DATA 606 - Final Project Proposal

Adam Douglas

10/30/2018

Data Preparation

The data comes from the American National Election Studies (ANES), a survey conducted by is a collaboration between Stanford University and the University of Michigan, with funding by the National Science Foundation.[1]

```
# Load data via a pipe-delimited file, using character type for all columns
rawData <- read_delim("anes_timeseries_2016_rawdata.txt","|",
                      col_types= paste(rep("c",1290),sep="",collapse=""))

# Empty data frame
survey <- data.frame(ID = as.character(rawData$V160001))

# Recode some initial variables
survey$surveyMethod <- fct_recode(rawData$V160501,
                                  "Face-To-Face" = "1", "Web" = "2")

survey$surveyType <- fct_recode(rawData$V160502,
                                "Only Pre" = "0", "Pre and Post" = "1")

survey$sex <- fct_recode(rawData$V161002,
                         "Male" = " 1", "Female" = " 2", NULL = "-1")

survey$payAttPol <- fct_recode(rawData$V161003,
                               "Always" = "1", "Most of the time" = "2",
                               "About half of the time" = "3",
                               "Some of the time" = "4", "Never" = "5")

survey$votePrim <- fct_recode(rawData$V161021,
                              "Yes" = " 1", "No" = " 2",
                              NULL = "-8", NULL = "-9")

survey$earlyVote <- fct_recode(rawData$V161022,
                               "Yes" = " 1", "No" = " 2",
                               NULL = "-1", NULL = "-9")

survey$clinton <- as.numeric(rawData$V161086)
survey$clinton[survey$clinton < 0] <- NA

survey$trump <- as.numeric(rawData$V161087)
survey$trump[survey$trump < 0] <- NA

survey <- survey %>% mutate(bothCandidates = trump - clinton)
```

[1] The American National Election Studies (www.electionstudies.org).

Research Question

My research question is: does a correlation exist between a person's exposure to certain news media outlets (web and traditional) and their stated attitudes towards the presidential candidates in the 2016 election?

Cases

Each case is a person surveyed. There are 4271 observations in the data set.

Data Collection

Data was collected by survey, either in-person or via the web. The larger data set has people surveyed both before and after the election, but we will look at questions answered just prior to the election.

Type of Study

This is an observational study

Data Source

Data was retreived from the ANES website

The American National Election Studies. These materials are based on work supported by the National Science Foundation under grant numbers SES 1444721, 2014-2017, the University of Michigan, and Stanford University. Any opinions, findings and conclusions, or recommendations expressed in these materials are those of the author(s) and do not necessarily reflect the views of the funding organizations.

Response

The response variable that we will be using is numeric and is the respondent's rating of then-candidate Donald Trump minus their rating of Hillary Clinton.

Explanatory

The explanatory variables are the respondent's use of various media (web, print, TV) and frequency (where given).

