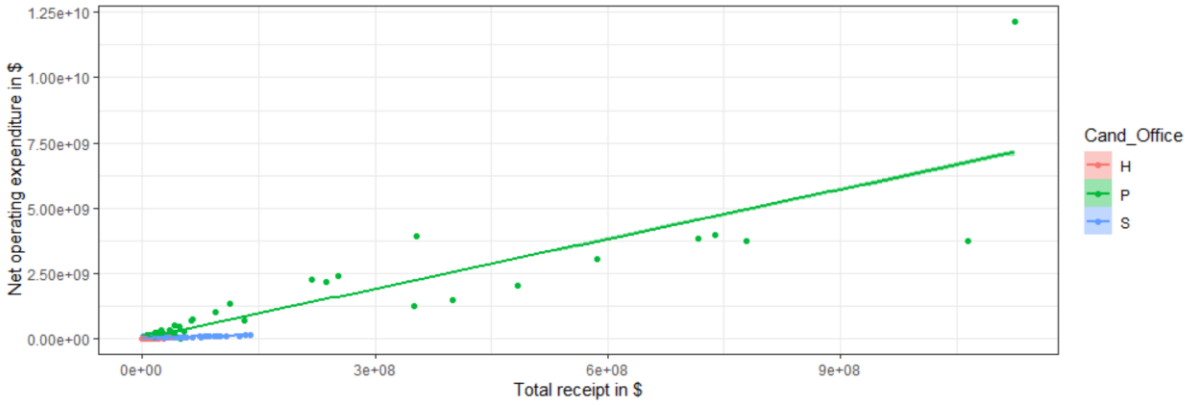
Assignment 3:

Part 1:

Figure 1.1.

Scatter plot of Total receipt VS net operating expenditure, with function fitting (95% CI envelope)



Total receipt and net operating expenditure have a linear positive correlation, overall president candidates have the highest amount of Total receipt, also highest net operating expenditure. The highest one overall is Obama (highest net operating expenditure and total receipt.)

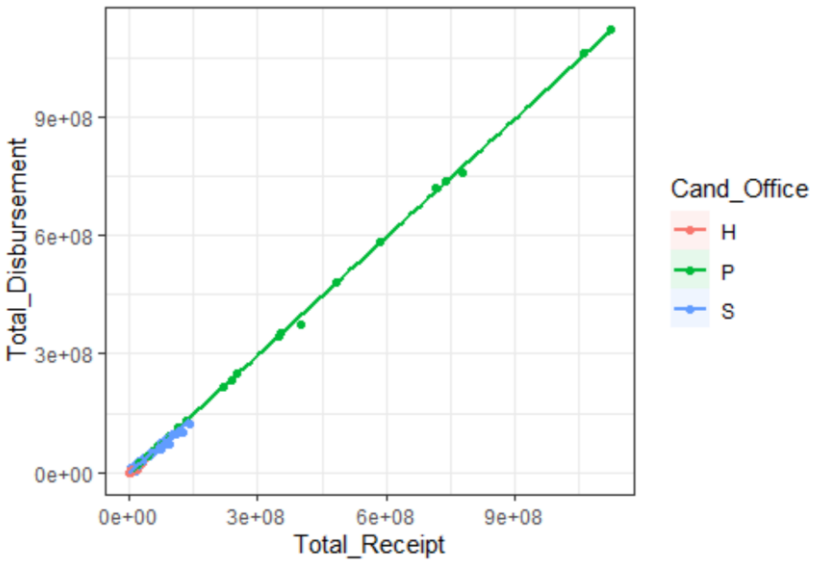
The slope is greater for president candidates, which means they spend more received money in their operating expenditure compared to senate and house representatives.

Here the ribbon around the fitted lines shows the 95% confidence interval. You need to zoom this picture to see the ribbon because the 95% CI is too small and can hardly be seen here (But it's actually exist!).

