Problemset-3

Juntong Lin

11/2/2020

Problemset-3

Juntong Lin

11/2/2020

Model

We are interested in predicting the popular vote outcome of the 2020 American federal election. 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.

Model Specifics

I will use a log regression model to predict the voters who will vote for Donald Trump. The log regression model I am using is:

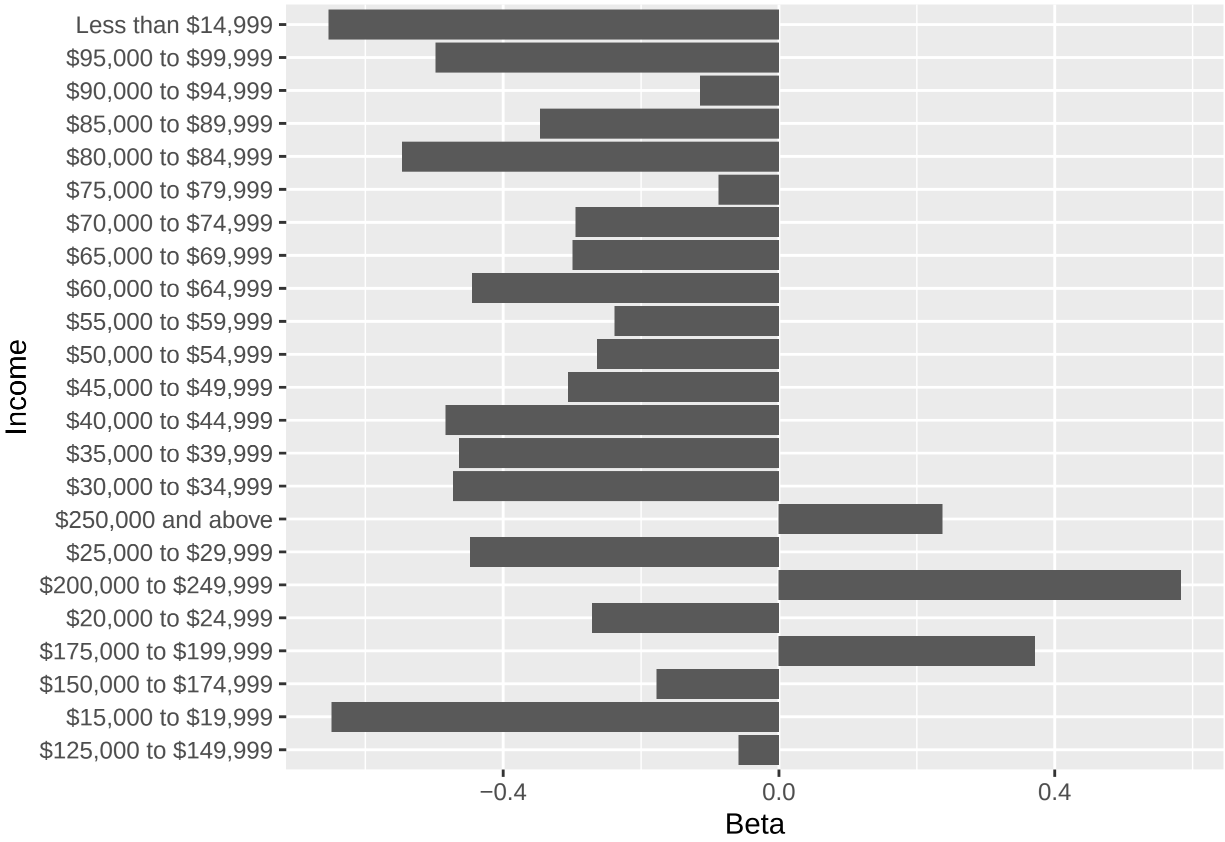
$$log(\frac{p}{1-p}) = \beta_0 + +\beta_1 x_{Age} + \beta_2 x_{employment} + \beta_3 x_{gender} + \beta_4 x_{Race} + \beta_5 x_{HHInc} + \beta_6 x_{Educ} + \beta_7 x_{State} + \epsilon$$

Where y is the proportion of voters who will vote for Donald Trump. There are seven variables in this model: Age, employment, gender, race, HHInc(household income), education and state. Take the age variable as an example, we expected change in log of probability of voting Trump with respect to a one-unit increase in voters' age. Similar to other variables.

