Kentrick Felix Kepawitono (ID: 56462267)

Truthful Story and Correct Visualization

Based on Figure 2, we can say that California and Texas are the two states that have the most participants in the election. No wonder that both California and Texas have the most participants since both states are the two most populous states in the US. As can be seen from the visualization, California has a darker color than Texas, which means that California has more election participants than Texas. Based on the legend, we can tell that California has more than 2000 participants, while texas has about 1000 to 1500 participants.

If we focus more on California (which have the most participants in the election), we can see that most of them (almost 3/4 of them) applied as a challenger (Figure 3). The second most applied position is open, then followed by incumbents. It is noticeable that very few people applied as incumbents, as compared to challenger and open. This is probably because more and more people are trying to work in the government sector.

Figure 4 is a visualization that shows the total receipts received and the total disbursement spent by a challenger in California. The smoothing is calculated using LOESS. It can be seen that there was stable money received and money spent from January 2020 to November 2020. However, after that, there was an increase in both money spent and received. This was probably because there was an increase in the total number of election participants after November 2020, as depicted in Figure 1.

Visualizations:

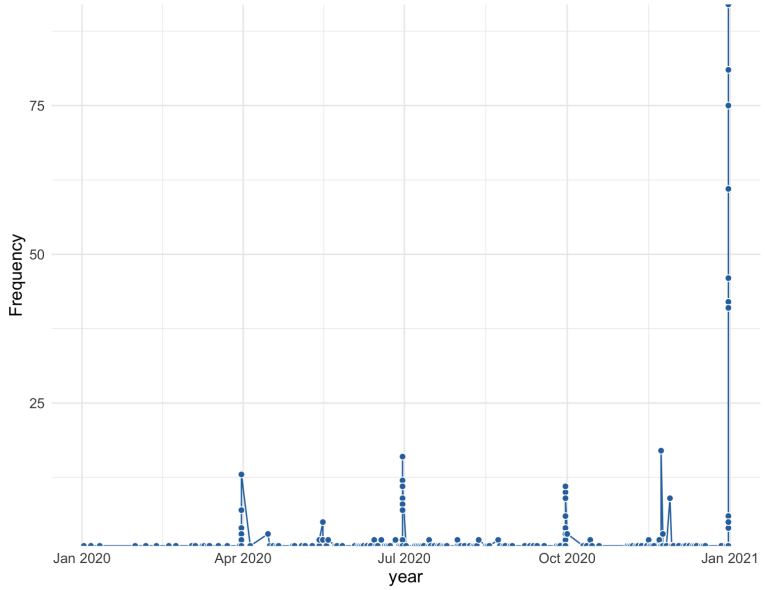

Figure 1

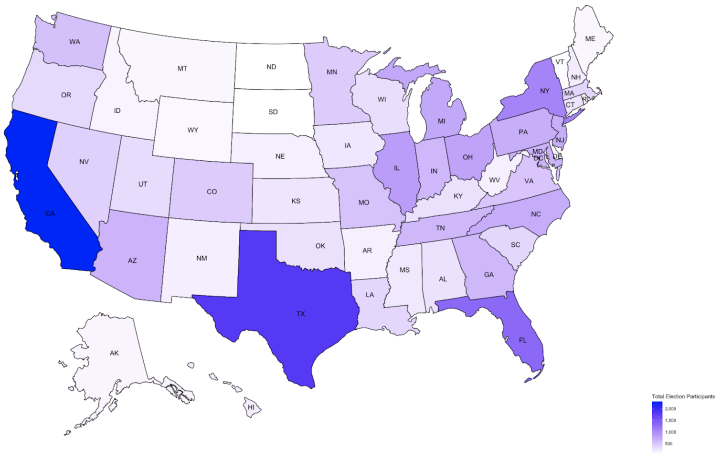

Figure 2

(P.S. I used usmap library and couldn't figure out how to make Alaska true size)

