

Your project title

Approximately Normal: Jenny Huang, Bella Larsen, Jessie Bierschenk, Aidan Gildea

28 October 2020

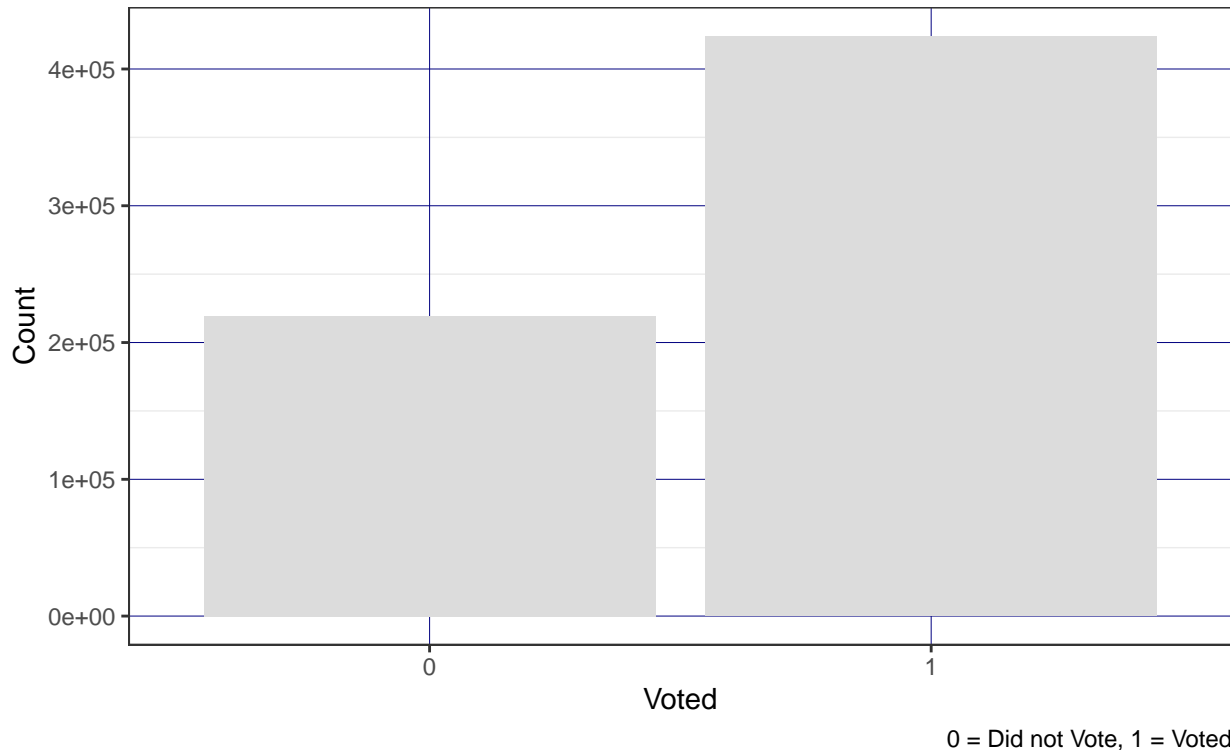
Introduction

We will begin our EDA by visualizing the relationship between the response variable `voted` and several of the other variables of particular interest.

We will begin by simply looking at the distribution of those who voted throughout the last 8 years of elections.

