Final Project_ US Presidential election analysis 2016

Radha Krishna Chitikela

May 5, 2017

1.Objective:

The main goal of this project is to analyse the effect of demographic features like education, race gender, immigration status, annual income of the family in determining the odds of voting for Donald trump compared to Hillary Clinton. All the other votes for independents and other parties were ignored. A weighted logistic regression model is fit with the objective to explore the resons that influenced the voters towards Donald Trump.

2.About the data:

The analysis is based on the Cooperative Congressional Election Study 2016, which is a 64000+ person national stratified sample survey administered by Polimetrix.

The survey consists of two waves in election years. In the pre-election phase, respondents answer two-thirds of the questionnaire. This segment of the survey asks about general political attitudes, various demographic factors, assessment of roll call voting choices, and political information. The pre-election phase is administered late September to late October and rolled out in three distinct time-periods, the end of September, the middle of October, and the end of October. Spacing of interviews across these intervals allows researchers to gauge the effects of campaign information and events on the state and district electorates. In the post-election phase, respondents answer the other third of the questionnaire, mostly consisting of items related to the election that just occurred. The post-election phase is administered in November.

3. Dataset description:

Overview of terminologies used in the dataset:

- tookpost: This variable says whether the person has taken a post election survey. All the persons with tookpost = 'Yes' are considered for the analysis.
- CC16_410a: The respondent's vote in the 2016 Presidential election. "NA" could mean they didn't vote or that they didn't take the post-election survey. The options include Donald Trump, Hillary Clinton and others.
- trump2: It is a binomial variable with "1" indicating a vote to Donald Trump and "0" indicating a vote to Hillary Clinton in 2016.
- commonweight_post: The weights for people who took the post election survey
- inputstate: The state in which the respondent is registered to vote.
- educ: A factor variable with six levels of education- No High School, High School Graduate, Some College but no degree, 2-year college degree, 4-year college degree.
- gender: Male or Female
- race: A factor variable indicating the race with levels: White, Black, Hispanic, Asian, Native American, Mixed, Other and Middle Eastern.
- ideo5: A variable describing the ideology of the respondent varying from very liberal to conservative.

This variable has 6 levels as Very liberal, liberal, moderate, conservative, very conservative and Not sure.

- immstat: A variable describing the immigration status. It has levels as Immigrant Citizen, Immigrant noncitizen, First Generation immigrant, i.e the respondent born in USA but his/her parents were immigrants, Second generation immigrant, Third generation immigrant.
- union: A factor variable indicating if the respondent is currently/formerly have been a member of a labor union.
- hadjob: Says whether the respondent has a job in the past 5 years.
- faminc: A factor variable with 17 levels indicating the annual family income of the respondent.
- healthins: A factor variable indicating if the respondent has health insurance or not.

4. Variable selction and analysis:

- We are considering only those who have participated in the post-election survey using the “tookpost” variable from the dataset.
- Also, post-stratification weights are used to to weight the opinions using “commonweight_post”, so effectively we are fitting a weighted logistic regression model.

```
library(alr4)
```

```
## Loading required package: car
```

```
## Loading required package: effects
```

```
##
## Attaching package: 'effects'
```

```
## The following object is masked from 'package:car':
##
##     Prestige
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.3.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:car':
##
##     recode
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

```r
library(faraway)
```

```
##
## Attaching package: 'faraway'
```

```
## The following objects are masked from 'package:alr4':
##
##     cathedral, pipeline, twins
```

```
## The following objects are masked from 'package:car':
##
##     logit, vif
```

```r
CCES= read.csv("CCES.csv")
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec =
## dec, : EOF within quoted string
```

```r
nrow(CCES[CCES$tookpost == 'Yes',])
```

```
## [1] 17101
```

```r
CCES_post  = filter(CCES,tookpost == 'Yes',CC16_410a == "Donald Trump (Republican)"
|CC16_410a == "Hillary Clinton (Democrat)"  )
dim(CCES_post)
```

```
## [1] 14187   517
```

```r
CCES_post = subset(CCES_post,!is.na(CCES_post$CC16_410a))
levels(CCES_post$CC16_410a)
```

