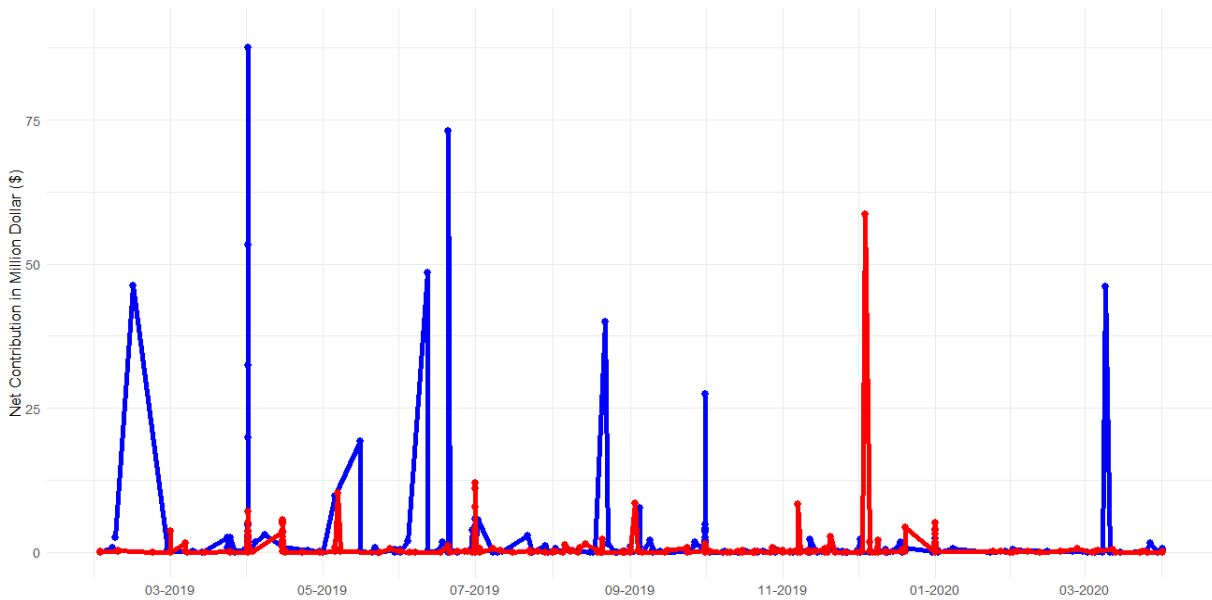


1. **HONEST STORY:** During the year 2019, right before the presidential election between Joe Biden vs. Donald Trump in 2020, there was a spike in the net contributions and total contributions among all Democrats and Republicans. After displaying the dataset for the year 2019, we conclude that the net contributions and total contributions did play a big role in the 2020 election results. Democrats leading Republicans in both the net contributions and total contributions resulted in the victory of president Joe Biden vs. the former president, Donald Trump.

