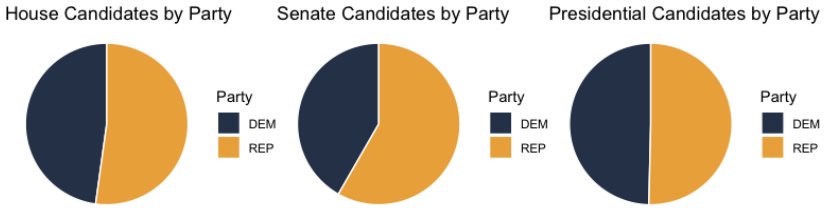
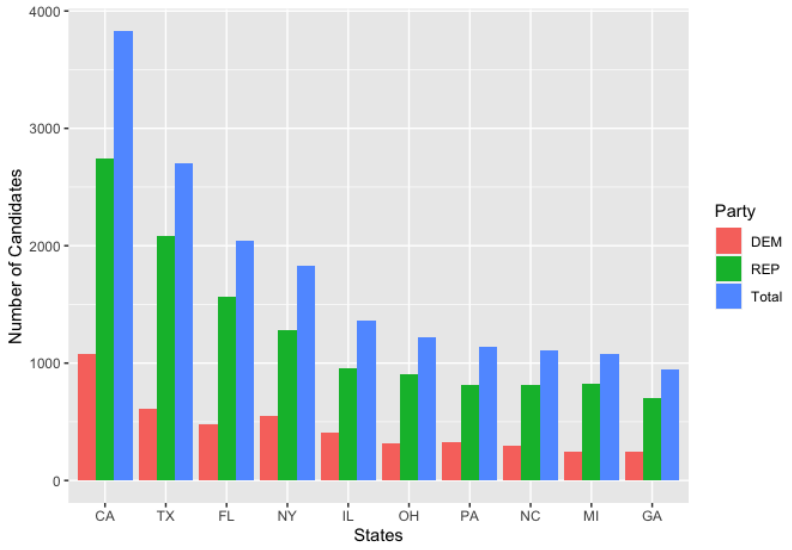
Part 1 - Visual Data Exploration

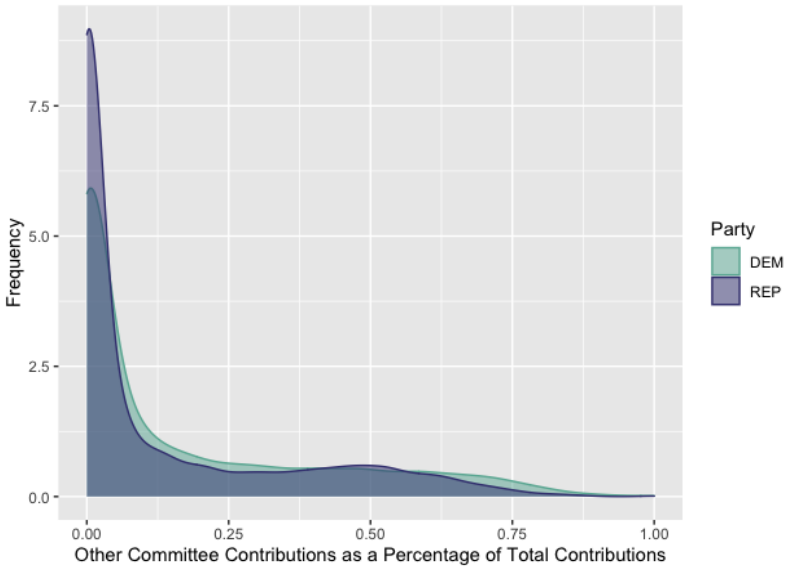
Disclaimer: Only data from Democrats and Republican candidates are included in these graphs except for the last graph, Visualization 6. Thus, 100 percent is equal to all Democratic and Republican candidates. I did this because the Democratic and Republican parties are the two major parties in America, making up the vast majority of politicians in the House, Senate, and Presidential positions. Other parties do not have a realistic path to winning these positions in current day politics.



