The affection for Trump

Weizhi Guo

2020/12/16

Estimating the factors that influence voting for Trump among the residence of the US

Name(s) of Author(s)

Date

#Abstract

The main objective of the study is to come up with parameter estimates for the appropriate regression model of factors that the influence the likelihood of one voting for Trump. The data set used has a total of 6067 observations and 18 variables. The variables are a mix of continous and categorical variables. Here we are interested in predicting the popular vote outcome of the 2020 American federal election Singh et al (2017). To do this we are employing a post-stratification technique. In the following sub-sections I will describe the model specifics and the post-stratification calculation. The data set used for the study is survey responses about individuals' demographic variables, opinion on political ideologies, whether an individual would particiapte in a general election as well employment status and whether an individual would vote for Trump.

```
# Loading in the cleaned survey Data
survey_data <- read.csv("/cloud/project/data/raw data.csv");head(survey_data)
```

```
##            interest   registration                    vote_2016
## 1  Some of the time     Registered                  Donald Trump
## 2  Some of the time     Registered Did not vote, but was eligible
## 3  Some of the time     Registered                  Donald Trump
## 4  Most of the time     Registered                  Donald Trump
## 5  Most of the time     Registered                  Donald Trump
## 6 Only now and then Not registered       Was not eligible to vote
##                   vote_intention               vote_2020            ideo5
## 1              Yes, I will vote             Donald Trump      Conservative
## 2              Yes, I will vote I am not sure/don't know      Conservative
## 3              Yes, I will vote             Donald Trump      Conservative
## 4              Yes, I will vote             Donald Trump      Conservative
## 5              Yes, I will vote             Donald Trump Very Conservative
## 6 No, I am not eligible to vote         I would not vote           Liberal
##                             employment      foreign_born gender census_region
## 1                  Full-time employed The United States Female       Midwest
## 2                  Full-time employed The United States Female         South
## 3                  Full-time employed The United States Female         South
## 4 Unemployed or temporarily on layoff The United States Female         South
## 5                             Retired The United States Female          West
```

```
## 6 Unemployed or temporarily on layoff The United States Female       Midwest
##       hispanic race_ethnicity     household_income
## 1 Not Hispanic          White   $75,000 to $79,999
## 2 Not Hispanic          White $100,000 to $124,999
## 3 Not Hispanic          White $175,000 to $199,999
## 4 Not Hispanic          White   $65,000 to $69,999
## 5 Not Hispanic          White    Less than $14,999
## 6 Not Hispanic          White    Less than $14,999
##                                   education state congress_district age
## 1                            Associate Degree    WI               WI04  49
## 2        College Degree (such as B.A., B.S.)    VA               VA08  39
## 3        College Degree (such as B.A., B.S.)    VA               VA09  46
## 4                         High school graduate    TX               TX10  75
## 5                         High school graduate    WA               WA05  52
## 6 Other post high school vocational training    OH               OH04  44
##   vote_trump
## 1          1
## 2          0
## 3          1
## 4          1
## 5          1
## 6          0
```

```
survey_data=na.omit(survey_data)
```

#Introduction The study uses survey data obtained using the link: https://www.census.gov/programs-surveys/acs to assess the factors their shape the voting pattern in the US presidential elections. The factors are analyzed as the whether they are demographic such as ethnicity, race or age, ideological among other variables. The data analysis process include exploratory and inferential analyses. The summary statistics are presented in the first part of the study and this is done using measures of spread and central tendencies and frequency tabulation through cross tabs. Survey is a good statistical tool in collection of data from people. The data collected from the survey conducted is analyzed using R-studio and findings presented as percentages in tabular forms. The inferential statistics mainly focuses on the use of the ordinal logistic regression model to assess the association between the ACS data set variables. The data set is considered appropriate for the study since it has the appropriate sample size of over 6067 observations which suffices for obtaining results that may be generalizable to the entire population.

#Methodology ## Model Specification The variables are a mix of continous and categorical variables. Here we are interested in predicting the popular vote outcome of the 2020 American federal election Singh et al (2017). To do this we are employing a post-stratification technique. In the following sub-sections the model specification is described as well as the post-stratification calculation. The data set used for the study is survey responses about individuals' demographic variables, opinion on political ideologies, whether an individual would particiapte in a general election as well employment status and whether an individual would vote for Trump. The binary logistic regression model will be used to model the proportion of voters who will vote for Donald Trump. This is a naive model, the age,foreign_born,gender,interest,registration+vote_2016,vote_2020, vote_intention, which is recorded as a numeric variable, to model the probability of voting for Donald Trump. The logistic regression model is appropriate since the study involves estimating the influence of several variables on the voting pattern which take binary outcomes. The interest will be estimating the odds of voters having trump as their preferred candidate. The general form of the model is represented as;