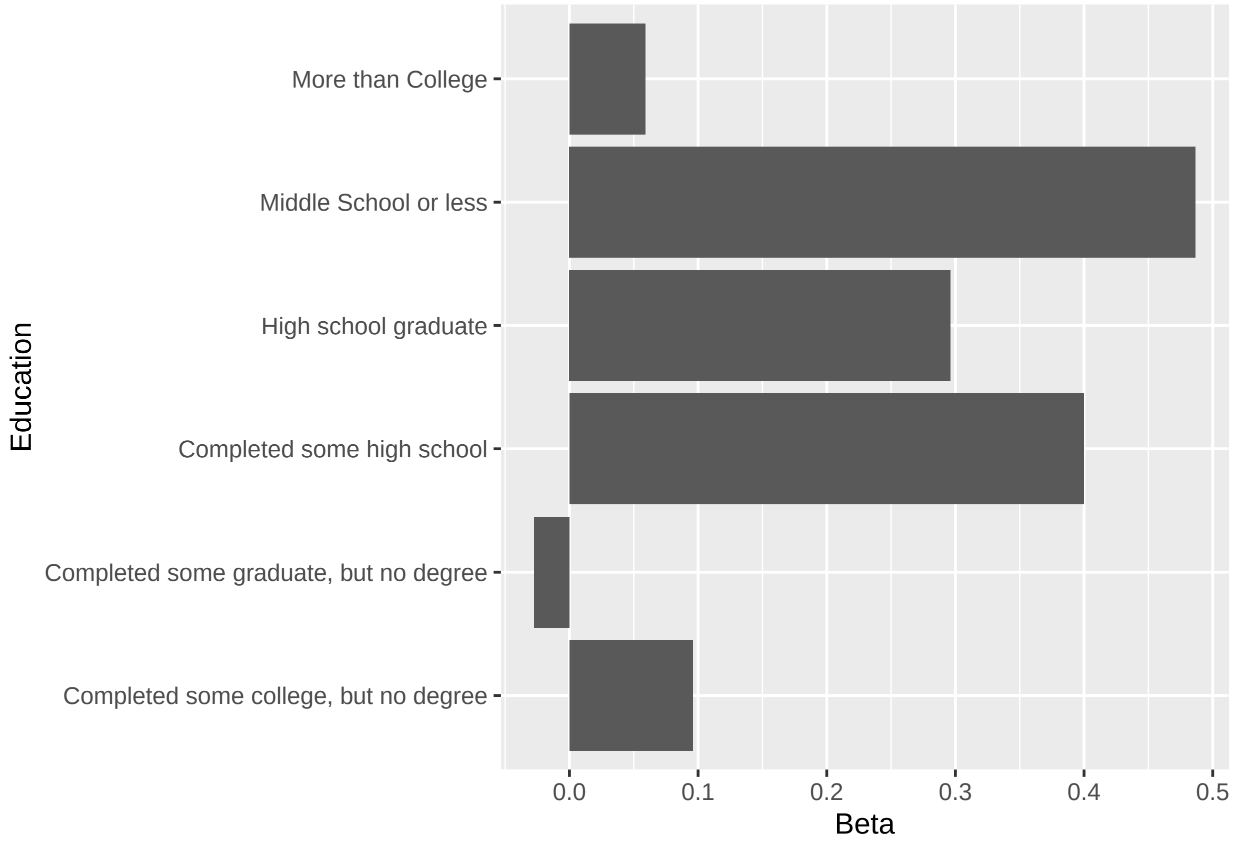
```r
# Creating the Model
model <- glm(vote_trump ~ .,
            data=survey_data, family= "binomial")

# Model Results (to Report in Results section)
summary(model)
```

```
##
## Call:
## glm(formula = vote_trump ~ ., family = "binomial", data = survey_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7604  -0.9991  -0.6000   1.1429   2.6291
##
## Coefficients:
##                                                 Estimate Std. Error z value
## (Intercept)                                     -1.43123    0.75556  -1.894
## age                                              0.01497    0.00190   7.881
## employment                                       0.22825    0.06577   3.470
## gender                                           0.39555    0.05805   6.814
## race                                             1.16405    0.07824  14.877
## household_income$125,000 to $149,999            -0.05849    0.15289  -0.383
## household_income$15,000 to $19,999              -0.64859    0.16479  -3.936
## household_income$150,000 to $174,999            -0.17713    0.18627  -0.951
## household_income$175,000 to $199,999             0.37136    0.22261   1.668
## household_income$20,000 to $24,999              -0.27086    0.16006  -1.692
## household_income$200,000 to $249,999             0.58313    0.20035   2.911
## household_income$25,000 to $29,999              -0.44785    0.15920  -2.813
## household_income$250,000 and above               0.23721    0.20930   1.133
## household_income$30,000 to $34,999              -0.47221    0.15777  -2.993
## household_income$35,000 to $39,999              -0.46377    0.16416  -2.825
## household_income$40,000 to $44,999              -0.48333    0.17216  -2.807
## household_income$45,000 to $49,999              -0.30553    0.16377  -1.866
## household_income$50,000 to $54,999              -0.26383    0.15513  -1.701
## household_income$55,000 to $59,999              -0.23839    0.19299  -1.235
## household_income$60,000 to $64,999              -0.44470    0.19353  -2.298
## household_income$65,000 to $69,999              -0.29893    0.21347  -1.400
## household_income$70,000 to $74,999              -0.29470    0.18713  -1.575
## household_income$75,000 to $79,999              -0.08704    0.18594  -0.468
## household_income$80,000 to $84,999              -0.54630    0.22709  -2.406
## household_income$85,000 to $89,999              -0.34618    0.24474  -1.415
## household_income$90,000 to $94,999              -0.11439    0.26285  -0.435
## household_income$95,000 to $99,999              -0.49768    0.20036  -2.484
## household_incomeLess than $14,999               -0.65286    0.13461  -4.850
## educationCompleted some college, but no degree   0.09593    0.08905   1.077
## educationCompleted some graduate, but no degree -0.02756    0.10005  -0.275
## educationCompleted some high school              0.39984    0.11729   3.409
## educationHigh school graduate                    0.29621    0.09051   3.273
## educationMiddle School or less                   0.48663    0.39235   1.240
## educationMore than College                       0.05899    0.10128   0.582
## stateAL                                         -0.30653    0.77562  -0.395
## stateAR                                         -0.16975    0.80611  -0.211
## stateAZ                                         -0.54530    0.76111  -0.716
## stateCA                                         -0.98061    0.74691  -1.313
## stateCO                                         -0.73529    0.77541  -0.948
## stateCT                                         -1.66923    0.79931  -2.088
## stateDC                                         -1.07189    0.88840  -1.207
## stateDE                                         -1.26381    0.87592  -1.443
## stateFL                                         -0.59413    0.74813  -0.794
## stateGA                                         -0.42490    0.75977  -0.559
## stateHI                                         -0.39464    0.86757  -0.455
## stateIA                                         -0.78868    0.79704  -0.990
## stateID                                         -0.22587    0.83323  -0.271
## stateIL                                         -0.94136    0.75355  -1.249
## stateIN                                         -0.71481    0.76822  -0.930
## stateKS                                         -0.17691    0.80352  -0.220
## stateKY                                         -0.48001    0.77484  -0.620
## stateLA                                         -0.43187    0.77990  -0.554
## stateMA                                         -1.56768    0.77768  -2.016
## stateMD                                         -0.94279    0.77708  -1.213
## stateME                                         -1.02241    0.87516  -1.168
## stateMI                                         -0.89687    0.76022  -1.180
## stateMN                                         -0.56964    0.78761  -0.723
## stateMO                                         -0.77143    0.76697  -1.006
## stateMS                                         -0.35624    0.81109  -0.439
## stateMT                                         -0.66615    0.90455  -0.736
## stateNC                                         -0.58966    0.75713  -0.779
## stateND                                         -0.47574    1.10743  -0.430
## stateNE                                         -0.65222    0.87004  -0.750
## stateNH                                         -0.93627    0.88662  -1.056
## stateNJ                                         -0.89150    0.75823  -1.176
## stateNM                                         -1.44380    0.88570  -1.630
## stateNV                                         -0.50354    0.78443  -0.642
## stateNY                                         -0.88843    0.74881  -1.186
## stateOH                                         -0.82596    0.75279  -1.097
## stateOK                                         -0.36260    0.78749  -0.460
## stateOR                                         -0.95370    0.77684  -1.228
## statePA                                         -0.70157    0.75338  -0.931
## stateRI                                         -1.20567    1.02852  -1.172
## stateSC                                         -0.31232    0.77045  -0.405
## stateSD                                         -0.30896    0.91636  -0.337
## stateTN                                         -0.33193    0.76815  -0.432
## stateTX                                         -0.46295    0.74810  -0.619
## stateUT                                         -0.82900    0.80297  -1.032
## stateVA                                         -0.97960    0.75901  -1.291
## stateVT                                         -2.84237    1.27685  -2.226
## stateWA                                         -0.96739    0.76798  -1.260
## stateWI                                         -1.08347    0.77047  -1.406
## stateWV                                         -0.37658    0.81151  -0.464
## stateWY                                         -2.23640    1.35080  -1.656
##                                                 Pr(>|z|)
## (Intercept)                                     0.058189 .
## age                                             3.24e-15 ***
## employment                                      0.000520 ***
## gender                                          9.49e-12 ***
## race                                             < 2e-16 ***
## household_income$125,000 to $149,999            0.702015
## household_income$15,000 to $19,999              8.29e-05 ***
## household_income$150,000 to $174,999            0.341634
## household_income$175,000 to $199,999            0.095271 .
## household_income$20,000 to $24,999              0.090593 .
## household_income$200,000 to $249,999            0.003607 **
## household_income$25,000 to $29,999              0.004906 **
## household_income$250,000 and above              0.257071
## household_income$30,000 to $34,999              0.002762 **
## household_income$35,000 to $39,999              0.004726 **
## household_income$40,000 to $44,999              0.004995 **
## household_income$45,000 to $49,999              0.062101 .
## household_income$50,000 to $54,999              0.089006 .
## household_income$55,000 to $59,999              0.216726
## household_income$60,000 to $64,999              0.021575 *
## household_income$65,000 to $69,999              0.161407
## household_income$70,000 to $74,999              0.115297
## household_income$75,000 to $79,999              0.639712
## household_income$80,000 to $84,999             0.016141 *  
## household_income$85,000 to $89,999             0.157210    
## household_income$90,000 to $94,999             0.663429    
## household_income$95,000 to $99,999             0.012994 *  
## household_incomeLess than $14,999              1.23e-06 ***
## educationCompleted some college, but no degree  0.281387    
## educationCompleted some graduate, but no degree 0.782977    
## educationCompleted some high school            0.000652 ***
## educationHigh school graduate                  0.001065 ** 
## educationMiddle School or less                 0.214860    
## educationMore than College                     0.560242    
## stateAL                                        0.692694    
## stateAR                                        0.833218    
## stateAZ                                        0.473715    
## stateCA                                        0.189216    
## stateCO                                        0.342996    
## stateCT                                        0.036769 *  
## stateDC                                        0.227607    
## stateDE                                        0.149066    
## stateFL                                        0.427108    
## stateGA                                        0.575991    
## stateHI                                        0.649199    
## stateIA                                        0.322415    
## stateID                                        0.786332    
## stateIL                                        0.211580    
## stateIN                                        0.352126    
## stateKS                                        0.825736    
## stateKY                                        0.535586    
## stateLA                                        0.579751    
## stateMA                                        0.043817 *  
## stateMD                                        0.225035    
## stateME                                        0.242706    
## stateMI                                        0.238101    
## stateMN                                        0.469526    
## stateMO                                        0.314504    
## stateMS                                        0.660509    
## stateMT                                        0.461463    
## stateNC                                        0.436095    
## stateND                                        0.667498    
## stateNE                                        0.453465    
## stateNH                                        0.290973    
## stateNJ                                        0.239687    
## stateNM                                        0.103076    
## stateNV                                        0.520924    
## stateNY                                        0.235442    
## stateOH                                        0.272553    
## stateOK                                        0.645192    
## stateOR                                        0.219569    
## statePA                                        0.351733    
## stateRI                                        0.241103    
## stateSC                                        0.685199    
## stateSD                                        0.735998    
## stateTN                                        0.665654    
## stateTX                                        0.536028    
## stateUT                                         0.301878
## stateVA                                         0.196834
## stateVT                                         0.026009 *
## stateWA                                         0.207796
## stateWI                                         0.159651
## stateWV                                         0.642614
## stateWY                                         0.097800 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8155.0  on 6110  degrees of freedom
## Residual deviance: 7410.3  on 6027  degrees of freedom
## AIC: 7578.3
##
## Number of Fisher Scoring iterations: 4
```

