Estimating the result of 2020 US presidential election

Jessica Glustien, Nick Callow, Olivia Bi, Min Zhang

2 November 2020

# Title of your Report

## Name(s) of Author(s)

## Date

# Model

## Model Specifics

#Trump or not trump

library(tidyverse)

## ── Attaching packages ───────────────────────────────────────────────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.2.1 ✓ purrr 0.3.3  
## ✓ tibble 2.1.3 ✓ dplyr 1.0.2  
## ✓ tidyr 1.0.0 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.4.0

## ── Conflicts ──────────────────────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

setwd("/Users/ZHANGMIN/Desktop")  
  
survey\_data <- read\_csv("survey\_data.csv")

## Parsed with column specification:  
## cols(  
## vote\_2016 = col\_character(),  
## vote\_2020 = col\_character(),  
## sex = col\_character(),  
## race = col\_character(),  
## household\_income = col\_character(),  
## education = col\_character(),  
## state = col\_character(),  
## age = col\_double(),  
## vote\_trump = col\_double(),  
## vote\_biden = col\_double(),  
## not\_sure = col\_double(),  
## someone\_else = col\_double()  
## )

census\_data <- read\_csv("census\_data.csv")

## Parsed with column specification:  
## cols(  
## age = col\_double(),  
## sex = col\_character(),  
## race = col\_character(),  
## education = col\_character(),  
## n = col\_double()  
## )

# Creating the Model: Trump or not Trump  
model\_1 <- glm(vote\_trump ~ age+as.factor(sex)+as.factor(race)+as.factor(education),   
 data=survey\_data, family= "binomial")  
summary(model\_1) # age >= 18

##   
## Call:  
## glm(formula = vote\_trump ~ age + as.factor(sex) + as.factor(race) +   
## as.factor(education), family = "binomial", data = survey\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5949 -1.0732 -0.5303 1.1591 2.3846   
##   
## Coefficients:  
## Estimate  
## (Intercept) -0.359671  
## age 0.011735  
## as.factor(sex)male 0.442305  
## as.factor(race)black/african american/negro -1.893814  
## as.factor(race)chinese -1.234244  
## as.factor(race)japanese -1.173824  
## as.factor(race)other asian or pacific islander -0.536201  
## as.factor(race)other race, nec -0.672148  
## as.factor(race)white 0.099697  
## as.factor(education)Associate Degree -0.846607  
## as.factor(education)College Degree (such as B.A., B.S.) -0.720801  
## as.factor(education)Completed some college, but no degree -0.705081  
## as.factor(education)Completed some graduate, but no degree -0.723320  
## as.factor(education)Completed some high school -0.447766  
## as.factor(education)Doctorate degree -0.189855  
## as.factor(education)High school graduate -0.560553  
## as.factor(education)Masters degree -0.562502  
## as.factor(education)Middle School - Grades 4 - 8 -0.779602  
## as.factor(education)Other post high school vocational training -0.367684  
## Std. Error  
## (Intercept) 0.692033  
## age 0.001733  
## as.factor(sex)male 0.055701  
## as.factor(race)black/african american/negro 0.263822  
## as.factor(race)chinese 0.379141  
## as.factor(race)japanese 0.607003  
## as.factor(race)other asian or pacific islander 0.277514  
## as.factor(race)other race, nec 0.257270  
## as.factor(race)white 0.233802  
## as.factor(education)Associate Degree 0.663550  
## as.factor(education)College Degree (such as B.A., B.S.) 0.659566  
## as.factor(education)Completed some college, but no degree 0.659785  
## as.factor(education)Completed some graduate, but no degree 0.671813  
## as.factor(education)Completed some high school 0.663084  
## as.factor(education)Doctorate degree 0.681241  
## as.factor(education)High school graduate 0.660456  
## as.factor(education)Masters degree 0.662505  
## as.factor(education)Middle School - Grades 4 - 8 0.817344  
## as.factor(education)Other post high school vocational training 0.667974  
## z value Pr(>|z|)  
## (Intercept) -0.520 0.60325  
## age 6.770 1.29e-11  
## as.factor(sex)male 7.941 2.01e-15  
## as.factor(race)black/african american/negro -7.178 7.05e-13  
## as.factor(race)chinese -3.255 0.00113  
## as.factor(race)japanese -1.934 0.05314  
## as.factor(race)other asian or pacific islander -1.932 0.05334  
## as.factor(race)other race, nec -2.613 0.00899  
## as.factor(race)white 0.426 0.66980  
## as.factor(education)Associate Degree -1.276 0.20200  
## as.factor(education)College Degree (such as B.A., B.S.) -1.093 0.27446  
## as.factor(education)Completed some college, but no degree -1.069 0.28523  
## as.factor(education)Completed some graduate, but no degree -1.077 0.28163  
## as.factor(education)Completed some high school -0.675 0.49950  
## as.factor(education)Doctorate degree -0.279 0.78048  
## as.factor(education)High school graduate -0.849 0.39603  
## as.factor(education)Masters degree -0.849 0.39585  
## as.factor(education)Middle School - Grades 4 - 8 -0.954 0.34017  
## as.factor(education)Other post high school vocational training -0.550 0.58201  
##   
## (Intercept)   
## age \*\*\*  
## as.factor(sex)male \*\*\*  
## as.factor(race)black/african american/negro \*\*\*  
## as.factor(race)chinese \*\*   
## as.factor(race)japanese .   
## as.factor(race)other asian or pacific islander .   
## as.factor(race)other race, nec \*\*   
## as.factor(race)white   
## as.factor(education)Associate Degree   
## as.factor(education)College Degree (such as B.A., B.S.)   
## as.factor(education)Completed some college, but no degree   
## as.factor(education)Completed some graduate, but no degree   
## as.factor(education)Completed some high school   
## as.factor(education)Doctorate degree   
## as.factor(education)High school graduate   
## as.factor(education)Masters degree   
## as.factor(education)Middle School - Grades 4 - 8   
## as.factor(education)Other post high school vocational training   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 8243.9 on 6100 degrees of freedom  
## Residual deviance: 7608.5 on 6082 degrees of freedom  
## AIC: 7646.5  
##   
## Number of Fisher Scoring iterations: 4

