Mark Hand

For Brian Roberts

Sep 2017

FEC Data Project

The Center for Responsive Politics publishes <http://OpenSecrets.org>, one of the most popular sources of FEC data. Their data, however, come from publicly available data published by the Federal Election Commission. This project attempts to reproduce some of the data presented on OpenSecrets.org using the primary source data provided by the FEC.

# Challenge One

Reproduce one of the summary tables found at: <https://www.opensecrets.org/overview/index.php?display=T&type=A&cycle=2016>.

1. Download the CandidateSummaryAction.csv from <http://classic.fec.gov/data/CandidateSummary.do?format=csv>
2. In Excel, convert all of the dollar amounts in columns 14-48 to numbers and eliminated the commas
3. Then import the data into R:

setwd("~/Google Drive/PhD/Money in Politics/FEC") # Set your working directory to you moved the data   
library(tidyverse) # If you get an error, first do install.packages("tidyverse")  
library(dplyr)  
library(WriteXLS)  
Candidates <- read\_csv("CandidateSummaryAction\_nodollars.csv")

According to the metadata from <http://classic.fec.gov/finance/disclosure/metadata/metadataforcandidatesummary.shtmloth_com_con>, use the following variables:

1. ind\_con: "Sum of the previous two fields (i.e. ind\_ite\_con and ind\_uni\_con): ind\_ite\_con: "Contributions must be itemized (i.e. listing specific information about the donor and the contribution) when the total given to the campaign from a single individual exceeds $200." ind\_uni\_con: "Total given to the campaign where specific information about the donors is not required because they have not given more than $200."
2. oth\_com\_con: "Mostly contributions from PACs, also includes contributions from other candidates. These also must be itemized regardless of amount"
3. cas\_on\_han\_clo\_of\_per: "Ending cash balance on the most recent filing."
4. tot\_con: "Sum of the previous 4 fields (i.e. ind\_con, par\_com\_con, oth\_com\_con and can\_con)"
5. tot\_dis: "Sum of all disbursement categories"

Now, attempting to rebuild the chart of contributions to 2016 House candidates:

House <- Candidates %>% filter(can\_off=="H")  
HouseGOPandDem <- House %>% filter(can\_par\_aff=="REP" | can\_par\_aff=="DEM")  
party <- c("DEM","REP")  
a <- as.data.frame(party)  
names(a)[1] <- "can\_par\_aff"  
b <- as.data.frame(rbind(nrow(HouseDem <- House %>% filter(can\_par\_aff=="DEM")), nrow(HouseGOP <- House %>% filter(can\_par\_aff=="REP"))))  
names(b)[1] <- "No\_of\_Cands"  
b <- cbind(a,b)  
HouseGOPandDem %>%   
 group\_by(can\_par\_aff) %>%   
 summarise(total\_raised=sum(tot\_con, na.rm=TRUE)) -> c  
HouseGOPandDem %>%   
 group\_by(can\_par\_aff) %>%   
 summarise(total\_spent=sum(tot\_dis, na.rm=TRUE)) -> d  
HouseGOPandDem %>%   
 group\_by(can\_par\_aff) %>%   
 summarise(cash\_on\_hand=sum(cas\_on\_han\_clo\_of\_per, na.rm=TRUE)) -> e  
HouseGOPandDem %>%   
 group\_by(can\_par\_aff) %>%   
 summarise(total\_from\_PACs=sum(oth\_com\_con, na.rm=TRUE)) -> f  
HouseGOPandDem %>%   
 group\_by(can\_par\_aff) %>%   
 summarise(total\_from\_Indivs=sum(ind\_con, na.rm=TRUE)) -> g  
h <- cbind(a,b,c,d,e,f,g)

Now that you're exhausted from trying to learn that much R that fast, export back to Excel to finish up:

library(WriteXLS)  
WriteXLS(h, ExcelFileName = "2016 House Candidates.xls")

And voila, our first chart!

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Party | Number of Candidates | Total Raised | Total Spent | Cash on Hand | Total from PACs, etc. | Total from Individuals |
| ALL | 1387 | $247,898,619 | $166,158,742 | $353,224,912 | $87,531,227 | $158,737,829 |
| DEM | 886 | $132,177,839 | $90,536,406 | $168,164,214 | $32,307,623 | $98,633,518 |
| REP | 501 | $115,720,780 | $75,622,336 | $185,060,698 | $55,223,604 | $60,104,311 |

This took me hours, not because of the FEC data but becuase I'm still learning how to use R. In general, this exercise underscored the point that the data reported by Open Secrets seems not to line up at all with the data available on the FEC website.

# Challenge Two

Reproduce the numbers from the top four "Super" PACs from the 2016 election cycle: <https://www.opensecrets.org/outsidespending/summ.php?cycle=2016&chrt=V&disp=O&type=S>.

1. Download the .csv file for 2016 Independent Expenditures. Go to this site: <http://classic.fec.gov/data/IndependentExpenditure.do?format=html&election_yr=2016>, click on the icon labeled "CSV."
2. Open that file in Excel and convert all of the dollar amounts in columns 14-48 to numbers and eliminated the commas.
3. Import that dataset to R:

Ind\_Expenditures <- read\_csv("independent-expenditure-nodollars.csv")

According to the metadata from <http://classic.fec.gov/finance/disclosure/metadata/metadataforindependentexpenditures.shtml>, use the following variables:

1. spe\_nam: "Name of committee, individual or group making expenditure"
2. exp\_amo: "Dollar amount of specific expenditure"
3. agg\_amo: "Total amount expended during the calendar year, per election, per office sought." A note on this: Given that this is the only other dollar amount present in the dataset that OpenSecrets could be calling "Amount Raised", we'll call this "maybe\_raised" until Monday.

... then export to Excel for final cleanup.

Ind\_Expenditures %>%   
 group\_by(spe\_nam) %>%   
 summarise(total\_spent = sum(exp\_amo, na.rm=TRUE),maybe\_raised = sum(agg\_amo, na.rm=TRUE)) -> Amounts   
  
Ind\_Expenditures %>%   
 group\_by(spe\_nam) %>%   
 select(spe\_nam, can\_par\_aff, sup\_opp) -> View  
  
Amounts %>% top\_n(5, total\_spent) -> TopAmounts  
  
WriteXLS(TopAmounts, ExcelFileName = "2016 Super PACs.xls")

Our second chart:

|  |  |  |
| --- | --- | --- |
| Super PAC Name | Total Spent | Total Raised, maybe? |
| NRCC | $137,349,828.84 | $9,977,753,194.82 |
| Priorities USA Action | $87,042,453.56 | $3,889,167,448.01 |
| Right to Rise USA | $85,994,327.66 | $1,422,306,061.52 |
| Senate Leadership Fund | $83,579,801.40 | $2,688,405,522.79 |
| Senate Majority PAC | $73,601,651.08 | $1,133,261,515.98 |

Some reflections: The expenditures here line up a *bit* better with the OpenSecrets data than those from the first assignment, but it's still unclear to me how OpenSecrets went from the FEC data to this chart; nor did it become clear what agg\_amo might mean.

Secondly, I noticed that OpenSecrets’ Super PAC charts are not a little misleading in their presentation of the purpose and views of the Super PACs. PACs listed multiple purposes, supported and opposed multiple candidates, and didn’t list ideology—only party.