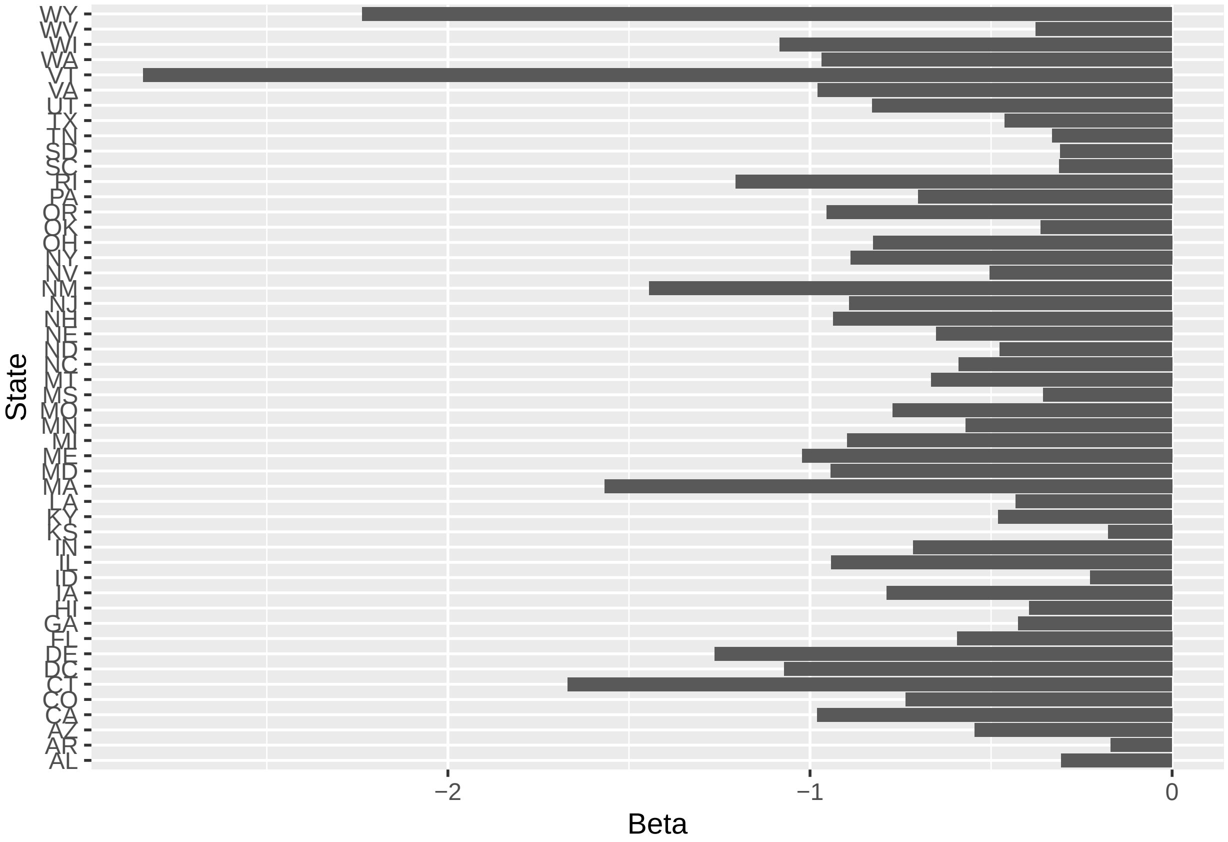
```r
# OR
# broom::tidy(model)

# Check beta for different categories
model$coefficients[6:28] %>%
  t() %>%
  as_tibble() %>%
  pivot_longer(cols = 1:23, names_to = "Income", values_to = "Beta") %>%
  mutate(Income = str_remove(Income, "household_income")) %>%
  ggplot(aes(Income, Beta)) +
  geom_bar(stat = "identity") +
  coord_flip()
```



```
# Check beta for different categories
model$coefficients[29:34] %>%
  t() %>%
  as_tibble() %>%
  pivot_longer(cols = 1:6, names_to = "Education", values_to = "Beta") %>%
  mutate(Education = str_remove(Education, "education")) %>%
  ggplot(aes(Education, Beta)) +
  geom_bar(stat = "identity") +
  coord_flip()
```



```
# Check beta for different categories
model$coefficients[35:84] %>%
  t() %>%
  as_tibble() %>%
  pivot_longer(cols = 1:50, names_to = "State", values_to = "Beta") %>%
  mutate(State = str_remove(State, "state")) %>%
  ggplot(aes(State, Beta)) +
  geom_bar(stat = "identity") +
  coord_flip()
```