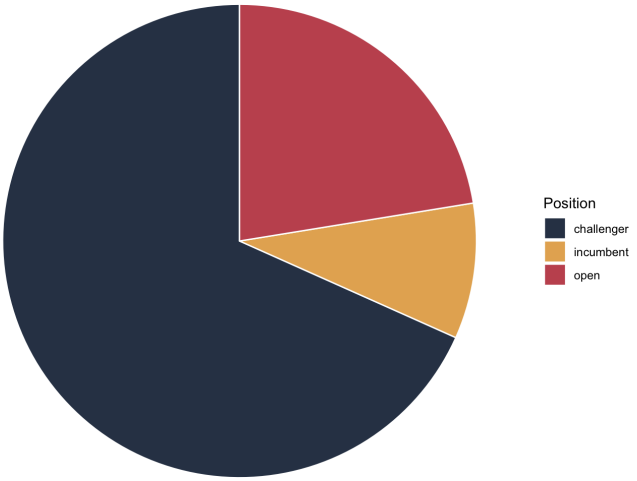

Figure 3

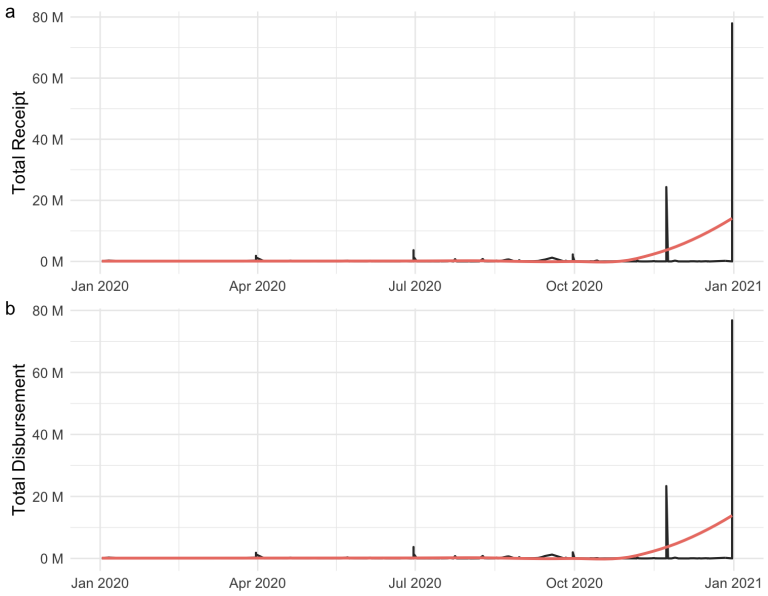
a
80 M
60 M
40 M
20 M
0 M
Total Receipt
Jan 2020
Apr 2020
Jul 2020
Oct 2020
Jan 2021
b
80 M
60 M
40 M
20 M
0 M
Total Disbursement
Jan 2020
Apr 2020
Jul 2020
Oct 2020
Jan 2021

Figure 4

Misleading Story and Misleading Visualizations

Based on Figure 6, we can say that California and Texas are the two states that have the least participants in the election. It is a little bit peculiar knowing that both California and Texas have the least participants since both states are the two most populous states in the US. As can be seen from the visualization, California has a lighter color than Texas, which means that California has fewer election participants than Texas. No legend is given so that we can assume that the darker colors are associated with higher intensities.

If we focus more on California, we can see that most of them (more than 1/2 of them) applied as a challenger (Figure 7). The second most applied position is open, followed by incumbents, which is half of the open participants. It is noticeable that very few people applied as incumbents, as compared to challenger and open. This is probably because more and more people are trying to work in the government sector.

Figure 8 is a visualization that shows the total receipts received and the total disbursement spent by a challenger in California. The smoothing is calculated using LOESS. It can be seen that there was stable money received from January 2020 to November 2020. However, after that, there was an increase in money received. On the other hand, the money spent was not that stable. There was a little bit of increase in money spent around October 2020 and December 2020. This was probably because there was an increase in the total number of election participants in October 2020 and November 2020, as depicted in Figure 5.

Visualizations:

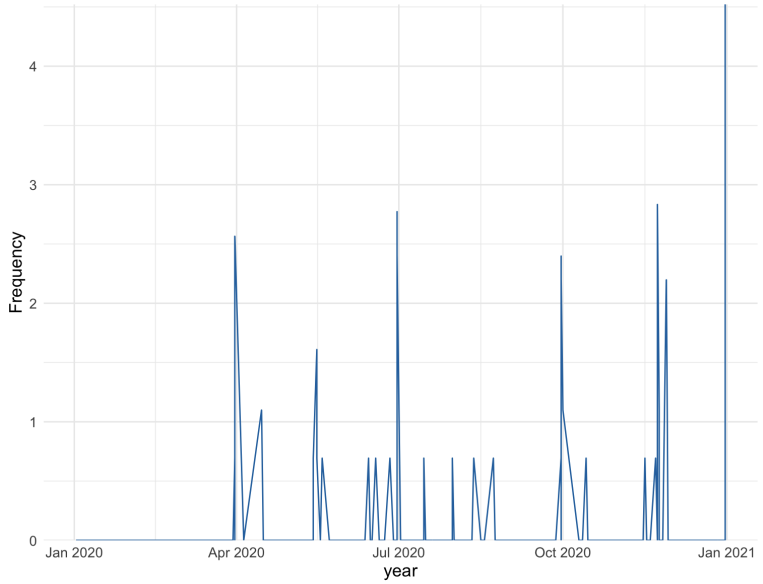

Figure 5

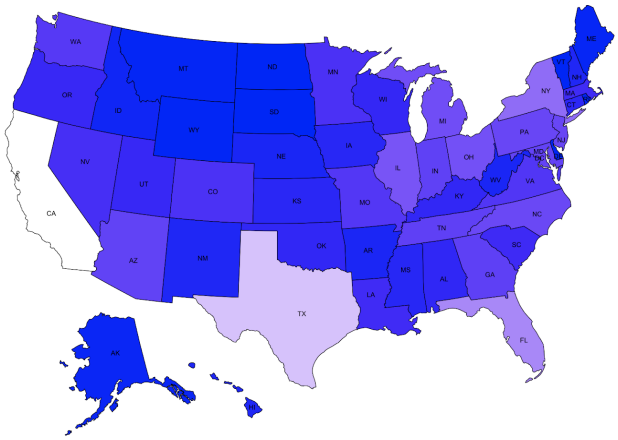

Figure 6

Figure 7

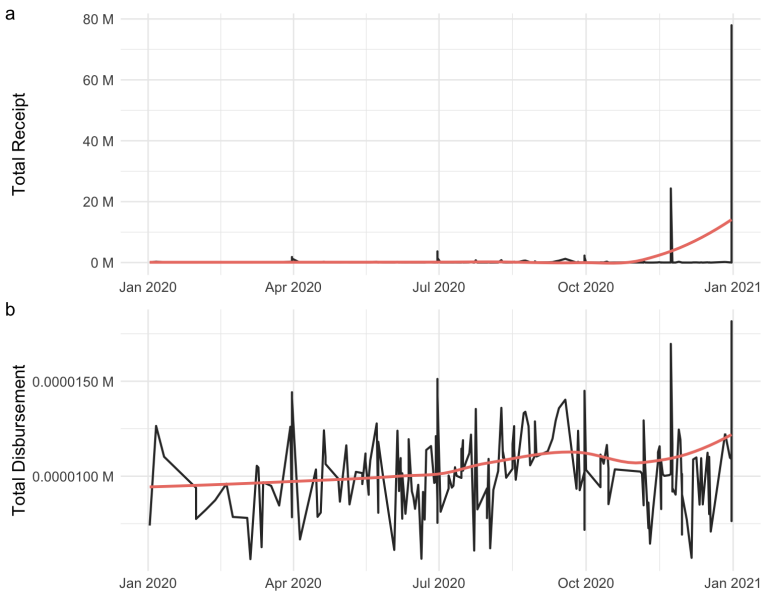

Figure 8

Comment on Misleading Story and Visualization

The fifth visualization uses logarithmic calculation for the frequency. This is definitely misleading for untrained eyes. The graph does not mention that the frequency is after the logarithm is applied. As a result, the viewer would think that the number of participants is as low as 4. Moreover, the points are also hidden. This is completely misleading because the amount of points are not large enough for it to have the points hidden.

The sixth visualization does not have a legend to indicate which color correspond to which number. Besides, the visualization also does not follow good practice. In fact, we tend to associate darker colors with higher intensities when the background of the figure is light. Without this good practice, untrained eyes would think that California and Texas are the states with the least number of participants. Moreover, Alaska is not in its true size, which also makes it misleading.

The seventh visualization is misleading because it is a stacked bar instead of a pie chart. This makes it hard for us to know the fraction of each party in the United States. With a pie chart, it will be easier for untrained eyes to emphasize simple fractions, such as 1/2, 1/3, and 1/4.

The eighth visualization compares total receipt and total disbursement; however, the total disbursement is applied with logarithmic calculation. This is misleading because the total receipt is not logged while the total disbursement is logged. This makes it incomparable. Besides, it also does not mention that the total disbursement is applied with logarithm. As a result, it seems like the total money spent is way lower than it should be.

