Part 1: Correct Story

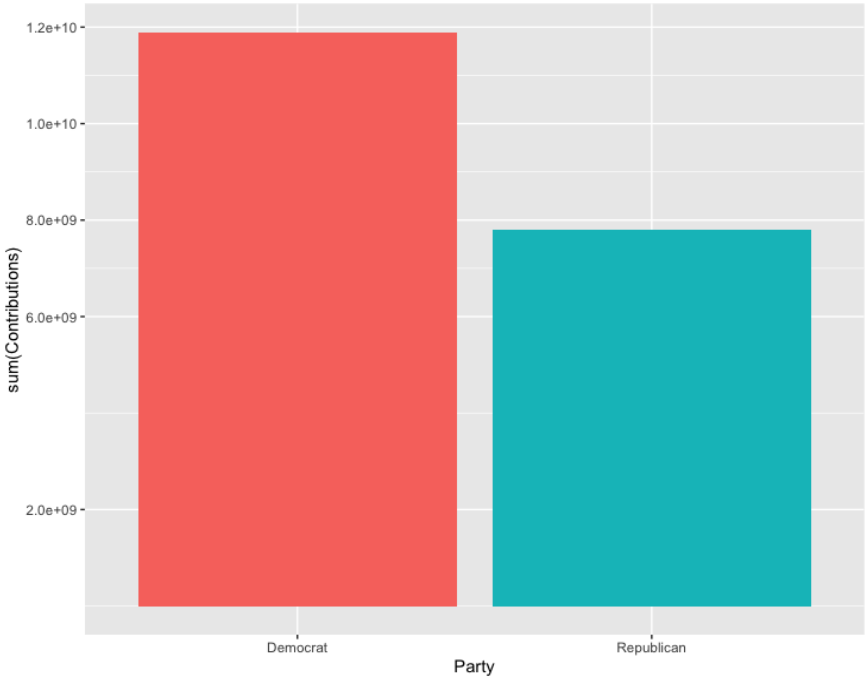
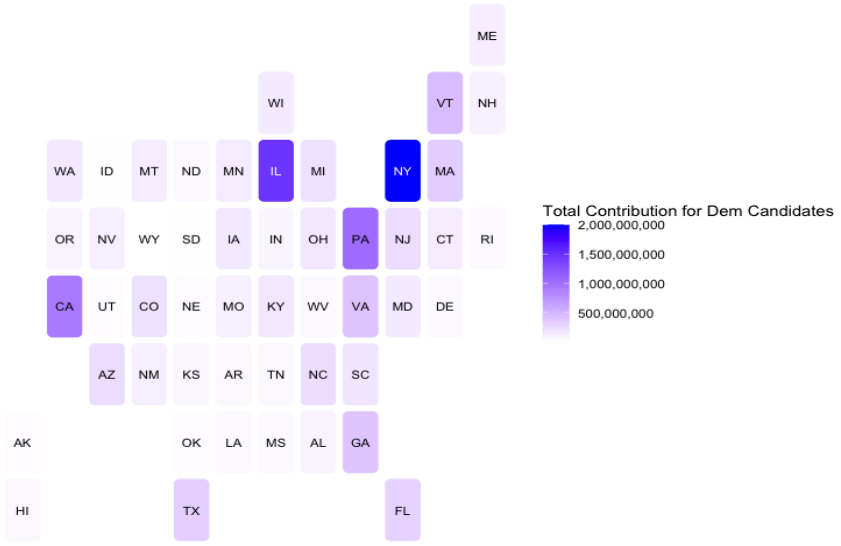


Figure 1: This bar graph shows the sum of all contributions to political candidates categorized by Party. As you can see, Democratic Candidates total contributions were almost $12 billion in total, while Republican Candidates total contributions were almost 8 billion in total. This shows that Democratic Candidates have more financial support than Republicans from a combination of Individual, Party, Committee, Candidate contributions.

Figure 2:
This Choropleth Heat Map shows the total contributions for Democratic Candidates across all 50 states. States with the highest amount of total contributions include New York, California, Illinois, and Pennsylvania. All of these states have high populations while New York, California, and Illinois are strong Democratic States. However, Pennsylvania is a swing state. States with the lowest amount of contributions include less populous states such as Idaho and Oklahoma or strong Republican States such as Tennessee and Mississippi.