## Post-Stratification

In order to estimate the proportion of voters who will vote for Donald Trump I need to perform a post-stratification analysis. Here I create cells based off different ages, race, eduction level. Using the model described in the previous sub-section I will estimate the proportion of voters in each age bin. I will then weight each proportion estimate (within each bin) by the respective population size of that bin and sum those values and divide that by the entire population size.

# Here I will perform the post-stratification calculation  
census\_data$logodds\_estimate <-  
 model\_1 %>%  
 predict(newdata = census\_data)  
  
census\_data$estimate <-  
 exp(census\_data$logodds\_estimate)/(1+exp(census\_data$logodds\_estimate))  
  
census\_data %>%  
 mutate(trump\_predict\_prop = estimate\*n) %>%   
 summarise(trump\_predict = sum(trump\_predict\_prop)/sum(n))

## # A tibble: 1 x 1  
## trump\_predict  
## <dbl>  
## 1 0.423

#biden or not biden

library(tidyverse)  
setwd("/Users/ZHANGMIN/Desktop")  
survey\_data <- read\_csv("survey\_data.csv")

## Parsed with column specification:  
## cols(  
## vote\_2016 = col\_character(),  
## vote\_2020 = col\_character(),  
## sex = col\_character(),  
## race = col\_character(),  
## household\_income = col\_character(),  
## education = col\_character(),  
## state = col\_character(),  
## age = col\_double(),  
## vote\_trump = col\_double(),  
## vote\_biden = col\_double(),  
## not\_sure = col\_double(),  
## someone\_else = col\_double()  
## )

census\_data <- read\_csv("census\_data.csv")

## Parsed with column specification:  
## cols(  
## age = col\_double(),  
## sex = col\_character(),  
## race = col\_character(),  
## education = col\_character(),  
## n = col\_double()  
## )