Figure 1.2. Time series of Net operating expenditure & Total receipt (for Democratic party and Republican party separately)

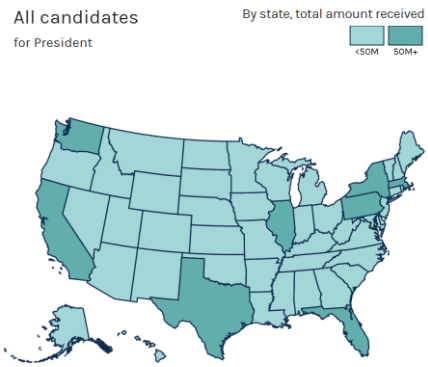
We can see that before 2008 REP candidates have received more and spend more money compared to DEM candidates. This situation changes at 2009 when the DEM raised more money than REP. Then the amount fluctuate a lot and they received and spend relatively same amount of funding from 2012 to 2016. The amount drops a lot in 2017. And recently we see an increase of the overall total receipt and net operating expenditure, and we can conclude that the political contribution is increase in both parties recently.

Figure 1.3 A scatter plot of total receipt vs total disbursement, subset by Cand_office



Here is a plot of total_receipt vs Total_Disbursement with a 95% confidence envelope as a ribbon around the fitted lines. We used fill here to make the ribbons the same colour as the lines. The fill is hardly seen here but it actually exists if you zoom in. Conclusions are similar as figure 1.1, Total receipt and total disbursement are positive linear correlated, and Presidents candidates have the highest value.

Figure 1.4. A Choropleth



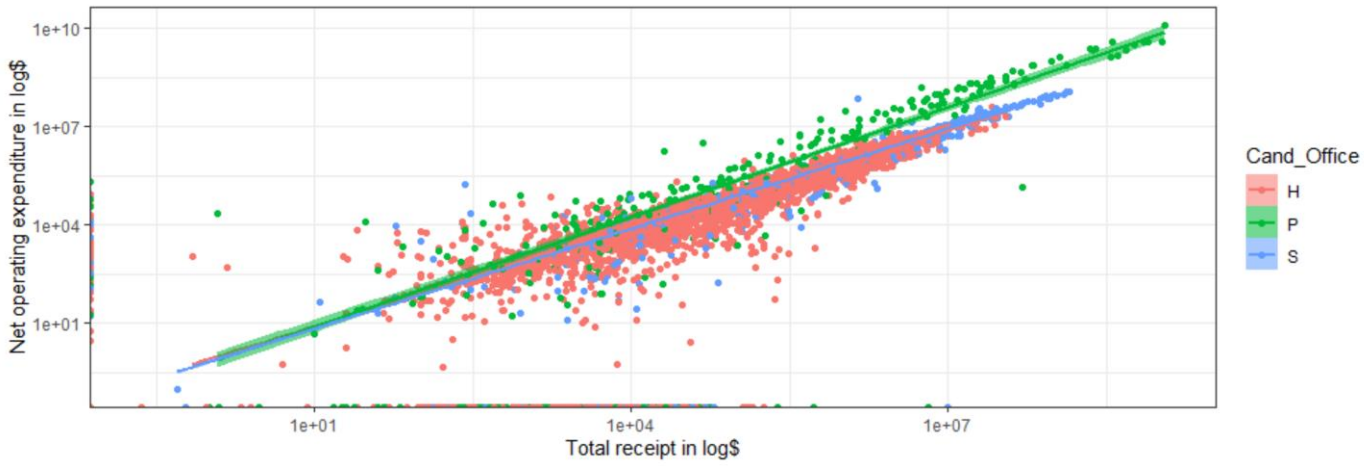
This figure shows the president candidates and their total amount received by state. As we can see there are only 8 states has a total amount received of 50M+. California, Washington, Texas, New York, Penn, Mass, Illinois has total received of 50M+.

Part 2&3:

Misleading visualization and information:

Figure 2.1

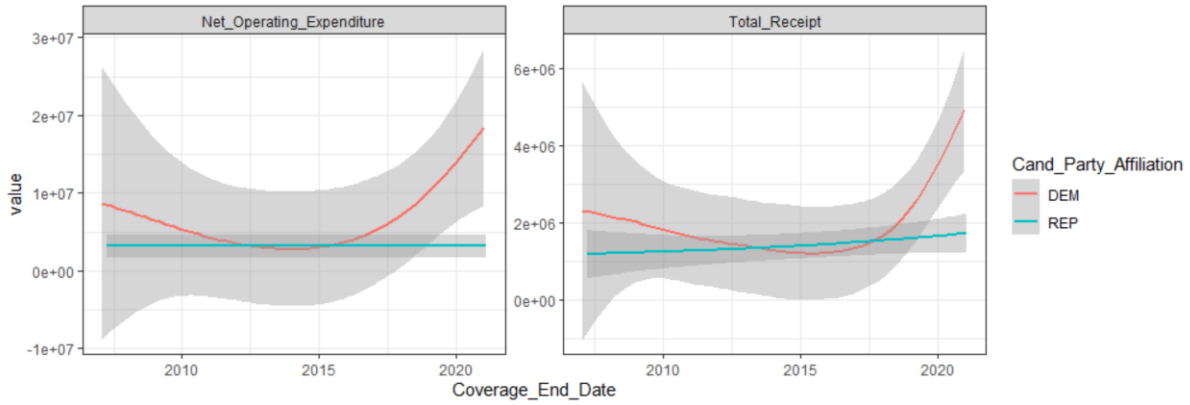
Scatter plot of Total receipt VS net operating expenditure, with function fitting (95% CI envelope)