```
## [1] ""                                 "Donald Trump (Republican)"
## [3] "Evan McMullin (Independent)"     "Gary Johnson (Libertarian)"
## [5] "Hillary Clinton (Democrat)"      "I'm not sure"
## [7] "I didn't vote in this election" "Jill Stein (Green)"
## [9] "Other"
```

```r
CCES_post$trump2 = ifelse(CCES_post$CC16_410a == "Donald Trump (Republican)",1,0)
table(CCES_post$trump2)
```

```
##
##    0    1
## 7358 6829
```

5. Data cleaning

```r
## take model variables
CCES_sub  = CCES_post[,c("trump2","inputstate","educ","gender","race","ideo5","comm
onweight_post","immstat","union","faminc","edloan")]
head(CCES_sub)
```

```
##   trump2    inputstate                 educ gender  race             ideo5
## 1      1 New Hampshire High school graduate Female White          Moderate
## 2      1     Louisiana High school graduate Female White          Moderate
## 3      0      Colorado               4-year Female White           Liberal
## 4      1         Texas High school graduate   Male White Very conservative
## 5      1       Georgia High school graduate   Male White      Conservative
## 6      1  Pennsylvania High school graduate Female White          Moderate
##   commonweight_post          immstat
## 1         0.7304500 Third generation
## 2         0.8928381 Third generation
## 3         1.0190072 Third generation
## 4         1.0095162 Third generation
## 5         1.8877565 Third generation
## 6         0.6824087 Third generation
##                                                      union
## 1 I am not now, nor have I been, a member of a labor union
## 2 I am not now, nor have I been, a member of a labor union
## 3 I am not now, nor have I been, a member of a labor union
## 4 I am not now, nor have I been, a member of a labor union
## 5 I am not now, nor have I been, a member of a labor union
## 6 I am not now, nor have I been, a member of a labor union
##              faminc edloan
## 1 Prefer not to say     No
## 2 $50,000 - $59,999     No
## 3 $60,000 - $69,999     No
## 4 $20,000 - $29,999     No
## 5 $30,000 - $39,999     No
## 6 $60,000 - $69,999     No
```

```r
summary(CCES_sub)
```

```
##      trump2              inputstate                        educ
##  Min.   :0.0000   Florida     :1169   2-year              :1622
##  1st Qu.:0.0000   California  :1152   4-year              :4004
##  Median :0.0000   Texas       : 888   High school graduate:2174
##  Mean   :0.4814   New York    : 745   No HS               : 155
##  3rd Qu.:1.0000   Pennsylvania: 715   Post-grad           :2687
##  Max.   :1.0000   Illinois    : 584   Some college        :3545
##                   (Other)     :8934
##     gender            race                        ideo5      commonweight_post
##  Female:7645   White   :11998                      :   0   Min.   : 0.0001
##  Male  :6542   Black   :  945   Conservative     :3679   1st Qu.: 0.4512
##                Hispanic:  437   Liberal          :2851   Median : 0.6762
##                Other   :  250   Moderate         :4541   Mean   : 0.8584
##                Mixed   :  249   Not sure         : 289   3rd Qu.: 0.9716
##                Asian   :  167   Very conservative:1409   Max.   :14.9983
##                (Other) :  141   Very liberal     :1418
##                   immstat
##                         :   0
##  First generation       :1096
##  Immigrant Citizen      : 510
##  Immigrant non-citizen:  34
##  Second generation      :3858
##  Third generation       :8655
##  NA's                   :  34
##                                                           union
##                                                              :   0
##  I am not now, nor have I been, a member of a labor union:9695
##  I formerly was a member of a labor union                :3484
##  Yes, I am currently a member of a labor union           :1003
##  NA's                                                    :   5
##
##
##                faminc      edloan
##  Prefer not to say:2000       :    0
##  $80,000 - $99,999:1345   No  :11839
##  $30,000 - $39,999:1324   Yes : 2324
##  $50,000 - $59,999:1243   NA's:   24
##  $20,000 - $29,999:1222
##  $40,000 - $49,999:1184
##  (Other)          :5869
```

```
dim(CCES_sub)
```

```
## [1] 14187    11
```

```
xtabs(~edloan+trump2, data= CCES_sub)
```

```
##       trump2
## edloan    0    1
##           0    0
##    No  5816 6023
##    Yes 1524  800
```

