

# 2012 US Campaign Finance Data Visualization

*Zhu Shen, Jingyi Liu, Sarangoo Chimgee*

*Due Date: 03/22/2018*

```
# Load the dataset
load("house_elections.rda")
load("candidates.rda")
load("committees.rda")
load("contributions.rda")

# Define the function that convert the state abbreviation and state's full name
state_name <- function(x) {
  state_df <- data.frame(
    abb = state.abb,
    full_name = tolower(state.name))
  if (x %in% state_df$abb){
    index <- as.numeric(which(state_df$abb == x))
    return(as.character(state_df[index,2]))
  }

  if (x %in% state_df$full_name){
    index <- as.numeric(which(state_df$full_name == x))
    return(as.character(state_df[index,1]))
  }
}

# Define the function of making the map
gplot <- function(input) {
  ggplot(data = states, mapping = aes(x = long, y = lat, group = group)) +
    geom_polygon(color = "white", fill = "white") +
    geom_polygon(data = input, aes(fill = input$state_donation)) +
    theme(axis.title = element_blank(),
          axis.text = element_blank(),
          axis.line = element_blank(),
          axis.ticks = element_blank(),
          panel.background = element_blank(),
          panel.grid = element_blank(),
          panel.border = element_blank(),
          text = element_text(size=15, face = "bold", color = "black"))
}

# Filter out the candidate data that is only for 2012 elections
# Change the "REP" and "DEM" to "R" and "D" respectively
candidate_2012 <- candidates %>%
  filter(cand_election_yr == 2012) %>%
  mutate(cand_party_affiliation = gsub("REP", "R", cand_party_affiliation),
         cand_party_affiliation = gsub("DEM", "D", cand_party_affiliation),
         cand_state = as.character(cand_state))

# Change the class of the transaction_date to be "Date" and the transaction_amt to be "Numeric"
contributions <- contributions %>%
  mutate(transaction_date = as.Date(transaction_dt, "%m%d%Y"),
```

```

transaction_amt = as.numeric(transaction_amt))

# Normally, presidential election starts the fundraising 15-20 months before the election.
# Conservatively we take all the donation two years before the election into account

# Filter out all committees' contribution that is "Independent expenditure advocating election of candi
# There are some 0 and negative amount and we will filter out those values and only consider the positi
contributions_2012 <- contributions %>%
  filter(transaction_date >= "2010-11-06" & transaction_date <= "2012-11-06") %>%
  filter(transaction_amt > 0 & transaction_type == "24E") %>%
  group_by(cand_id) %>%
  summarize(total_donation = sum(transaction_amt, na.rm =TRUE))

# Full join the contribution and candidate data
candidate_contributions_2012 <- contributions_2012 %>%
  full_join(candidate_2012, by="cand_id")

```

For this project, we analyzed the campaign finance data for Republican and Democratic parties across states in the 2012 election cycle. We only considered transactions that fall into the category of independent expenditure advocating the election of a candidate (24E) from the Contribution by Committee dataset. Since the candidate normally prepares for the election 15- 20 months before the election date, only contributions that took place between November 6, 2010 to November 6, 2012 are considered. By grouping the total amount of the contribution by states and parties, we manage to visualize the campaign finance discrepancy between parties and distributions among states.

