

# Final Project

*Bridget Silha*

*11/7/2019*

```
require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages ----- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1    v purrr  0.3.3
## v tibble  2.1.3    v dplyr  0.8.3
## v tidyr   1.0.0    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
tinytex::install_tinytex()
```

```
## Warning: Detected an existing tlmgr at /Users/bridgetsilha/Library/TinyTeX/
## bin/x86_64-darwin/tlmgr. It seems TeX Live has been installed (check
## tinytex::tinytex_root()). You are recommended to uninstall it, although
## TinyTeX should work well alongside another LaTeX distribution if a LaTeX
## document is compiled through tinytex::latexmk().
```

```
## The directory /usr/local/bin is not writable. I recommend that you make it writable. See https://git
```

```
## TinyTeX installed to /Users/bridgetsilha/Library/TinyTeX
```

```
#Read CVS data in
```

```
total_fec_committee_party_and_pac <- read_csv("committees.csv")
```

```
## Parsed with column specification:
## cols(
##   name = col_character(),
##   cycles = col_character(),
##   treasurer_name = col_character(),
##   committee_type = col_character(),
##   committee_type_full = col_character(),
##   filing_frequency = col_character(),
##   designation = col_character(),
##   designation_full = col_character(),
##   organization_type = col_character(),
##   organization_type_full = col_character(),
##   affiliated_committee_name = col_character(),
##   party = col_character(),
```

```
## party_full = col_character(),
## state = col_character(),
## committee_id = col_character(),
## candidate_ids = col_character(),
## first_file_date = col_date(format = ""),
## last_file_date = col_date(format = ""),
## last_f1_date = col_date(format = "")
## )
```

```
pac_financial_info_2018 <- read_csv("committee_summary_2018.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Link_Image = col_character(),
##   CMTE_ID = col_character(),
##   CMTE_NM = col_character(),
##   CMTE_TP = col_character(),
##   CMTE_DSGN = col_character(),
##   CMTE_FILING_FREQ = col_character(),
##   CMTE_ST1 = col_character(),
##   CMTE_ST2 = col_character(),
##   CMTE_CITY = col_character(),
##   CMTE_ST = col_character(),
##   CMTE_ZIP = col_character(),
##   TRES_NM = col_character(),
##   CAND_ID = col_character(),
##   ITEM_REF_REB_RET = col_logical(),
##   SUBTTL_REF_REB_RET = col_logical(),
##   UNITEM_REF_REB_RET = col_logical(),
##   ITEM_OTHER_REF_REB_RET = col_logical(),
##   UNITEM_OTHER_REF_REB_RET = col_logical(),
##   SUBTTL_OTHER_REF_REB_RET = col_logical(),
##   ITEM_OTHER_INCOME = col_logical()
##   # ... with 12 more columns
## )
```

```
## See spec(...) for full column specifications.
```

```
## Warning: 28 parsing failures.
```

```
##   row      col      expected  actual      file
## 3227 EXP_SUBJECT_LIMITS 1/0/T/F/TRUE/FALSE -2205.84 'committee_summary_2018.csv'
## 3227 ITEM_CONVN_EXP_DISB 1/0/T/F/TRUE/FALSE -2235.84 'committee_summary_2018.csv'
## 3227 ITEM_OTHER_DISB    1/0/T/F/TRUE/FALSE 501053.65 'committee_summary_2018.csv'
## 3227 SUBTTL_CONVN_EXP_DISB 1/0/T/F/TRUE/FALSE -2205.84 'committee_summary_2018.csv'
## 3227 UNITEM_CONVN_EXP_DISB 1/0/T/F/TRUE/FALSE 30      'committee_summary_2018.csv'
## ....
## See problems(...) for more details.
```

```
##Clean Data
```

```
##https://www.fec.gov/campaign-finance-data/committee-type-code-descriptions/
```

```

#Join on Committee ID
new_data_2018 <- right_join(total_fec_committee_party_and_pac, pac_financial_info_2018, by=c("committee_id", "party"))

#Cut down variables
new_data_2018 <- new_data_2018 %>%
  select(-cycles, -treasurer_name, -filing_frequency:-affiliated_committee_name)

new_data_2018 <- new_data_2018 %>%
  select (-Link_Image, -CMTE_ST1:-CMTE_CITY, -CMTE_ZIP:-CAND_ID, -INDV_CONTB:-OTH_CMTE_CONTB, -TRANF_FRM)

skimr::skim(new_data_2018)