```r
##check for rows with NA's and remove them
row.has.na <- apply(CCES_sub, 1, function(x){any(is.na(x))})
sum(row.has.na)
```

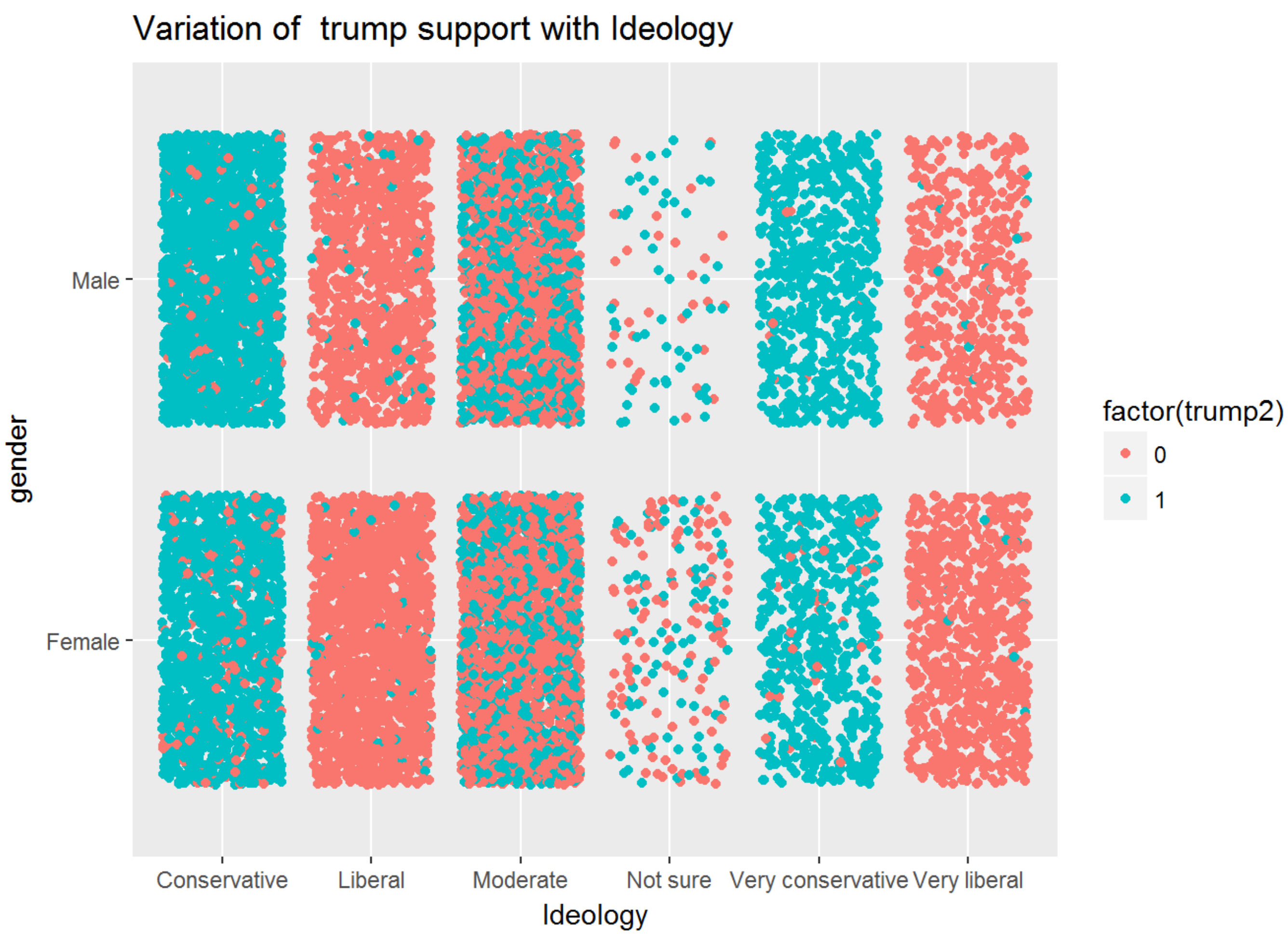
```
## [1] 63
```