In the code above, I create the model and some plots. The plots are calculating the regression coefficient. We can see when controling the other factors, how a certain variable would affect the probability of voting Trump. In the first graph, we can the high income group have positive beta while others are negative. In the second graph, we can see an education level below high school, especially the below middle school group, has large beta. In the third graph, we can see there is a large negative beta for the state VT (Vermont) and WY (Wyoming). The age, gender, employment, and race do not perform a bar plot, but we can find the results in the model summary outcome. When setting the basic age as 18, basic employment as "full time employment", basic gender as male, and basic race as whitle, the beta performs positive.

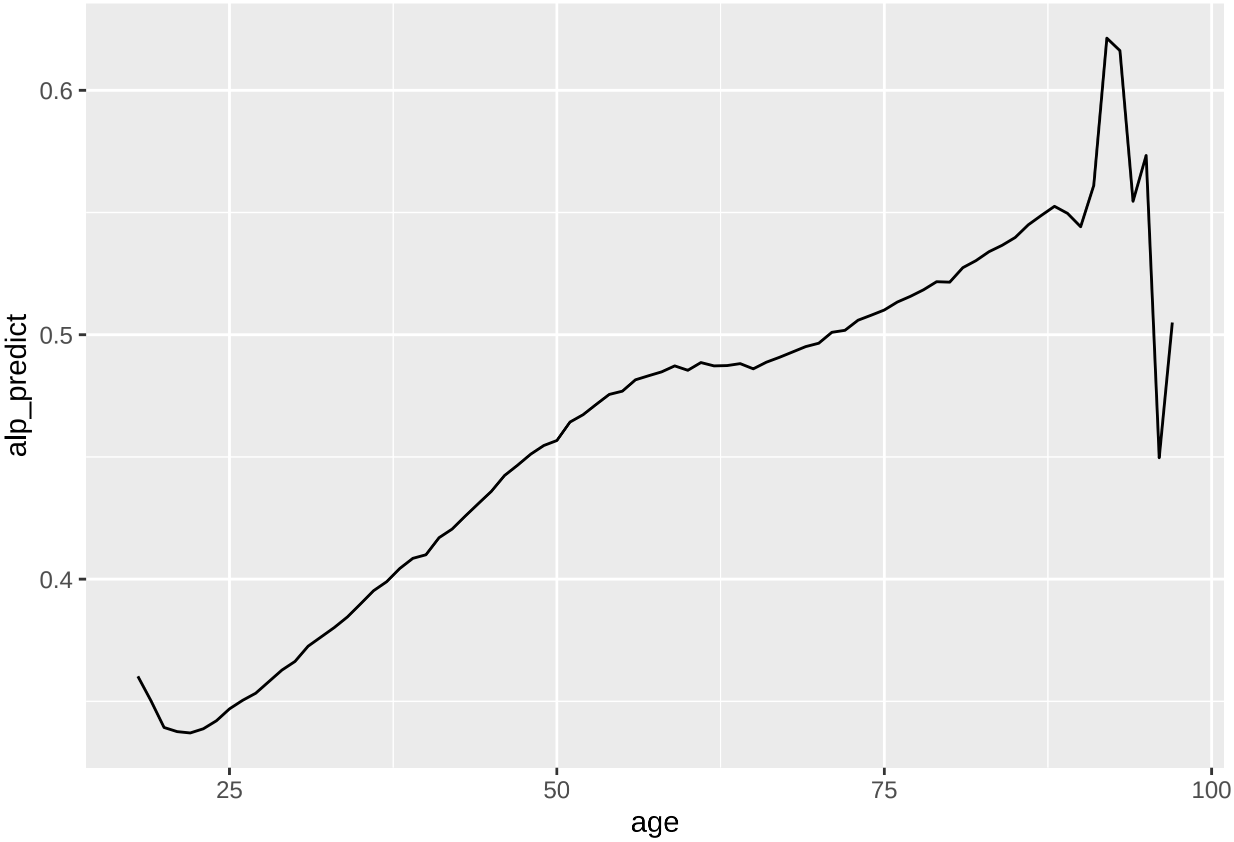
Post-Stratification

In order to estimate the proportion of voters who will vote for Donald Trump I need to perform a post-stratification analysis. I perform both a overall probability for all the data and a seperate probability for each group.