Appendix (The Code Supporting the Visualizations)

Figure 1:

```
CA_challenger <- subset(fec, Cand_Incumbent_Challenger_Open_Seat ==
"CHALLENGER" &
                Cand_Office_St == "CA" &
                !Coverage_End_Date=="" &
                Total_Receipt > 0)

CA_challenger_each_month <- count(CA_challenger, Coverage_End_Date)


ggplot(CA_challenger_each_month, aes(as.Date(Coverage_End_Date, format =
"%m/%d/%y"), n)) +
  geom_line(size = 0.5, color = "#0072B2") +
  geom_point(color = "white", fill = "#0072B2", shape = 21, size = 2) +
  scale_y_continuous(expand = c(0, 0),
                     name = "Frequency") +
  scale_x_date(name = "year") +
  theme_minimal() +
  theme(text = element_text(size = 13))
```

Figure 2:

```
#Cite: https://youtu.be/Hi3fXGRBCMA
participants <- count(fec, Cand_Office_St)
participants <- subset(participants, (!Cand_Office_St == "US") &
                         (!Cand_Office_St == "DC") &
                         (!Cand_Office_St == "GU") &
                         (!Cand_Office_St == "VI") &
                         (!Cand_Office_St == "MP") &
                         (!Cand_Office_St == "AS"))
colnames(participants)[1] <- "state"

plot_usmap(data = participants,
           values = "n",
           labels = TRUE) +
  scale_fill_continuous(low = "white",
                        high = "blue",
                        label = scales::comma,
                        name = "Total Election Participants") +
  theme(legend.position = "right")
```

Figure 3:

```
CA_cha_inc_open <- subset(fec, (Cand_Incumbent_Challenger_Open_Seat ==
"CHALLENGER" |
                                Cand_Incumbent_Challenger_Open_Seat ==
"INCUMBENT" |
                                Cand_Incumbent_Challenger_Open_Seat ==
"OPEN") &
                           Cand_Office_St == "CA" &
                           Operating_Expenditure > 0 &
                           Total_Contribution > 0)

incumbent <-
data.frame(incumbent=unique(CA_cha_inc_open$Cand_Id[which(CA_cha_inc_open$C
and_Incumbent_Challenger_Open_Seat == "INCUMBENT")]))
challenger <-
data.frame(challenger=unique(CA_cha_inc_open$Cand_Id[which(CA_cha_inc_open$
Cand_Incumbent_Challenger_Open_Seat == "CHALLENGER")]))
open <-
data.frame(open=unique(CA_cha_inc_open$Cand_Id[which(CA_cha_inc_open$Cand_I
ncumbent_Challenger_Open_Seat == "OPEN")]))

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

ggplot(CA_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()
```

Figure 4:

```
CA_challenger_each_month <- count(CA_challenger, Coverage_End_Date)

CA_Total_Receipt <- data.frame(aggregate(CA_challenger$Total_Receipt,
by=list(Coverage_End_Date=CA_challenger$Coverage_End_Date), FUN=sum))
CA_Total_Disbursement <-
data.frame(aggregate(CA_challenger$Total_Disbursement,
by=list(Coverage_End_Date=CA_challenger$Coverage_End_Date), FUN=sum))
```

```
#Total Receipt
p1 <- ggplot(CA_Total_Receipt, aes(as.Date(Coverage_End_Date, format =
"%m/%d/%y"), x)) +
  geom_line(color = "grey20", size = 0.75) +
  geom_smooth(aes(color = "smooth"), method="loess", size = 1, na.rm =
TRUE, se = FALSE) +
  xlab(NULL) + ylab("Total Receipt") +
  theme_minimal() + theme(text = element_text(size=13), legend.position =
"none") +
  scale_y_continuous(labels = label_number(suffix = " M", scale = 1e-6))

#Total Disbursement
p2 <- ggplot(CA_Total_Disbursement, aes(as.Date(Coverage_End_Date, format =
"%m/%d/%y"), x)) +
  geom_line(color = "grey20", size = 0.75) +
  geom_smooth(aes(color = "smooth"), method="loess", size = 1, na.rm =
TRUE, se = FALSE) +
  xlab(NULL) + ylab("Total Disbursement") +
  theme_minimal() + theme(text = element_text(size=13), legend.position =
"none") +
  scale_y_continuous(labels = label_number(suffix = " M", scale = 1e-6))

cowplot::plot_grid(p1, p2, ncol = 1, align = "v",
                   labels = "auto", label_fontface = "plain", hjust = 0,
vjust = 1)
```