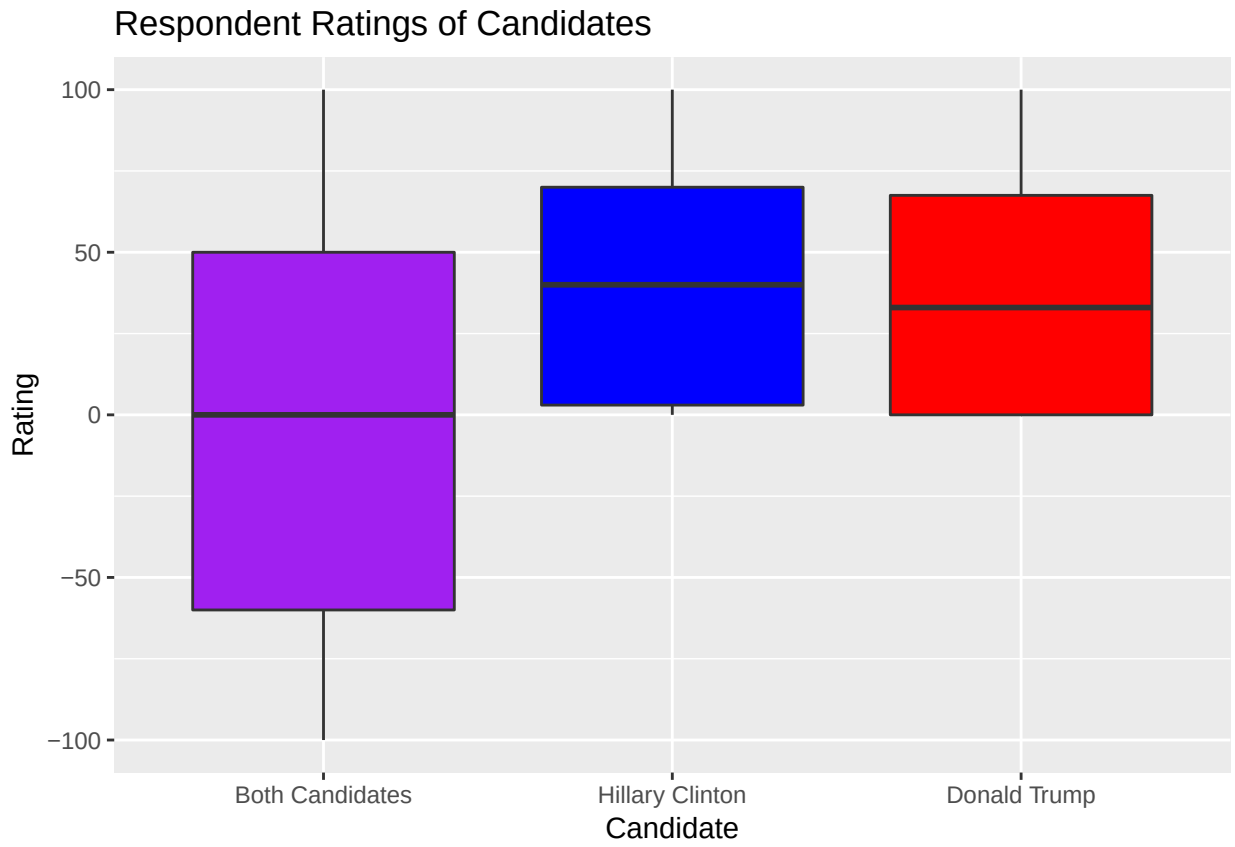
Relevant summary statistics

```
survey %>% filter(!is.na(clinton), !is.na(trump)) %>% summary(.)
```