Visualizing the Distribution of Voting Status

More people reportedly voted than did not vote

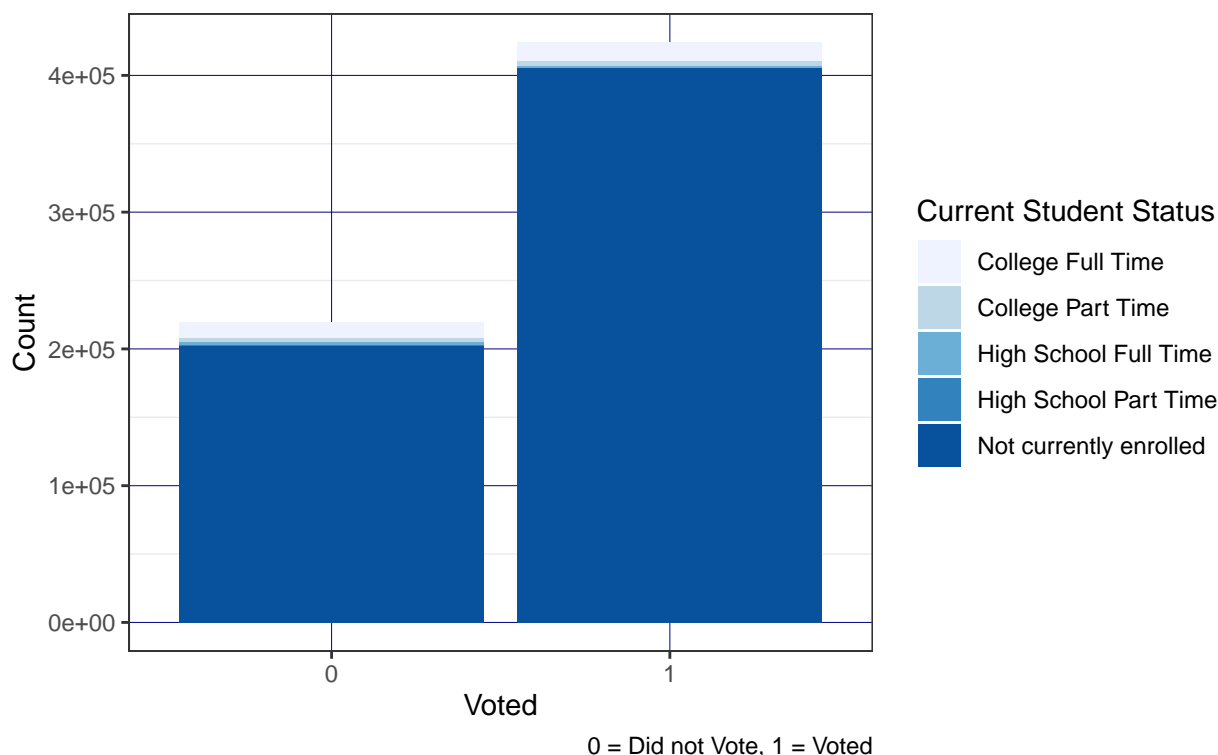


From the barplot above, it is clear the more individuals in the data set voted (voted = 1) than did not (voted = 0).

As college students ourselves, we want to analyze whether or not being a student influences the frequency of voting. We will explore this preliminarily by visualizing the distribution of if school aged individuals (18-24) voted or not – categorized by their current student level. This is seen in the bar plot below.

Voting Distribution of Population of 16–24 Year Olds

Examining relationship between student status and voting



From the bar plot, it is evident that a majority of these individuals were not currently enrolled. This may be a result of a general national trend, but we want to investigate if it is the result of a larger proportion of older individuals within the range of ages between 16-24. We will investigate this by analyzing those who are not currently enrolled in school within this age range.

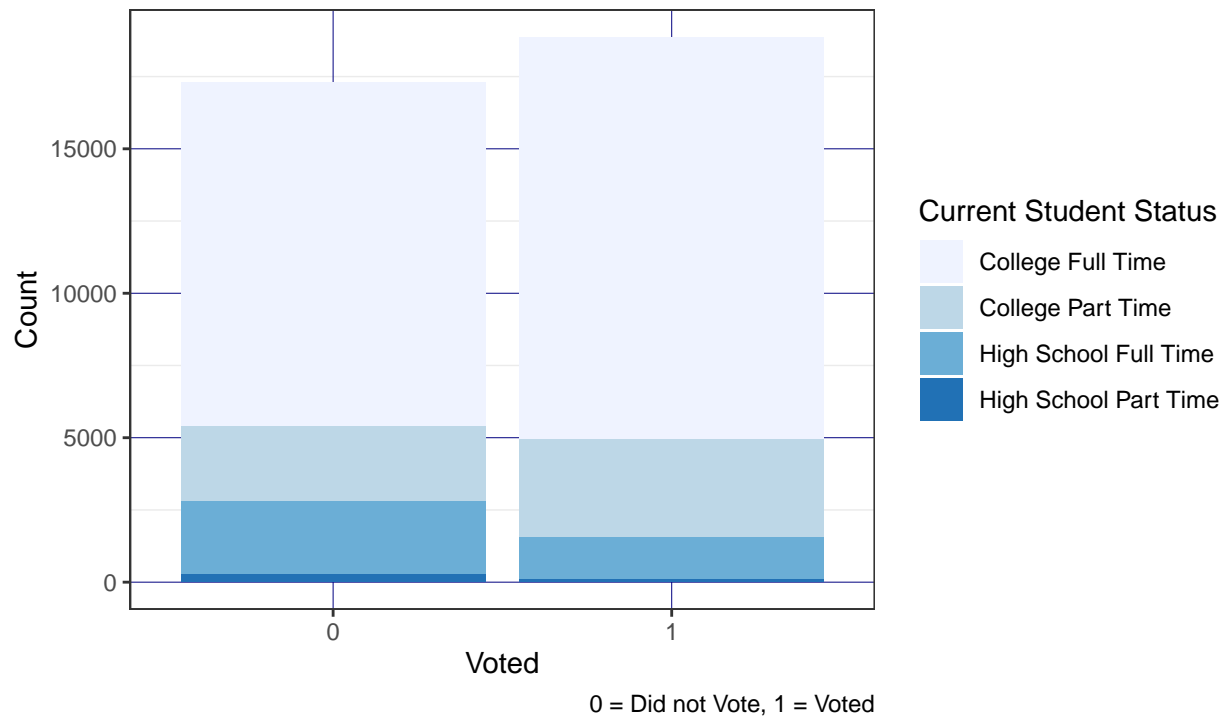
AGE	n	prop
18	2327	0.065
19	3671	0.102
20	4367	0.122
21	4760	0.133
22	5955	0.166
23	6991	0.195
24	7791	0.217