$$ln\left(\frac{P}{1-P}\right)$$

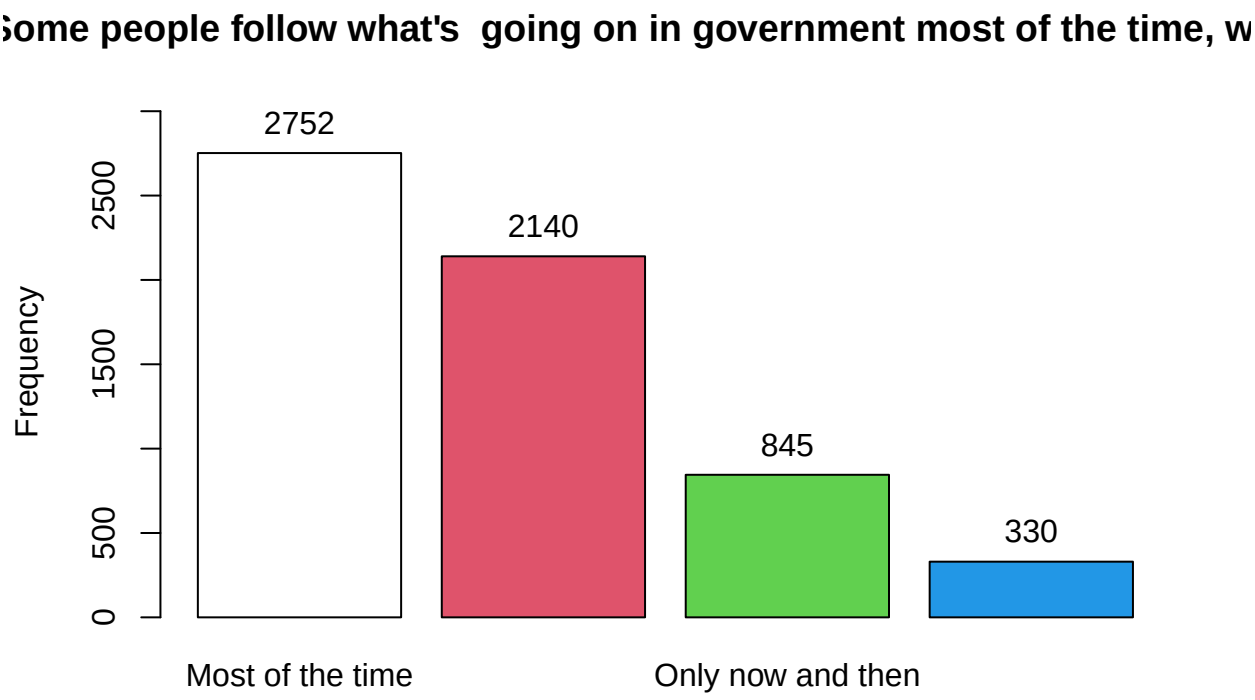
where we model the log odds of the event, where p represents the probability of the event.

$$Z_i = ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 x_1 + .. + \beta_n x_n$$

Where y represents the proportion of voters who will vote for Donald Trump. Similarly, β_0 represents the intercept of the model, and is the probability of voting for Donald Trump at age 0. Additionally, β_1 represents the slope of the model. So, for everyone one unit increase in age, we expect a β_1 increase in the probability of voting for Donald Trump. The above equation can be modeled using the glm() by setting the family argument to "binomial". But we are more interested in the probability of the event, than the log odds of the event. The odds of an events presents the relative risk or tendency of the desired outcome occurring given certain measures or values of the independent variables.The log odds of the event, can be converted to probability of event as follows:

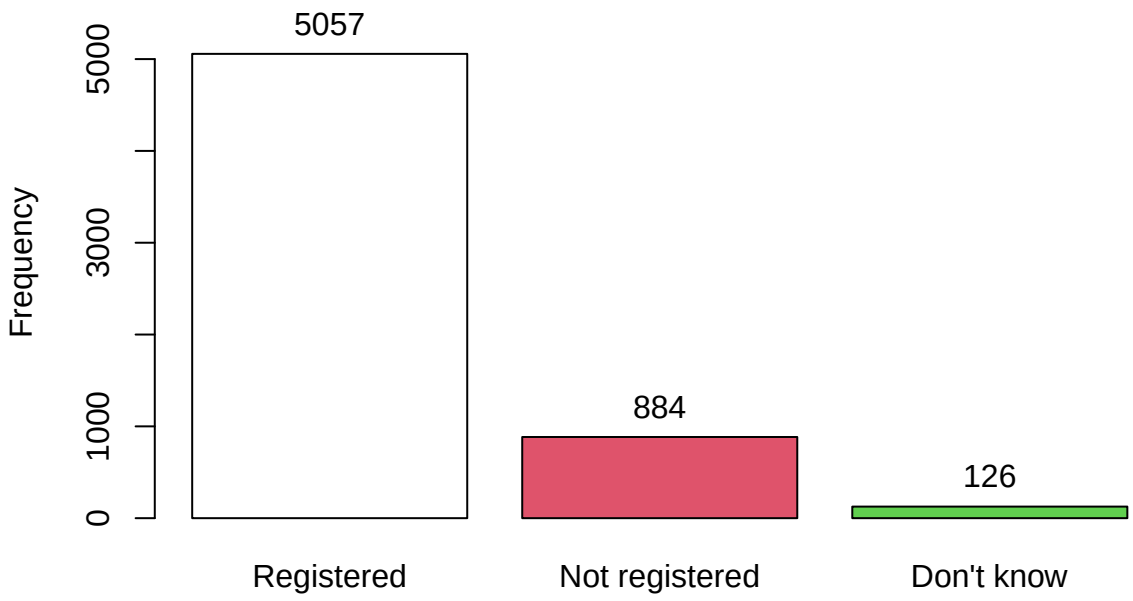
$$P_i = 1 - \left(\frac{1}{1 + e_i^z}\right)$$

```
tab1(survey_data$interest, sort.group = "decreasing", cum.percent = TRUE,main = "Some people follow wha
```



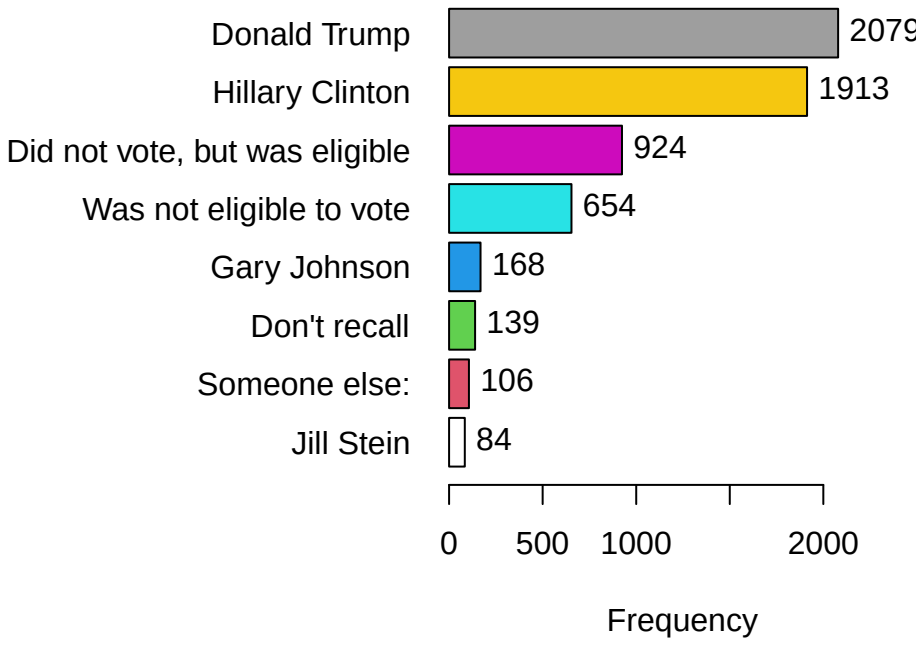

```
## survey_data$interest :
##                   Frequency Percent Cum. percent
## Most of the time       2752    45.4         45.4
## Some of the time       2140    35.3         80.6
## Only now and then       845    13.9         94.6
## Hardly at all           330     5.4        100.0
##   Total                6067   100.0        100.0
```

```
tab1(survey_data$registration, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of reg
```



