Donald Trump May Win His Re-election Campaign?

Prediction on 2020 US Presidential Election

Yijie Zhao, Yifan Xu, Yuyan Xu & Yuze Kang

1 November 2020

Abstract

First sentence. Second sentence. Third sentence. Fourth sentence.

Keywords: Forecasting; US 2020 Election; Trump; Biden; Multilevel Regression with Post-stratification.

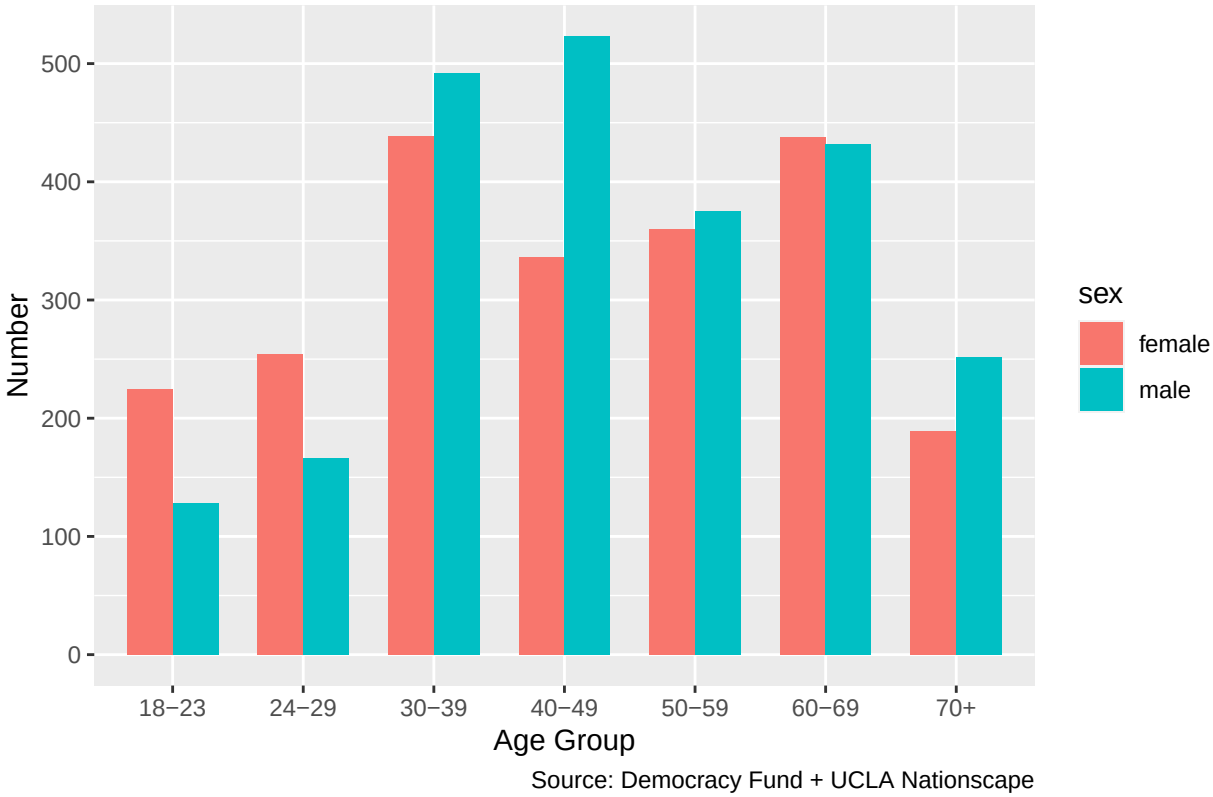
Abstract and Keywords

Introduction

Data

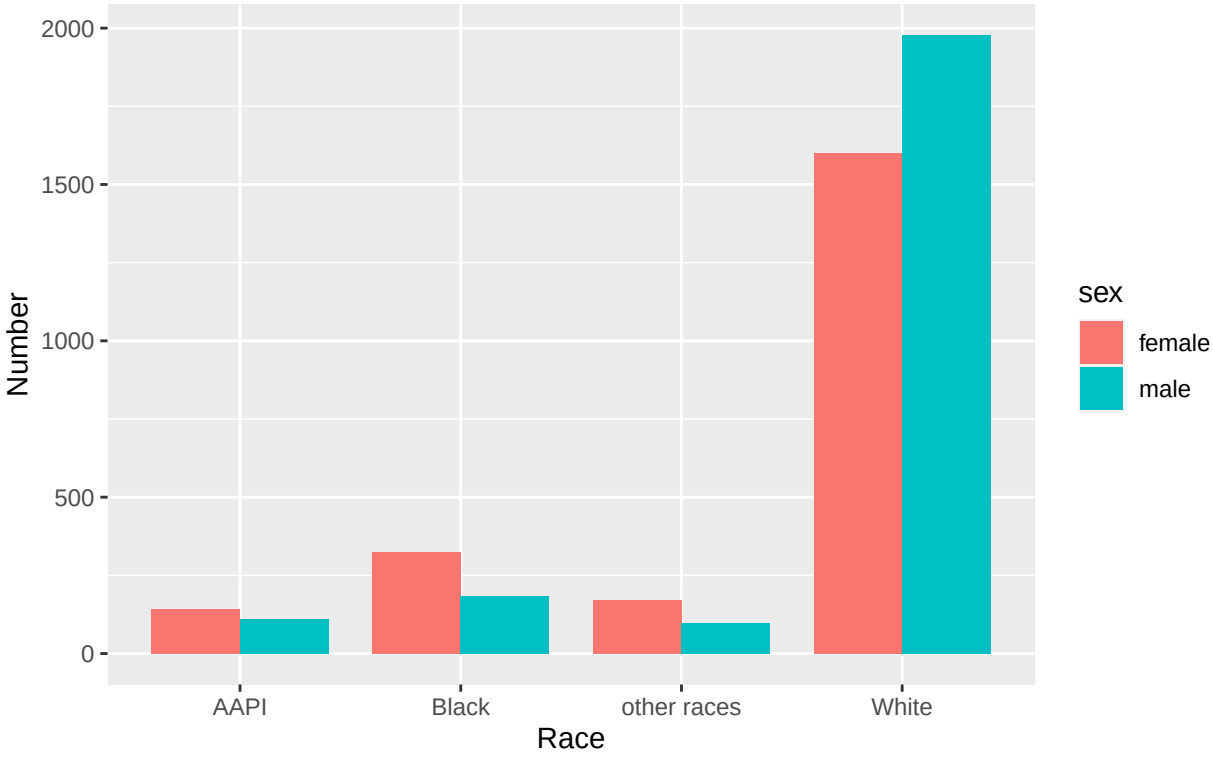
Survey Data

Figure 1. Age Distribution for Male and Female



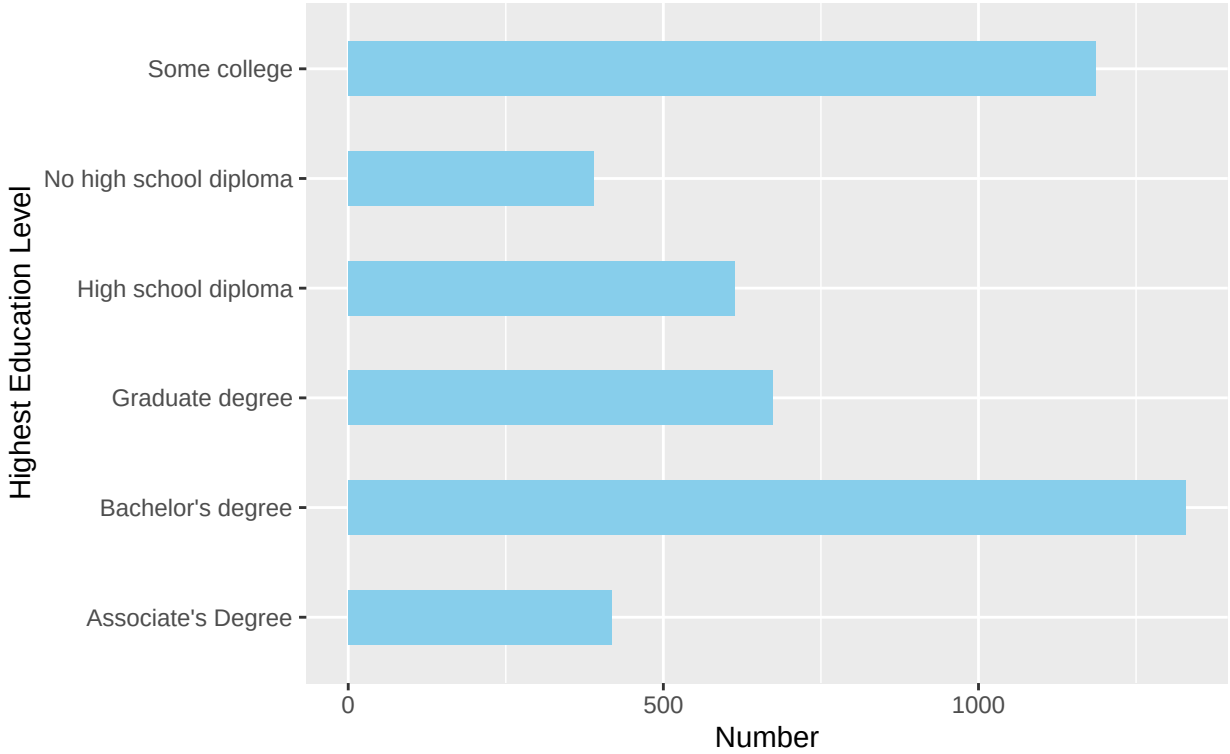
Source: Democracy Fund + UCLA Nationscape