```

```

## Skim summary statistics
## n obs: 12320
## n variables: 23
##
## -- Variable type:character -----
##      variable missing complete      n min max empty n_unique
##      candidate_ids      1    12319 12320    2 161    0      3826
##      CMTE_DSGN          1    12319 12320    1   1    0         6
##      CMTE_FILING_FREQ    0    12320 12320    1   1    0         5
##      CMTE_NM            0    12320 12320    3 178    0     12218
##      CMTE_ST            1    12319 12320    2   2    0         61
##      CMTE_TP            0    12320 12320    1   1    0         15
##      committee_id       0    12320 12320    9   9    0     12298
##      committee_type      1    12319 12320    1   1    0         15
##      committee_type_full  1    12319 12320    5  48    0         15
##      name                1    12319 12320    3 178    0     12217
##      ORG_TP             8975     3345 12320    1   1    0          8
##      party              7784     4536 12320    1   3    0         44
##      party_full         7847     4473 12320    4  26    0         26
##      state              3     12317 12320    2   2    0         63
##
## -- Variable type:Date -----
##      variable missing complete      n      min      max      median
##      first_file_date      3    12317 12320 1972-04-14 2019-07-28 2014-12-12
##      last_f1_date        276    12044 12320 1982-01-26 2019-11-12 2017-11-29
##      last_file_date      3    12317 12320 2004-10-08 2019-11-12 2019-07-30
##      n_unique
##      5168
##      2658
##      758
##
## -- Variable type:logical -----
##      variable missing complete      n mean      count
##      COH_BOY    12320      0 12320   NaN      12320
##      COH_COY    12317      3 12320    0 NA: 12317, FAL: 3
##
## -- Variable type:numeric -----
##      variable missing complete      n      mean      sd      p0
##      FEC_ELECTION_YR      0    12320 12320    2018    0      2018
##      TTL_CONTB          101    12219 12320 708433.51 1.2e+07 -2695
##      TTL_DISB          295    12025 12320 832470.86 1.3e+07 -12685.64

```

```
##      TTL_RECEIPTS      292      12028 12320 859095.91 1.3e+07 -36441.47
##      p25      p50      p75      p100      hist
## 2018      2018      2018      2018
##      25      24275      188490.77      1.2e+09
## 2155      32687.63 221684.4      1.3e+09
##      604.43 32433.1 224135.04      1.3e+09
```

*#Read CVS data in*

```
total_fec_committee_party_and_pac <- read_csv("committees.csv")
```

```
## Parsed with column specification:
## cols(
##   name = col_character(),
##   cycles = col_character(),
##   treasurer_name = col_character(),
##   committee_type = col_character(),
##   committee_type_full = col_character(),
##   filing_frequency = col_character(),
##   designation = col_character(),
##   designation_full = col_character(),
##   organization_type = col_character(),
##   organization_type_full = col_character(),
##   affiliated_committee_name = col_character(),
##   party = col_character(),
##   party_full = col_character(),
##   state = col_character(),
##   committee_id = col_character(),
##   candidate_ids = col_character(),
##   first_file_date = col_date(format = ""),
##   last_file_date = col_date(format = ""),
##   last_f1_date = col_date(format = "")
## )
```

```
pac_financial_info_2020 <- read_csv("committee_summary_2020.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Link_Image = col_character(),
##   CMTE_ID = col_character(),
##   CMTE_NM = col_character(),
##   CMTE_TP = col_character(),
##   CMTE_DSGN = col_character(),
##   CMTE_FILING_FREQ = col_character(),
##   CMTE_ST1 = col_character(),
##   CMTE_ST2 = col_character(),
##   CMTE_CITY = col_character(),
##   CMTE_ST = col_character(),
##   CMTE_ZIP = col_character(),
##   TRES_NM = col_character(),
##   CAND_ID = col_character(),
##   ITEM_REF_REB_RET = col_logical(),
##   SUBTTL_REF_REB_RET = col_logical(),
```

```
## UNITEM_REF_REB_RET = col_logical(),
## ITEM_OTHER_REF_REB_RET = col_logical(),
## UNITEM_OTHER_REF_REB_RET = col_logical(),
## SUBTTL_OTHER_REF_REB_RET = col_logical(),
## ITEM_OTHER_INCOME = col_logical()
## # ... with 13 more columns
## )
```

```
## See spec(...) for full column specifications.
```

```
## Warning: 2 parsing failures.
## row col expected actual file
## 3806 TTL_COMMUNICATION_COST 1/0/T/F/TRUE/FALSE 3476.82 'committee_summary_2020.csv'
## 5428 TTL_COMMUNICATION_COST 1/0/T/F/TRUE/FALSE 52648.65 'committee_summary_2020.csv'
```

```
##Clean Data
```

```
#https://www.fec.gov/campaign-finance-data/committee-type-code-descriptions/
```

```
#Join on Committee ID
```

```
new_data_2020 <- right_join(total_fec_committee_party_and_pac, pac_financial_info_2020, by=c("committee_id", "committee_type"))
```

```
#Cut down new variables
```

```
new_data_2020 <- new_data_2020 %>%
  select(-cycles, -treasurer_name, -filing_frequency, -affiliated_committee_name)
```

```
new_data_2020 <- new_data_2020 %>%
```

```
  select (-Link_Image, -CMTE_ST1:-CMTE_CITY, -CMTE_ZIP:-CAND_ID, -INDV_CONTB:-OTH_CMTE_CONTB, -TRANF_FRM)
```

```
skimr::skim(new_data_2020)
```

```
## Skim summary statistics
```

```
## n obs: 9962
```

```
## n variables: 23
```

```
##
```

```
## -- Variable type:character -----
```