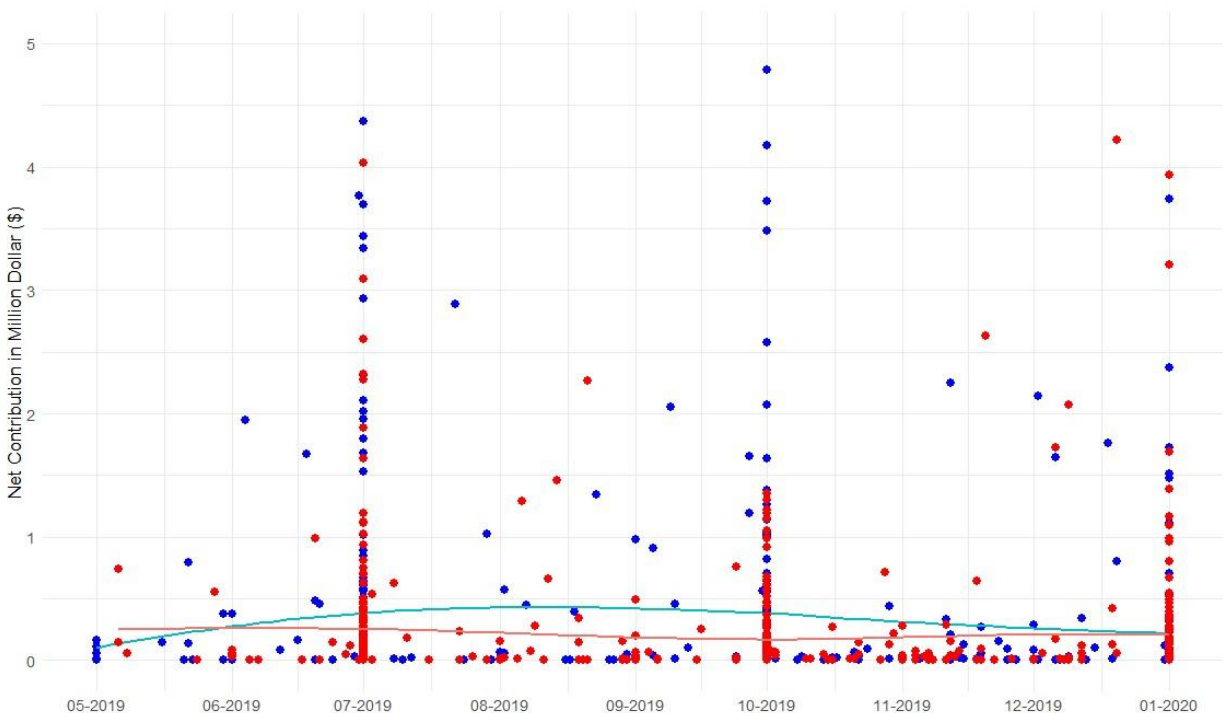
Time Series Plot



The time series plot displays the net contribution data that each and every Democrat and Republican candidate received during the time period from February 2019 to April 2020.

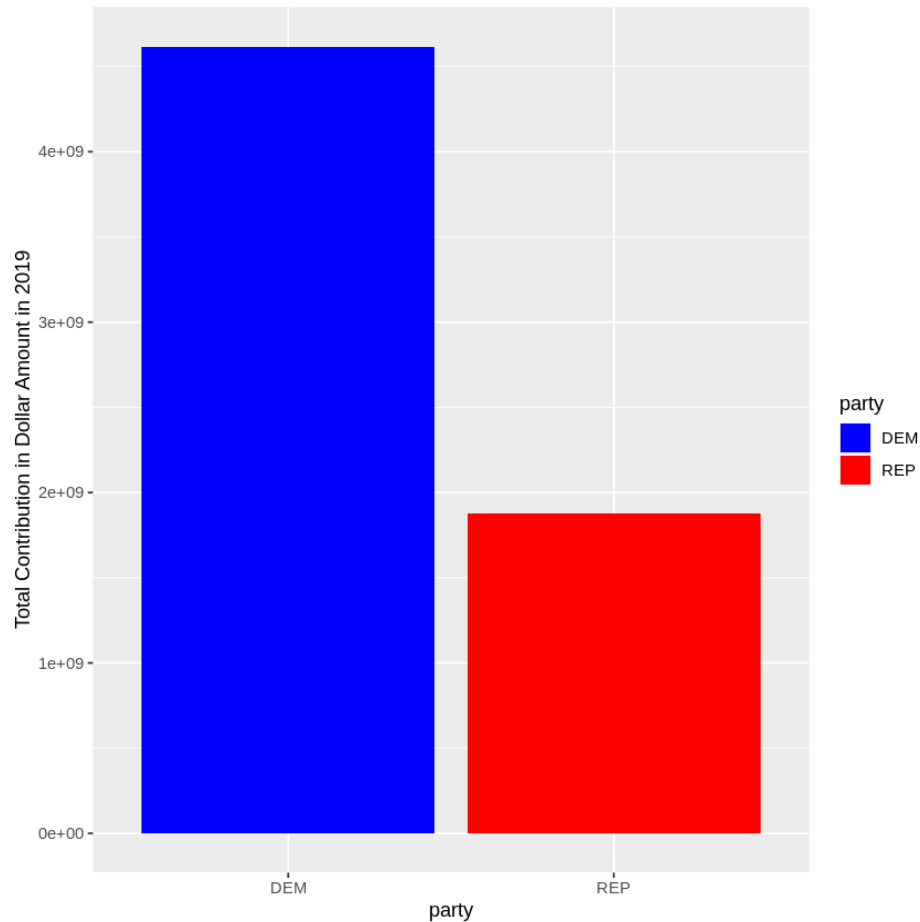
There are more Democratic candidates with larger amounts of net contributions.