Visualization 1: In this group of pie charts, the proportions of Democratic Party candidates and Republican Party candidates running for House, Senate, and Presidential positions are shown. In the first pie chart, it can be seen that Republican candidates are just over 50% of House candidates while Democratic candidates are just under 50%. In the second pie chart, Republican candidates are about 60% of Senate candidates while Democratic candidates are 40%. In the last pie chart, Republican and Democratic candidates for President are almost split evenly, with Republican candidates comprising slightly more than 50%, while Democratic candidates comprise slightly less than 50%.

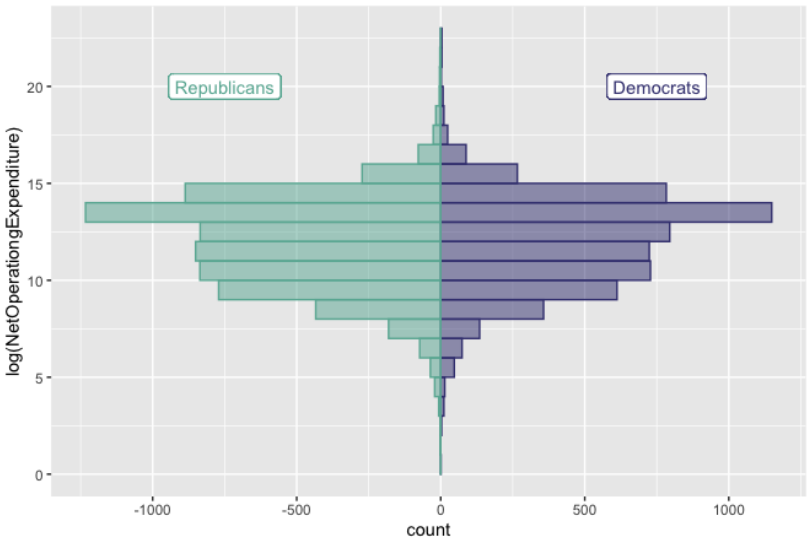


Visualization 2: This graph shows the number of Democratic, Republican, and total number of political candidates for House, Senate, and Presidential positions in the top 10 most populous states in the US. As you can see, Republican Candidates greatly outnumber Democratic candidates in these states. In California, the most populous state, Republican candidates outnumber Democratic Candidates about 2750 to 1000. In Georgia, the 10th most populous state, Republican candidates outnumber Democratic Candidates about 800 to 200.