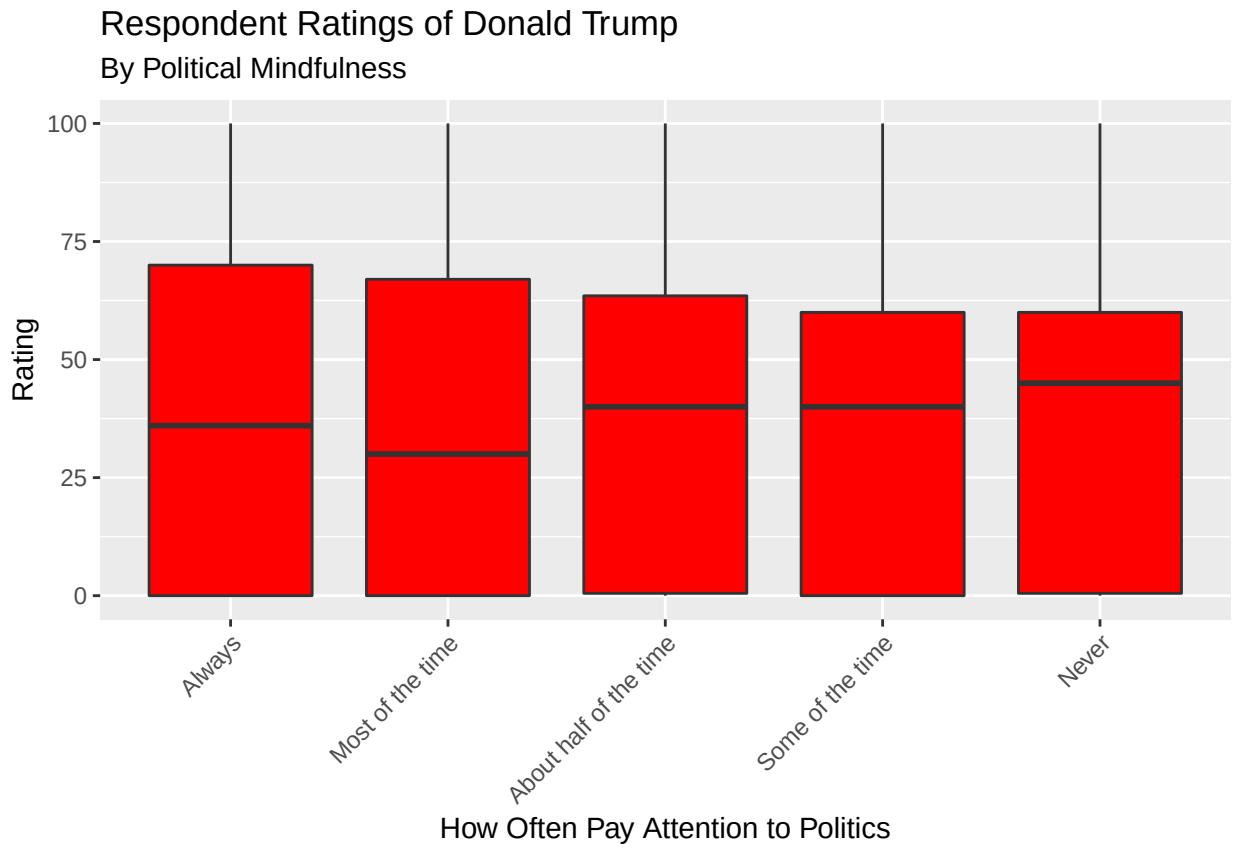
```
##        ID             surveyMethod         surveyType       sex
##  300001 :   1   Face-To-Face:1175   Only Pre    : 611   Male  : 558
##  300002 :   1   Web         :3034   Pre and Post:3598   Female: 617
##  300003 :   1                                           NA's  :3034
##  300004 :   1
##  300006 :   1
##  300007 :   1
##  (Other):4203
```

```
##                   payAttPol    votePrim    earlyVote      clinton
##  Always                : 852   Yes :1864   Yes : 185   Min.   :  0.00
##  Most of the time      :1484   No  :2340   No  :3433   1st Qu.:  2.00
##  About half of the time: 872   NA's:   5   NA's: 591   Median : 40.00
##  Some of the time      : 923                           Mean   : 42.08
##  Never                 :  78                           3rd Qu.: 70.00
##                                                        Max.   :100.00
##
##      trump        bothCandidates
##  Min.   :  0.00   Min.   :-100.000
##  1st Qu.:  0.00   1st Qu.: -65.000
##  Median : 30.00   Median :  -9.000
##  Mean   : 36.98   Mean   :  -5.104
##  3rd Qu.: 70.00   3rd Qu.:  55.000
##  Max.   :100.00   Max.   : 100.000
##
```