Figure 2. Race Distribution for Male and Female



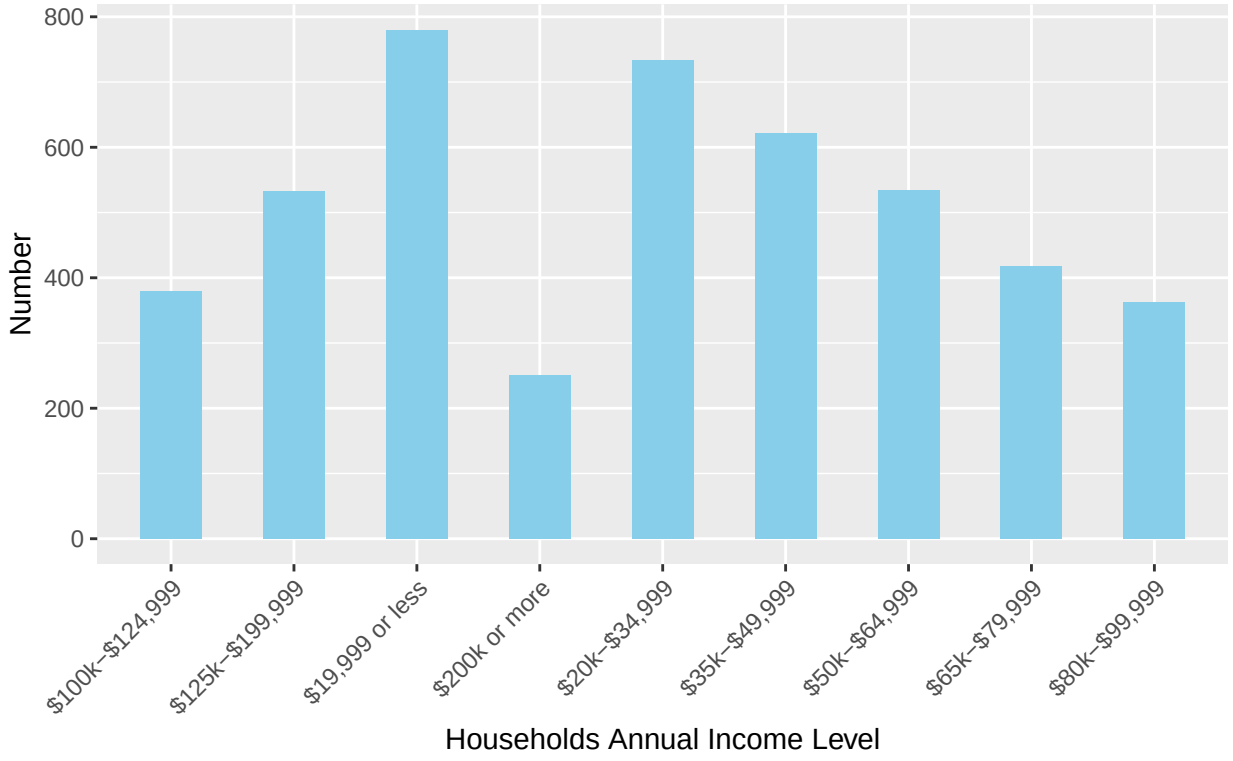
Source: Democracy Fund + UCLA Nationscape

Figure 3. Distribution of Education Level



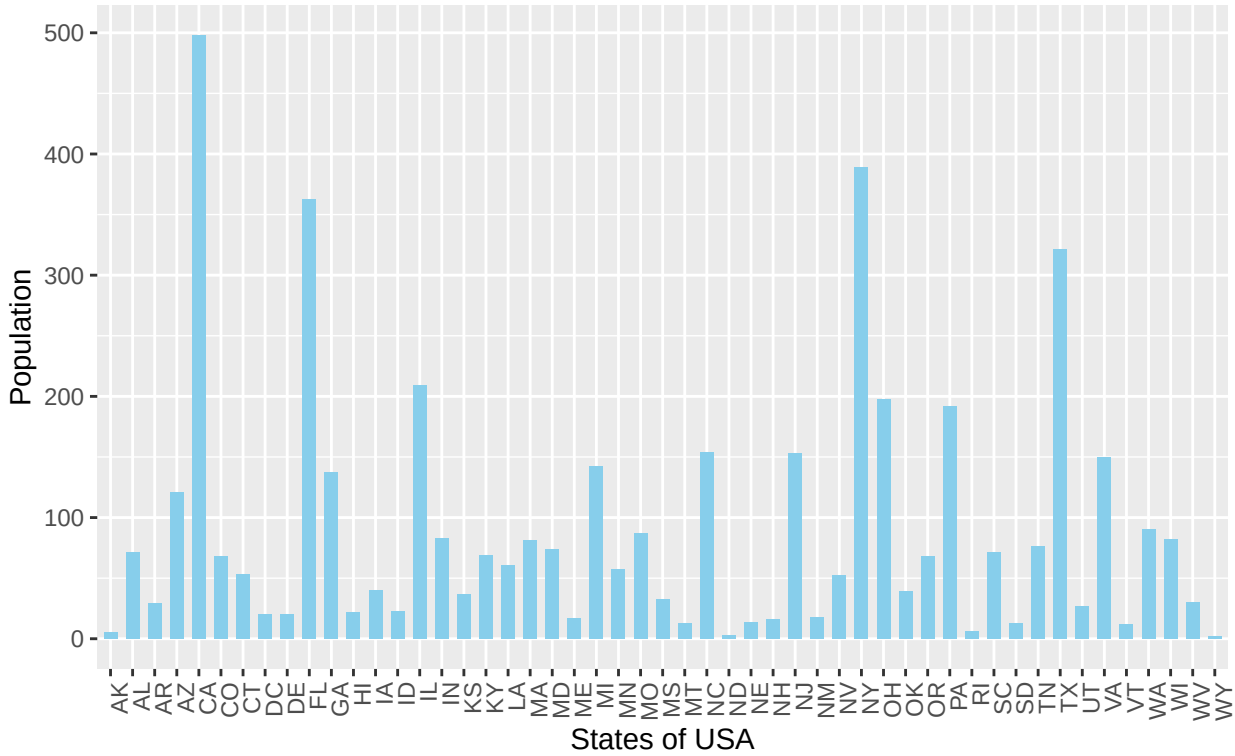
Source: Democracy Fund + UCLA Nationscape.