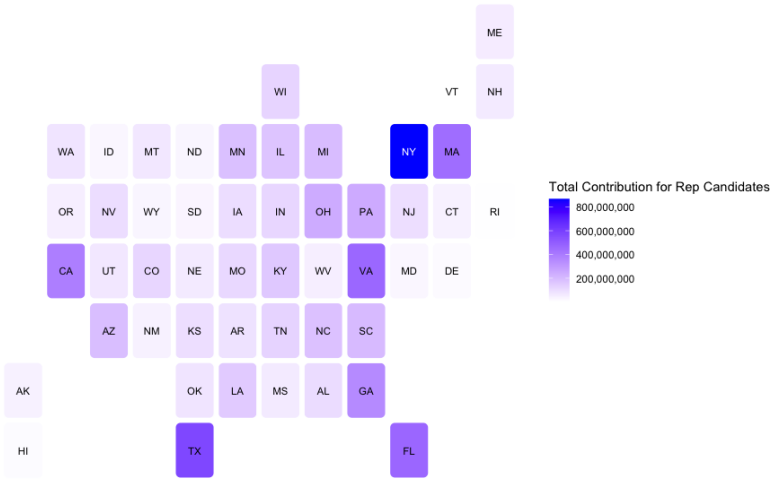


Figure 3:
This Choropleth Heat Map shows the total contributions for Republican Candidates across all 50 states. States with the highest amount of total contributions include New York, Texas, Florida, Massachusetts and Virginia. All of these states have high populations while Texas, Florida, and Virginia are strong republican states. However, New York and Massachusetts are strong republican states. States with the least total contributions are either strong Democratic states such as Oregon and Connecticut or low population states such as Wyoming and West Virginia. The general pattern shown for these two choropleths is that candidates in states that lean strongly to that candidate's party and/or have a high population have a higher chance to recieve high amounts of contributions, while states that lean away from the candidate's party or have a low population have a higher chance to recieve low amounts of contributions. As shown by the scales for total Contribution numbers, the highest level of contributions for Republican candidates in a state is only around 800 million while it is around 2 billion for Democratic Candidates. Although both parties have the highest amount of contributions in New York, there is a huge difference in the contribution totals (around $1 million), which is consistent with the conclusion from figure 1.

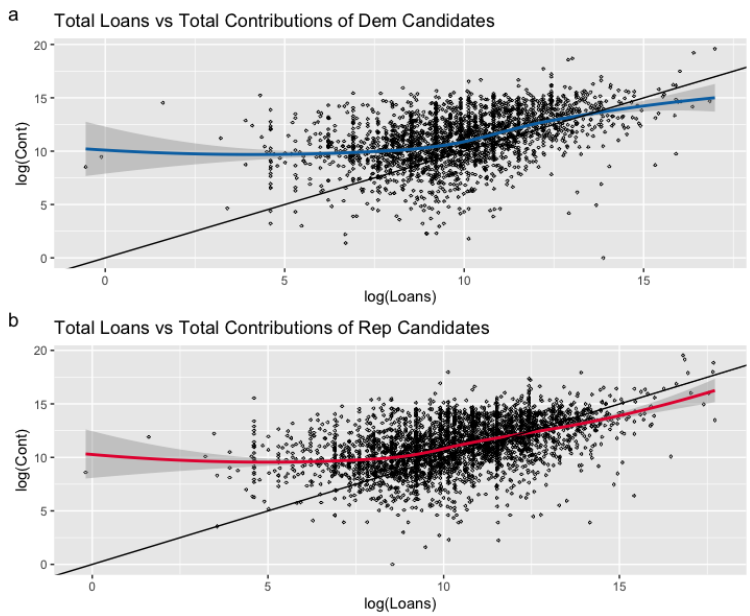


Figure 4:
These scatter plots show the distribution of log(Total Loans) compared to log(Total Contributions) separated by party. As you can see from both graphs, the majority of candidates in both parties have a greater amount of Contributions than loans, as most points lie above the y = x lines on the graphs. Both distributions are also quite similar as shown by their local regression curves, and both distributions have point clutter in the (5-15, 5-15) range on the graph.