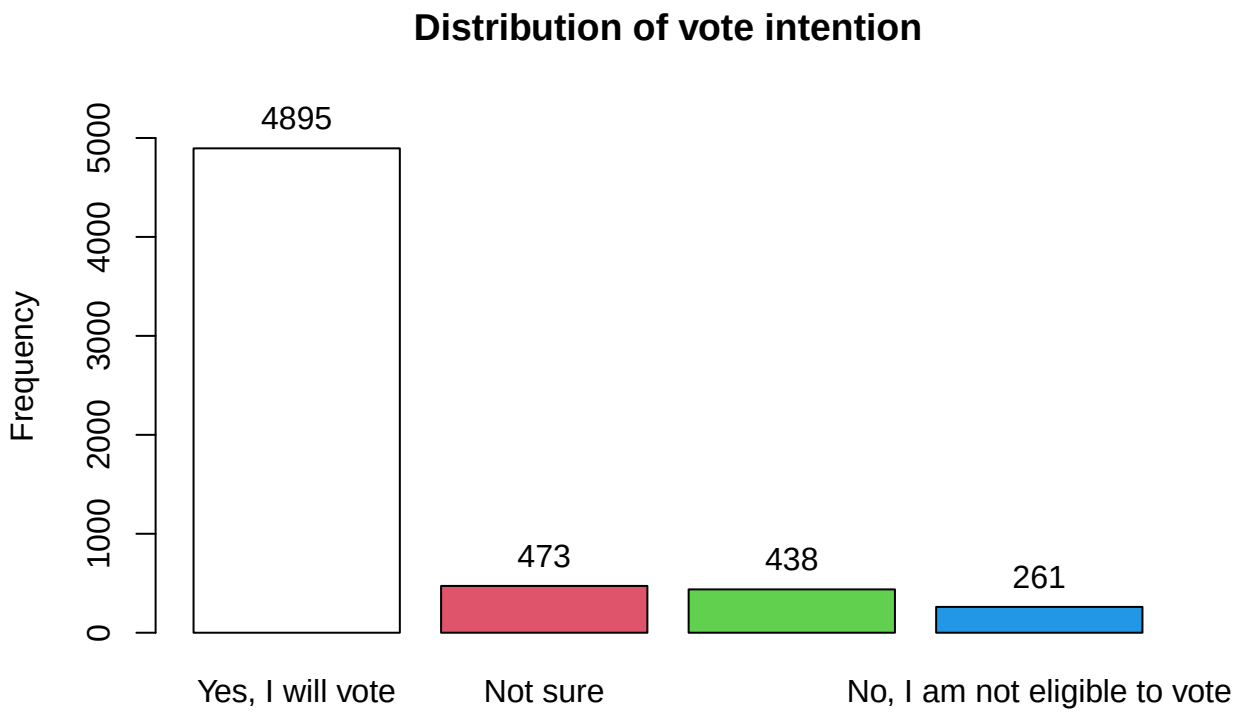

```
## survey_data$registration :
##                Frequency Percent Cum. percent
## Registered          5057    83.4         83.4
## Not registered       884    14.6         97.9
## Don't know           126     2.1        100.0
##   Total             6067   100.0        100.0
```

```
attach(survey_data)

tab1(survey_data$vote_2016, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of 2016 v
```



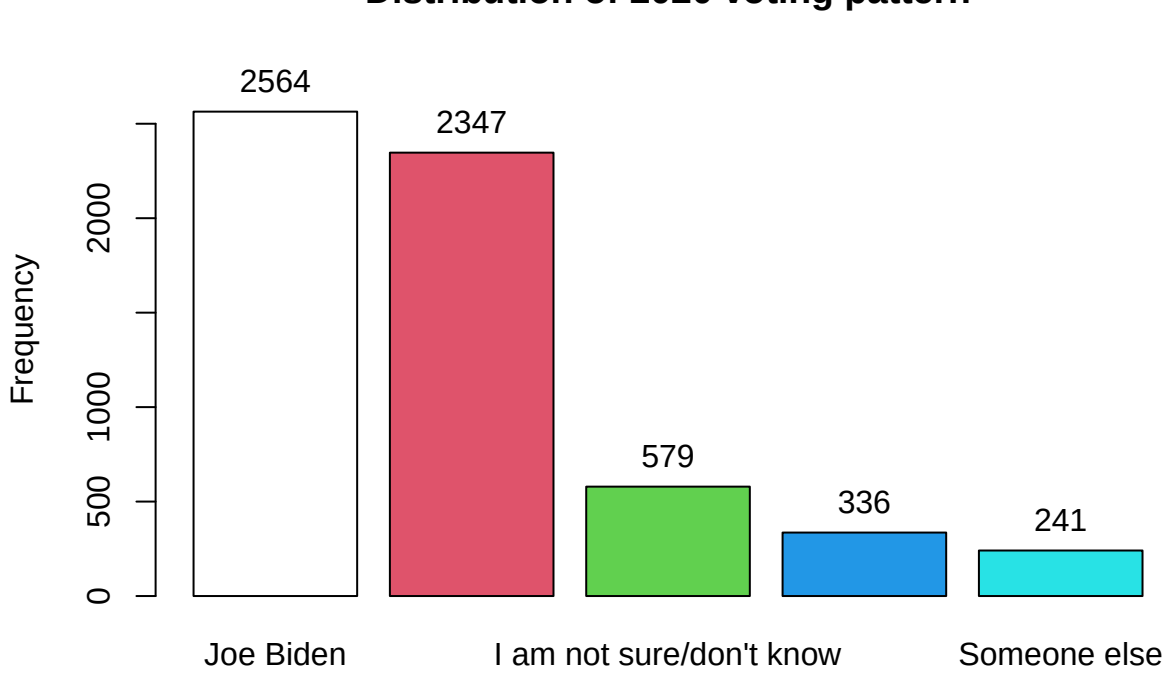

```
## survey_data$vote_2016 :
##                                Frequency Percent Cum. percent
## Donald Trump                        2079    34.3         34.3
## Hillary Clinton                     1913    31.5         65.8
## Did not vote, but was eligible       924    15.2         81.0
## Was not eligible to vote             654    10.8         91.8
## Gary Johnson                         168     2.8         94.6
## Don't recall                         139     2.3         96.9
## Someone else:                        106     1.7         98.6
## Jill Stein                            84     1.4        100.0
##   Total                             6067   100.0        100.0
```

```
tab1(survey_data$vote_intention, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of
```