Figure 4. Distribution of Households Total Income



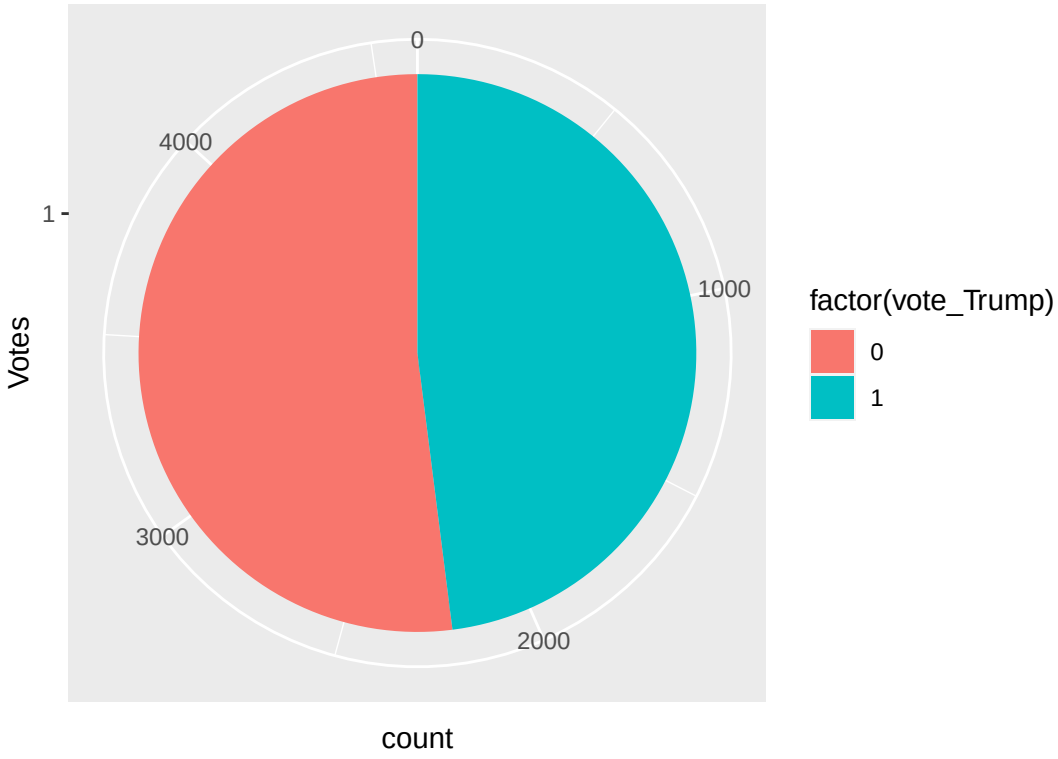
Source: Democracy Fund + UCLA Nationscape.

Figure 5. Population Distribution Among USA States



Source: Democracy Fund + UCLA Nationscape.

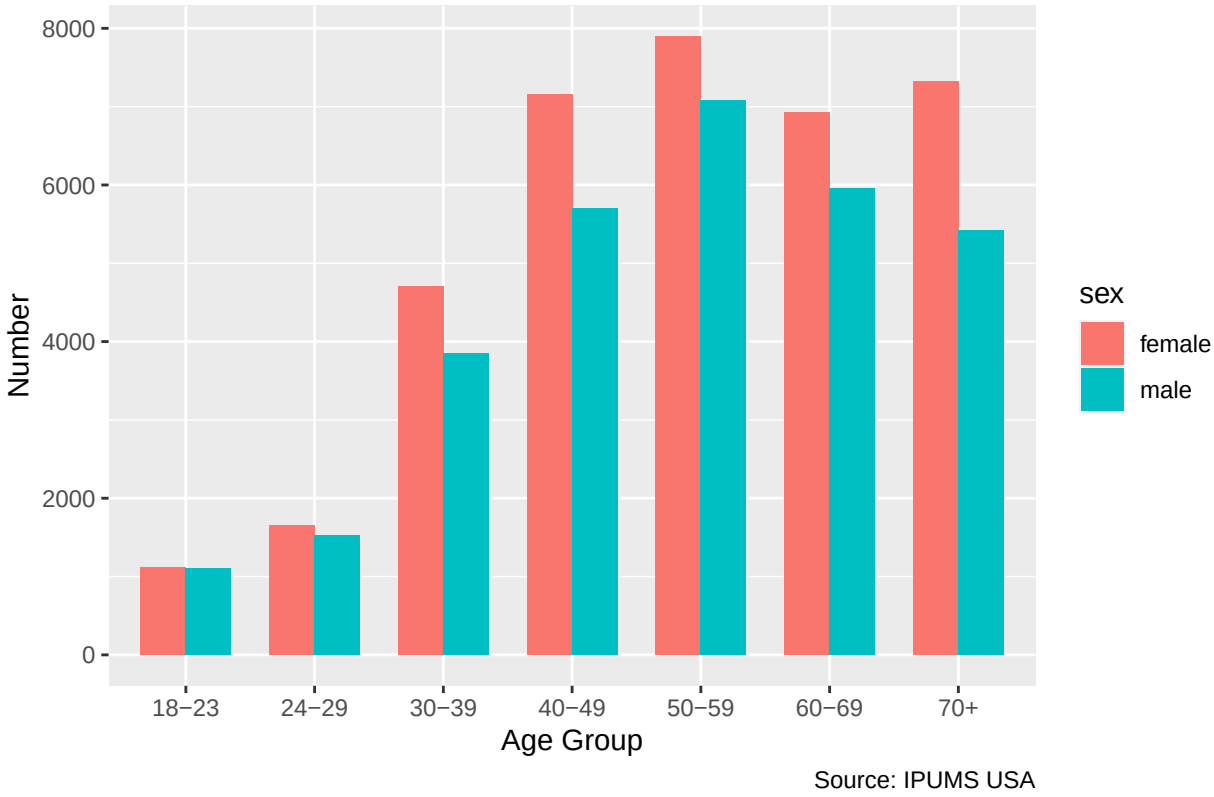
Figure 6. Blue − Vote Trump v.s. Vote Biden − Red