Here we can see the net operating expenditure and total receipt have a linear positive correlation, senator and president candidates are more concentrated along the correlation line, and house candidates are more dispersive. The total receipt and net operating expenditures of senator & presidents candidates are **similar**.

Figure 2.2

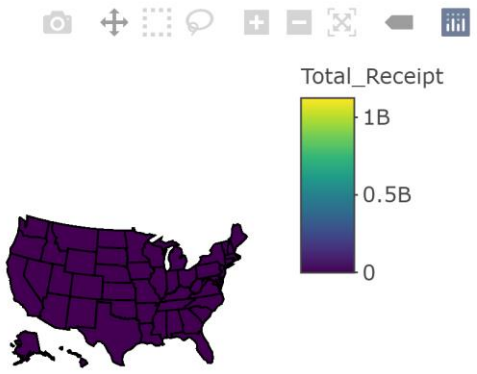
Net_operating_expenditure and total_receipt time series, distinct by party affiliation



The REP net operating expenditures are always same, and total receipt see a slight increase.

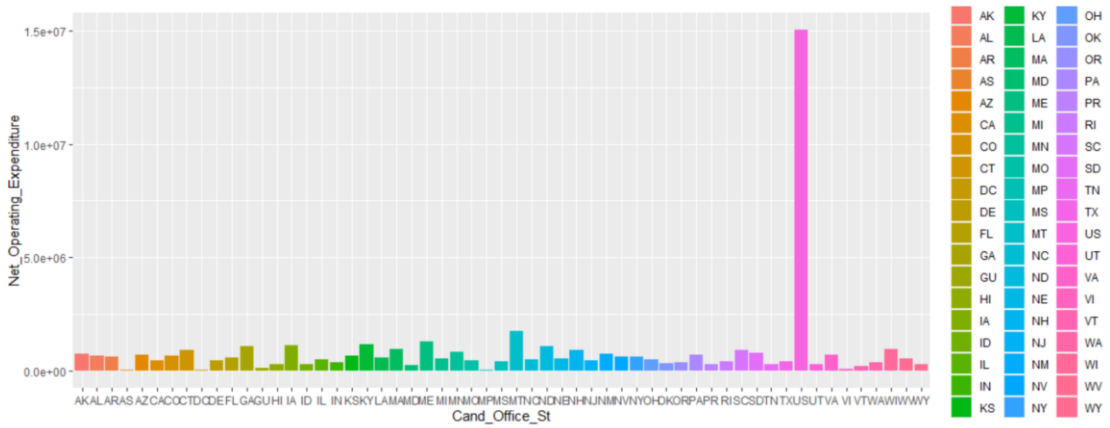
The DOM net operating expenditures and total receipt see a decrease till 2017 then a increase until now.

Figure 2.3 Total receipt grouped by states in 2020.



All the states have similar amount of total receipt political contribution, all under 0.2Billion US dollar.

Figure 2.4 Net operating expenditure grouped by states.



This figure shows the AS states DO, MP states has the lowest net operating expenditure, is almost 0. And US state has the highest net operating expenditure as 1.5e+07.

2.1 Transform the x-scale and y-scale into log10. This makes the difference among presidents, house and senator smaller, and the interpretation can be subtle and hard to notice. This makes the charts more readable too but distort the real difference among the candidates office.

2.2 Directly plot the real number of expenditure and total receipt makes the REP almost not changing. This is not true, because the time series of REP is relatively small compared to DEM but doesn't mean the REP didn't change. By converting to the log scale by 1.2, the difference and fluctuation is more easily to notice by the reader.