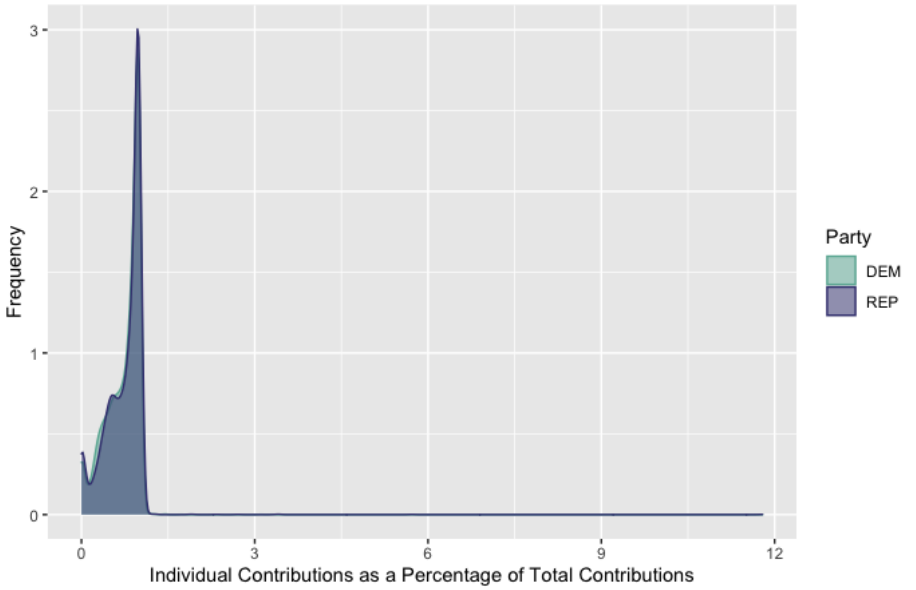


Visualization 3: This graph shows the distribution of "Other Committee Contributions" as a percentage of Total Contributions to Democratic and Republicans. Democrats and Republicans are separated into two distributions which overlay each other. "Other Committee Contributions " include mostly contributions from Political Action Committees, but also other candidates. Candidates with a total contribution of 0$ were removed from the data.

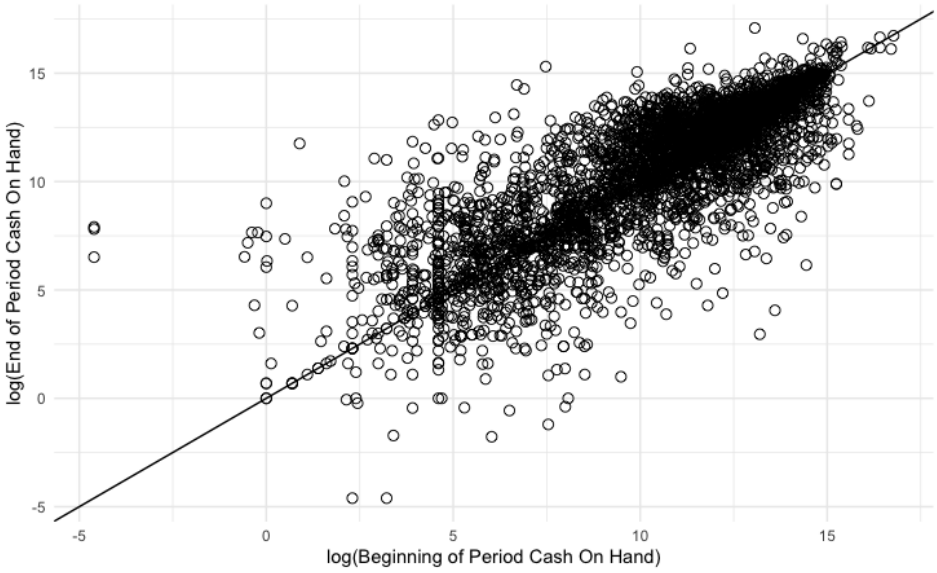
As is shown by the graph, before around 3-4% Republicans have a higher frequency of candidates than Democrats. For the rest of the graph, Democrats have a slightly higher frequency than Republicans except for a small blip between around 40-55% where the Republican distribution slightly overtakes the Democratic distribution. This shows that Democrats are more likely to receive a greater percentage of their total contributions from non Party Committees compared to Republicans, as the Republican distribution is higher only on the very left of the graph and slightly higher for a small blip in the middle.



Visualization 4: This graph shows the distributions of the log of the Net Operating Expenditure for Republican and Democratic candidates, with the Republican distribution on the left and the Democratic distribution on the right. The Net Operating Expenditure is the sum of all operating expenditures minus any offsets to those expenditures. Zero values were dropped from the graph so the log of Net Operating Expenditures could be graphed. The Republican and Democratic distributions nearly mirror each other with no significant changes in terms of changes in the distribution from bin to bin. However, the bins for the Republican distribution for bins 8 to 15 are noticeably greater than the bins in the Democratic distribution. For example, bin 14 in the Republican distribution has a value of almost 1250, while the value of the same bin in the Democratic distribution is just over 1075. Most of the other bins have a similar value for both distributions. As the majority of data points are in bins 8 to 15 across both distributions, there was no relationship showing no major difference between the Expenditures for Republican and Democratic Candidates, at least in Candidates with an Net Expenditure above 0.