```r
CCES_sub = na.omit(CCES_sub)
```

- The data is cleaned before fitting the model and missing values are removed.

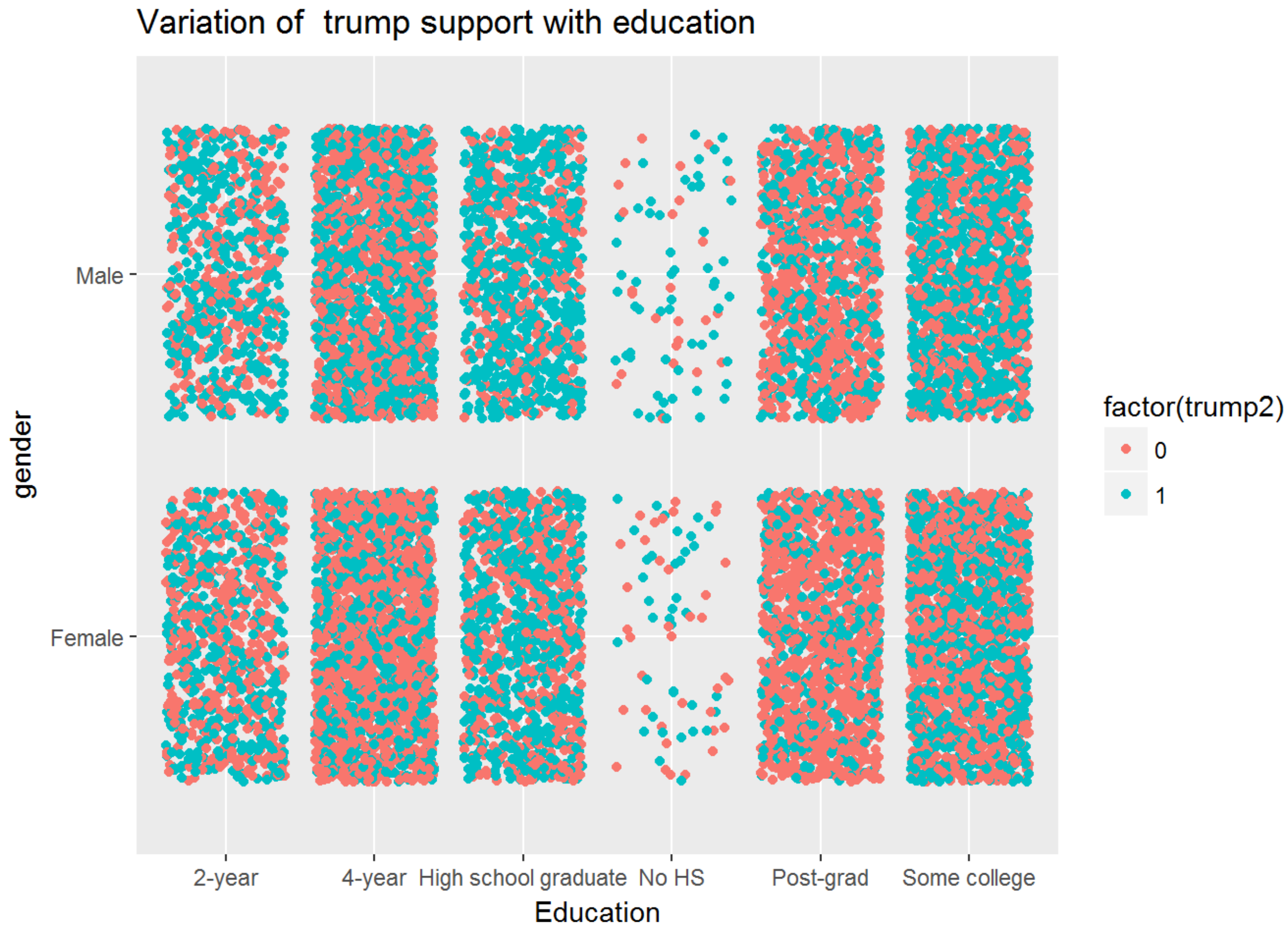
6. Exploratory data analysis:

```r
library(ggplot2)
ggplot(CCES_sub,aes(x=ideo5, y = gender, color = factor(trump2)))+geom_point()+geo
m_jitter()+
  labs(x= "Ideology", title = "Variation of  trump support with Ideology" )
```



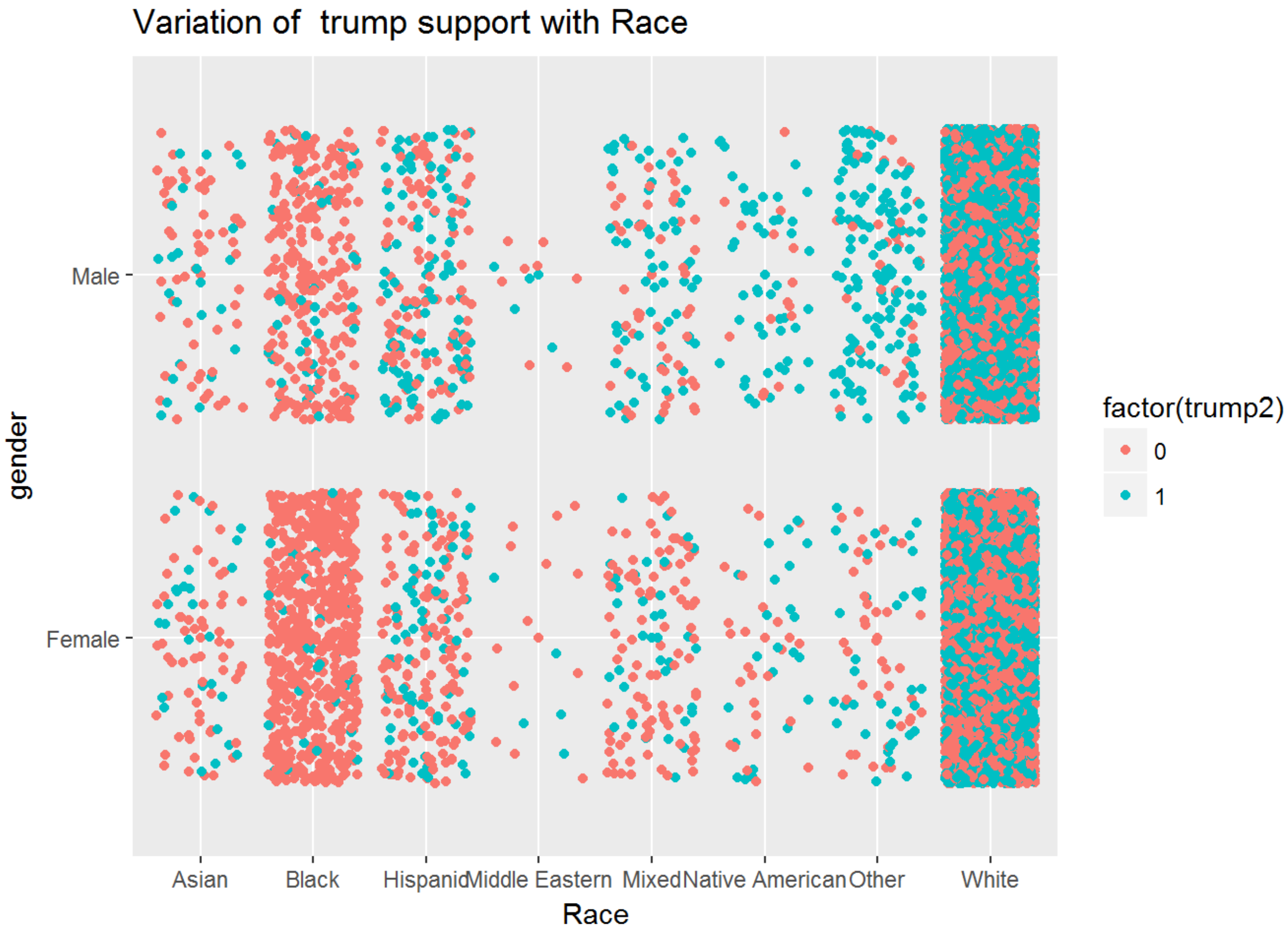
- We notice that the support for trump is very low in respondents those who identified them with moderate to very liberal ideology. As the ideology moves towardss more conservative, the support to trump astronomically increased both in female and male respondents.