Scatter Plot with Smoothing



The scatter plot with smoothing displays an overall trend of Democrats leading the Republican in net contributions across the period from May 2019 to January 2020.

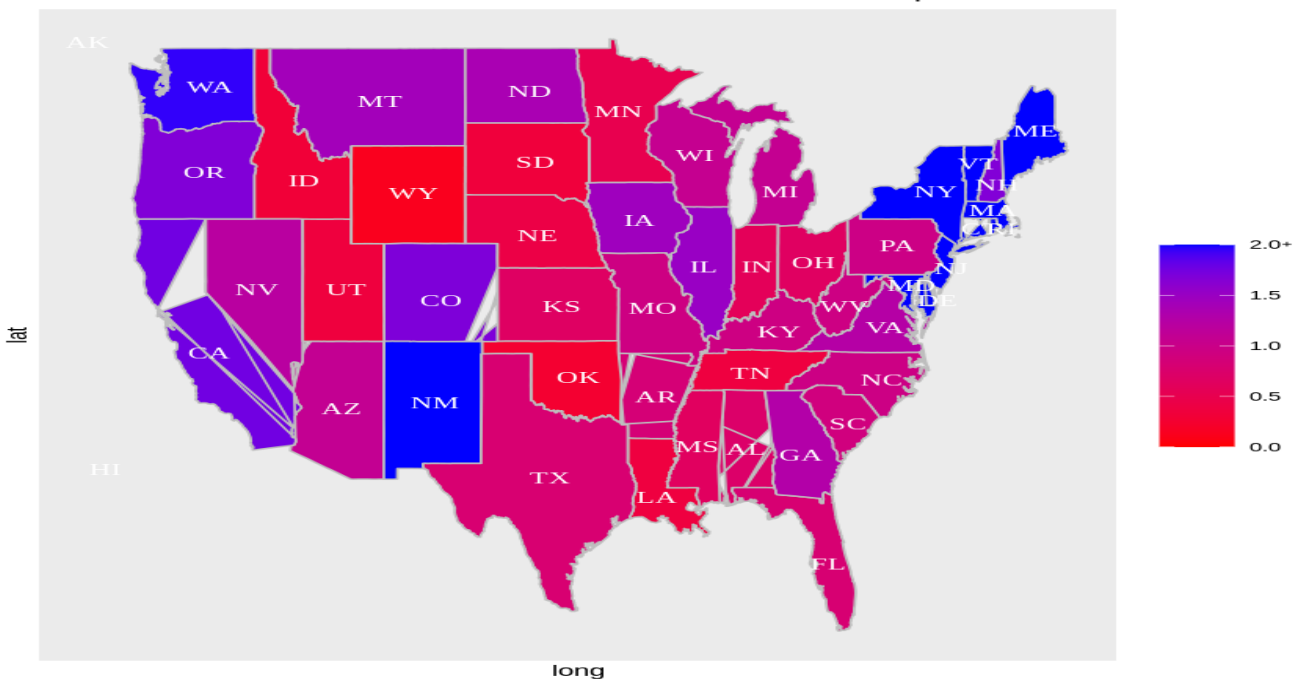
Bar chart



The bar chart shows the Democratic Party's and the Republican Party's total contributions among all of its presidential candidates during the whole year of 2019. Democrat's total contributions were more than double that of the Republican's in 2019.

Mapping Visualization (Choropleth)