Source: Democracy Fund + UCLA Nationscape.

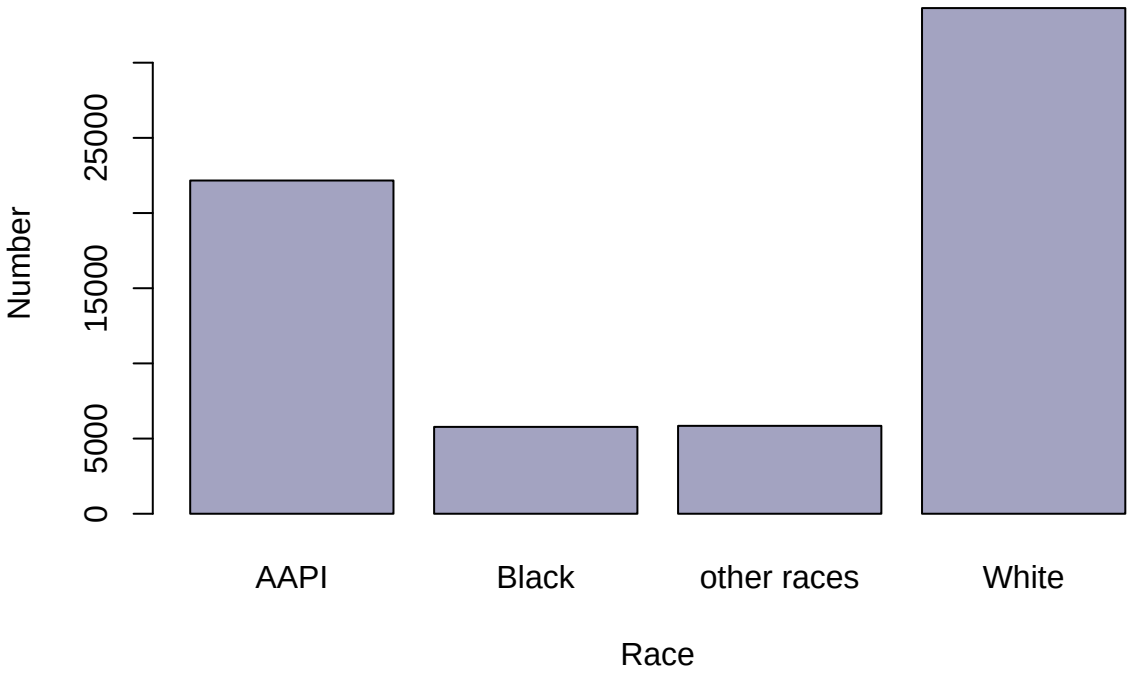
Post-stratification Data

Figure 7. Age Distribution for Male and Female




```
## `summarise()` ungrouping output (override with `.groups` argument)
```

Figure 8. Race Distribution. Source from IPUMS USA



```
## `summarise()` ungrouping output (override with `.groups` argument)
```