From the kable above, it is apparent that more than 40% of those not currently enrolled in school are 23-24 years old. This could be a potential reason for why this age range includes so many who are not currently enrolled as a student.

To more meaningfully analyze the relationship between being a student and if they vote or not, we adjusted our visualization to only include those currently enrolled in some level of education. This is seen in the visualization below.

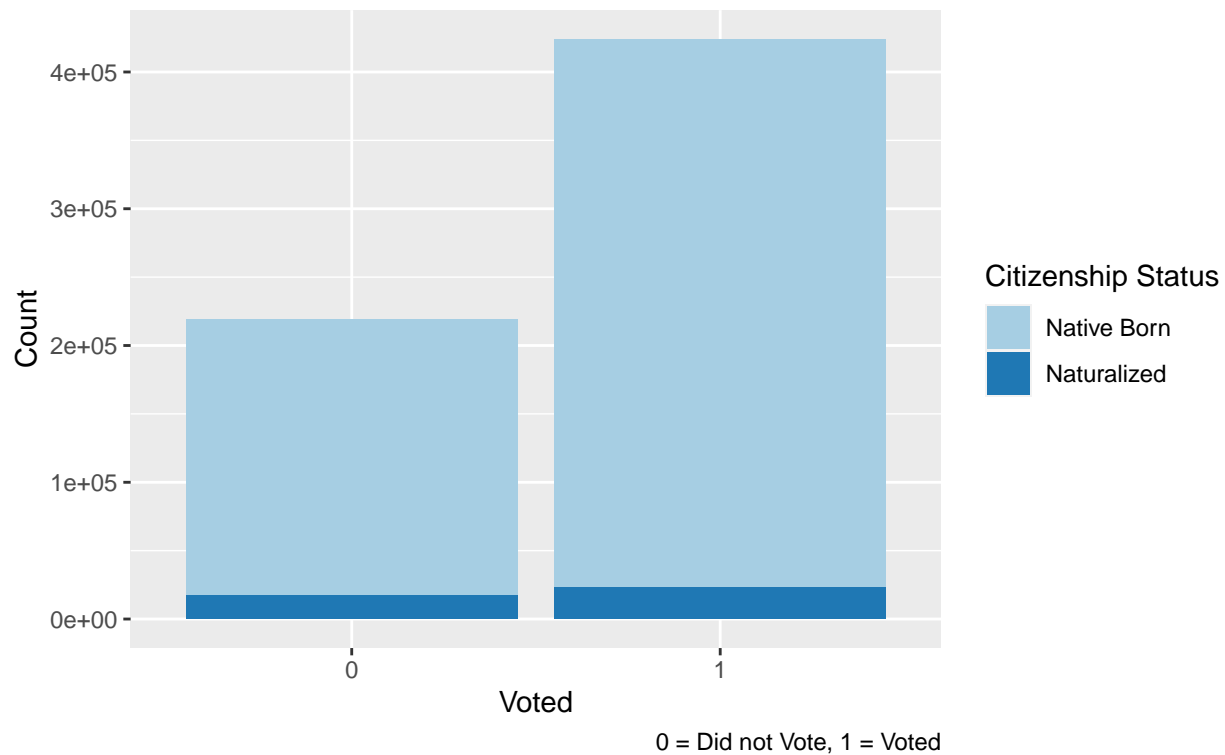
Voting Distribution of Population of 16–24 Year Olds Enrolled in School