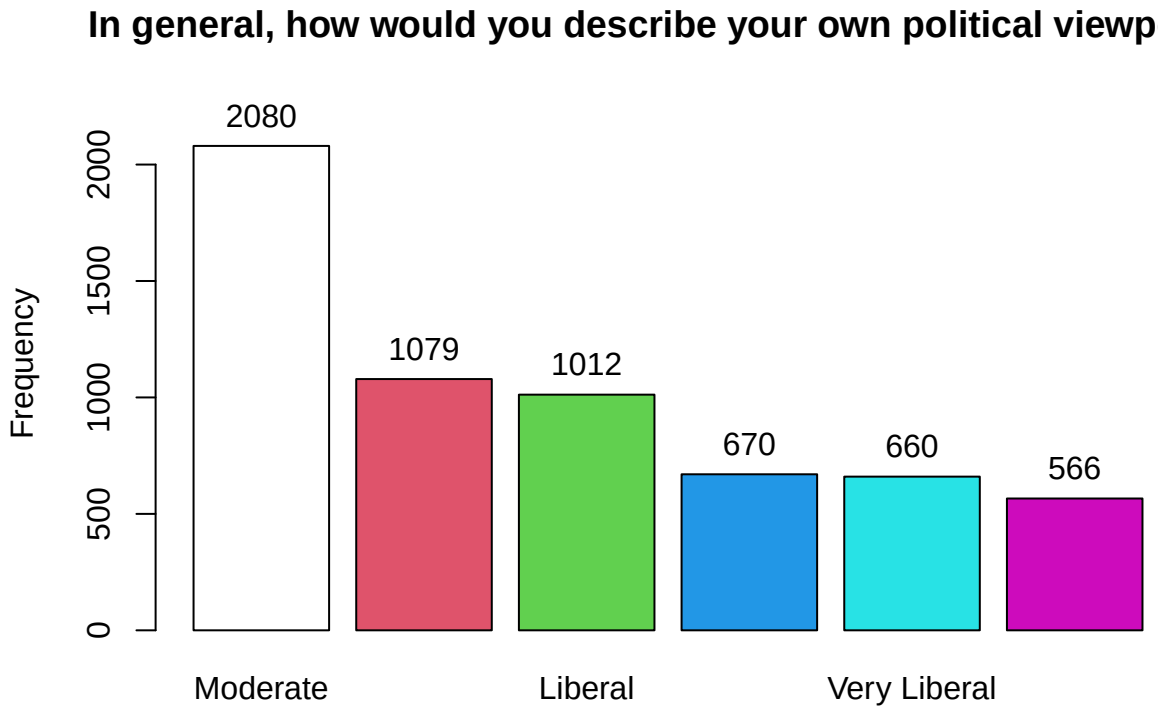

```
## survey_data$vote_intention :
##                                       Frequency Percent Cum. percent
## Yes, I will vote                           4895    80.7         80.7
## Not sure                                    473     7.8         88.5
## No, I will not vote but I am eligible       438     7.2         95.7
## No, I am not eligible to vote               261     4.3        100.0
##   Total                                    6067   100.0        100.0
```

```
tab1(survey_data$vote_2020, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of 2020 v
```



```
## survey_data$vote_2020 :
##                            Frequency Percent Cum. percent
## Joe Biden                       2564    42.3         42.3
## Donald Trump                    2347    38.7         80.9
## I am not sure/don't know         579     9.5         90.5
## I would not vote                 336     5.5         96.0
## Someone else                     241     4.0        100.0
##   Total                         6067   100.0        100.0
```

```
tab1(survey_data$ideo5, sort.group = "decreasing", cum.percent = TRUE,main = "In general, how would you
```