Figure 9. Education Distribution. Source from IPUMS USA

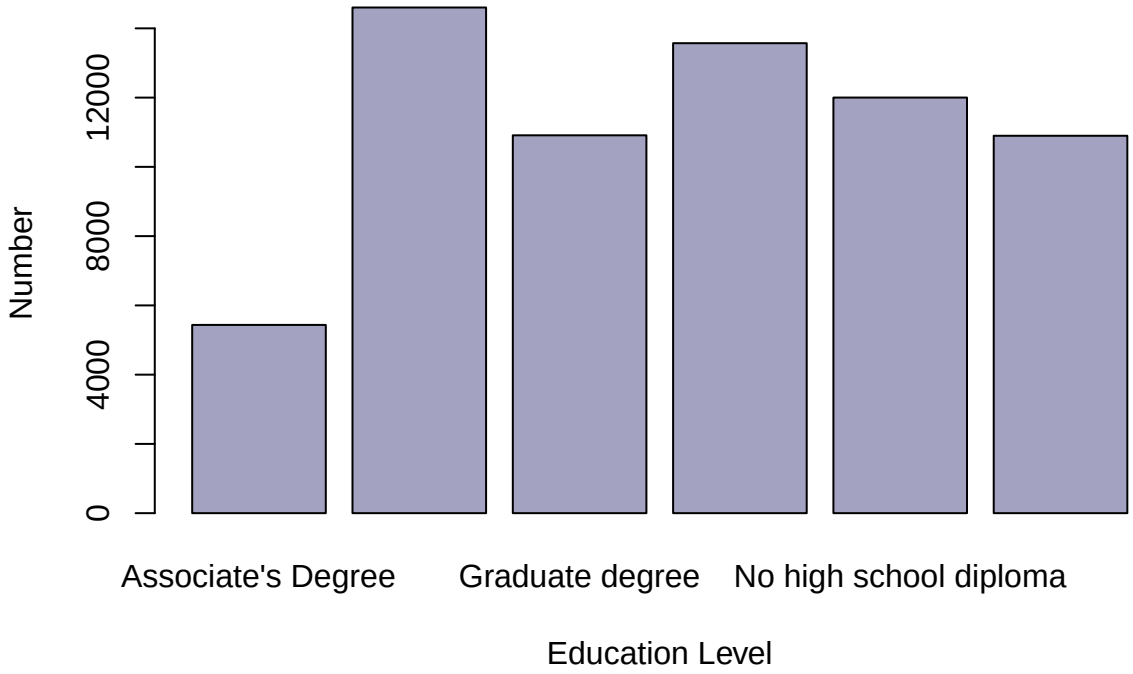
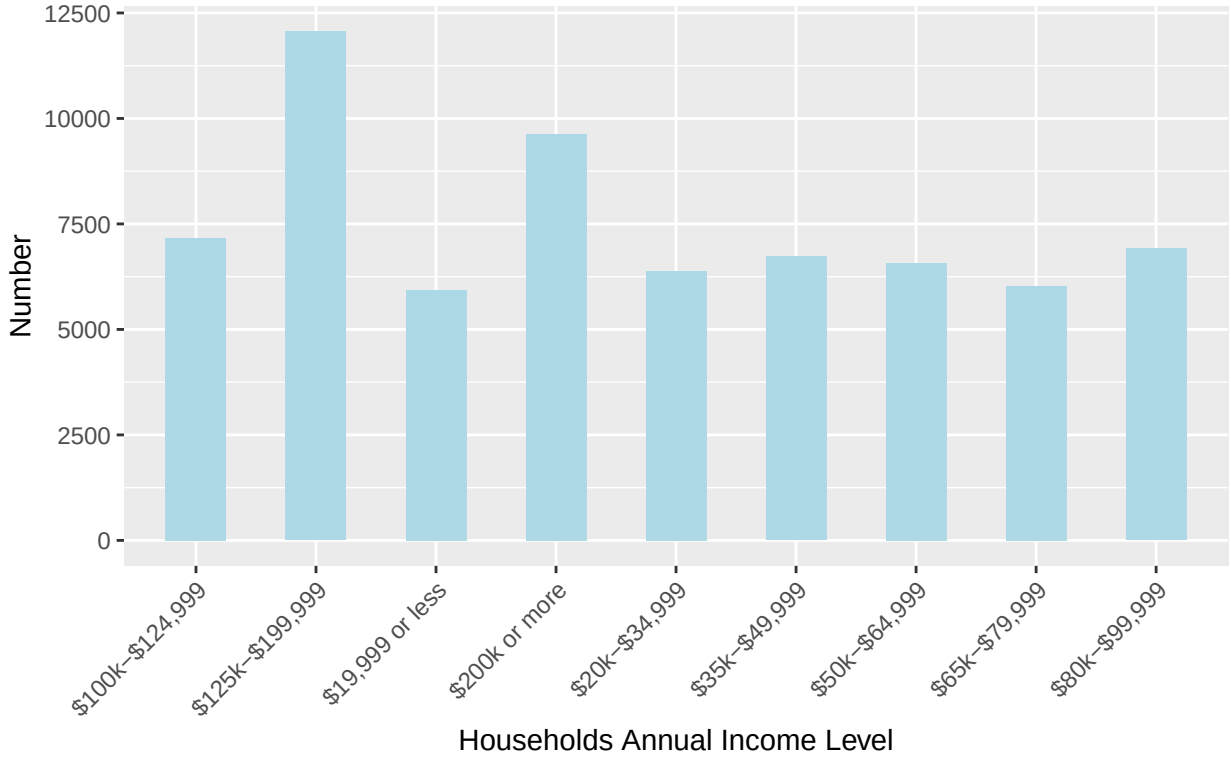
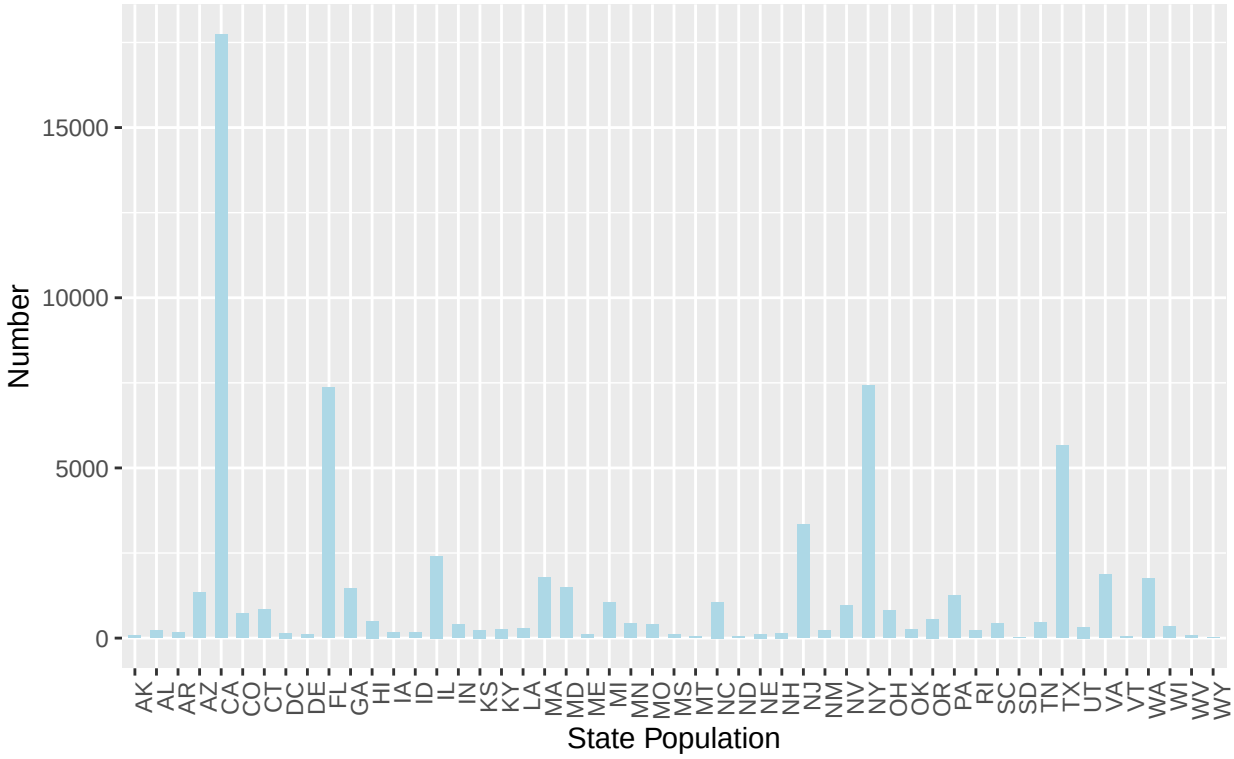