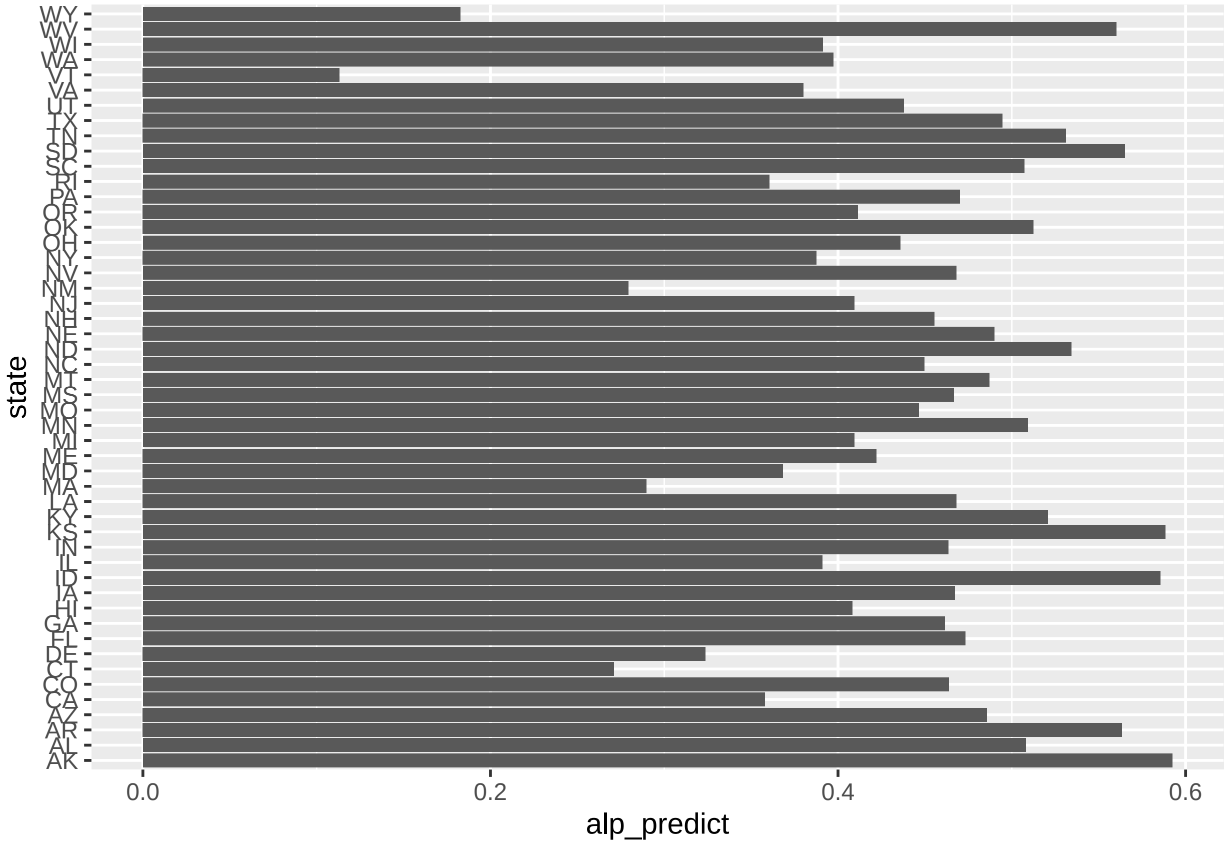
```r
# Here I will perform the post-stratification calculation
census_data$logodds_estimate <-
  model %>%
  predict(newdata = census_data)

census_data$estimate <-
  exp(census_data$logodds_estimate)/(1+exp(census_data$logodds_estimate))

# Total Popular vote (everyone)
census_data %>%
  mutate(alp_predict_prop = estimate*n) %>%
  summarise(alp_predict = sum(alp_predict_prop)/sum(n))
```

```
## # A tibble: 1 x 1
##   alp_predict
##         <dbl>
## 1       0.434
```

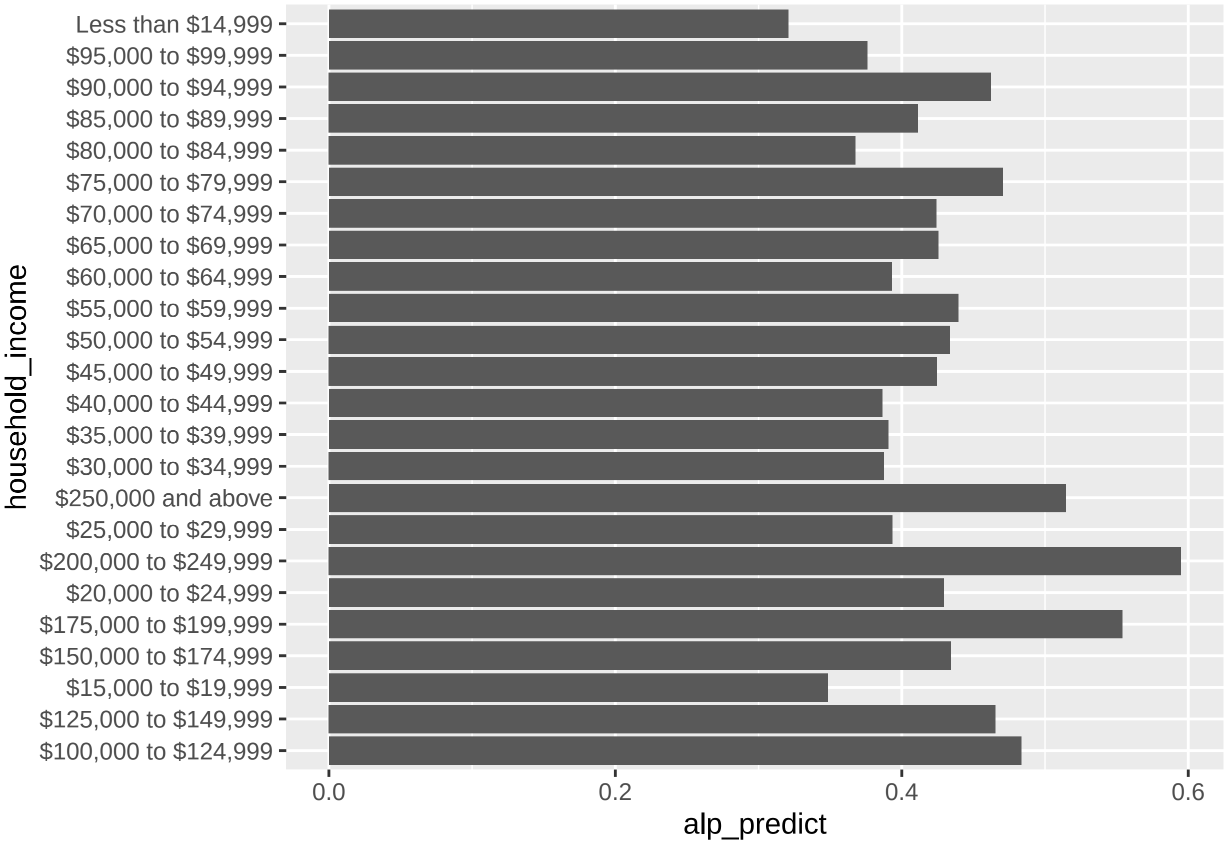
```
# different group age vote for trump odds
census_data %>%
  mutate(alp_predict_prop = estimate*n) %>%
  group_by(age) %>%
  summarise(alp_predict = sum(alp_predict_prop)/sum(n)) %>%
  ggplot(aes(age, alp_predict)) +
  geom_line()
```