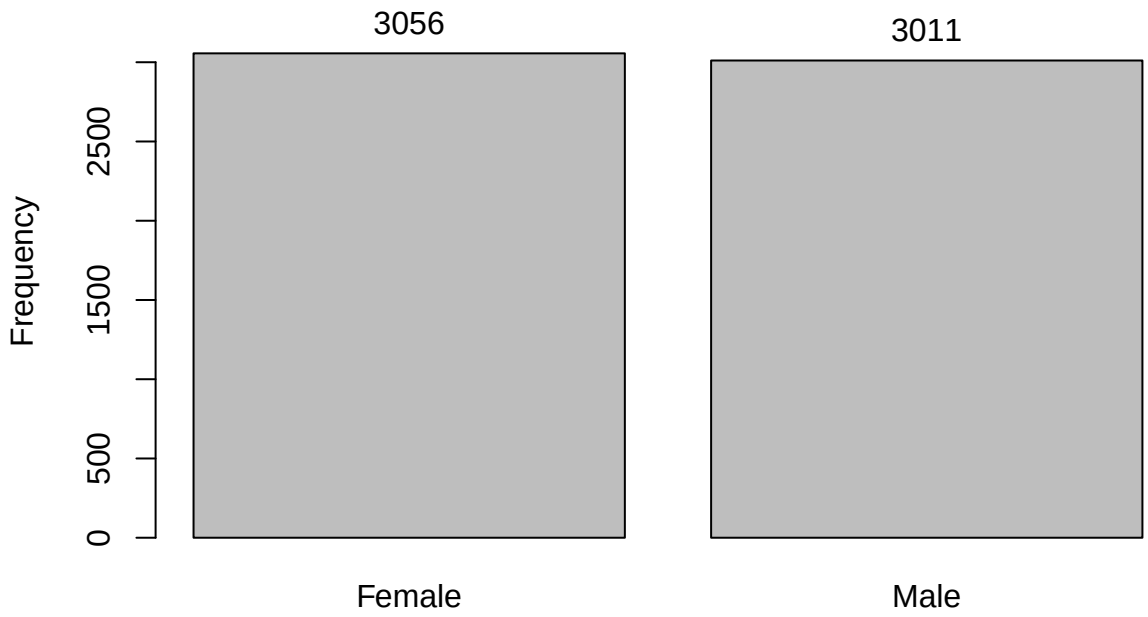



```
## survey_data$ideo5 :
##                      Frequency Percent Cum. percent
## Moderate                  2080    34.3         34.3
## Conservative              1079    17.8         52.1
## Liberal                   1012    16.7         68.7
## Very Conservative          670    11.0         79.8
## Very Liberal               660    10.9         90.7
## Not Sure                   566     9.3        100.0
##   Total                   6067   100.0        100.0
```

```
tab1(survey_data$employment, sort.group = "decreasing", cum.percent = TRUE,main = "Describe your curren
```


```
## survey_data$employment :
##                                    Frequency Percent Cum. percent
## Full-time employed                      2459    40.5         40.5
## Retired                                 1038    17.1         57.6
## Unemployed or temporarily on layoff      620    10.2         67.9
## Part-time employed                       550     9.1         76.9
## Self-employed                            398     6.6         83.5
## Homemaker                                346     5.7         89.2
## Permanently disabled                     310     5.1         94.3
## Student                                  266     4.4         98.7
## Other:                                    80     1.3        100.0
##   Total                                 6067   100.0        100.0
```

```
tab1(survey_data$gender, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of responde
```




```
## survey_data$gender :
##         Frequency Percent Cum. percent
## Female       3056    50.4         50.4
## Male         3011    49.6        100.0
##   Total      6067   100.0        100.0
```