## Graphs

```

# Only consider the transaction for election year 2012 and all committees' contribution that is "Indepe
contributions_state <- contributions %>%
  filter(transaction_date >= "2011-11-06" & transaction_dt <= "2012-11-06") %>%
  filter(transaction_amt > 0 & transaction_type == "24E")

# Filter out the candidates from Democratic and Republican and for better visualization, we only consid
# Because some of the contribution amount is extremely large, we normalize the amount for better compar
candidate_contributions_state <- contributions_state %>%
  inner_join(candidate_2012, by="cand_id") %>%
  filter(cand_party_affiliation == "R" | cand_party_affiliation == "D") %>%
  group_by(cand_state) %>%
  filter(n_distinct(cand_party_affiliation) == 2) %>%
  group_by(cand_state, cand_party_affiliation) %>%
  summarize(total_donation = sum(transaction_amt)) %>%
  filter(cand_state != "") %>%
  mutate(norm_total_donation = sqrt(total_donation)) %>%
  mutate(norm_total_donation = ifelse(cand_party_affiliation == "R", -1*norm_total_donation, norm_total.

# Creating the tornado graph with the data, x axis is the stae, y axis is the normalized donation and t
gg <- ggplot(data = candidate_contributions_state, aes(x = cand_state, y = norm_total_donation, fill = c
  geom_bar(stat="identity", position="identity") +
  theme_bw() +
  theme(axis.line = element_blank(),
        panel.border = element_blank(),
        text = element_text(size=10, face = "bold", color = "black")) +

