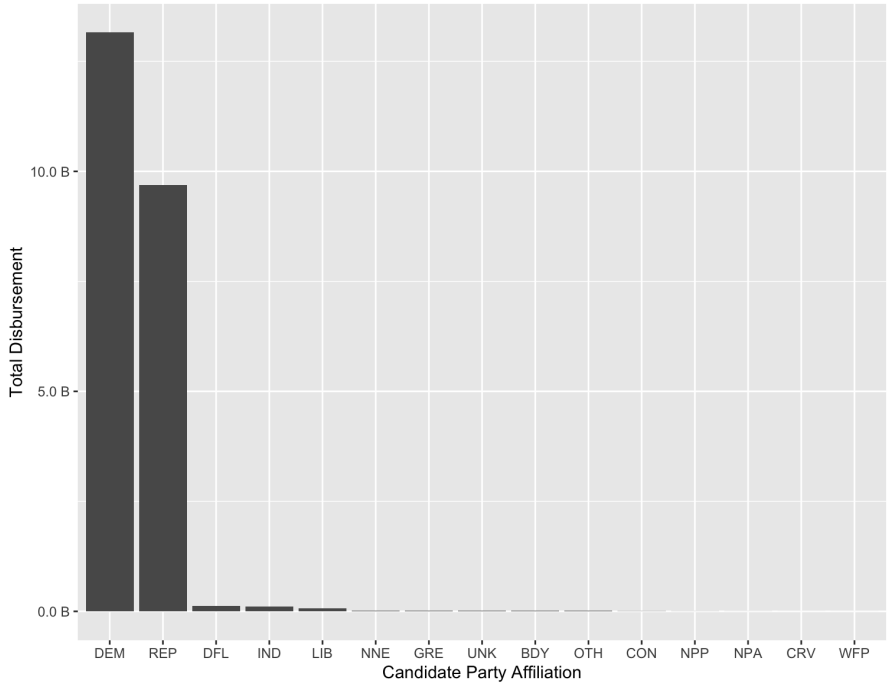
Visualization 4:




```
#Cite: https://www.roelpeters.be/scale-ggplot-y-axis-millions-or-thousands-r/
ggplot(fec, aes(x=reorder(Cand_Office_St, Total_Disbursement, FUN = sum),
y=Total_Disbursement)) +
  geom_bar(stat = "identity") +
  coord_flip() + labs(x = "", y = "Total Disbursement") +
  scale_y_continuous(labels = label_number(suffix = " B", scale = 1e-9))
```

The above visualization is a bar plot. It shows the total disbursement of each state, including the whole state (US), spent on campaigns. From the visualization, we can see that the US has the highest total disbursement spent for campaigns. California comes as the first state that spent the most, followed by Texas and Florida. Virgin Islands (VI), American Samoa (AS), and Northern Mariana Islands (MP) are three territories that spent the least amount of money when it comes to campaigns.

Visualization 5:



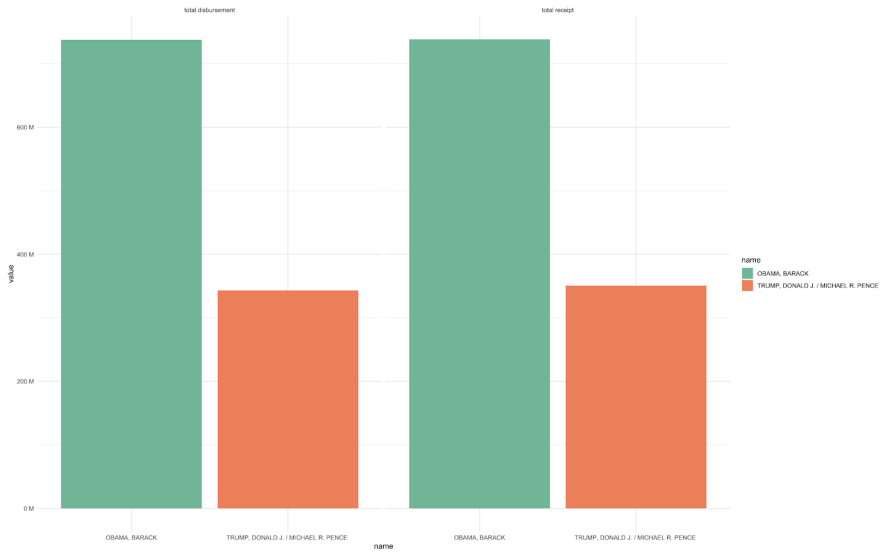

```
#Cite: https://stackoverflow.com/questions/1660124/how-to-sum-a-variable-by-group
#Cite:
https://stackoverflow.com/questions/32141332/select-and-plot-top-frequencies-with-d
plyr
#Cite: https://www.roelpeters.be/scale-ggplot-y-axis-millions-or-thousands-r/

Party_exp <- data.frame(aggregate(fec$Total_Disbursement,
by=list(Cand_Party_Affiliation=fec$Cand_Party_Affiliation), FUN=sum))

top_n(Party_exp, n=15, x) %>%
  ggplot(., aes(x=reorder(Cand_Party_Affiliation, -x), y=x))+
  geom_bar(stat='identity') +
  scale_y_continuous(labels = label_number(suffix = " B", scale = 1e-9)) +
  labs(x = "Candidate Party Affiliation", y = "Total Disbursement")
```

The above visualization is a bar plot. The visualization shows how the top 15 party affiliations spent their money on campaigns. In this visualization, we want to focus on two popular party affiliations: Democratic Party and Republican Party. The Democratic Party seems to spend more than the Republican Party on campaigns. The difference in the spending is quite large; the Democratic Party spent more than $12.5 billion, while the Republican Party spent under $10 billion.