	variable	missing	complete	n	min	max	empty	n_unique
##	candidate_ids	0	9962	9962	2	161	0	2929
##	CMTE_DSGN	0	9962	9962	1	1	0	6
##	CMTE_FILING_FREQ	0	9962	9962	1	1	0	4
##	CMTE_NM	0	9962	9962	3	178	0	9922
##	CMTE_ST	0	9962	9962	2	2	0	60
##	CMTE_TP	0	9962	9962	1	1	0	13
##	committee_id	0	9962	9962	9	9	0	9955
##	committee_type	0	9962	9962	1	1	0	13
##	committee_type_full	0	9962	9962	5	48	0	13
##	name	0	9962	9962	3	178	0	9922
##	ORG_TP	6893	3069	9962	1	1	0	7
##	party	6344	3618	9962	1	3	0	41
##	party_full	6399	3563	9962	4	26	0	26
##	state	0	9962	9962	2	2	0	60

```
##
```

```
## -- Variable type:Date -----
```

```
##      variable missing complete      n      min      max      median
## first_file_date      0      9962 9962 1972-04-14 2019-10-22 2014-03-06
## last_f1_date      30      9932 9962 1982-01-26 2019-11-12 2018-07-30
## last_file_date      0      9962 9962 2016-11-03 2019-11-12 2019-10-04
## n_unique
##      4902
##      2318
##      231
##
## -- Variable type:logical -----
## variable missing complete      n mean count
## COH_BOY      9962      0 9962  NaN  9962
## COH_COY      9962      0 9962  NaN  9962
##
## -- Variable type:numeric -----
##      variable missing complete      n      mean      sd      p0 p25
## FEC_ELECTION_YR      0      9962 9962      2020      0      2020 2020
## TTL_CONTB      49      9913 9962 279960.25 4162515.41 -1321.54  0
## TTL_DISB      54      9908 9962 250830.24 4219255.96 -17439.72 80
## TTL_RECEIPTS      54      9908 9962 333284.55 4661473.35 -4713.24  0
##      p50      p75      p100      hist
## 2020      2020      2020
## 4422.36 58564.59      3.4e+08
## 6101.02 55070.25      3.4e+08
## 6293.6  70747.47      3.4e+08
```

```
#Bind 2018 and 2020 datasets together
```

```
new_data <- bind_rows(new_data_2018, new_data_2020)
```

```
#Filter for party type
```

```
new_data <- new_data %>%
```

```
  filter(party == "DEM" | party == "REP" | party == "CRV" | party == "LIB" | party == "IAP" | party ==
```

```
#Filter for committee type
```

```
new_data <- new_data %>%
```

```
  filter(CMTE_TP == "N" | CMTE_TP == "O" | CMTE_TP == "Q" | CMTE_TP == "V" | CMTE_TP == "W")
```

```
#split into training and test data
```

```
range(new_data$last_file_date)
```

```
## [1] "2017-02-07" "2019-11-12"
```