# Creating the Model: Biden or not Biden  
model\_2 <- glm(vote\_biden ~ age+as.factor(sex)+as.factor(race)+as.factor(education),   
 data=survey\_data, family= "binomial")  
summary(model\_2) # age >= 18

##   
## Call:  
## glm(formula = vote\_biden ~ age + as.factor(sex) + as.factor(race) +   
## as.factor(education), family = "binomial", data = survey\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8201 -1.0123 -0.8427 1.2306 1.6805   
##   
## Coefficients:  
## Estimate  
## (Intercept) -0.253113  
## age -0.002589  
## as.factor(sex)male -0.336820  
## as.factor(race)black/african american/negro 1.727758  
## as.factor(race)chinese 1.190857  
## as.factor(race)japanese 1.491502  
## as.factor(race)other asian or pacific islander 0.741609  
## as.factor(race)other race, nec 0.761511  
## as.factor(race)white 0.259811  
## as.factor(education)Associate Degree 0.024439  
## as.factor(education)College Degree (such as B.A., B.S.) 0.019878  
## as.factor(education)Completed some college, but no degree -0.183820  
## as.factor(education)Completed some graduate, but no degree -0.063433  
## as.factor(education)Completed some high school -0.470429  
## as.factor(education)Doctorate degree -0.367980  
## as.factor(education)High school graduate -0.473198  
## as.factor(education)Masters degree 0.039954  
## as.factor(education)Middle School - Grades 4 - 8 -0.096572  
## as.factor(education)Other post high school vocational training -0.322146  
## Std. Error  
## (Intercept) 0.687049  
## age 0.001690  
## as.factor(sex)male 0.054378  
## as.factor(race)black/african american/negro 0.256378  
## as.factor(race)chinese 0.342880  
## as.factor(race)japanese 0.544385  
## as.factor(race)other asian or pacific islander 0.278683  
## as.factor(race)other race, nec 0.261167  
## as.factor(race)white 0.244191  
## as.factor(education)Associate Degree 0.654559  
## as.factor(education)College Degree (such as B.A., B.S.) 0.650909  
## as.factor(education)Completed some college, but no degree 0.651104  
## as.factor(education)Completed some graduate, but no degree 0.662700  
## as.factor(education)Completed some high school 0.654489  
## as.factor(education)Doctorate degree 0.673086  
## as.factor(education)High school graduate 0.651954  
## as.factor(education)Masters degree 0.653956  
## as.factor(education)Middle School - Grades 4 - 8 0.788178  
## as.factor(education)Other post high school vocational training 0.659524  
## z value Pr(>|z|)  
## (Intercept) -0.368 0.712570  
## age -1.532 0.125501  
## as.factor(sex)male -6.194 5.86e-10  
## as.factor(race)black/african american/negro 6.739 1.59e-11  
## as.factor(race)chinese 3.473 0.000514  
## as.factor(race)japanese 2.740 0.006148  
## as.factor(race)other asian or pacific islander 2.661 0.007788  
## as.factor(race)other race, nec 2.916 0.003548  
## as.factor(race)white 1.064 0.287343  
## as.factor(education)Associate Degree 0.037 0.970217  
## as.factor(education)College Degree (such as B.A., B.S.) 0.031 0.975637  
## as.factor(education)Completed some college, but no degree -0.282 0.777698  
## as.factor(education)Completed some graduate, but no degree -0.096 0.923744  
## as.factor(education)Completed some high school -0.719 0.472281  
## as.factor(education)Doctorate degree -0.547 0.584581  
## as.factor(education)High school graduate -0.726 0.467952  
## as.factor(education)Masters degree 0.061 0.951283  
## as.factor(education)Middle School - Grades 4 - 8 -0.123 0.902483  
## as.factor(education)Other post high school vocational training -0.488 0.625229  
##   
## (Intercept)   
## age   
## as.factor(sex)male \*\*\*  
## as.factor(race)black/african american/negro \*\*\*  
## as.factor(race)chinese \*\*\*  
## as.factor(race)japanese \*\*   
## as.factor(race)other asian or pacific islander \*\*   
## as.factor(race)other race, nec \*\*   
## as.factor(race)white   
## as.factor(education)Associate Degree   
## as.factor(education)College Degree (such as B.A., B.S.)   
## as.factor(education)Completed some college, but no degree   
## as.factor(education)Completed some graduate, but no degree   
## as.factor(education)Completed some high school   
## as.factor(education)Doctorate degree   
## as.factor(education)High school graduate   
## as.factor(education)Masters degree   
## as.factor(education)Middle School - Grades 4 - 8   
## as.factor(education)Other post high school vocational training   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 8385.6 on 6100 degrees of freedom  
## Residual deviance: 7961.7 on 6082 degrees of freedom  
## AIC: 7999.7  
##   
## Number of Fisher Scoring iterations: 4