Figure 5:

```
#Cite:
https://stackoverflow.com/questions/3445590/subset-filter-rows-in-a-data-fr
ame-based-on-a-condition-in-a-column
CA_challenger <- subset(fec, Cand_Incumbent_Challenger_Open_Seat ==
"CHALLENGER" &
                          Cand_Office_St == "CA" &
                          !Coverage_End_Date=="" &
                          Total_Receipt > 0)

CA_challenger_each_month <- count(CA_challenger, Coverage_End_Date)


ggplot(CA_challenger_each_month, aes(as.Date(Coverage_End_Date, format =
"%m/%d/%y"), log(n))) +
  geom_line(size = 0.5, color = "#0072B2") +
```

```
  scale_y_continuous(expand = c(0, 0),
                     name = "Frequency") +
  scale_x_date(name = "year") +
  theme_minimal() +
  theme(text = element_text(size = 13))
```

Figure 6:

```
#Cite: https://youtu.be/Hi3fXGRBCMA
participants <- count(fec, Cand_Office_St)
participants <- subset(participants, (!Cand_Office_St == "US") &
                         (!Cand_Office_St == "DC") &
                         (!Cand_Office_St == "GU") &
                         (!Cand_Office_St == "VI") &
                         (!Cand_Office_St == "MP") &
                         (!Cand_Office_St == "AS"))
colnames(participants)[1] <- "state"

plot_usmap(data = participants,
           values = "n",
           labels = TRUE) +
  scale_fill_continuous(low = "blue",
                        high = "white",
                        label = scales::comma) +
  theme(legend.position = "")
```

Figure 7:

```
CA_cha_inc_open <- subset(fec, (Cand_Incumbent_Challenger_Open_Seat ==
"CHALLENGER" |
                                  Cand_Incumbent_Challenger_Open_Seat ==
"INCUMBENT" |
                                  Cand_Incumbent_Challenger_Open_Seat ==
"OPEN") &
                            Cand_Office_St == "CA" &
                            Operating_Expenditure > 0 &
                            Total_Contribution > 0)

incumbent <-
data.frame(incumbent=unique(CA_cha_inc_open$Cand_Id[which(CA_cha_inc_open$C
and_Incumbent_Challenger_Open_Seat == "INCUMBENT")]))
challenger <-
data.frame(challenger=unique(CA_cha_inc_open$Cand_Id[which(CA_cha_inc_open$
```

```
Cand_Incumbent_Challenger_Open_Seat == "CHALLENGER")]))
open <-
data.frame(open=unique(CA_cha_inc_open$Cand_Id[which(CA_cha_inc_open$Cand_I
ncumbent_Challenger_Open_Seat == "OPEN")]))

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

ggplot(CA_inc_cha_open, aes(x = 1, y = value, fill = factor(Position,
levels = rev(Position)))) +
  geom_col(position = "stack", color = "white") +
  scale_fill_manual("Party",values = c("#edae49", "#d1495b", "#2e4057")) +
  theme_void()
```

Figure 8:

```
CA_challenger_each_month <- count(CA_challenger, Coverage_End_Date)

CA_Total_Receipt <- data.frame(aggregate(CA_challenger$Total_Receipt,
by=list(Coverage_End_Date=CA_challenger$Coverage_End_Date), FUN=sum))
CA_Total_Disbursement <-
data.frame(aggregate(CA_challenger$Total_Disbursement,
by=list(Coverage_End_Date=CA_challenger$Coverage_End_Date), FUN=sum))


#Total Receipt
p1 <- ggplot(CA_Total_Receipt, aes(as.Date(Coverage_End_Date, format =
"%m/%d/%y"), x)) +
  geom_line(color = "grey20", size = 0.75) +
  geom_smooth(aes(color = "smooth"), method="loess", size = 1, na.rm =
TRUE, se = FALSE) +
  xlab(NULL) + ylab("Total Receipt") +
  theme_minimal() + theme(text = element_text(size=13), legend.position =
"none") +
  scale_y_continuous(labels = label_number(suffix = " M", scale = 1e-6))

#Total Disbursement
p2 <- ggplot(CA_Total_Disbursement, aes(as.Date(Coverage_End_Date, format =
"%m/%d/%y"), log(x))) +
  geom_line(color = "grey20", size = 0.75) +
  geom_smooth(aes(color = "smooth"), method="loess", size = 1, na.rm =
```

```
TRUE, se = FALSE) +
  xlab(NULL) + ylab("Total Disbursement") +
  theme_minimal() + theme(text = element_text(size=13), legend.position =
"none") +
  scale_y_continuous(labels = label_number(suffix = " M", scale = 1e-6))

cowplot::plot_grid(p1, p2, ncol = 1, align = "v",
                   labels = "auto", label_fontface = "plain", hjust = 0,
vjust = 1)
```