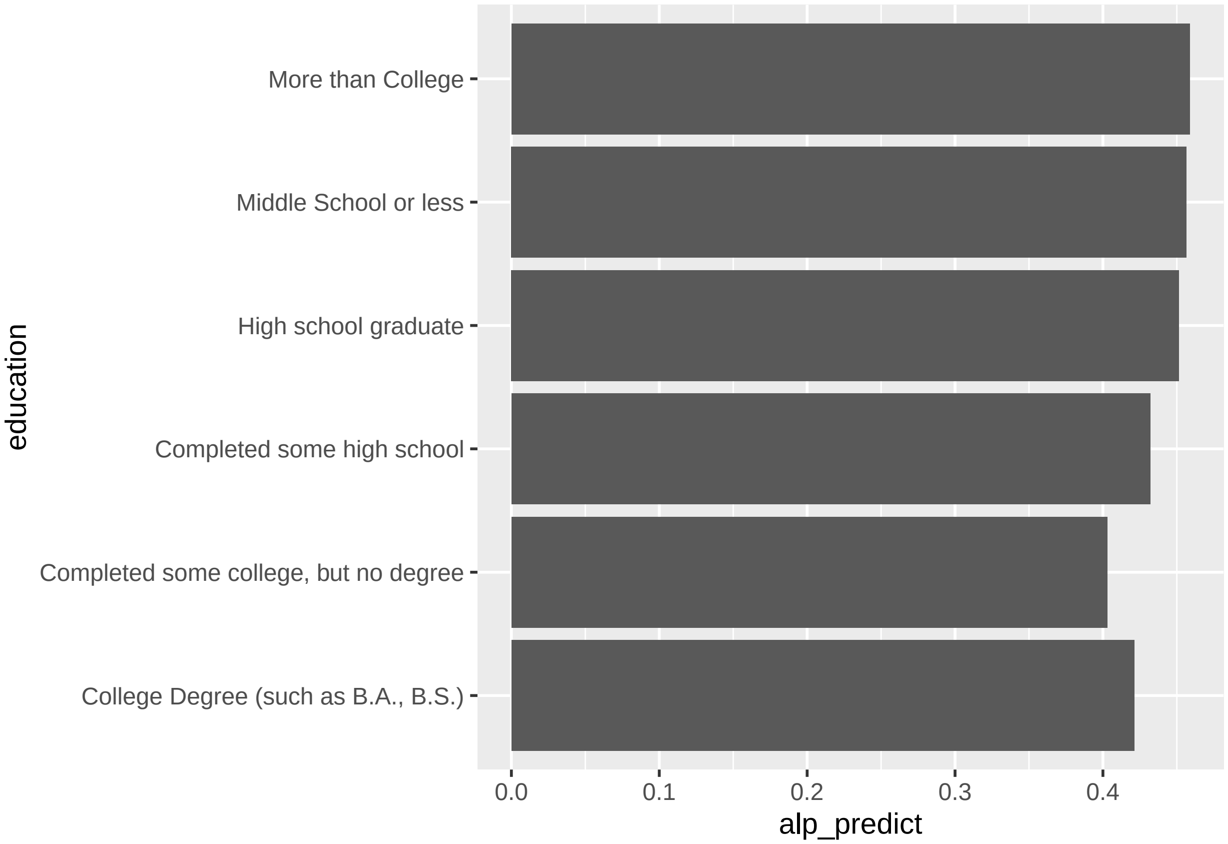
```
# different group state vote for trump odds
census_data %>%
  mutate(alp_predict_prop = estimate*n) %>%
  group_by(state) %>%
  summarise(alp_predict = sum(alp_predict_prop)/sum(n)) %>%
  ggplot(aes(state, alp_predict)) +
  geom_bar(stat = "identity") +
  coord_flip()
```