Figure 10. Distribution of Households Total Income



Source: IPUMS USA.

Figure 11. Population Distribution of USA States



Source: IPUMS USA.

Model

```
#Create cells
survey_data$cell<-paste(survey_data$age_group,survey_data$sex,survey_data$race)
ACS2018_data$cell<-paste(ACS2018_data$age_group,ACS2018_data$sex,ACS2018_data$race)

#Create census_data
state_population<-ACS2018_data %>% group_by(state) %>% summarise(sum_state=n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
census_data<-ACS2018_data %>% group_by(state, age_group,sex,race) %>% summarise(count=n())
```

```
## `summarise()` regrouping output by 'state', 'age_group', 'sex' (override with `.groups` argument)
```

```
num_state<-nrow(state_population)

num_cell_state<-nrow(census_data)

for (i in 1:num_state) {
  for(j in 1:num_cell_state){
    if(state_population$state[i] == census_data$state[j]){
      census_data$cell_prob_per_state[j] <- as.numeric(census_data$count[j])/as.numeric(state_populatio
    }
  }
}
```

```
## Warning: Unknown or uninitialised column: `cell_prob_per_state`.
```

```
###Calculate a raw probability that a person will support Trump(only based on survey data)
survey_data %>% summarise(Support_T_raw_prob = sum(vote_Trump)/nrow(survey_data))
```

```
## # A tibble: 1 x 1
##   Support_T_raw_prob
##                <dbl>
## 1              0.480
```

```
#Build a linear regression model using survey data
model_survey <- lm(vote_Trump ~ age_group+sex+race, data = survey_data)
census_data$estimate<-model_survey %>% predict(newdata = census_data)
head(census_data)
```

```
## # A tibble: 6 x 7
## # Groups:   state, age_group, sex [5]
##   state age_group sex    race  count cell_prob_per_state estimate
##   <chr> <chr>     <chr>  <chr> <int>               <dbl>    <dbl>
## 1 AK    18-23     female AAPI      1              0.0120    0.199
## 2 AK    18-23     male   White     1              0.0120    0.460
## 3 AK    24-29     female AAPI      2              0.0241    0.251
## 4 AK    24-29     male   White     1              0.0120    0.513
## 5 AK    30-39     female AAPI      6              0.0723    0.331
## 6 AK    30-39     female Black     1              0.0120    0.101
```

```
census_data %>% mutate(predict_prob_perS = estimate*cell_prob_per_state) %>% group_by(state) %>% summari
```