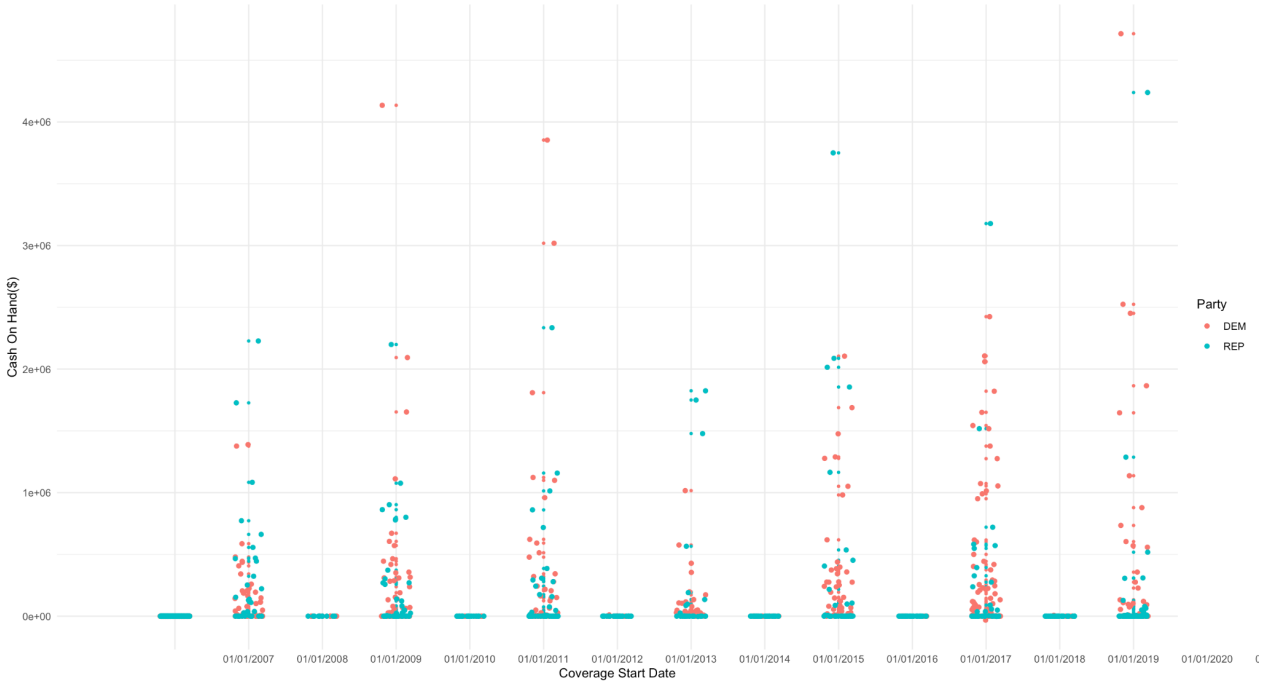


Figure 5:
This graph shows the amount of Cash on Hand candidates have based on when their Coverage date started and their information started to be tracked. There are two colors of points which represent the Democratic and Republican parties. As you can see from the graph, most candidates are clustered near the bottom, with a low relative contribution compared to a few candidates with a high amount of Cash on Hand. These outliers can be assumed to be serious candidates that have a lot of support. Finally, most people's coverage starts non-election years, showing that they start preparing for their campaigns and advertising early on.

Story:

1. Democratic candidates receive more Contributions in total than Republican candidates by a 2/3 factor.
2. Candidate contributions for Democratic and Republican candidates are high in states that the Party is strong in or in highly populous states.
3. Candidate contributions for Democratic and Republican candidates are low in states that the Party is weak in or in less populous states.
4. Most candidates have a greater amount of contributions compared to loans regardless of party.
5. The vast majority of candidate's loans and contributions lie within a similar range regardless of party.
6. Candidates mostly start preparing for campaigns on non election years
7. A small percentage of Candidates have a much greater number of candidates than the rest. These can be assumed to be serious candidates with a lot of support.

Part 2: Misleading Story, Part 3: Annotations

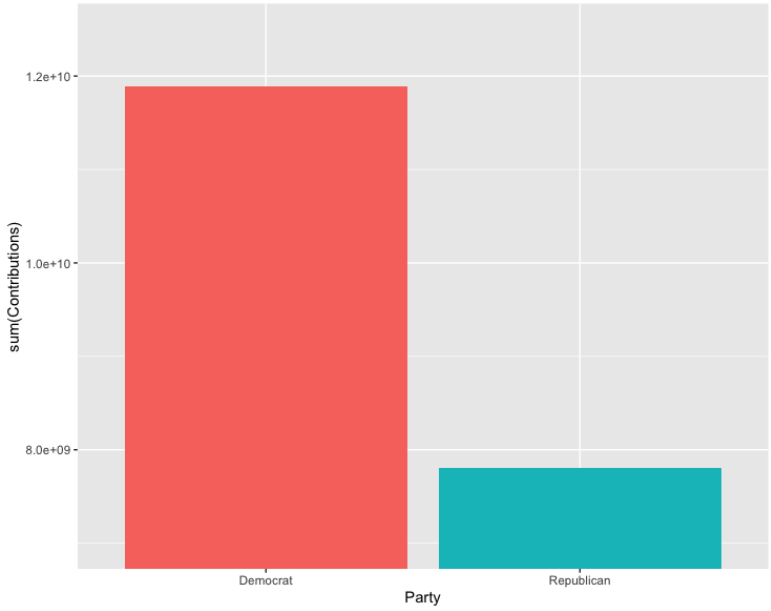


Figure 6:
This graph uses the same information as figure 1 above. It shows the sum of total contributions to candidates by party. However, instead of starting at 0, the y axis starts at $6 billion. This gives the illusion that all Democratic candidates have up to 4 or 5 times the total contributions of Republican candidates, but the reality is that total Republican contributions total about 2/3 of total Democratic contributions.

