

DATA 606: Project Proposal

Walt Wells, Fall 2016

Contents

```
# load data
camp <- read.csv("FEC_campaign_summary.csv")
pac <- read.csv("FEC_independent-expenditure.csv")
```

Research question

You should phrase your research question in a way that matches up with the scope of inference your dataset allows for.

In this project we will examine campaign expenditures for the 2016 US presidential election. We will concentrate on the two major party nominees, and will look at expenditures by the month, starting in Feb 2016.

Because of the nuances of campaign finance law, while we will be able to see the intake for the campaigns, we will not be able to see fund intake for the independent PACs. As a result, we will look at expenditures.

The summary stats by month for each campaign are already available. We will tidy the independant PAC data to create monthly summary expenditures for groups that favor or oppose the two major party nominees.

Our goal is to compare monthly expenditures for the campaign and for PACs across the two candidates. We will further attempt to use the “sup_opp” variable in the PAC data to divide PAC expenditures across “negative” or “positive” efforts.

NOTE: Before the project is due, there should be another month release of the data. We will update our collected dataset accordingly. The data cleaning methods we build are expected to work well with new observations. Three cheers for reproduceable research!

NOTE2: A secondary study I may choose to conduct will be to further explore PAC data by the “pur” or “purpose of expenditure” category and review what PACs were spending their money on. This would require some superior data cleaning efforts, as this var currently has 1362 levels with items as varied as “Robocalls” and “Taxi from event”.

Cases

What are the cases, and how many are there?

In our raw data, there are of the campaign 207 data and 44666 of the PAC data.

Really, we’ll only want the monthly observations for our two candidates so once the data is cleaned and merged into a long format dataset, I would expect the number of cases = n months * 6, where 6 represents ClintonCampaign, ProClintonPAC, AntiClintonPAC, TrumpCampaign, ProTrumpPAC, AntiTrumpPAC.

Estimate of about ~40.

Data collection

Describe the method of data collection.

The data is delivered to the FEC and redistributed to the public as part of disclosure law.

“The Federal Election Commission (FEC) is an independent regulatory agency established in 1975 to administer and enforce the Federal Election Campaign Act (FECA). That statute limits the sources and amounts of the contributions used to finance federal elections, requires public disclosure of campaign finance information and—in tandem with the Primary Matching Payment Act and the Presidential Election Campaign Fund Act—provides for the public funding of Presidential elections. For more information on the FEC’s role in regulating federal elections, see the brochure “The FEC and the Federal Campaign Finance Law.” - From the FEC Site

Type of study

What type of study is this (observational/experiment)?

This is an observational data study.

Data Source

If you collected the data, state self-collected. If not, provide a citation/link.

Data is source from is publicly available datasets from the Federal Election Commission. The data we will examine includes summary data on the 2016 US Presidential election from both the campaigns and independant groups or PACs.

Response

What is the response variable, and what type is it (numerical/categorical)?

The response variable will be the expenditure by the month. This will be a numerical variable.

Explanatory

What is the explanatory variable, and what type is it (numerical/categorical)?

We expect to have a few explanatory variables, candidate (Categorical), whether it is campaign or PAC (Categorical), and month (Categorical). As we delve deeper, we may wish to add a few additional variables.

Relevant summary statistics

Provide summary statistics relevant to your research question. For example, if you’re comparing means across groups provide means, SDs, sample sizes of each group. This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

```
if (!require('dplyr')) install.packages('dplyr')
if (!require('stringr')) install.packages('stringr')
if (!require('ggplot2')) install.packages('ggplot2')
```