Visualization 5: This graph shows the distribution of aggregated Individual Contributions to Democratic and Republican candidates as a percentage of total contributions to candidates. Candidate data with a Total Contribution value of zero were removed, as dividing by zero is an error. Also, even if the value were considered 100%, this would be considered outlier data. The distribution for Democratic and Republican candidates are overlaid over each other, but both distributions match almost identically, showing no significant difference in the Individual Contributions as a percentage of total contributions across the two parties' candidates. For both parties, the distribution has an extremely strong peak at about 1%, with the vast majority of values lying between about zero and one percent.



Visualization 6: (This visualization includes all parties of candidates)
This graph is a scatter plot of points with the x value representing the log of candidates' cash on hand for their campaign at the start of the two year period in which data is collected on them. The y value represents the log of candidates' cash on hand at the end of the period. As you can see from the graph, the points are heavily concentrated in a line from the bottom left to the top right of the graph, showing a positive correlation between the two values. The values follow fairly closely the y = x line on the graph, except that in the region between around 9 and 15 on the x axis, there appears to be a concentration of values right above the line. However, this needs to be investigated more due to the amount of points on the graph. Assuming the relation is true, candidates with high amounts of cash on hand tended to increase their cash on hand ever so slightly after two years. Otherwise, candidates tend to keep around the same amount of cash on hand for their campaigns, barring a few outliers.

Part 2 - Hypothesis

Hypothesis 1 - My first hypothesis is that there are many Republican candidates with 0$ of total contributions. In Visualization 1, which shows the proportion of Democratic and Republican candidates running for different positions, it is clear that there are many more Republican candidates than Democratic candidates, as there is a greater percentage of Republican candidates running for each position. This assertion is supported by Visualization 2, which shows the amount of Republican, Democratic, and total candidates in the 10 most populous states, in which Republican candidates outnumbered Democratic candidates greatly in all 10 of the most populous states. However, in Visualization 3, the distribution of Other Committee Contributions as a percentage of total contributions by party and Visualization 5, the distribution of Individual Contributions as a percentage of total contributions by party, the area under the curve of Democratic and Republican distributions is about the same. This can be explained by the removal of candidates whose total contributions were 0$ in both visualizations. Therefore, more Republican candidates must have been removed compared to Democratic candidates, otherwise, the Republican distributions would have noticeably more area.

Hypothesis 2 - My second hypothesis is that among candidates that have more than 0$ of total contributions, for most candidates, most of their contributions come from either Party Committees or Candidates themselves. As can be seen in Visualization 3 and Visualization 5, the proportion of total contributions for candidates from individuals and other committees for both Republicans and Democrats skew low. In Visualization 3, both proportion distributions peak at the left of the graph, then quickly drop and plateau for the rest of the graph. The Republican distribution around 9, dropping to 1 at around the 25% mark. The drop in the Democratic distribution is less drastic, starting at about 6 and dropping to 1 around the 25% mark. It can be safely assumed that for both distributions, the majority of values lie at 25% or below. In Visualization 5, both proportion distributions start low, increasing quickly to a high peak at 1%, then dropping steeply and quickly to the point where the Democratic distribution is a line extremely close to zero. The Republican distribution can not be seen at this point, so it can be assumed to either end before or match closely the Democratic distribution, leaving the same conclusion either way. Thus, the vast majority of values lie around 1% or less for both distributions. Only other two types of contributions, from Party Committees or Candidates themselves comprise the total amount of contributions, insinuating that most candidates must receive a very high proportion of their contributions from a combination of these two types of contributions.