Figure 7:
This Chloropleth uses the same information as Figure 3 above. It shows the total contributions for Republican Candidates across the 50 states. The graph gives the impression of low values in New York and Texas due to them being lighter and our perceiving dark colors as having a higher value against a light background. However, lighter colors in this graph represent a higher value and darker colors represent a lower value, which could serve to confuse viewers. Additionally, the sizes of the states gives the viewer a distortion of visual importance among the states, as they are not equal sizes as in the heat map in Figure 3 above. Higher values are not given their visual importance when it comes to lighter colors and vice versa.

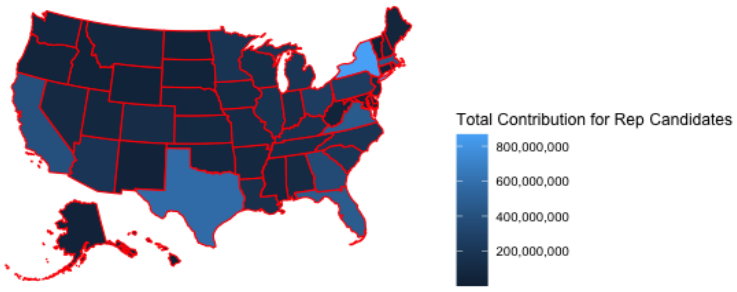
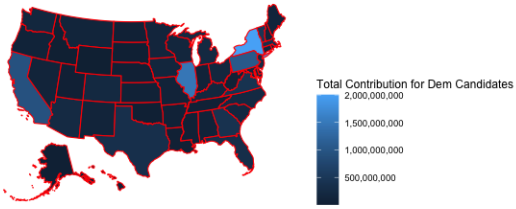


Figure 8:
Same explanation as figure 7, except with Democratic Candidates.



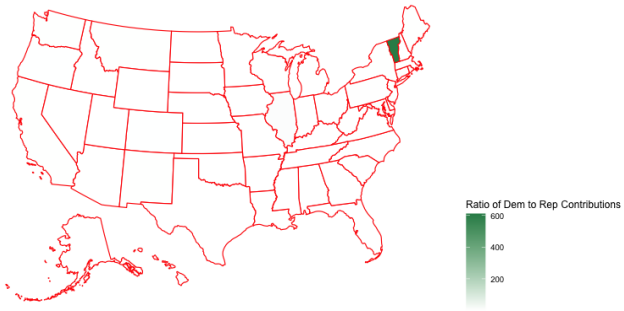
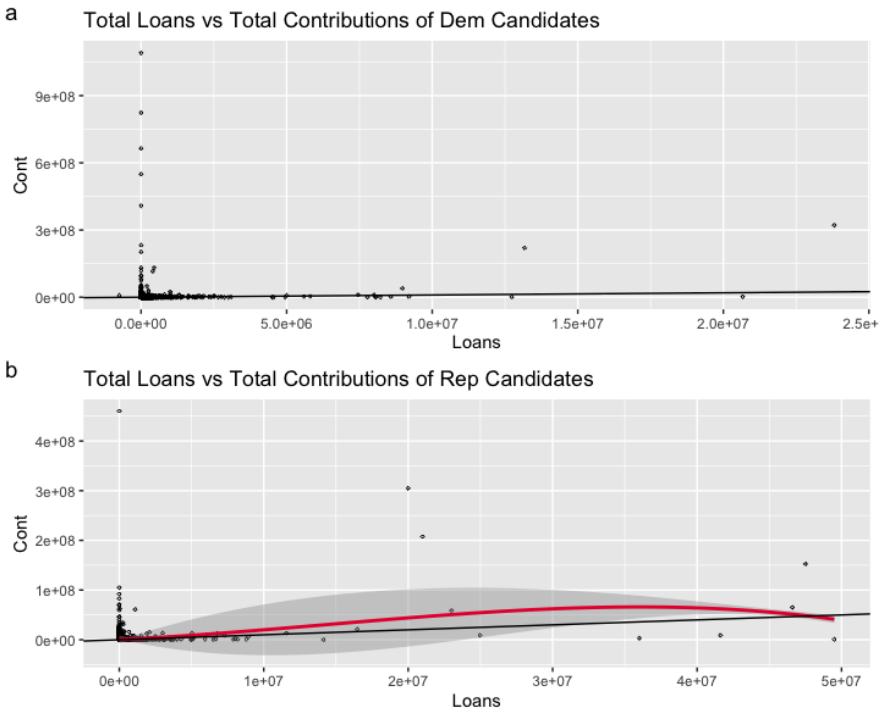