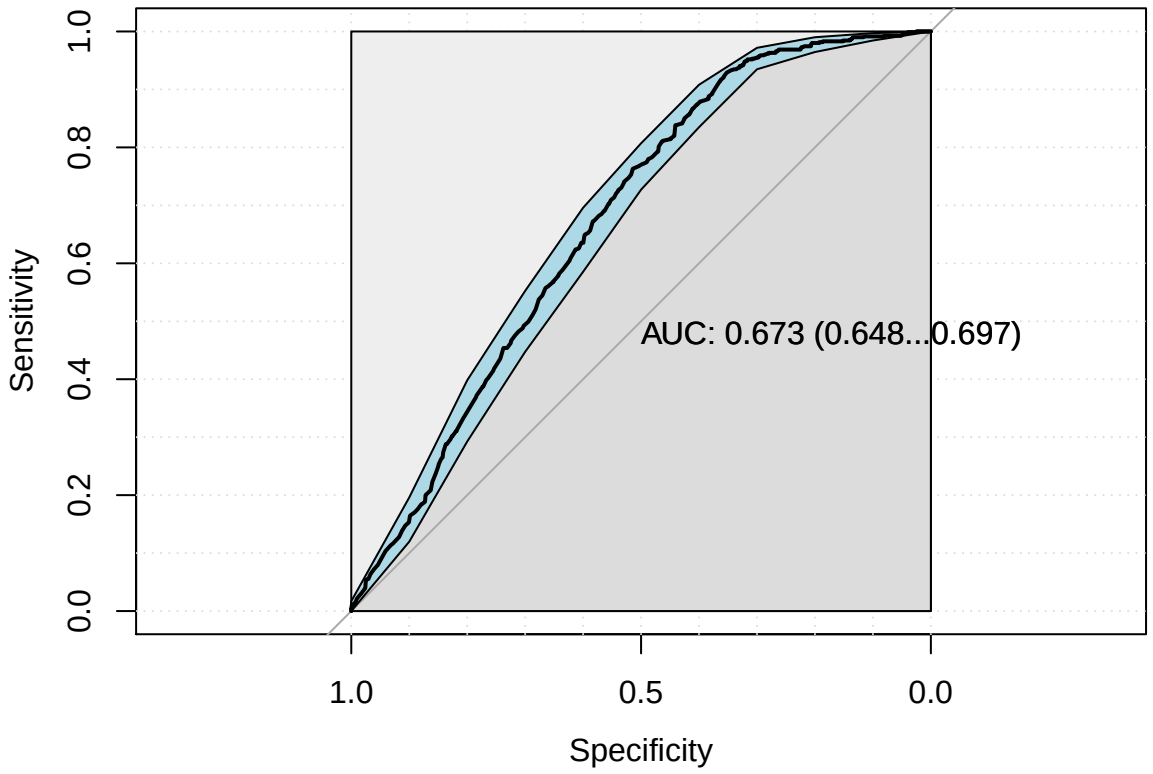
```
# Creating the Model
model <- lm(vote_trump ~ age+gender+ race_ethnicity, data=survey_data);#summary(model)



predicted <- plogis(predict(model, testData))  # predicted scores
# or
predicted <- predict(model, testData, type="response")

pROC_obj=roc(testData$vote_trump, predicted,smoothed = TRUE,
             # arguments for ci
             ci=TRUE, ci.alpha=0.9, stratified=FALSE,
             # arguments for plot
             plot=TRUE, auc.polygon=TRUE, max.auc.polygon=TRUE, grid=TRUE,
             print.auc=TRUE, show.thres=TRUE)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
sens.ci <- ci.se(pROC_obj)
plot(sens.ci, type="shape", col="lightblue")
```

```
## Warning in plot.ci.se(sens.ci, type = "shape", col = "lightblue"): Low
## definition shape.
```




```
# Model Results (to Report in Results section)
# summary(model)
# OR
broom::tidy(model)
```

```
## # A tibble: 17 x 5
##    term                                    estimate std.error statistic  p.value
##    <chr>                                      <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                              0.226    0.0527       4.28  1.89e- 5
##  2 age                                      0.00260  0.000375     6.94  4.33e-12
##  3 genderMale                               0.100    0.0121       8.29  1.37e-16
##  4 race_ethnicityAsian (Asian Indian)      -0.106    0.0697      -1.52  1.28e- 1
##  5 race_ethnicityAsian (Chinese)           -0.222    0.0748      -2.96  3.06e- 3
##  6 race_ethnicityAsian (Filipino)           0.0132   0.0870       0.151 8.80e- 1
##  7 race_ethnicityAsian (Japanese)          -0.145    0.124       -1.18  2.40e- 1
##  8 race_ethnicityAsian (Korean)            -0.0632   0.138       -0.456 6.48e- 1
##  9 race_ethnicityAsian (Other)             -0.129    0.0954      -1.36  1.75e- 1
## 10 race_ethnicityAsian (Vietnamese)        -0.0877   0.144       -0.611 5.41e- 1
## 11 race_ethnicityBlack, or African Americ~ -0.276    0.0533      -5.17  2.39e- 7
## 12 race_ethnicityPacific Islander (Guaman~  0.634    0.468        1.36  1.75e- 1
## 13 race_ethnicityPacific Islander (Native~ -0.0585   0.155       -0.376 7.07e- 1
## 14 race_ethnicityPacific Islander (Other)  -0.350    0.183       -1.91  5.56e- 2
## 15 race_ethnicityPacific Islander (Samoan) -0.279    0.273       -1.02  3.07e- 1
## 16 race_ethnicitySome other race           -0.115    0.0554      -2.08  3.77e- 2
## 17 race_ethnicityWhite                      0.0543   0.0511       1.06  2.88e- 1
```

```
tibble::tibble(m1)
```

```
## # A tibble: 17 x 1
##       m1
```

```
##    <dbl>
##  1 1.25
##  2 1.00
##  3 1.11
##  4 0.899
##  5 0.801
##  6 1.01
##  7 0.865
##  8 0.939
##  9 0.879
## 10 0.916
## 11 0.759
## 12 1.89
## 13 0.943
## 14 0.705
## 15 0.756
## 16 0.891
## 17 1.06
```