census\_data$logodds\_est <-  
 model\_2 %>%  
 predict(newdata = census\_data)  
  
census\_data$estimate <-  
 exp(census\_data$logodds\_est)/(1+exp(census\_data$logodds\_est))  
  
census\_data <-  
 census\_data %>%   
 mutate(biden\_predict\_prop = estimate\*n) %>%   
 summarise(biden\_predict = sum(biden\_predict\_prop)/sum(n))  
view(census\_data)

#Trump or Biden:

library(tidyverse)  
setwd("/Users/ZHANGMIN/Desktop")  
survey\_data <- read\_csv("survey\_data\_2.csv") # data only contain voters who will vote for Biden or Trump.

## Parsed with column specification:  
## cols(  
## vote\_2016 = col\_character(),  
## vote\_2020 = col\_character(),  
## sex = col\_character(),  
## race = col\_character(),  
## household\_income = col\_character(),  
## education = col\_character(),  
## state = col\_character(),  
## age = col\_double(),  
## vote\_trump = col\_double(),  
## vote\_biden = col\_double(),  
## not\_sure = col\_double(),  
## someone\_else = col\_double(),  
## biden\_or\_trump = col\_double()  
## )

census\_data <- read\_csv("census\_data.csv")

## Parsed with column specification:  
## cols(  
## age = col\_double(),  
## sex = col\_character(),  
## race = col\_character(),  
## education = col\_character(),  
## n = col\_double()  
## )

# Creating the Model: Trump or Biden  
model\_3 <- glm(biden\_or\_trump ~ age+as.factor(sex)+as.factor(race)+as.factor(education),   
 data=survey\_data, family= "binomial")  
summary(model\_3) # age >= 18