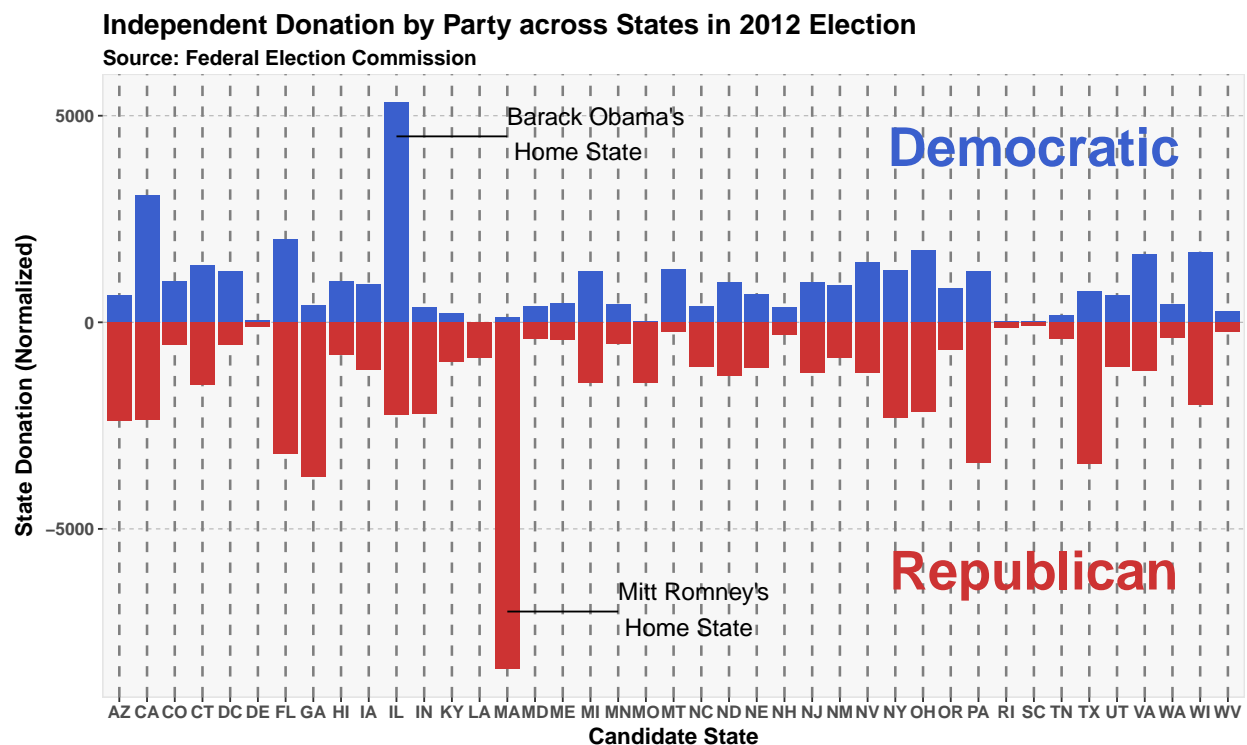
```

```

labs(title = "Independent Donation by Party across States in 2012 Election",
      subtitle = "Source: Federal Election Commission",
      x = "Candidate State", y = "State Donation (Normalized)") +
scale_fill_manual(breaks = c("Democratic", "Republican"),
                  values=c("royalblue3","brown3")) +
theme(text = element_text(colour="black", size=13, face="bold"),
      panel.background = element_rect(fill = "grey97",
                                      colour = "grey90",
                                      size = 0.5, linetype = "solid"),
      panel.grid.minor.y = element_blank(),
      panel.grid.minor.x = element_blank(),
      panel.grid.major.x = element_line(size=.7, color="grey50", linetype = "dashed"),
      panel.grid.major.y = element_line(size=.3, color="grey70", linetype = "dashed")) +
annotate(geom="text", x = c("SC", "SC"), y = c(4250, -6000),
        label = c("Democratic", "Republican"),
        size = 11, color = c("royalblue3","brown3"), fontface = 'bold') +
annotate(geom = "segment", x = "IL", xend = "MA", y = 4500, yend = 4500, size = 0.5) +
annotate(geom = "text", x = "MA", y = 5200, label = "Barack Obama's\n Home State", hjust = 0, vjust = "top") +
annotate(geom = "segment", x = "MA", xend = "MN", y = -7000, yend = -7000, size = 0.5) +
annotate(geom = "text", x = "MN", y = -6300, label = "Mitt Romney's\n Home State", hjust = 0, vjust = "bottom")