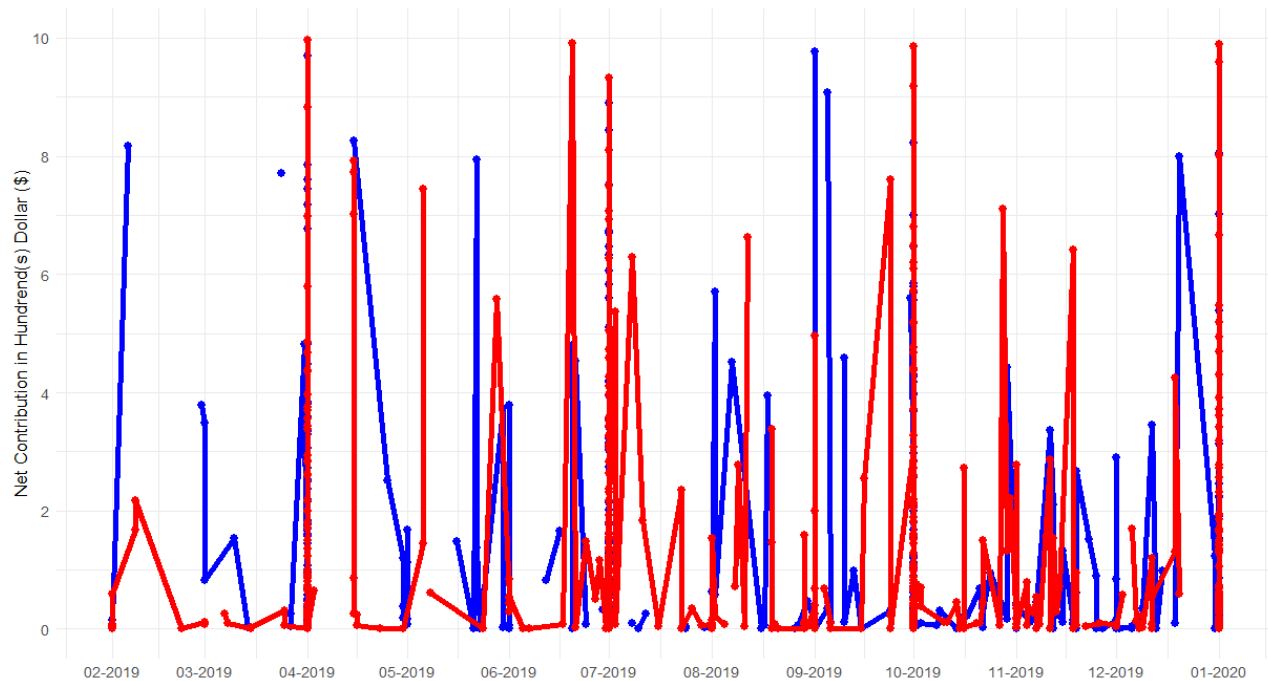
Ratio of total contributions of democrats candidates to republican candidates



This mapping visualization shows the ratio of contributions of democratic candidates to that of republican candidates for each state in 2019. It suggests that there are a couple of states with a ratio more than 2, which means democrat candidates had significant total contributions than republican candidates did. Plus, the overall color is mostly pink or purple.

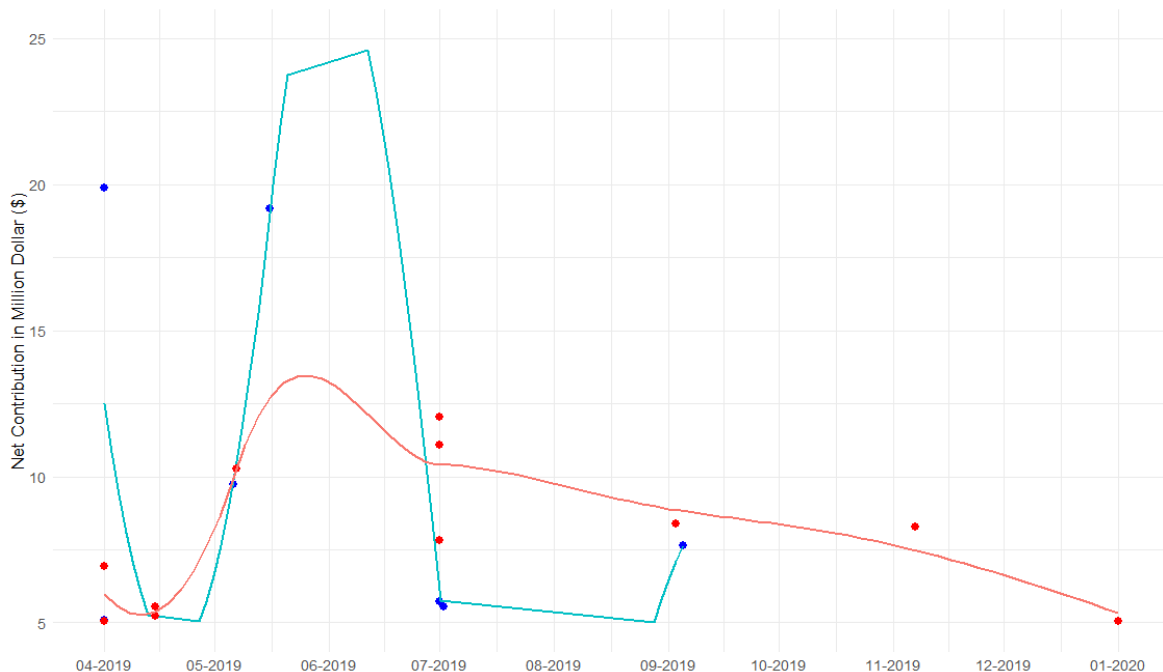
2. MISLEADING STORY: During the year 2019, right before the presidential election between Joe Biden vs. Donald Trump in 2020, there was a spike in the net contributions and total contributions among all Democrats and Republicans. After displaying the dataset for the year 2019, we conclude that the net contributions and total contributions in 2019 did not play a big role in the 2020 election results. Democrats and Republicans had almost the same amount of net contributions and total contributions. No correlations to Joe Biden winning.