Figure 9:
This Chloropleth shows the Ratio of Democratic Contributions to Republican Candidates in all states. It gives the impression that the amount of contributions that are given to Democratic and Republican candidates across most states are similar, which is not the case. Because Vermont is an extreme outlier, with Democratic Contributions outnumbering Republican Contributions by around 800 times, it looks like there is no noticable difference in Contributions across all other states.

Figure 10:
These graphs use the same information at Figure 4. However these graphs do not log transform the Total Loans and Total Contributions. This leads to it being hard to read patterns in the graphs due to the high variance and outliers. Without log transforming the data, more importance is given to these outliers. Most points lie in the bottom left of the graph, but their features and relationships are indistinguishable. In fact, the local regression in the top graph cannot even be seen. Finally, due to the x axis being wider than the y axis, it seems like most candidates have a greater amount in total loans compared to total contributions when it's actually the other way around.



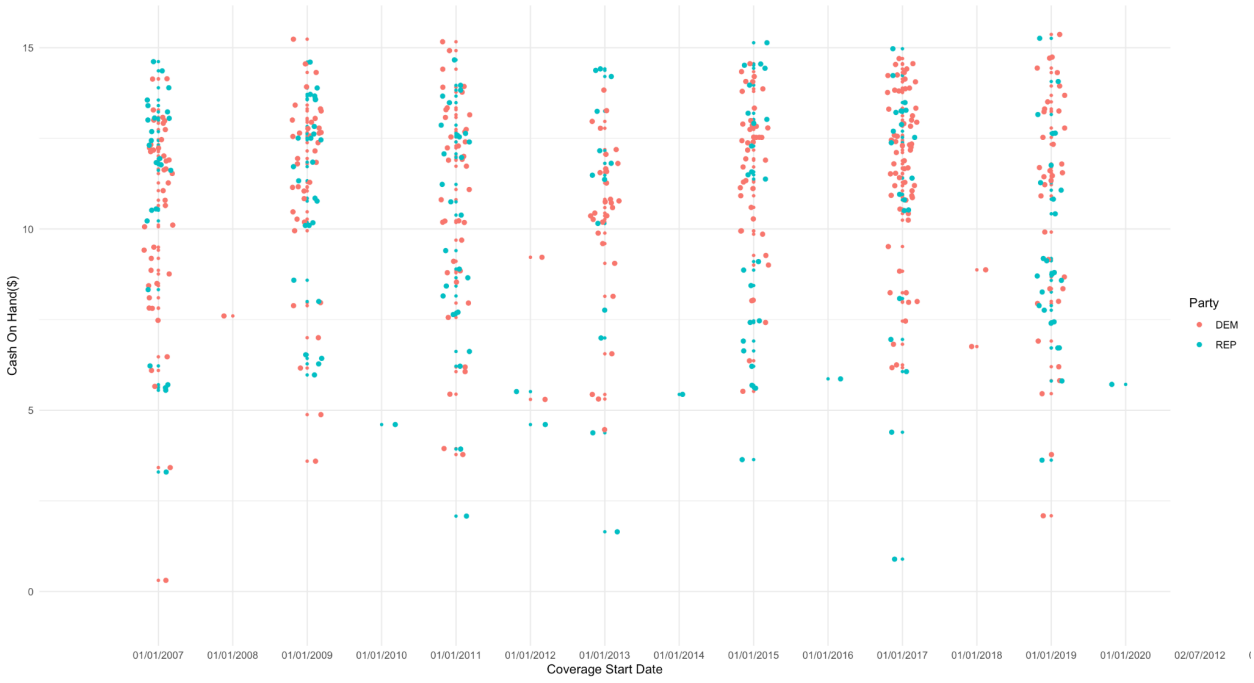


Figure 11:

This scatter plot uses the same information in Figure 5, except the y values (Cash on Hand) are log transformed. This leads to the impression that there are no outliers or candidates that have a much higher amount of contributions. However, in this case, it is useful to see the actual values to confirm the conclusion that most candidates have a relatively low amount of contributions compared to a few “serious candidates” that have a relatively high amount of contributions. What would be most important to see in this scatter plot would be the values and parties of the outliers in my opinion, as there is no clear relationship between Democratic and Republican candidates in this comparison.