```

gg



The bar graph shows the contributions for two parties across the states in 2012, which only includes states with contribution amounts from both parties for comparison. The extreme discrepancy of the bar length for both parties in the same state shows that Georgia, Indiana, Massachusetts and Missouri are mostly Republican states, and Illinois and Montana are Democratic states. As the home states of two candidates for presidency, Massachusetts inclined mostly to Republican and Illinois inclined mostly to the Democratic. This result is unexpected because Massachusetts has a long-standing Democratic preference. Some explanations for this unusual pattern include:

- Massachusetts had more support for Republican candidate Romney because it is his home state.
- The Massachusetts governor in 2012 was a Republican, so people were a lot more inclined to support moderate Republicans.
- Republican candidate got a lot more donations but donations did not correlate to voters sentiments or leaning. Another interesting state is Florida, with long bars for both parties, is known to have an extremely diverse political preference.

```
# split the data of independent expenditure in the election cycle of 2012 by parties
# Sum up the contribution by state
candidate_contributions_2012_rep <- candidate_contributions_2012 %>%
  filter(cand_party_affiliation == "R") %>%
  group_by(cand_state) %>%
  summarize(state_donation = sum(total_donation, na.rm =TRUE))

candidate_contributions_2012_dem <- candidate_contributions_2012 %>%
  filter(cand_party_affiliation == "D") %>%
  group_by(cand_state) %>%
  summarize(state_donation = sum(total_donation, na.rm =TRUE))

# Apply the conversion function of state abbreviation to state full name to each of the party data set
candidate_contributions_2012_rep$cand_state <- lapply(candidate_contributions_2012_rep$cand_state, FUN =
candidate_contributions_2012_rep$cand_state <- as.character(candidate_contributions_2012_rep$cand_state)
candidate_contributions_2012_dem$cand_state <- lapply(candidate_contributions_2012_dem$cand_state, FUN =
candidate_contributions_2012_dem$cand_state <- as.character(candidate_contributions_2012_dem$cand_state)

# Change the 0 values in the data file to be 1 because later we will use the log transition. We convert
candidate_contributions_2012_rep$state_donation[candidate_contributions_2012_rep$state_donation == 0] <- 1
candidate_contributions_2012_dem$state_donation[candidate_contributions_2012_dem$state_donation == 0] <- 1

# get the states data from the map_data package
states <- map_data("state")

# Add another column named cand_state with same element as the column region
states$cand_state <- states$region

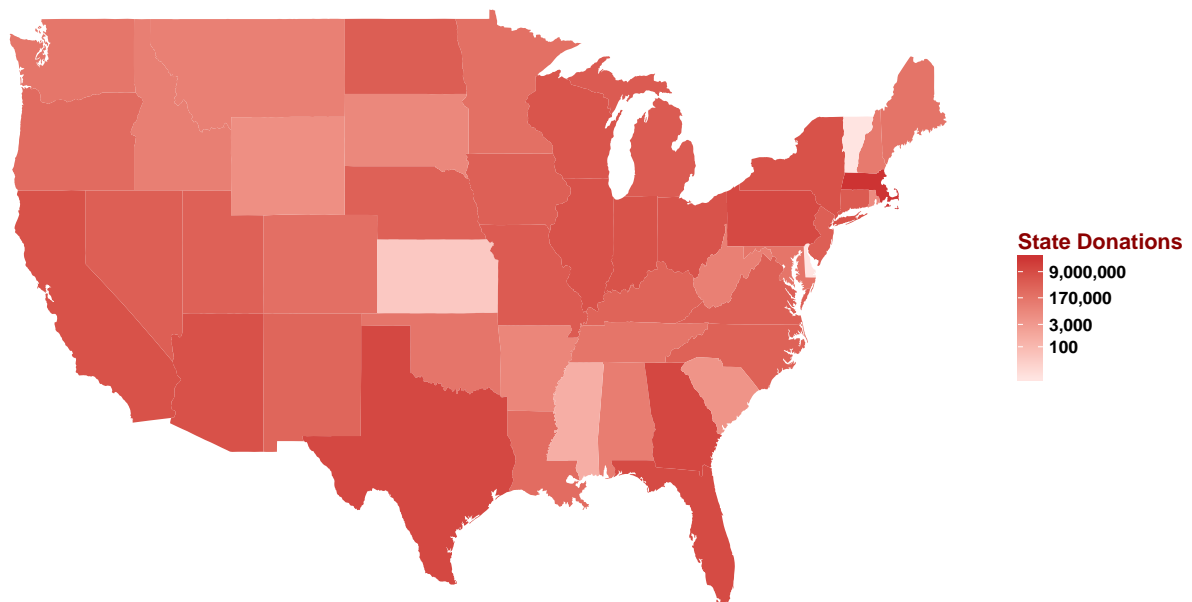
# Join the states data and contribution data files by cand_state
map_rep <- inner_join(states, candidate_contributions_2012_rep, by = "cand_state")
map_dem <- inner_join(states, candidate_contributions_2012_dem, by = "cand_state")

# Create the map for Republican
rep <- ggplot(map_rep) +
  labs(title = "Republican Donations in US 2012 Election (By States)",
        subtitle = "Source: Federal Election Commission") +
  scale_fill_gradient(low="mistyrose1", high="brown3", trans = "log",
                      name = "State Donations", breaks = c(100,3000, 170000, 9000000),
                      labels = c("100","3,000", "170,000","9,000,000"))+
  theme(legend.title = element_text(colour="red4", size=15, face="bold"))

rep
```