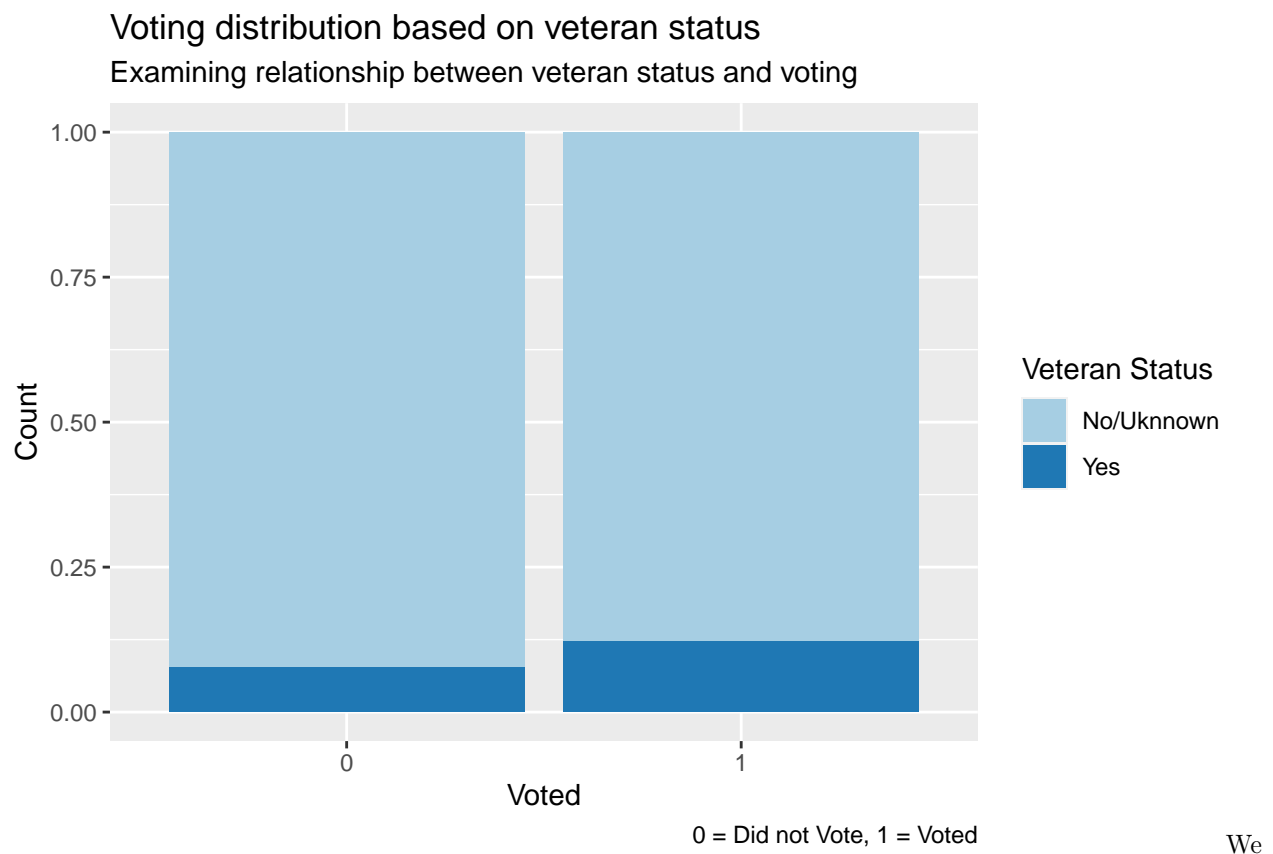
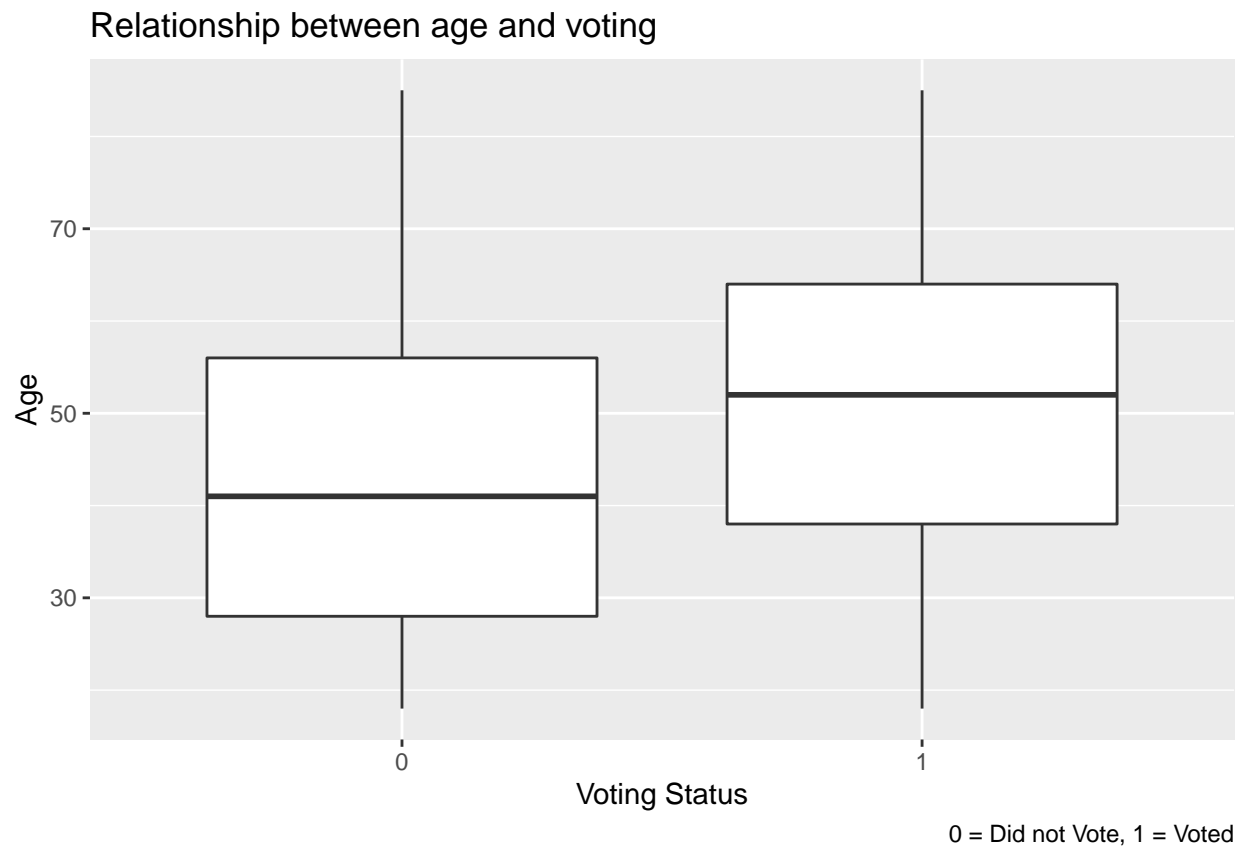
Examining relationship between student status and voting