Campaign Data Initial Review

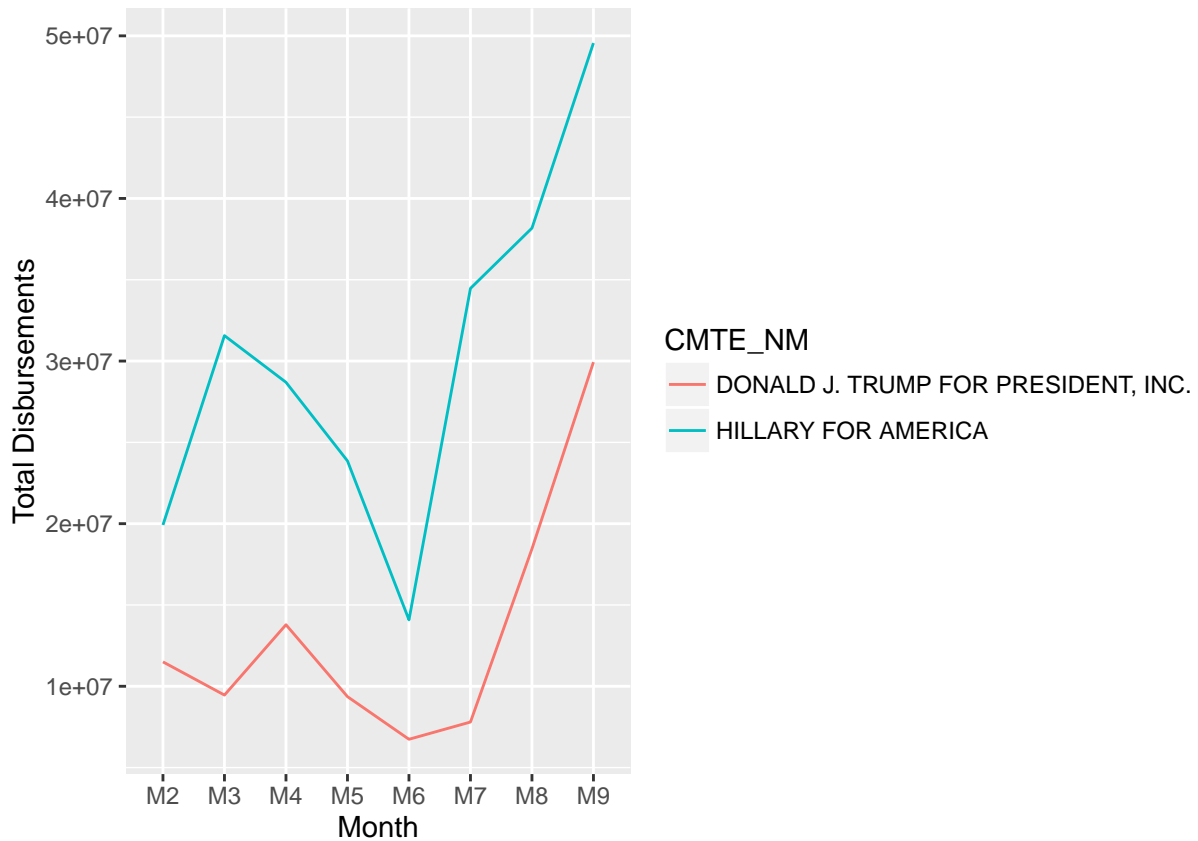
```
campsub <- filter(camp, CMTE_NM == "HILLARY FOR AMERICA" | CMTE_NM == "DONALD J. TRUMP FOR PRESIDENT, INC")
  filter(str_detect(RPT_TP, "M")) %>%
  arrange(CMTE_NM, RPT_TP) %>%
  select(CMTE_NM, RPT_TP, TTL_DISB_SUM_PAGE_PER) %>%
```

```
droplevels()
```

```
knitr::kable(campsub)
```

CMTE_NM	RPT_TP	TTL_DISB_SUM_PAGE_PER
DONALD J. TRUMP FOR PRESIDENT, INC.	M2	11500706
DONALD J. TRUMP FOR PRESIDENT, INC.	M3	9458274
DONALD J. TRUMP FOR PRESIDENT, INC.	M4	13785551
DONALD J. TRUMP FOR PRESIDENT, INC.	M5	9361528
DONALD J. TRUMP FOR PRESIDENT, INC.	M6	6739944
DONALD J. TRUMP FOR PRESIDENT, INC.	M7	7800249
DONALD J. TRUMP FOR PRESIDENT, INC.	M8	18450447
DONALD J. TRUMP FOR PRESIDENT, INC.	M9	29933921
HILLARY FOR AMERICA	M2	19919632
HILLARY FOR AMERICA	M3	31564541
HILLARY FOR AMERICA	M4	28694882
HILLARY FOR AMERICA	M5	23859389
HILLARY FOR AMERICA	M6	14085977
HILLARY FOR AMERICA	M7	34466402
HILLARY FOR AMERICA	M8	38177394
HILLARY FOR AMERICA	M9	49554328

```
ggplot(campsub, aes(x=RPT_TP, y=TTL_DISB_SUM_PAGE_PER,  
                    group=CMTE_NM, colour=CMTE_NM)) +  
  geom_line() + ylab("Total Disbursements") +  
  xlab("Month")
```



PAC Data Initial Review

This data set will require some effort to tidy and clean and normalize, including steps to summarize PACs by candidate, summarize expenditures by months, etc.

We can see this quickly by attempting to simply review the values we will use to calculate our monthly sums.

```
pac$expenditure <- str_replace_all(as.character(pac$exp_amo), fixed("$"), "")
pac$expenditure <- as.numeric(str_replace_all(pac$expenditure, ',', ''))

summary(pac$expenditure)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	-5000	17	34	8656	52	50000000