Misleading Time Series



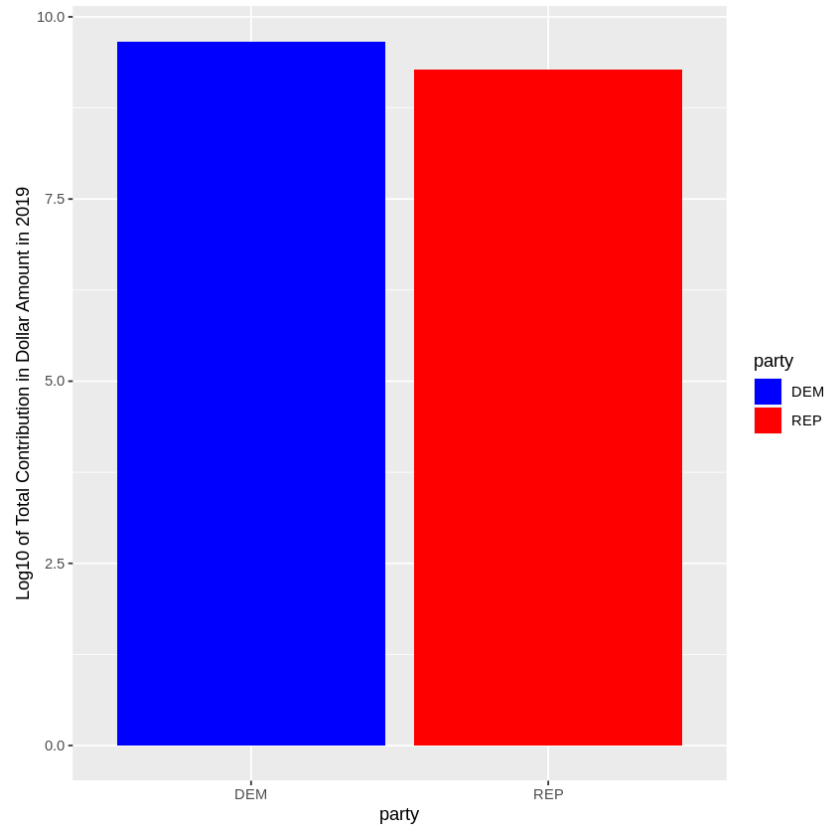
The time series plot displays an almost equal amount of net contributions across Democrat candidates and Republican candidates during the whole year of 2019.

Misleading Scatter Plot with Smoothing



The scatter plot with smoothing displays an overall trend of Democrat candidates leading in the early of 2019 and then Republican candidates leading near the end of 2019.