Visualization 6:




```
#Cite: https://www.roelpeters.be/scale-ggplot-y-axis-millions-or-thousands-r/
#We want to get the data of Obama's and Trump's latest presidential term only (2012
for Obama and 2016 for Trump)
obama_trump <- data.frame(fec[fec$Link_Image ==
"https://www.fec.gov/data/candidate/P80003338/?cycle=2012" |
                                  fec$Link_Image ==
"https://www.fec.gov/data/candidate/P80001571/?cycle=2016", ])

#Make a new table to make a side by side bars
revised_obama_trump <- data.frame(
  name = c(obama_trump$Cand_Name[which(obama_trump$Cand_Id == "P80003338")],
           obama_trump$Cand_Name[which(obama_trump$Cand_Id == "P80003338")],
           obama_trump$Cand_Name[which(obama_trump$Cand_Id == "P80001571")],
           obama_trump$Cand_Name[which(obama_trump$Cand_Id == "P80001571")]),
  payment_type = c("total disbursement", "total receipt",
                   "total disbursement", "total receipt"),
  value = c(obama_trump$Total_Disbursement[which(obama_trump$Cand_Id ==
"P80003338")],
            obama_trump$Total_Receipt[which(obama_trump$Cand_Id == "P80003338")],
            obama_trump$Total_Disbursement[which(obama_trump$Cand_Id ==
"P80001571")],
            obama_trump$Total_Receipt[which(obama_trump$Cand_Id == "P80001571")])
)

ggplot(revised_obama_trump, aes(x = name, y = value, fill = name)) +
  geom_col() +
  facet_wrap(~payment_type) +
  scale_fill_brewer(palette = "Set2") +
  scale_y_continuous(labels = label_number(suffix = " M", scale = 1e-6)) +
  theme_minimal()
```

The above visualization is a side-by-side bar plot, which is intended to make it easier to compare the total disbursement and total receipt of Barack Obama and Donald Trump. The data acquired from their most recent term of office (2012 for Obama and 2016 for Trump). The green represents Obama and the orange represents Trump. From this visualization, we can tell that Obama received and spent more money on the campaign than Trump (doubles the amount that Trump received and spent).

Kentrick Felix Kepawitono (ID: 56462267)

Hypotheses

Hypothesis 1

A lot of candidates preferred to sign up as challengers because funding should be bigger for the challengers for starting the campaign from scratch and gain supporters; therefore, there will be fairness around the candidates.

Visualization 1 and 2 refute the claim partially, while visualization 3 supports a part of the claim. The claim mentions that challengers would get more funding to support their campaign. However, according to visualization 2, the majority of challengers get lower funding as compared to the incumbents. Visualization 2 shows the dots representing the challengers spread around $2,500 to $12,500, while the dots representing the incumbents are around $12,500 to $17,500. Visualization 1 shows that the top two candidates who receive the most funding are challengers. However, we can ignore these because they are outliers since they do not represent the majority of the challengers.

Visualization 3, on the other hand, supports the first part of the claim. Visualization 3 supports the claim that says: “there are more candidates apply as challengers.” Visualization 3 shows that more than half of the pie chart is represented by challengers. The open comes next, taking over more than a quarter of the pie chart. The incumbent comes the last.

Hypothesis 2

The higher mean of total receipt of the Democratic Party as compared to the Republican Party shows that the Democratic Party would tend to spend more money for their campaigns.

```
> #Mean of the total receipt from Democratic Party
> mean(fec$Total_Receipt[which(fec$Cand_Party_Affiliation == "DEM")])
[1] 1464536
>
> #Mean of the total receipt from Republican Party
> mean(fec$Total_Receipt[which(fec$Cand_Party_Affiliation == "REP")])
[1] 964994
```

The claim is supported by visualization 1, 5, and 6. Visualization 1 shows that there is a positive correlation between total receipt and total disbursement; therefore, the more the total receipt of the Democratic Party, the more total disbursement of the Democratic Party. In this case, we can see that the mean of the total receipt from the Democratic Party is higher than the total receipt of the Republican Party; therefore, we can deduce that the Democratic Party also has higher total disbursement than the Republican Party.

Visualization 2 also supports the claim. Visualization 2 (the bar plot) shows that Democratic Party has higher total disbursement of the campaigns than Republican Party. There is a difference between the two parties; the Democratic Party spent more than $12.5 billion, while the Republican Party spent under $10 billion. As an example, I also included visualization 6. Barack Obama, who is part of the Democratic Party, spent almost twice the amount that Donald Trump, who is part of the Republican Party, spent.