## Republican Donations in US 2012 Election (By States)

Source: Federal Election Commission

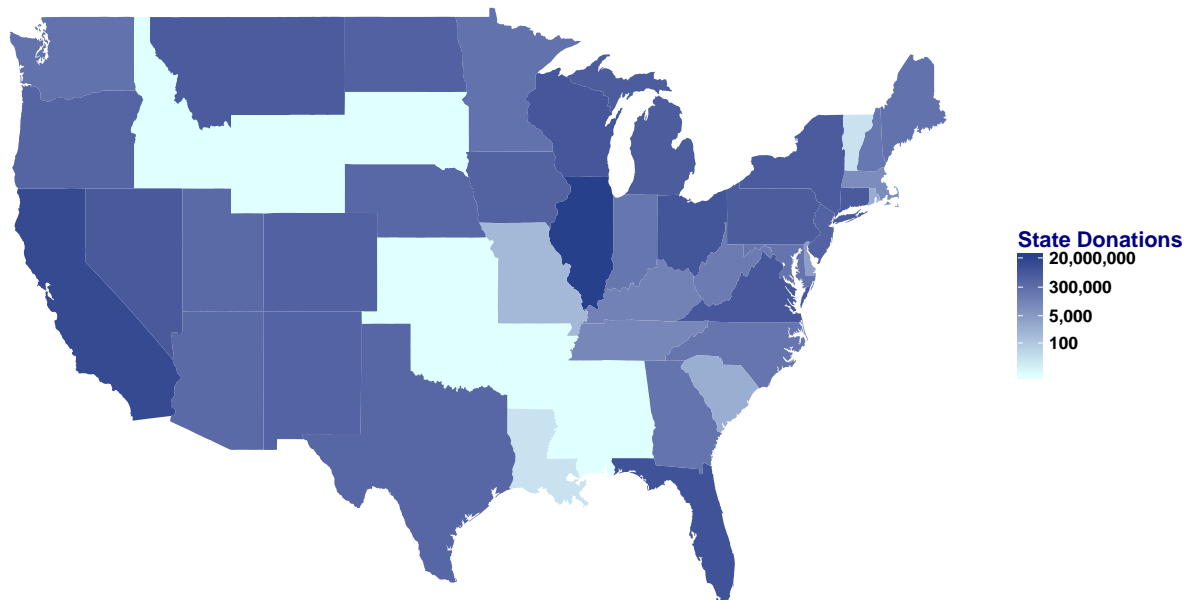


```
# Create the map for Democratic
dem <- gplot(map_dem) +
  labs(title = "Democratic Donations in US 2012 Election (By States)",
        subtitle = "Source: Federal Election Commission") +
  scale_fill_gradient(low="lightcyan", high="royalblue4", trans = "log",
                      name = "State Donations", breaks = c(100,5000, 300000,20000000),
                      labels = c("100","5,000","300,000","20,000,000"))+
  theme(legend.title = element_text(colour="navy", size=15, face="bold"))

dem
```

## Democratic Donations in US 2012 Election (By States)

Source: Federal Election Commission



These two US states maps show the geographical distribution of the contribution across states; the darker the color, the more contribution to the candidates from that state. In the map for Republican candidates, Massachusetts, Pennsylvania, Florida and Texas are the darkest, which means these states contributed the most while Kansas and Mississippi are the lightest, showing that these states contributed the least. The total contribution from Alaska, Delaware, Vermont is missing from the dataset. Similarly, in the map for Democratic candidates, Illinois, California, and Florida contributed the most while Louisiana, Vermont, and Missouri contributed the least, even lower than \$500. In the map, the contribution for nine states is missing.

Why there are no records of independent expenditure supporting the candidate from these states? Checking these states one by one, we can conclude several reasons for these missing records.

- When the Transaction Does Not Fall into the Category of 24E

If a candidate has only 24E transactions and no 24K transactions, the candidate is most likely to be preparing for their next election cycle by transferring their excess contributions to later elections. Don Young, a U.S Representative for Alaska's at-large congressional district since 1973, had no sign of 24E transactions mainly because of two reasons: his popularity and the inactive political participation of Alaska. This explains the reason for no 24K transactions because he has had the overwhelming majority in every campaign since 2010 and the apolitical stance of Alaska shows how Alaskan citizens have no incentive to act politically.

- Only Candidate in the Election

If a candidate is the only nominee in the election, then people will no incentive to contribute money to this candidate because the candidate will be elected for sure. For example, Terri Sewell was the only candidate to file for the Democratic nomination in 2012 in Alabama. Even though we can find her name and ID in the candidate data, there is no record of the independent expenditure for her in the contribution data.

- Term Limit

If a candidate has already served for two consecutive terms in the previous years, the candidate could not seek re-election due to the term limit. An example is Kenneth Corn, who served in the Oklahoma State Senate, representing District 4 from 2002 to 2010, but was not re-elected due to the term limit.

- Other Reasons\*

There are many other possible reasons. For example, the candidate is so unpopular that all the contribution comes from personal saving; or the state is apolitical that citizens did not care much about the political election; or there are some missing records in the data files when the data was recorded initially, etc.

## **Conclusion**

We've learned that the overcomplexity and the immense size of the data doesn't excuse the data to be missing; instead, it entails that in the real world and the normality of having data such size, conditions and special cases like the above shows that the assumed "missing" data is not actually missing, but it's a part of how data works and requires to interpret it in a different way.

Github