Code

Visualization 1:

```
library(ggplot2)
library(cowplot)

Data <- read.csv("Downloads/fec_2008-2022.csv")

House <- data.frame(Cand = Data$Cand_Id[which(Data$Cand_Office == "H")], Party =
Data$Cand_Party_Affiliation[which(Data$Cand_Office == "H")])

Senate <- data.frame(Cand = Data$Cand_Id[which(Data$Cand_Office == "S")], Party =
Data$Cand_Party_Affiliation[which(Data$Cand_Office == "S")])

President <- data.frame(Cand = Data$Cand_Id[which(Data$Cand_Office == "P")], Party =
Data$Cand_Party_Affiliation[which(Data$Cand_Office == "P")])
House1 <- House

House["Party"][House["Party"] != "DEM"] <- NA

House1["Party"][House1["Party"] != "REP"] <- NA

House2 <- rbind(House, House1)

House3 <- House2[complete.cases(House2), ]

House_table <- table(House3['Party'])

House_final_data <- data.frame(Party = c("DEM", "REP"), Value = c(House_table['DEM'],
House_table['REP']))

p1 <- ggplot(House_final_data, aes(x="", y=Value, fill=Party)) +
        geom_bar(stat="identity", width=1, color="white") +
        coord_polar("y", start=0) +
        scale_fill_manual(values = c("#2e4057", "#edae49")) +
        ggtitle("House Candidates by Party") +
        theme_void()

Senate1 <- Senate

Senate["Party"][Senate["Party"] != "DEM"] <- NA

Senate1["Party"][Senate1["Party"] != "REP"] <- NA

Senate2 <- rbind(Senate, Senate1)
```

```
Senate3 <- Senate2[complete.cases(Senate2), ]

Senate_table <- table(Senate3['Party'])

Senate_final_data <- data.frame(Party = c("DEM", "REP"), Value = c(Senate_table['DEM'],
Senate_table['REP']))

p2 <- ggplot(Senate_final_data, aes(x="", y=Value, fill=Party)) +
        geom_bar(stat="identity", width=1, color="white") +
        coord_polar("y", start=0) +
        scale_fill_manual(values = c("#2e4057", "#edae49")) +
        ggtitle("Senate Candidates by Party") +
        theme_void()

President1 <- President

President["Party"][President["Party"] != "DEM"] <- NA

President1["Party"][President1["Party"] != "REP"] <- NA

President2 <- rbind(President, President1)

President3 <- President2[complete.cases(President2), ]

President_table <- table(President3['Party'])

President_final_data <- data.frame(Party = c("DEM", "REP"), Value = c(President_table['DEM'],
President_table['REP']))
print(President_table)

p3 <- ggplot(President_final_data, aes(x="", y=Value, fill=Party)) +
        geom_bar(stat="identity", width=1, color="white") +
        coord_polar("y", start=0) +
        scale_fill_manual(values = c("#2e4057", "#edae49")) +
        ggtitle("Presidential Candidates by Party") +
        theme_void()

cowplot::plot_grid(p1, p2, p3, nrow = 1, align = 'h', labels = 'auto', label_fontface = "plain", hjust
= 1, vjust = 0)
```