##   
## Call:  
## glm(formula = biden\_or\_trump ~ age + as.factor(sex) + as.factor(race) +   
## as.factor(education), family = "binomial", data = survey\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.326 -1.071 0.434 1.157 1.654   
##   
## Coefficients:  
## Estimate  
## (Intercept) 0.034005  
## age -0.009278  
## as.factor(sex)male -0.450867  
## as.factor(race)black/african american/negro 2.292946  
## as.factor(race)chinese 1.480360  
## as.factor(race)japanese 1.583816  
## as.factor(race)other asian or pacific islander 0.810221  
## as.factor(race)other race, nec 0.923620  
## as.factor(race)white 0.146219  
## as.factor(education)Associate Degree 0.559432  
## as.factor(education)College Degree (such as B.A., B.S.) 0.443056  
## as.factor(education)Completed some college, but no degree 0.331587  
## as.factor(education)Completed some graduate, but no degree 0.455633  
## as.factor(education)Completed some high school 0.006608  
## as.factor(education)Doctorate degree -0.051267  
## as.factor(education)High school graduate 0.078934  
## as.factor(education)Masters degree 0.376423  
## as.factor(education)Middle School - Grades 4 - 8 0.353172  
## as.factor(education)Other post high school vocational training 0.028228  
## Std. Error  
## (Intercept) 0.706687  
## age 0.001854  
## as.factor(sex)male 0.059748  
## as.factor(race)black/african american/negro 0.295399  
## as.factor(race)chinese 0.406271  
## as.factor(race)japanese 0.627031  
## as.factor(race)other asian or pacific islander 0.310784  
## as.factor(race)other race, nec 0.291388  
## as.factor(race)white 0.268223  
## as.factor(education)Associate Degree 0.668679  
## as.factor(education)College Degree (such as B.A., B.S.) 0.664186  
## as.factor(education)Completed some college, but no degree 0.664541  
## as.factor(education)Completed some graduate, but no degree 0.677944  
## as.factor(education)Completed some high school 0.668736  
## as.factor(education)Doctorate degree 0.688887  
## as.factor(education)High school graduate 0.665802  
## as.factor(education)Masters degree 0.667291  
## as.factor(education)Middle School - Grades 4 - 8 0.847986  
## as.factor(education)Other post high school vocational training 0.673983  
## z value Pr(>|z|)  
## (Intercept) 0.048 0.961621  
## age -5.004 5.61e-07  
## as.factor(sex)male -7.546 4.48e-14  
## as.factor(race)black/african american/negro 7.762 8.35e-15  
## as.factor(race)chinese 3.644 0.000269  
## as.factor(race)japanese 2.526 0.011540  
## as.factor(race)other asian or pacific islander 2.607 0.009133  
## as.factor(race)other race, nec 3.170 0.001526  
## as.factor(race)white 0.545 0.585658  
## as.factor(education)Associate Degree 0.837 0.402805  
## as.factor(education)College Degree (such as B.A., B.S.) 0.667 0.504730  
## as.factor(education)Completed some college, but no degree 0.499 0.617800  
## as.factor(education)Completed some graduate, but no degree 0.672 0.501532  
## as.factor(education)Completed some high school 0.010 0.992116  
## as.factor(education)Doctorate degree -0.074 0.940676  
## as.factor(education)High school graduate 0.119 0.905628  
## as.factor(education)Masters degree 0.564 0.572681  
## as.factor(education)Middle School - Grades 4 - 8 0.416 0.677056  
## as.factor(education)Other post high school vocational training 0.042 0.966592  
##   
## (Intercept)   
## age \*\*\*  
## as.factor(sex)male \*\*\*  
## as.factor(race)black/african american/negro \*\*\*  
## as.factor(race)chinese \*\*\*  
## as.factor(race)japanese \*   
## as.factor(race)other asian or pacific islander \*\*   
## as.factor(race)other race, nec \*\*   
## as.factor(race)white   
## as.factor(education)Associate Degree   
## as.factor(education)College Degree (such as B.A., B.S.)   
## as.factor(education)Completed some college, but no degree   
## as.factor(education)Completed some graduate, but no degree   
## as.factor(education)Completed some high school   
## as.factor(education)Doctorate degree   
## as.factor(education)High school graduate   
## as.factor(education)Masters degree   
## as.factor(education)Middle School - Grades 4 - 8   
## as.factor(education)Other post high school vocational training   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 7197.8 on 5199 degrees of freedom  
## Residual deviance: 6590.7 on 5181 degrees of freedom  
## AIC: 6628.7  
##   
## Number of Fisher Scoring iterations: 4

census\_data$logodds\_est <-  
 model\_3 %>%  
 predict(newdata = census\_data)  
  
census\_data$estimate <-  
 exp(census\_data$logodds\_est)/(1+exp(census\_data$logodds\_est))  
  
election\_result <-  
 census\_data %>%   
 mutate(biden\_win = estimate\*n) %>%   
 summarise(biden\_or\_trump = sum(biden\_win)/sum(n))  
view(election\_result)

# Results

# Discussion

Here you will summarize the previous sections and discuss conclusions drawn from the results. Make sure to elaborate and connect your analysis to the goal of the study.

## Weaknesses

Here we discuss weaknesses of the study, data, analysis, etc. You can also discuss areas for improvement.

## Next Steps

Here you discuss subsequent work to be done after this report. This can include next steps in terms of statistical analysis (perhaps there is a more efficient algorithm available, or perhaps there is a caveat in the data that would allow for some new technique). Future steps should also be specified in terms of the study setting (eg. including a follow-up survey on something, or a subsequent study that would complement the conclusions of your report).

# References