Voting distribution based on citizenship status

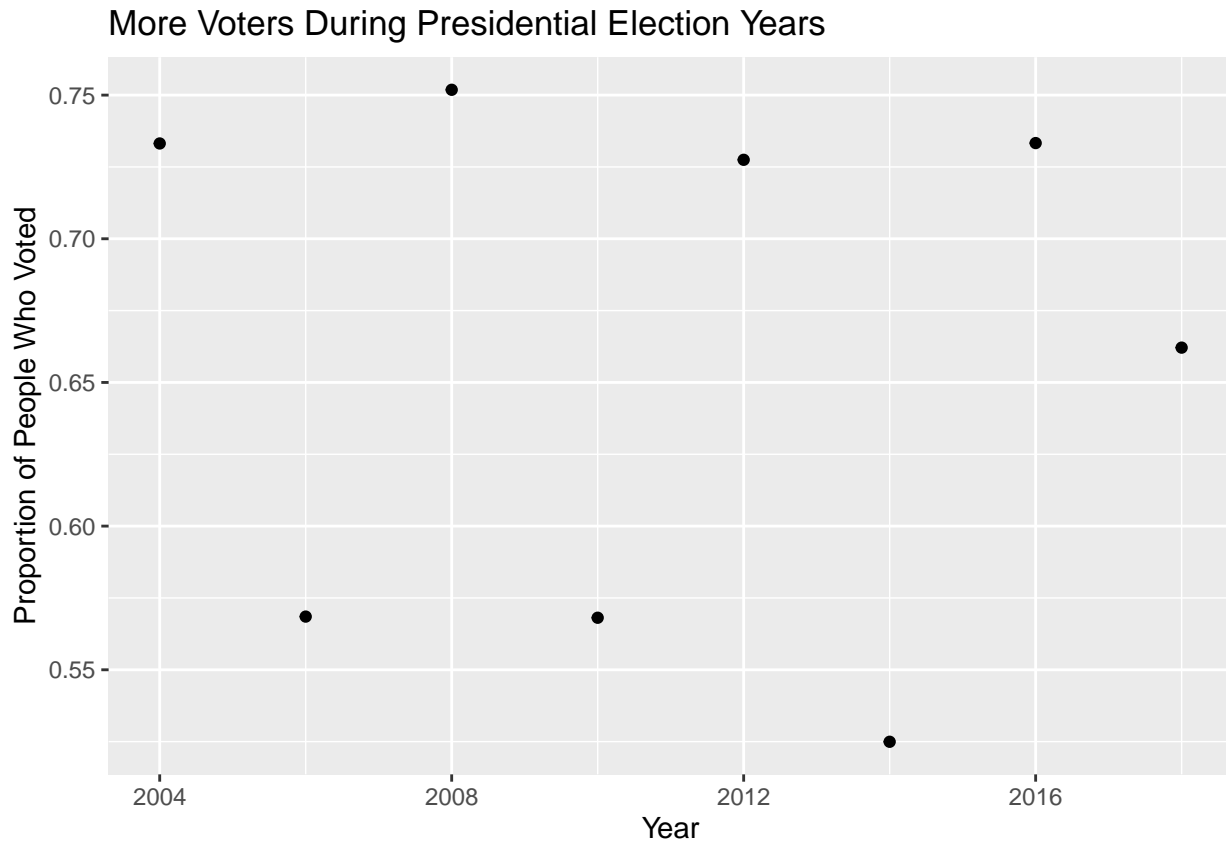
Examining relationship between student status and voting





are also interested in looking at how voter turnout has changed over the years.

We notice that the proportion of people who voted fluctuates depending on whether the year falls on a presidential election. In the trend of the proportion of voting over time, we see a clear divide between the years when there is a presidential elections versus when there is not. In the future, we may decide to add the variable “Election Year” as an interaction term with year as a divide between years that fall on an election.