Visualization 2:

```
library(ggplot2)

Data <- read.csv("Downloads/fec_2008-2022.csv")

All_States <- data.frame(State = Data$Cand_State, Party = Data$Cand_Party_Affiliation)
All_States1 <- All_States

All_States["Party"][All_States["Party"] != "DEM"] <- NA

All_States1["Party"][All_States1["Party"] != "REP"] <- NA

All_States <- All_States[complete.cases(All_States), ]

All_States1 <- All_States1[complete.cases(All_States), ]

All_States_Table1 <- table(All_States['State'])
All_States_Table2 <- table(All_States1['State'])

df <- data.frame(Party = c("DEM", "REP", "Total"),
                 State = c('CA','CA', 'CA',
                           'TX', 'TX', 'TX',
                           'FL', 'FL', 'FL',
                           'NY', 'NY', 'NY',
                           'PA', 'PA', 'PA',
                           'IL', 'IL', 'IL',
                           'OH', 'OH', 'OH',
                           'GA', 'GA', 'GA',
                           'NC', 'NC', 'NC',
                           'MI', 'MI', 'MI'),
                 Counts = c(All_States_Table1['CA'], All_States_Table2['CA'],
                            All_States_Table1['CA'] + All_States_Table2['CA'],
                            All_States_Table1['TX'], All_States_Table2['TX'],
                            All_States_Table1['TX'] + All_States_Table2['TX'],
                            All_States_Table1['FL'], All_States_Table2['FL'],
                            All_States_Table1['FL'] + All_States_Table2['FL'],
                            All_States_Table1['NY'], All_States_Table2['NY'],
                            All_States_Table1['NY'] + All_States_Table2['NY'],
                            All_States_Table1['PA'], All_States_Table2['PA'],
                            All_States_Table1['PA'] + All_States_Table2['PA'],
                            All_States_Table1['IL'], All_States_Table2['IL'],
                            All_States_Table1['IL'] + All_States_Table2['IL'],
                            All_States_Table1['OH'], All_States_Table2['OH'],
                            All_States_Table1['OH'] + All_States_Table2['OH'],
```

```
                    All_States_Table1['GA'], All_States_Table2['GA'],
                    All_States_Table1['GA'] + All_States_Table2['GA'],
                    All_States_Table1['NC'], All_States_Table2['NC'],
                    All_States_Table1['NC'] + All_States_Table2['NC'],
                    All_States_Table1['MI'], All_States_Table2['MI'],
                    All_States_Table1['MI'] + All_States_Table2['MI']))

ggplot(df, aes(x=reorder(State, -Counts), y=Counts, fill=Party)) +
        geom_bar(position="dodge", stat = "identity") +
        labs(x = "States", y = "Number of Candidates")
```

Visualization 3:

```
library(ggplot2)

Data <- read.csv("Downloads/fec_2008-2022.csv")

Dem_Cont <- data.frame(CandCommitteeContribution =
        Data$Other_Committee_Contribution[which(Data$Cand_Party_Affiliation == "DEM")],
        Party = Data$Cand_Party_Affiliation[which(Data$Cand_Party_Affiliation == "DEM")],
        TotalContribution = Data$Total_Contribution[which(Data$Cand_Party_Affiliation ==
        "DEM")])

Rep_Cont <- data.frame(CandCommitteeContribution =
        Data$Other_Committee_Contribution[which(Data$Cand_Party_Affiliation == "REP")],
        Party = Data$Cand_Party_Affiliation[which(Data$Cand_Party_Affiliation == "REP")],
        TotalContribution = Data$Total_Contribution[which(Data$Cand_Party_Affiliation ==
        "REP")])

All_Cont = rbind(Dem_Cont, Rep_Cont)

All_Cont["TotalContribution"][All_Cont["TotalContribution"] == 0.0] <- NA

All_Cont <- All_Cont[complete.cases(All_Cont), ]

All_Cont$OtherCommitteeContProportion =
All_Cont$CandCommitteeContribution/All_Cont$TotalContribution

ggplot(All_Cont, aes(x=OtherCommitteeContProportion, fill=Party, color=Party)) +
        geom_density(alpha=0.5) + labs(x = "Other Committee Contributions as a Percentage of
        Total Contributions", y = "Frequency") +
        scale_fill_manual(values=c("#69b3a2", "#404080")) +
        scale_color_manual(values=c("#69b3a2", "#404080"))
```