Story:

1. Democratic candidates receive much more in total contributions in total than Republican candidates by a factor of 5
2. States with lower total contributions for Democratic and Republican candidates seem like they have higher total contributions and vice versa.
3. Larger states seem like they have larger values due to their visual importance and smaller states seem like they have smaller values.
4. States have no noticeable difference in total contributions to Democratic and Republican except in Vermont, where Democratic candidates have about 600 times higher amounts in total contributions compared to Republican candidates.
5. The majority of candidates have a greater amount in total loans compared to their total contributions.
6. There is no large difference between all candidate’s contribution amounts.

Figure 1

```
library(ggplot2)

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

demContributions <- data.frame(Cont =
Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "DEM")])
repContributions <- data.frame(Cont =
Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "REP")])

print(sum(demContributions$Cont))
print(sum(repContributions$Cont))

df <- data.frame(Party = c("Democrat", "Republican"), Sum = c(sum(demContributions$Cont),
sum(repContributions$Cont)))

ggplot(df, aes(x = Party, y = Sum, fill = c("#143c80", "#ba2f25"))) +
  scale_y_continuous(breaks = c(2000000000, 40000000000, 6000000000,
                                8000000000, 10000000000, 12000000000)) +
  xlab("Party") + ylab("sum(Contributions)") +
  theme(legend.position="none") +
  geom_bar(stat = "identity")
```

Figure 2

```
"https://rdrr.io/cran/statebins/man/statebins.html"

library(ggplot2)
library(usmap)
library(usmapdata)

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

demContributions <- data.frame(state = Data$Cand_State[which(Data$Cand_Party_Affiliation
== "DEM")], Cont = Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "DEM")])

demContributions <- aggregate(demContributions$Cont, by=list(demContributions$state),
FUN=sum)

demContributions <- demContributions[-c(1, 5, 10, 14, 29, 44, 52, 58),]

demContributions$state <- demContributions$Group.1
demContributions$Cont <- demContributions$x
```

```
demContributions <- subset(demContributions, select = -c(x, Group.1))

demContributions$state <- as.character(demContributions$state)

ggplot(demContributions, aes(state=state, fill=Cont)) +
  geom_statebins() +
  scale_fill_continuous(name = "Total Contribution for dem Candidates", low = "white", high =
"blue", label = scales::comma) +
  theme_void()
```

Figure 3

"https://rdrr.io/cran/statebins/man/statebins.html"

```
library(ggplot2)
library(usmap)
library(usmapdata)

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

repContributions <- data.frame(state = Data$Cand_State[which(Data$Cand_Party_Affiliation ==
"REP")], Cont = Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "REP")])

repContributions <- aggregate(repContributions$Cont, by=list(repContributions$state),
FUN=sum)

repContributions <- repContributions[-c(1, 4, 6, 11, 15, 30, 44, 46, 54, 60),]

repContributions$state <- repContributions$Group.1
repContributions$Cont <- repContributions$x

repContributions <- subset(repContributions, select = -c(x, Group.1))

'df <- data.frame(state = c("CA", "HI", "AK"), cont = c(5, 7, 9))'

'plot_usmap(data = repContributions, values = "Cont", color = "red") +
  scale_fill_continuous(name = "Total Contribution for Rep Candidates",
                        low = "white", high = "blue", label = scales::comma) +
  theme(legend.position = "right")'
repContributions$state <- as.character(repContributions$state)

ggplot(repContributions, aes(state=state, fill=Cont)) +
  geom_statebins() +
```

```
  scale_fill_continuous(name = "Total Contribution for Rep Candidates", low = "white", high =
"blue", label = scales::comma) +
  theme_void()
```