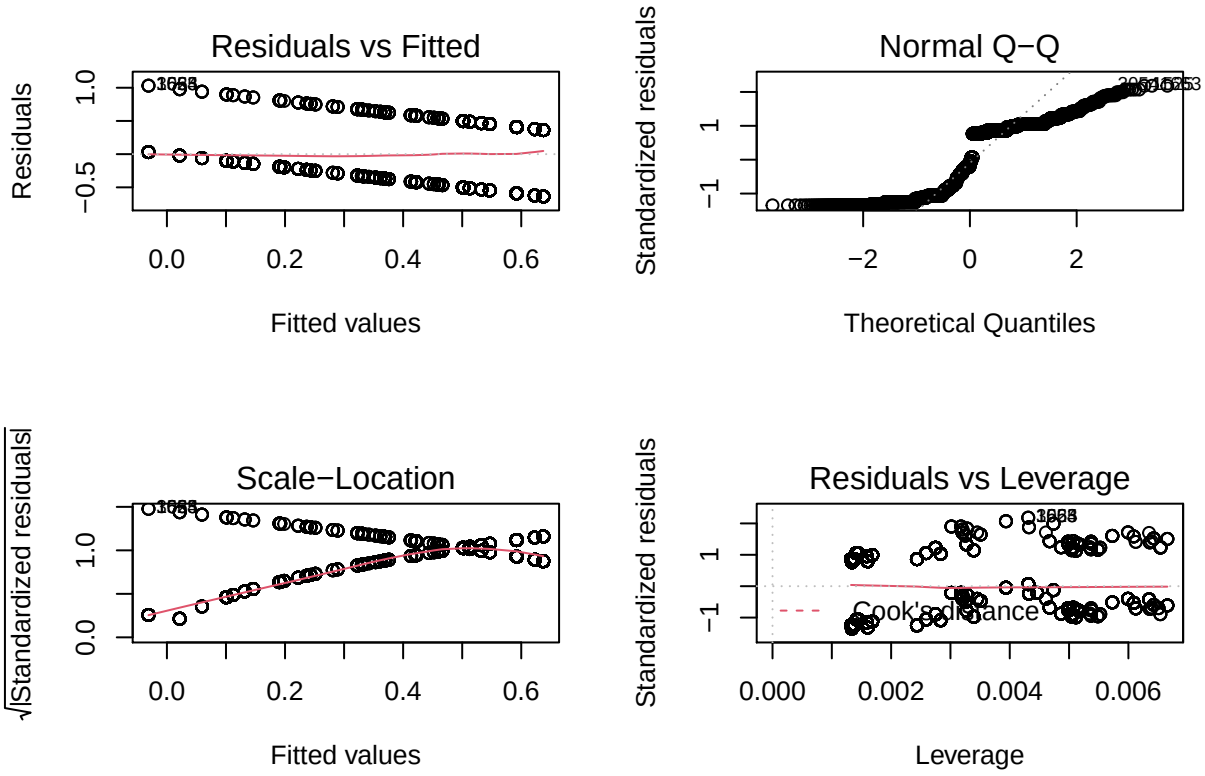
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 51 x 2
##    state predict_prob
##    <chr>        <dbl>
##  1 AK           0.426
##  2 AL           0.433
##  3 AR           0.464
##  4 AZ           0.479
##  5 CA           0.445
##  6 CO           0.475
##  7 CT           0.449
##  8 DC           0.409
##  9 DE           0.413
## 10 FL           0.468
## # ... with 41 more rows
```

```
#VIF value
car::vif(model_survey)
```

```
##               GVIF Df GVIF^(1/(2*Df))
## age_group 1.075793  6        1.006107
## sex       1.038100  1        1.018872
## race      1.077028  3        1.012444
```

```
# Model diagnostics
par(mfrow=c(2,2))
plot(model_survey)
```




```
# Model statistics
summary(model_survey)
```

```
##
## Call:
## lm(formula = vote_Trump ~ age_group + sex + race, data = survey_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.63732 -0.50115 -0.05932  0.40851  1.03124
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.198567   0.037734   5.262 1.49e-07 ***
## age_group24-29   0.052830   0.034227   1.544 0.122776
## age_group30-39   0.132518   0.029876   4.436 9.40e-06 ***
## age_group40-49   0.177323   0.030378   5.837 5.67e-09 ***
## age_group50-59   0.162933   0.031000   5.256 1.54e-07 ***
## age_group60-69   0.131711   0.030412   4.331 1.52e-05 ***
## age_group70+     0.131498   0.034450   3.817 0.000137 ***
## sexmale          0.090556   0.014220   6.368 2.10e-10 ***
## raceBlack       -0.229804   0.036507  -6.295 3.36e-10 ***
## raceother races -0.008205   0.041586  -0.197 0.843601
## raceWhite        0.170872   0.031151   5.485 4.35e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4736 on 4598 degrees of freedom
## Multiple R-squared:  0.1036, Adjusted R-squared:  0.1016
```

```
## F-statistic: 53.12 on 10 and 4598 DF,  p-value: < 2.2e-16
```

Results

Discussion

Reference