Results

In order to estimate the proportion of voters who will vote for Donald Trump, the post-stratification analysis is performed. Here celss are created based on different ages. Using the model described in the previous sub-section, an estimate of the proportion of voters in each age bin is obtained. From the findings above most of the people of the united states are not considering to vote for Donald Trump in the 2020 general election. Only 33% of the people that participated in the survey are willing to vote for Donald Trump in 2020 general election. 84% of those who voted for Trump in 2016 are considering to vote for him again in the 2020 general election. Of the sample surveyed the white, males, those of age 65 years and above, republican and those with very conservative ideology consider voting for Donald Trump in 2020 general election. At least 30 % of the sample in each census region are willing to vote for Trump in the coming election. 8% of the democrats are also considering voting for trump while 88% of the democrats would not be voting for him. The Black race are not considering voting for trump. This is also evident in the youths who are aged 18-29 years; only 22 % of the sample showed interest in voting for Trump. 42 % of those who earn income of above 100k are willing to vote in trump in the 2020 general election whereas those of liberal ideology are not considering voting for trump, only 9% show an interest in him. Even before fitting the model, it was clear from the frequency tabulation that most of the individuals would not vote for trump, up to 61.3%(3720) stated that they were against Trump's bid. From the sample, only 38.7%(2347) of the indicated they would vote for Trump. The results of the model indicated that the age of individuals, intention to vote a was significant in explaining the election outcome. As the age of an individual increases, the likelihood of that individual voting for trump decreases, this is shown by the negative age coefficient estimate.

Discussion

The survey intended to establish how favorable is Donald Trump in the US. The survey sample findings show that 21% of the sample population consider Trump to be very favorable while 42% consider him very unfavorable, 6 % haven't heard enough about him. The 21% that consider him very favorable are those with very conservative ideology (63%), those who voted for him in the 2016 general elections, and the republicans. Those who consider Trump to be very unfavorable are those with liberal ideology, those who voted for Clinton and Jill in the 2016 general elections, the blacks and the Hispanic, the female some whites. 11% of the blacks haven't heard enough about Trump. Generally, Trump is considered unfavorable as can be inferred from the findings.

#Conclusion

The study was carried out with the objective of determining the factors that may influence the election outcome for Trump in the general election. using both the descriptive statisitics and the inferantial analysis, the following deduction can be made; Majority of the people of the united states are not considering voting for Donald Trump in the 2020 general election. Just about 33% of the participants would actually vote for Donald Trump in 2020 general election. The loyalty for Trump has declined among those who voted for him in 2016 to now 84% who stated they would still support him in the 2020 general election. The whites, males, those of age 65 years and above, republican and those with very conservative ideology consider voting for Donald Trump in 2020 general election. At least 30 % of the sample in each census region are willing to vote for Trump in the coming election. Trump recieves the least support from democrats, with only 8% of them willing to vote fro him. Trump has little favour among the Black race who are not considering supporting his bid. This is also evident among the youths who are aged 18-29 years; only 22 % of the sample showed interest in voting for Trump. 42 % of those who earn income of above 100k are willing to vote in trump in the 2020 general election whereas those of liberal ideology would not vote for trump, only 9% show an interest in him. Even before fitting the model, it was clear from the frequency tabulation that most of the individuals would not vote for trump, up to 61.3%(3720) stated that they were against Trump's bid. From the sample, only 38.7%(2347) of the indicated they would vote for Trump. The results of the model indicated that the age of individuals, intention to vote a was significant in explaining the election outcome. As the age of an individual increases, the likelihood of that individual voting for trump decreases, this is shown by the negative age coefficient estimate.

Weaknesses

in the process of conducting the analysis, it was noted that the analysis was highly impacted by presence of inconsistent observations such as missing values. A significant effort was undertaken trying to format the data in a manner would make it workable. Future procedure in data collection should be more rigorous to limit the chances of errors and inconsistencies in the data.

Next Steps

Subsequent works related to the study should consider inclusion of more variables in the model. it would also help using other classification techniques such as the random forest model and the artificial neural network models and compare their performance with the linear regression models.

References

Singh, P., Sawhney, R. S., & Kahlon, K. S. (2017, November). Forecasting the 2016 US presidential elections using sentiment analysis. In Conference on e-Business, e-Services and e-Society (pp. 412-423). Springer, Cham.

Survey data source; https://www.voterstudygroup.org/publication/nationscape-data-set Acs census data, IPUMS: https://usa.ipums.org/usa/index.shtml