Figure 4

```
library(ggplot2)

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

demContsLoans <- data.frame(Cont =
Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "DEM")],
                            Loans = Data$Total_Loan[which(Data$Cand_Party_Affiliation == "DEM")])

demContsLoans["Cont"][demContsLoans["Cont"] <= 0] <- NA

demContsLoans["Loans"][demContsLoans["Loans"] <= 0] <- NA

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

demContsLoans$Cont <- log(demContsLoans$Cont)

demContsLoans$Loans <- log(demContsLoans$Loans)

g1 <- ggplot(demContsLoans, aes(x = Loans, y = Cont)) +
  geom_point(size=0.5, shape=23) +
  geom_smooth(method = "loess", formula = y ~ x,  color = "#0072B2") +
  ggtitle("Total Loans vs Total Contributions of Dem Candidates") +
  xlab("log(Loans)") +
  ylab("log(Cont)") +
  geom_abline(slope = 1, intercept = 0)


repContsLoans <- data.frame(Cont =
Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "REP")],
                            Loans = Data$Total_Loan[which(Data$Cand_Party_Affiliation == "REP")])

repContsLoans["Cont"][repContsLoans["Cont"] <= 0] <- NA

repContsLoans["Loans"][repContsLoans["Loans"] <= 0] <- NA

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

repContsLoans$Cont <- log(repContsLoans$Cont)
```

```
repContsLoans$Loans <- log(repContsLoans$Loans)

g2 <- ggplot(repContsLoans, aes(x = Loans, y = Cont)) +
  geom_point(size=0.5, shape=23) +
  geom_smooth(method = "loess", formula = y ~ x,  color = "#e61c3d") +
  ggtitle("Total Loans vs Total Contributions of Rep Candidates") +
  xlab("log(Loans)") +
  ylab("log(Cont)") +
  geom_abline(slope = 1, intercept = 0)


cowplot::plot_grid(g1, g2, ncol = 1, align = 'h', labels = 'auto',
                   label_fontface = "plain", hjust = 0, vjust = 1)
```

Figure 5

```
"https://stackoverflow.com/questions/48197502/r-how-to-delete-rows-with-certain-dates"
library(ggplot2)

Data <- read.csv("Downloads/fec_2008-2022.csv")
Data$date <- as.Date(Data$Coverage_Start_Date, format = "%m/%d/%Y")

df <- data.frame(BeginCash = Data$Cash_On_Hand_BOP[which(Data$Cand_State == "CA")],
                 BeginDate = Data$Coverage_Start_Date[which(Data$Cand_State == "CA")],
                 Party =  Data$Cand_Party_Affiliation[which(Data$Cand_State == "CA")])

df1 <- df

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

df1["Party"][df1["Party"] != "REP"] <- NA

df2 <- rbind(df, df1)

df2 <- na.omit(df2)

ggplot(df2, aes(x=BeginDate, y=BeginCash, color = Party)) +
  geom_point(size = 0.05, alpha = 0.05) +
  coord_cartesian(xlim=c(0, 14)) +
  xlab("Coverage Start Date") +
  ylab("Cash On Hand($)") +
  geom_jitter(width = 0.3) +
  theme_minimal()
```

Figure 6

```
library(ggplot2)

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

demContributions <- data.frame(Cont =
Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "DEM")])
repContributions <- data.frame(Cont =
Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "REP")])

print(sum(demContributions$Cont))
print(sum(repContributions$Cont))

df <- data.frame(Party = c("Democrat", "Republican"), Sum = c(sum(demContributions$Cont),
sum(repContributions$Cont)))

ggplot(df, aes(x = Party, y = Sum, fill = c("#143c80", "#ba2f25"))) +
  scale_y_continuous(breaks = c(8000000000, 10000000000, 12000000000)) +
  xlab("Party") + ylab("sum(Contributions)") +
  theme(legend.position="none") +
  coord_cartesian(ylim = c(7000000000, 12500000000)) +
  geom_bar(stat = "identity")
```

Figure 7

```
library(ggplot2)
library(usmap)
library(usmapdata)

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

repContributions <- data.frame(state = Data$Cand_State[which(Data$Cand_Party_Affiliation ==
"REP")], Cont = Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "REP")])

repContributions <- aggregate(repContributions$Cont, by=list(repContributions$state),
FUN=sum)

repContributions <- repContributions[-c(1, 4, 6, 11, 15, 30, 44, 46, 54, 60),]

print(repContributions)

repContributions$state <- repContributions$Group.1
repContributions$Cont <- repContributions$x
```

```
repContributions <- subset(repContributions, select = -c(x, Group.1))

'df <- data.frame(state = c("CA", "HI", "AK"), cont = c(5, 7, 9))'

plot_usmap(data = repContributions, values = "Cont", color = "red") +
  scale_fill_continuous(name = "Total Contribution for Rep Candidates", label = scales::comma)
+
  theme(legend.position = "right")
```

Figure 8

```
library(ggplot2)
library(usmap)
library(usmapdata)

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

demContributions <- data.frame(state = Data$Cand_State[which(Data$Cand_Party_Affiliation
== "DEM")], Cont = Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "DEM")])

demContributions <- aggregate(demContributions$Cont, by=list(demContributions$state),
FUN=sum)

demContributions <- demContributions[-c(1, 5, 10, 14, 29, 44, 52, 58),]

demContributions$state <- demContributions$Group.1
demContributions$Cont <- demContributions$x

demContributions <- subset(demContributions, select = -c(x, Group.1))

'df <- data.frame(state = c("CA", "HI", "AK"), cont = c(5, 7, 9))'

plot_usmap(data = demContributions, values = "Cont", color = "red") +
  scale_fill_continuous(name = "Total Contribution for Dem Candidates", label = scales::comma)
+
  theme(legend.position = "right")
```