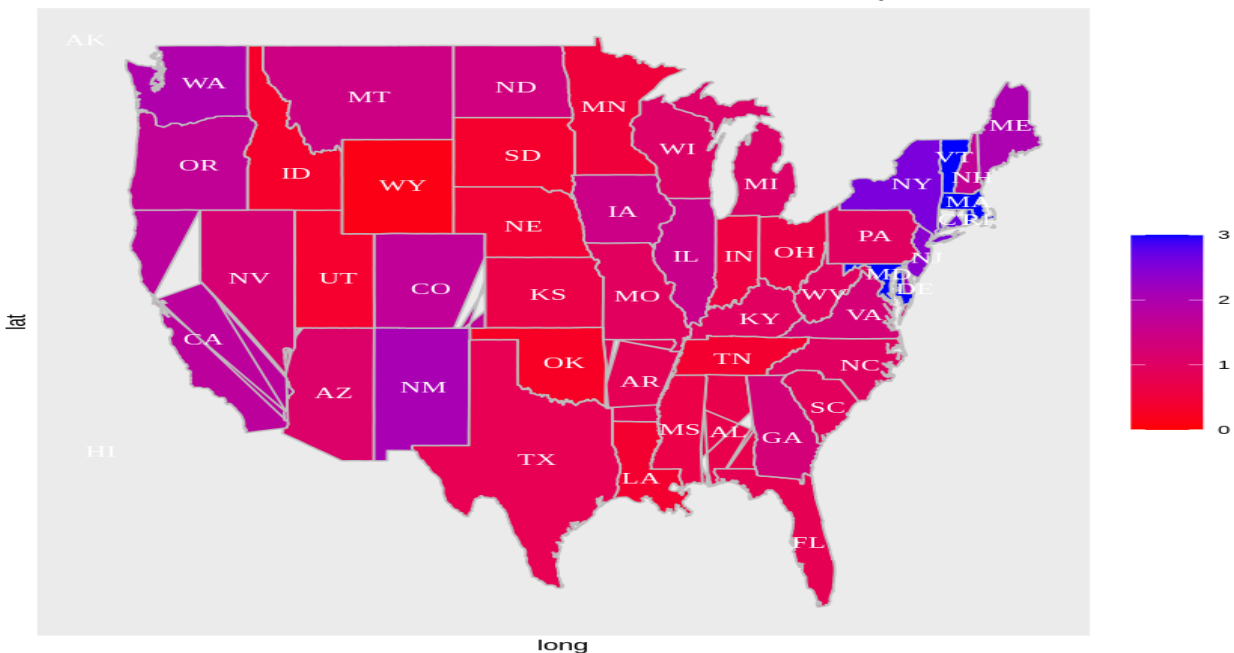
Misleading Bar chart



The bar chart displays the total amount of contributions of democratic candidates and republican candidates in 2019. There seems to be a slight difference in the total contributions among both Democrats and Republicans. Very insignificant difference.

Misleading Mapping Visualization (Choropleth)

Ratio of total contributions of democratics candidates to republican candidates



This mapping visualization (Choropleth) illustrates the ratio of contributions of democratic candidates to that of republican candidates for each state in 2019. The overall color of the map is reddish and purple.

3. Annotation

For the misleading time series plot, it is misleading by zooming in and discarding outliers data.

For the misleading scatter plot with smoothing, it is misleading by starting with the y axis at 5 instead of 0 and limiting the range from 5 to 25 to hide outliers data points.

For the misleading bar chart, we applied log10 transformation to the data, which makes the difference seem a lot less significant.

For the misleading mapping visualization (Choropleth), it is misleading because it did not apply the appropriate color scales. It made all states seem reddish than they actually are.

#code and comments

#PART 1: Honest Story

#Time Series Plot

```
library(tidyverse)
```

```
myData <- read.csv('C:/Users/Thanh Van/Documents/R/R/fec_2008-2022.csv')
```

```
myData$Coverage_Start_Date <- as.Date(myData$Coverage_Start_Date,  
format="%m/%d/%Y")
```

```
ggplot(subset(myData, Net_Contribution > 0, rm.na = TRUE), aes(Coverage_Start_Date,  
(Net_Contribution/1000000))) +  
  scale_y_continuous(limits = c(0,90)) +  
  geom_point(data = subset(myData, Cand_Party_Affiliation == "DEM"), color = "blue", size  
= 2.5) +  
  geom_line(data = subset(myData, Cand_Party_Affiliation == "DEM", rm.na = TRUE), size  
= 2, color = 'blue') +  
  geom_point(data = subset(myData, Cand_Party_Affiliation == "REP"), color = "red", size =  
2.5) +  
  geom_line(data = subset(myData, Cand_Party_Affiliation == "REP", rm.na = TRUE), size =  
2, color = 'red') +  
  scale_x_date(limits = c(as.Date("2019-02-01"),  
as.Date("2020-01-01")),  
date_breaks = "1 month", date_labels = "%m-%Y") +  
  xlab(NULL) + ylab("Net Contribution in Million Dollar ($)") +  
  theme_minimal() + theme(text = element_text(size=13))
```