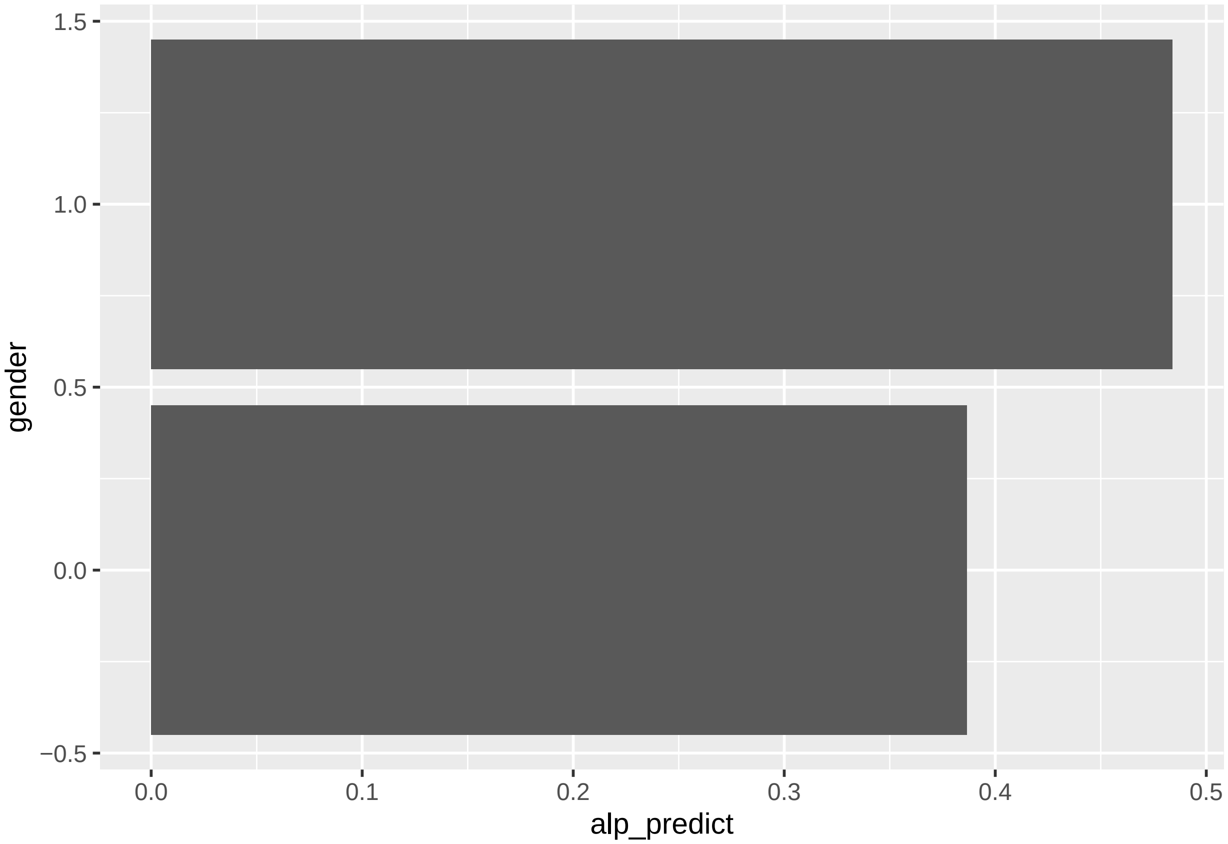
```
# different group income vote for trump odds
census_data %>%
  mutate(alp_predict_prop = estimate*n) %>%
  group_by(household_income) %>%
  summarise(alp_predict = sum(alp_predict_prop)/sum(n)) %>%
  ggplot(aes(household_income, alp_predict)) +
  geom_bar(stat = "identity") +
  coord_flip()
```



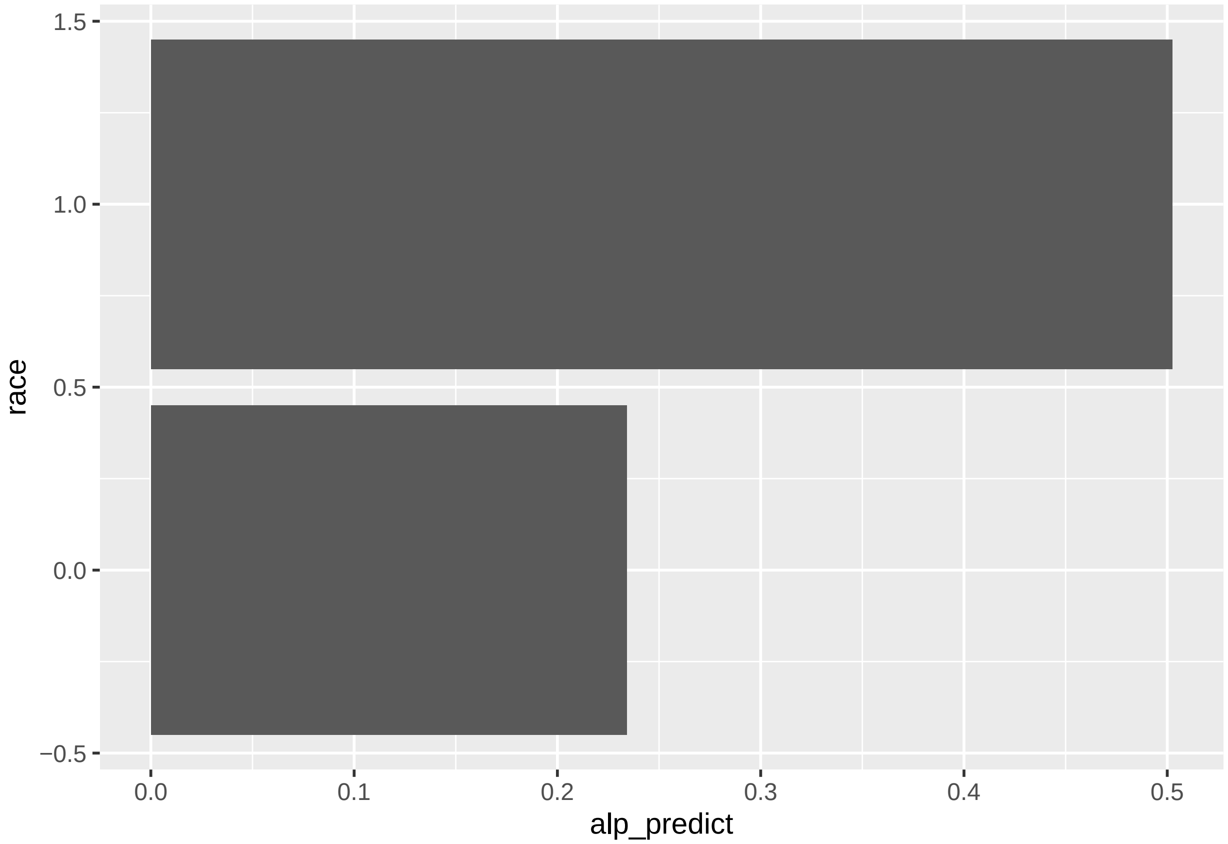
```r
# different group education vote for trump odds
census_data %>%
  mutate(alp_predict_prop = estimate*n) %>%
  group_by(education) %>%
  summarise(alp_predict = sum(alp_predict_prop)/sum(n)) %>%
  ggplot(aes(education, alp_predict)) +
  geom_bar(stat = "identity") +
  coord_flip()
```