Figure 9

```
library(ggplot2)
library(usmap)
library(usmapdata)

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

```
demContributions <- data.frame(state = Data$Cand_State[which(Data$Cand_Party_Affiliation
== "DEM")], Cont = Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "DEM")])

demContributions <- aggregate(demContributions$Cont, by=list(demContributions$state),
FUN=sum)

demContributions <- demContributions[-c(1, 5, 10, 14, 29, 44, 52, 58),]

demContributions$state <- demContributions$Group.1
demContributions$Cont <- demContributions$x

demContributions <- subset(demContributions, select = -c(x, Group.1))

repContributions <- data.frame(state = Data$Cand_State[which(Data$Cand_Party_Affiliation ==
"REP")], Cont = Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "REP")])

repContributions <- aggregate(repContributions$Cont, by=list(repContributions$state),
FUN=sum)

repContributions <- repContributions[-c(1, 4, 6, 11, 15, 30, 44, 46, 54, 60),]

repContributions$state <- repContributions$Group.1
repContributions$Cont <- repContributions$x

repContributions <- subset(repContributions, select = -c(x, Group.1))

totalContributions <- merge(demContributions, repContributions, by = "state")

totalContributions$Ratio <- totalContributions$Cont.x/totalContributions$Cont.y

totalContributions <- subset(totalContributions, select = -c(Cont.x, Cont.y))

plot_usmap(data = totalContributions, values = "Ratio", color = "red") +
  scale_fill_continuous(name = "Ratio of Dem to Rep Contributions", low = "white", high =
"seagreen", label = scales::comma) +
  theme(legend.position = "right")
```

Figure 10

```
library(ggplot2)

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

```
demContsLoans <- data.frame(Cont =
Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "DEM")],
                        Loans = Data$Total_Loan[which(Data$Cand_Party_Affiliation == "DEM")])

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

g1 <- ggplot(demContsLoans, aes(x = Loans, y = Cont)) +
  geom_point(size=0.5, shape=23) +
  geom_smooth(method = "loess", formula = y ~ x,  color = "#0072B2") +
  ggtitle("Total Loans vs Total Contributions of Dem Candidates") +
  xlab("log(Loans)") +
  ylab("log(Cont)") +
  geom_abline(slope = 1, intercept = 0)


repContsLoans <- data.frame(Cont =
Data$Total_Contribution[which(Data$Cand_Party_Affiliation == "REP")],
                        Loans = Data$Total_Loan[which(Data$Cand_Party_Affiliation == "REP")])

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

g2 <- ggplot(repContsLoans, aes(x = Loans, y = Cont)) +
  geom_point(size=0.5, shape=23) +
  geom_smooth(method = "loess", formula = y ~ x,  color = "#e61c3d") +
  ggtitle("Total Loans vs Total Contributions of Rep Candidates") +
  xlab("log(Loans)") +
  ylab("log(Cont)") +
  geom_abline(slope = 1, intercept = 0)


cowplot::plot_grid(g1, g2, ncol = 1, align = 'h', labels = 'auto',
                   label_fontface = "plain", hjust = 0, vjust = 1)
```

Figure 11

```
"https://stackoverflow.com/questions/48197502/r-how-to-delete-rows-with-certain-dates"
library(ggplot2)

Data <- read.csv("Downloads/fec_2008-2022.csv")
Data$date <- as.Date(Data$Coverage_Start_Date, format = "%m/%d/%Y")

df <- data.frame(BeginCash = Data$Cash_On_Hand_BOP[which(Data$Cand_State == "CA")],
                 BeginDate = Data$Coverage_Start_Date[which(Data$Cand_State == "CA")],
                 Party =  Data$Cand_Party_Affiliation[which(Data$Cand_State == "CA")])
```

```
df["BeginCash"][df["BeginCash"] <= 0] <- NA

df1 <- df

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

df$BeginCash = log(df$BeginCash)


df1["Party"][df1["Party"] != "REP"] <- NA

df1$BeginCash = log(df1$BeginCash)

df2 <- rbind(df, df1)

df2 <- na.omit(df2)


ggplot(df2, aes(x=BeginDate, y=BeginCash, color = Party)) +
  geom_point(size = 0.75) +
  coord_cartesian(xlim=c(0, 14)) +
  xlab("Coverage Start Date") +
  ylab("Cash On Hand($)") +
  geom_jitter(width = 0.2) +
  theme_minimal()
```