#Scatter Plot with Smoothing

```
library(tidyverse)
```

```
myData <- read.csv('C:/Users/Thanh Van/Documents/R/R/fec_2008-2022.csv')
```

```

myData$Coverage_Start_Date <- as.Date(myData$Coverage_Start_Date,
format="%m/%d/%Y")
myData$Net_Contribution <- myData$Net_Contribution/1000000

ggplot(subset(myData, Net_Contribution > 0, rm.na = TRUE), aes(Coverage_Start_Date,
Net_Contribution)) +
  scale_y_continuous(limits = c(0,5),
                     breaks = seq(0, 5, by = 1)) +
  geom_point(data = subset(myData, Cand_Party_Affiliation == "DEM"), color = "blue", size =
2.5) +
  geom_smooth(data = subset(myData, Cand_Party_Affiliation == "DEM"), aes(color = "red"),
method="loess",
             size = 1, na.rm = TRUE, se = FALSE) +
  geom_point(data = subset(myData, Cand_Party_Affiliation == "REP"), color = "red", size =
2.5) +
  geom_smooth(data = subset(myData, Cand_Party_Affiliation == "REP"), aes(color = "blue"),
method="loess",
             size = 1, na.rm = TRUE, se = FALSE) +
  scale_x_date(limits = c(as.Date("2019-05-01"),
as.Date("2020-01-01")),
              date_breaks = "1 month", date_labels = "%m-%Y") +
  xlab(NULL) + ylab("Net Contribution in Million Dollar ($)") +
  theme_minimal() + theme(text = element_text(size=13))

```

#Mapping Visualization (Choropleth)

```

library(tidyverse)
data_set2 <- read.csv('C:/Users/Thanh Van/Documents/R/R/fec_2008-2022.csv')

dataDEM <- data_set2[data_set2$Cand_Party_Affiliation=='DEM',]
dataREP <- data_set2[data_set2$Cand_Party_Affiliation=='REP',]
dCount <- c()
rCount <- c()
stateslong <- c()
for(i in 1:length(states)){
  dCount <- c(dCount,
sum(dataDEM$Total_Contribution[which(dataDEM$Cand_Office_St==states[i])]))
  rCount <- c(rCount,
sum(dataREP$Total_Contribution[which(dataREP$Cand_Office_St==states[i])]))
  stateslong <- c(stateslong,tolower(state.name[grep(states[i], state.abb)]))
}
pCount<-c()
for(i in 1:length(states)){
  pCount <- c(pCount,dCount[i]/rCount[i])
}
mapDf = data.frame(region=stateslong, values=pCount,abb=states)
all_states<-map_data("state")
total<-merge(all_states,mapDf,y.by="region")
centroids <- data.frame(region=tolower(state.name), long=state.center$x, lat=state.center$y)
centroids$abb<-state.abb[match(centroids$region,tolower(state.name))]

```

```
p<-ggplot(total, aes(long, lat, group=group, fill=values)) +
  geom_polygon(color="grey")+labs(fill = "")+ ggtitle("Ratio of total contributions of
democratics candidates to republican candidates")+
  # hide ticks on x and y axis
  scale_y_continuous(breaks=c())+
  scale_x_continuous(breaks=c()) +
  with(centroids,
    annotate(geom="text", x = long, y=lat, label = abb,
      size = 4,color="white",family="Times")
  )
p2<-p+scale_fill_continuous(
  low = "red",high = "#0000FF",limits=c(0, 2),
  guide=guide_colorbar(barwidth = 2,barheight = 10))
```