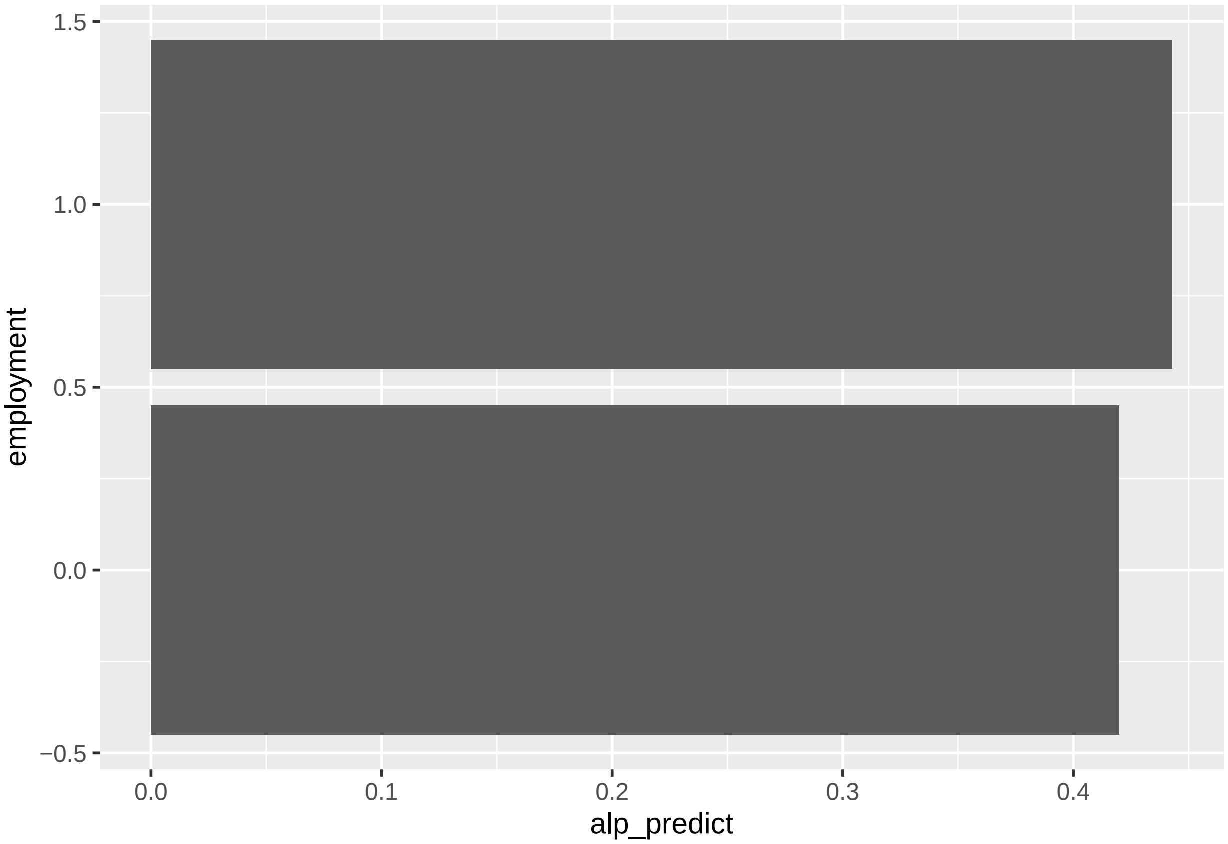
```
# different group gender vote for trump odds
census_data %>%
  mutate(alp_predict_prop = estimate*n) %>%
  group_by(gender) %>%
  summarise(alp_predict = sum(alp_predict_prop)/sum(n)) %>%
  ggplot(aes(gender, alp_predict)) +
  geom_bar(stat = "identity") +
  coord_flip()
```



```r
# different group race vote for trump odds
census_data %>%
  mutate(alp_predict_prop = estimate*n) %>%
  group_by(race) %>%
  summarise(alp_predict = sum(alp_predict_prop)/sum(n)) %>%
  ggplot(aes(race, alp_predict)) +
  geom_bar(stat = "identity") +
  coord_flip()
```



```
# different group employment vote for trump odds
census_data %>%
  mutate(alp_predict_prop = estimate*n) %>%
  group_by(employment) %>%
  summarise(alp_predict = sum(alp_predict_prop)/sum(n)) %>%
  ggplot(aes(employment, alp_predict)) +
  geom_bar(stat = "identity") +
  coord_flip()
```



Results

From the post-Stratification above, I got a overall probability and seven plots. Overall, Trump has around 43.4% to win the election. Plot 1 is about the age group, which we can see the age group above 65 exceed the 50 percent. Plot 2 is for the state, which we can see only 13 states exceed the 50 percent. Plot 3 is the household income, which we can see three high income group exceed the 50 percent. Plot 4 is the education degree, we can see the below high school education groups exceed the 50 percent. Plot 5 is the gender, which we can see males are more likely to vote Trump. Plot 6 is race, which white people would more likely to vote Trump. Plot 7 is employment, which the two groups are about the same, but the full employment group is a little more likely to vote Trump.

Discussion

From the results above, we can have several conclusions: higher age groups would more likely to vote Trump; 13 states are more likely to vote Trump; high income groups are more likely to vote Trump; low level education groups are more likely to vote Trump; males are more likely to vote Trump; the white are more likely to vote Trump; the employment may be not important, but full employment group is more likely to vote Trump. There are many things interesting to discuss from these results. The age question is hard to discuss, maybe that the policies Biden make are more popular in the young group. For those 13 states, many of them are the traditional states that would support repubilcan party. For the income groups, since Biden is going to take wealth tax, the wealth groups are less likely to support him. Therefore, the high income groups tend to support Trump. And for gender, since Biden has a female vice president, female are more likely to vote him while males for another. For the race, since it's said that Trump is a racist, the non white group is less likely to vote him. Overall, if only base on this study above, the overall probability that Trump would win is 43.4%, which means Biden's overall probability to win is 56.6%. Biden would have a higher probability to win this election than Trump.

Weaknesses

In this study, I only discuss the age, employment, gender, race household income, education and state as varibles that will affect the election. However, there are also other important variables, such as vote_2016 and ideo level. I did not discuss them because I failed to clean their data. I need to improve my ability of programming in order to provide a more precise study report.

Next Steps

In the next study, I would try to clean the data of vote_2016 and ideo5 to see if there is new findings. Also, I would keep up with the voting process and find out whether there is any new changes.

References

Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek. IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D010.V10.0

Tausanovitch, Chris and Lynn Vavreck. 2020. Democracy Fund + UCLA Nationscape, October 10-17, 2019 (version 20200814). Retrieved from https://www.voterstudygroup.org/