Visualization 4:

```
library(ggplot2)

Data <- read.csv("Downloads/fec_2008-2022.csv")

DEM <- data.frame(NetContribution =
Data$Net_Contribution[which(Data$Cand_Party_Affiliation == "DEM")],
            NetOperatingExpenditure =
Data$Net_Operating_Expenditure[which(Data$Cand_Party_Affiliation == "DEM")],
            Party = Data$Cand_Party_Affiliation[which(Data$Cand_Party_Affiliation == "DEM")])

REP <- data.frame(NetContribution =
Data$Net_Contribution[which(Data$Cand_Party_Affiliation == "REP")],
            NetOperatingExpenditure =
Data$Net_Operating_Expenditure[which(Data$Cand_Party_Affiliation == "REP")],
            Party = Data$Cand_Party_Affiliation[which(Data$Cand_Party_Affiliation == "REP")])

DEM[DEM == 0] = NA
REP[REP == 0] = NA

DEM$NetOperatingExpenditure = log(DEM$NetOperatingExpenditure)
REP$NetOperatingExpenditure = log(REP$NetOperatingExpenditure)

DEM <- DEM[complete.cases(DEM), ]
REP <- REP[complete.cases(REP), ]

myBreaks <- seq(from=0, to=log(max(Data$Net_Contribution, na.rm=TRUE)), by=1)

ggplot(DEM, aes(x=NetOperatingExpenditure)) +
        geom_histogram( aes(x = NetOperatingExpenditure, y = ..count..), fill="#404080",
                color="#404080", alpha=0.5, breaks=myBreaks) +
        geom_label( aes(x=20, y=750, label="Democrats"), color="#404080") +
        geom_histogram(data=REP, aes(x = NetOperatingExpenditure, y = -..count..),
                fill="#69b3a2", color="#69b3a2", alpha=0.5,
                breaks=myBreaks) +
        geom_label( aes(x=20, y=-750, label="Republicans"), color="#69b3a2") +
        labs(x = "log(NetOperatingExpenditure)")
                coord_flip()
```

Visualization 5:

```
library(ggplot2)

Data <- read.csv("Downloads/fec_2008-2022.csv")

Dem_Cont <- data.frame(IndividualContribution =
Data$Individual_Contribution[which(Data$Cand_Party_Affiliation == "DEM")],
                       Party = Data$Cand_Party_Affiliation[which(Data$Cand_Party_Affiliation ==
"DEM")],
                       TotalContribution = Data$Total_Contribution[which(Data$Cand_Party_Affiliation
== "DEM")])

Rep_Cont <- data.frame(IndividualContribution =
Data$Individual_Contribution[which(Data$Cand_Party_Affiliation == "REP")],
                       Party = Data$Cand_Party_Affiliation[which(Data$Cand_Party_Affiliation ==
"REP")],
                       TotalContribution = Data$Total_Contribution[which(Data$Cand_Party_Affiliation
== "REP")])

All_Cont = rbind(Dem_Cont, Rep_Cont)

All_Cont["TotalContribution"][All_Cont["TotalContribution"] == 0.0] <- NA

All_Cont <- All_Cont[complete.cases(All_Cont), ]

All_Cont$IndividualContProportion = All_Cont$IndividualContribution/All_Cont$TotalContribution

ggplot(All_Cont, aes(x=IndividualContProportion, fill=Party, color=Party)) +
        geom_density(alpha=0.5) + labs(x = "Individual Contributions as a Percentage of Total
        Contributions", y = "Frequency") +
        scale_fill_manual(values=c("#69b3a2", "#404080")) +
        scale_color_manual(values=c("#69b3a2", "#404080"))
```

Visualization 6:

```
library(ggplot2)

Data <- read.csv("Downloads/fec_2008-2022.csv")

df <- data.frame(BeginCash = Data$Cash_On_Hand_BOP, EndCash =
Data$Cash_On_Hand_COP)

df[df <= 0] = NA

df <- df[complete.cases(df), ]

df$BeginCash = log(df$BeginCash)

df$EndCash = log(df$EndCash)

ggplot(df, aes(x = BeginCash, y = EndCash)) +
        geom_point(pch = 21, color = "black", size = 2.5) +
        scale_x_continuous(name = "log(Beginning of Period Cash On Hand)") +
        scale_y_continuous(name = "log(End of Period Cash On Hand)") +
        geom_abline(slope = 1, intercept = 0) +
        theme_minimal()
```