#Bar Chart

```
library(tidyverse)
data_set2 <- read.csv('C:/Users/Thanh Van/Documents/R/R/fec_2008-2022.csv')

data2019 <- data_set2[data_set2$Coverage_Start_DateYear==2019,]
dem2019 <- data2019[data2019$Cand_Party_Affiliation=='DEM',]
rep2019 <- data2019[data2019$Cand_Party_Affiliation=='REP',]
dtotal <- sum(dem2019$Total_Contribution)
rTotal <- sum(rep2019$Total_Contribution)
dfgood <-data.frame(party=c('DEM','REP'), total=c(dtotal,rTotal))
ggplot(dfgood,aes(y=total,x=party,fill=party))+geom_bar(stat="identity")+labs(y="Total
Contribution in Dollar Amount in 2019")+scale_fill_manual(values=c("blue", "red"))
```

#PART 2: Misleading Story

#Time Series Plot (misleading)

```
library(tidyverse)
myData <- read.csv('C:/Users/Thanh Van/Documents/R/R/fec_2008-2022.csv')

myData$Coverage_Start_Date <- as.Date(myData$Coverage_Start_Date,
format="%m/%d/%Y")

ggplot(subset(myData, Net_Contribution > 0, rm.na = TRUE), aes(Coverage_Start_Date,
(Net_Contribution/100000))) +
  scale_y_continuous(limits = c(0,10),
    breaks = seq(0, 10, by = 2)) +
  geom_point(data = subset(myData, Cand_Party_Affiliation == "DEM"), color = "blue", size =
2.5) +
  geom_line(data = subset(myData, Cand_Party_Affiliation == "DEM", rm.na = TRUE), size =
2, color = 'blue') +
  geom_point(data = subset(myData, Cand_Party_Affiliation == "REP"), color = "red", size =
2.5) +
  geom_line(data = subset(myData, Cand_Party_Affiliation == "REP", rm.na = TRUE), size = 2,
color = 'red') +
  scale_x_date(limits = c(as.Date("2019-02-01"),
as.Date("2020-01-01")),
```

```

date_breaks = "1 month", date_labels = "%m-%Y") +
xlab(NULL) + ylab("Net Contribution in Hundrend(s) Dollar ($)") +
theme_minimal() + theme(text = element_text(size=13))

```

#Scatter Plot with Smoothing (misleading)

```

library(tidyverse)
myData <- read.csv('C:/Users/Thanh Van/Documents/R/R/fec_2008-2022.csv')

myData$Coverage_Start_Date <- as.Date(myData$Coverage_Start_Date,
format="%m/%d/%Y")
myData$Net_Contribution <- myData$Net_Contribution/1000000

ggplot(subset(myData, Net_Contribution > 5, Net_Contribution > 25, rm.na = TRUE),
aes(Coverage_Start_Date, Net_Contribution)) +
  scale_y_continuous(limits = c(5,25),
                     breaks = seq(5, 25, by = 5)) +
  geom_point(data = subset(myData, Cand_Party_Affiliation == "DEM"), color = "blue", size =
2.5) +
  geom_smooth(data = subset(myData, Cand_Party_Affiliation == "DEM"), aes(color = "red"),
method="loess",
             size = 1, na.rm = TRUE, se = FALSE) +
  geom_point(data = subset(myData, Cand_Party_Affiliation == "REP"), color = "red", size =
2.5) +
  geom_smooth(data = subset(myData, Cand_Party_Affiliation == "REP"), aes(color = "blue"),
method="loess",
             size = 1, na.rm = TRUE, se = FALSE) +
  scale_x_date(limits = c(as.Date("2019-04-01"),
as.Date("2020-01-01")),
              date_breaks = "1 month", date_labels = "%m-%Y") +
xlab(NULL) + ylab("Net Contribution in Million Dollar ($)") +
theme_minimal() + theme(text = element_text(size=13))

```

#Code for bar chart (misleading)

```

df100 <-data.frame(party=c('DEM','REP'), total=c(log(dttotal,10),log(rTotal,10)))
ggplot(df100,aes(y=total,x=party,fill=party))+geom_bar(stat="identity")+labs(y="Log10 of Total
Contribution in Dollar Amount in 2019")+scale_fill_manual(values=c("blue", "red"))

```

#Code for map (misleading)

```

p+scale_fill_continuous(
  low = "red",high = "#0000FF",limits=c(0,3),
  guide=guide_colorbar(barwidth = 2,barheight = 10))
)

```