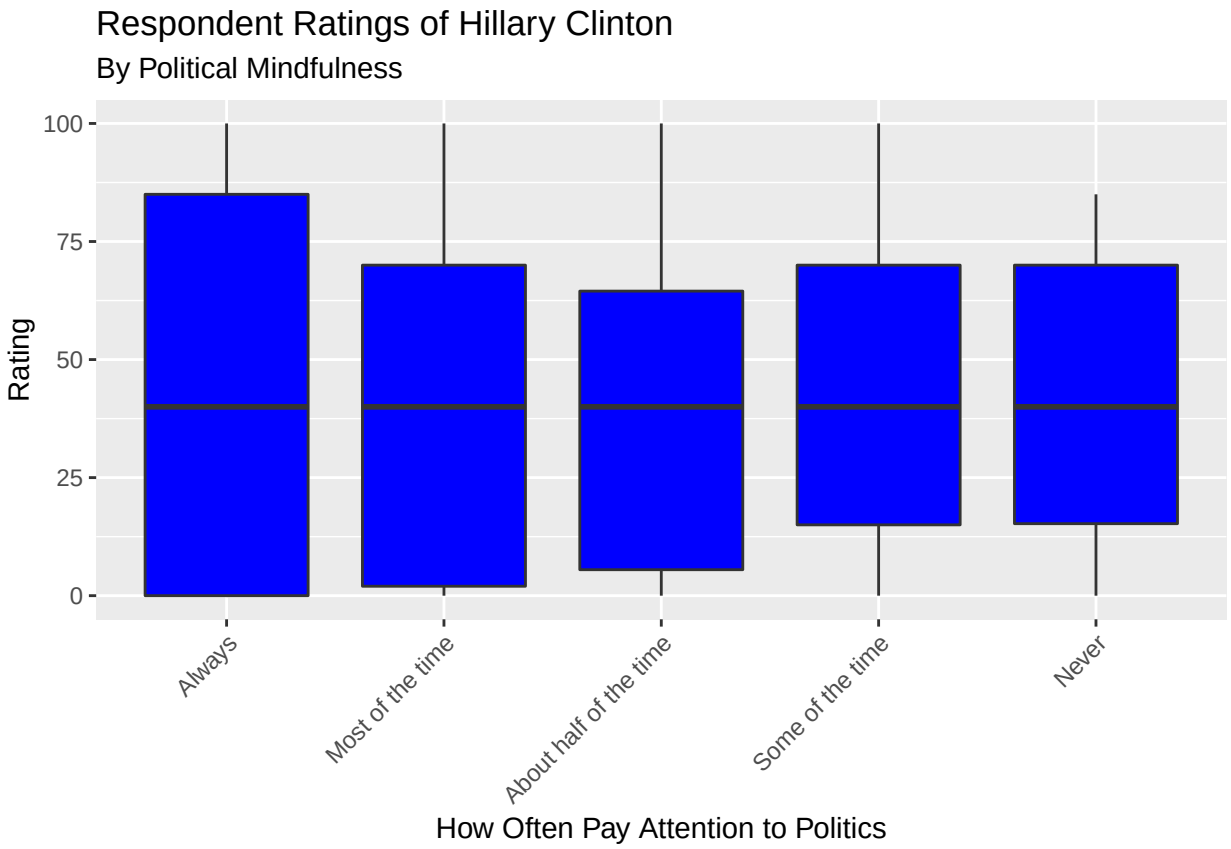
```r
# Tidy variables for some summary statistics
ratings <- survey %>% gather(key = "candidate", value = "rating",
                             clinton, trump, bothCandidates) %>%
  filter(earlyVote == "No", votePrim == "No") %>%
  select(payAttPol, candidate, rating)

# Box plot of individual candidate ratings
ratings %>% filter(!is.na(rating)) %>%
  ggplot(aes(x=candidate, y=rating, fill=candidate)) +
  geom_boxplot(fill=c("purple","blue","red")) +
  ggtitle("Respondent Ratings of Candidates") +
  xlab("Candidate") +
  ylab("Rating") +
  scale_x_discrete(labels=c("Both Candidates","Hillary Clinton","Donald Trump"))
```




```
# Box plot of Trump ratings by political attentiveness
ratings %>% filter(candidate=="trump", !is.na(rating)) %>%
  ggplot(aes(x=payAttPol, y=rating)) +
  geom_boxplot(fill="red") +
  ggtitle("Respondent Ratings of Donald Trump","By Political Mindfulness") +
  xlab("How Often Pay Attention to Politics") +
  ylab("Rating") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# Box plot of Clinton ratings by political attentiveness
ratings %>% filter(candidate=="clinton", !is.na(rating)) %>%
  ggplot(aes(x=payAttPol, y=rating)) +
  geom_boxplot(fill="blue") +
  ggtitle("Respondent Ratings of Hillary Clinton","By Political Mindfulness") +
  xlab("How Often Pay Attention to Politics") +
  ylab("Rating") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Respondent Ratings of Hillary Clinton
By Political Mindfulness
Rating
100
75
50
25
0
Always
Most of the time
About half of the time
Some of the time
Never
How Often Pay Attention to Politics