2.3 Total receipt grouped by states in 2020, this is misleading because the total receipt all under 0.5B and can't differentiate among each states.

2.4 is misleading because the order of magnitude different is very large. It's very hard to see the AS, DO, MP's expenditure number. But that doesn't mean it's near 0.

Appendix: R code

```
library(ggplot2)

library(dplyr)

library(TIMP)

library(nlme)
```

```
library(lubridate)
library(tidyr)
theme_set(theme_bw())

election <- read.csv("C:\\ information visualization-UCI\\fec_2008-2022_data.csv")
View(election)

election$tenure <- difftime(as.POSIXct(election$Coverage_End_Date,
format='%m-%d-%Y %H:%M:%S'),as.POSIXct(election$Coverage_Start_Date,
format='%m-%d-%Y %H:%M:%S'))

plot(Net_Operating_Expenditure~Total_Receipt, data=election)
# I tried fitmodel here, but there is an error
#f = fitModel(Net_operating_expenditure~A*Total_Receipt+B,data=election)
#try to use ggplot2 explore the coff of total_receipt and net_operating_expenditure.
# Here I tried multiple ggplot charts to make a more beautiful charts. Best were selected for the assignments.
ggplot(election, aes(x=Total_Receipt, y=Net_Operating_Expenditure,color=Cand_Office))+
  geom_point()+
  geom_smooth(method = "lm", aes(fill = Cand_Office))+
  labs(y= "Net operating expenditure in $", x = "Total receipt in $")

ggplot(election, aes(x=Total_Receipt, y=Total_Disbursement,color=Cand_Office))+
  geom_point()+
  geom_smooth(method = "lm",alpha = 0.1, aes(fill = Cand_Office))
#misleading one, changed the scale into log10.
ggplot(election, aes(x=Total_Receipt, y=Net_Operating_Expenditure,color=Cand_Office))+
  geom_point()+
  geom_smooth(method = "lm",alpha = .55, aes(fill = Cand_Office))+
  scale_x_log10()+
```

```
  scale_y_log10()+
  labs(y= "Net operating expenditure in log$", x = "Total receipt in log$")

# Secound & third figure, time series of the total receipt and net expenditure.
df <- election %>%
  dplyr::filter(Cand_Party_Affiliation %in% c("DEM", "REP"))
# distinct(Cand_Party_Affiliation)

output_ <-
  df %>%
  select(Coverage_End_Date,
         c('Total_Receipt', 'Net_Operating_Expenditure'),
         Cand_Party_Affiliation
  ) %>%
  mutate(Coverage_End_Date = lubridate::mdy(Coverage_End_Date))

output_ %>%
  gather(key, value, -Coverage_End_Date, -Cand_Party_Affiliation) %>%
  ggplot(
    aes(Coverage_End_Date, value, col = Cand_Party_Affiliation, group =
Cand_Party_Affiliation)
  ) +
  # geom_line()+
  geom_smooth() +
  scale_y_log10() +
  facet_wrap( ~ key, scales = "free_y")

output_ %>% #for the part 2, this changed the scale to the normal value, and makes the
changes for REP hard to notice.
  gather(key, value, -Coverage_End_Date, -Cand_Party_Affiliation) %>%
```

```
  ggplot(
    aes(Coverage_End_Date, value, col = Cand_Party_Affiliation, group =
Cand_Party_Affiliation)
  ) +
  # geom_line()+
  geom_smooth() +
  facet_wrap( ~ key, scales = "free_y")


# Next I want to plot a map visualization
# Download the shapefile. (note that I store it in a folder called DATA. You have to change that
if needed.)
download.file("http://thematicmapping.org/downloads/TM_WORLD_BORDERS_SIMPL-0.3.zip" ,
destfile="C:/code/world_shape_file.zip")


system("unzip C:/code/world_shape_file.zip")


library(rgdal)
library(ggplot2)
library(plotly)
library(readr)
states_contr = plot_geo(election,
                        locationmode = 'USA-states',
                        frame = ~Coverage_End_Date)%>%
  add_trace(locations = ~Cand_Office_St,
            z = ~Total_Receipt,
            zmin = 0,
            zmax = 1124592945,
            color = ~Total_Receipt)%>%
  layout(geo = list(scope = 'usa'))
states_contr
```