Model Selection

```
## Start: AIC=79968.19
## voted ~ metro + sex + marst + veteran + citizen + hispanic_status +
##     employed + highest_education + current_student + race + AGE
##
##           Df Deviance   AIC
## <none>          79916 79968
## - veteran       1    79921 79971
## - metro          1    79930 79980
## - hispanic_status 1    79945 79995
## - citizen        1    79987 80037
## - sex            1    80020 80070
## - employed       1    80238 80288
## - current_student 4    80369 80413
## - race           4    80412 80456
## - marst          2    80807 80855
## - AGE            1    82534 82584
## - highest_education 8    86009 86045
## # A tibble: 26 x 5
```

```
##      term                                estimate std.error statistic    p.value
##      <chr>                                <dbl>      <dbl>      <dbl>      <dbl>
## 1 (Intercept)                          -0.312      0.0908      -3.44 5.85e- 4
## 2 metroNot Metro/Unknown                -0.0797     0.0218      -3.65 2.66e- 4
## 3 sexMale                               -0.188      0.0184     -10.2 2.60e- 24
## 4 marstMarried                           0.640      0.0258      24.8 2.20e-135
## 5 marstNot Married/Other                 0.145      0.0278       5.22 1.79e- 7
## 6 veteranYes                             0.0774     0.0347       2.23 2.59e- 2
## 7 citizenNaturalized                    -0.324     0.0380      -8.51 1.67e- 17
## 8 hispanic_statusNot Hispanic/Unknown   0.165      0.0304       5.42 5.83e- 8
## 9 employedYes                           0.385      0.0215      17.9 7.76e- 72
## 10 highest_educationBachelors Degree    0.638      0.0360      17.7 3.85e- 70
## # ... with 16 more rows
```

```
## Start: AIC=79968.19
```

```
## voted ~ metro + sex + marst + veteran + citizen + hispanic_status +
##      employed + highest_education + current_student + race + AGE
```

```
##
##      Df Deviance    AIC
## <none>                79916 79968
## - veteran             1    79921 79971
## - metro                1    79930 79980
## - hispanic_status      1    79945 79995
## - citizen              1    79987 80037
## - sex                  1    80020 80070
## - employed             1    80238 80288
## - current_student      4    80369 80413
## - race                  4    80412 80456
## - marst                2    80807 80855
## - AGE                  1    82534 82584
## - highest_education    8    86009 86045
```

```
## # A tibble: 26 x 5
```

```
##      term                                estimate std.error statistic    p.value
##      <chr>                                <dbl>      <dbl>      <dbl>      <dbl>
## 1 (Intercept)                          -0.312      0.0908      -3.44 5.85e- 4
## 2 metroNot Metro/Unknown                -0.0797     0.0218      -3.65 2.66e- 4
## 3 sexMale                               -0.188      0.0184     -10.2 2.60e- 24
## 4 marstMarried                           0.640      0.0258      24.8 2.20e-135
## 5 marstNot Married/Other                 0.145      0.0278       5.22 1.79e- 7
## 6 veteranYes                             0.0774     0.0347       2.23 2.59e- 2
## 7 citizenNaturalized                    -0.324     0.0380      -8.51 1.67e- 17
## 8 hispanic_statusNot Hispanic/Unknown   0.165      0.0304       5.42 5.83e- 8
## 9 employedYes                           0.385      0.0215      17.9 7.76e- 72
## 10 highest_educationBachelors Degree    0.638      0.0360      17.7 3.85e- 70
## # ... with 16 more rows
```

```
## Start: AIC=88057.71
```

```
## voted ~ metro + marst + race + AGE
```

```
##
##      Df Deviance    AIC
## <none>                88040 88058
## - metro             1    88147 88163
## - race               4    88493 88503
## - marst              2    89619 89633
```

```
## - AGE      1      89618 89634

## # A tibble: 9 x 5
##   term                                estimate std.error statistic    p.value
##   <chr>                                <dbl>      <dbl>      <dbl>      <dbl>
## 1 (Intercept)                       -0.346    0.0704      -4.92 8.66e- 7
## 2 metroNot Metro/Unknown             -0.207    0.0201     -10.3 8.02e- 25
## 3 marstMarried                        0.713    0.0243      29.4 7.66e-190
## 4 marstNot Married/Other              0.0550    0.0258       2.13 3.30e- 2
## 5 raceAsian or Pacific Islander     -0.366    0.0724      -5.06 4.23e- 7
## 6 raceBlack                          0.424    0.0673       6.31 2.85e- 10
## 7 raceNative American               -0.513    0.0892      -5.75 8.83e- 9
## 8 raceWhite                         0.259    0.0627       4.14 3.52e- 5
## 9 AGE                               0.0189    0.000487     38.8 0.

## Start:  AIC=88057.71
## voted ~ metro + marst + race + AGE
##
##           Df Deviance   AIC
## <none>      88040 88058
## - metro    1      88147 88163
## - race      4      88493 88503
## - marst     2      89619 89633
## - AGE       1      89618 89634

## # A tibble: 9 x 5
##   term                                estimate std.error statistic    p.value
##   <chr>                                <dbl>      <dbl>      <dbl>      <dbl>
## 1 (Intercept)                       -0.346    0.0704      -4.92 8.66e- 7
## 2 metroNot Metro/Unknown             -0.207    0.0201     -10.3 8.02e- 25
## 3 marstMarried                        0.713    0.0243      29.4 7.66e-190
## 4 marstNot Married/Other              0.0550    0.0258       2.13 3.30e- 2
## 5 raceAsian or Pacific Islander     -0.366    0.0724      -5.06 4.23e- 7
## 6 raceBlack                          0.424    0.0673       6.31 2.85e- 10
## 7 raceNative American               -0.513    0.0892      -5.75 8.83e- 9
## 8 raceWhite                         0.259    0.0627       4.14 3.52e- 5
## 9 AGE                               0.0189    0.000487     38.8 0.
```

Checking Model Conditions

keep in mind: citizenship and registration exclusion for the model