```r
ggplot(CCES_sub,aes(x= educ, y = gender, color = factor(trump2)))+geom_point()+geo
m_jitter()+
  labs(x= "Education", title = "Variation of  trump support with education" )
```



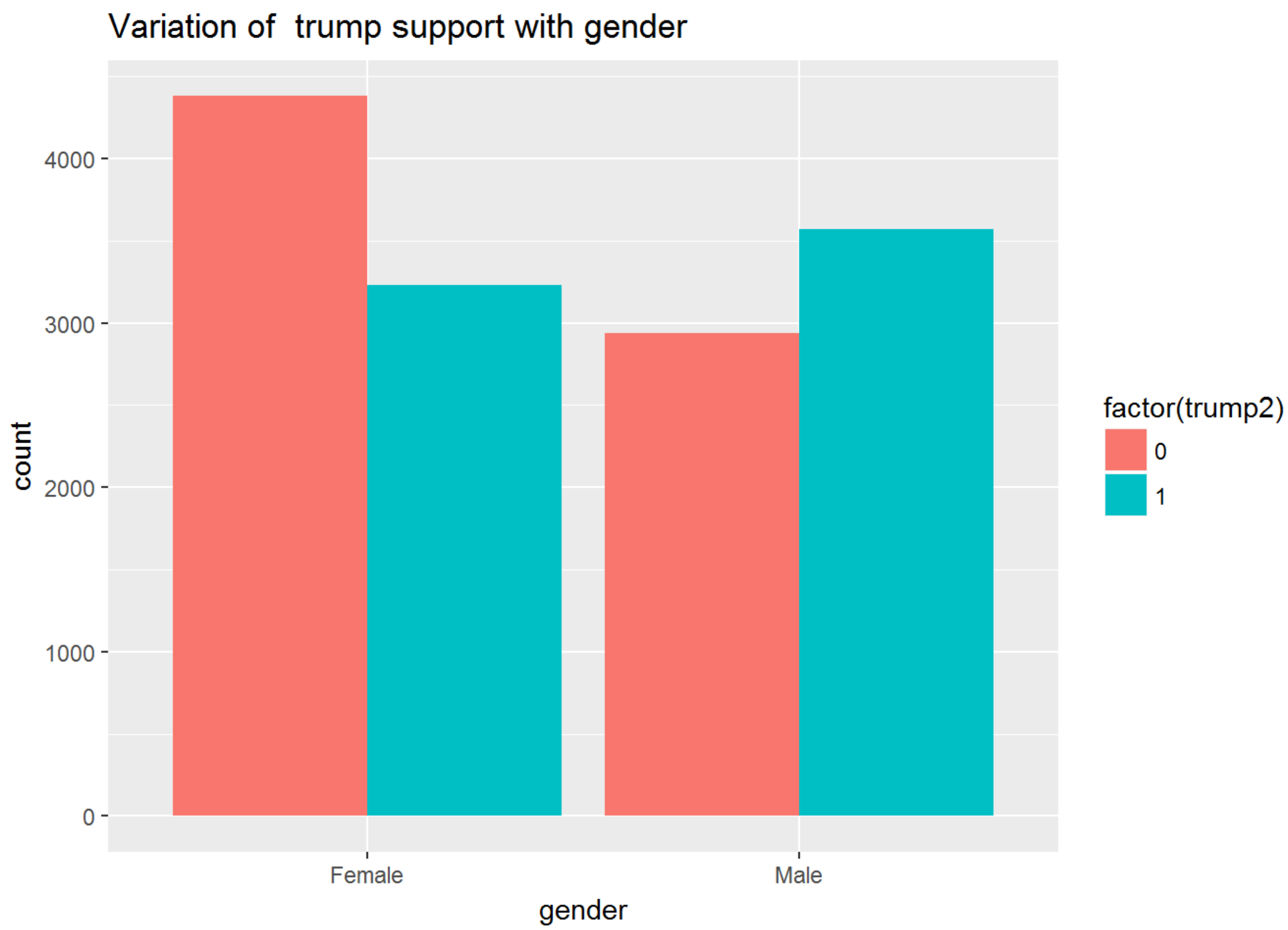
- We notice that support for trump is high in people with High School or lesser education. Both in male and female respondents, support for trump kept on decreasing with increase in level of education.

```
ggplot(CCES_sub,aes(x=race, y = gender, color = factor(trump2)))+geom_point()+geom
_jitter()+
  labs(x= "Race", title = "Variation of  trump support with Race" )
```

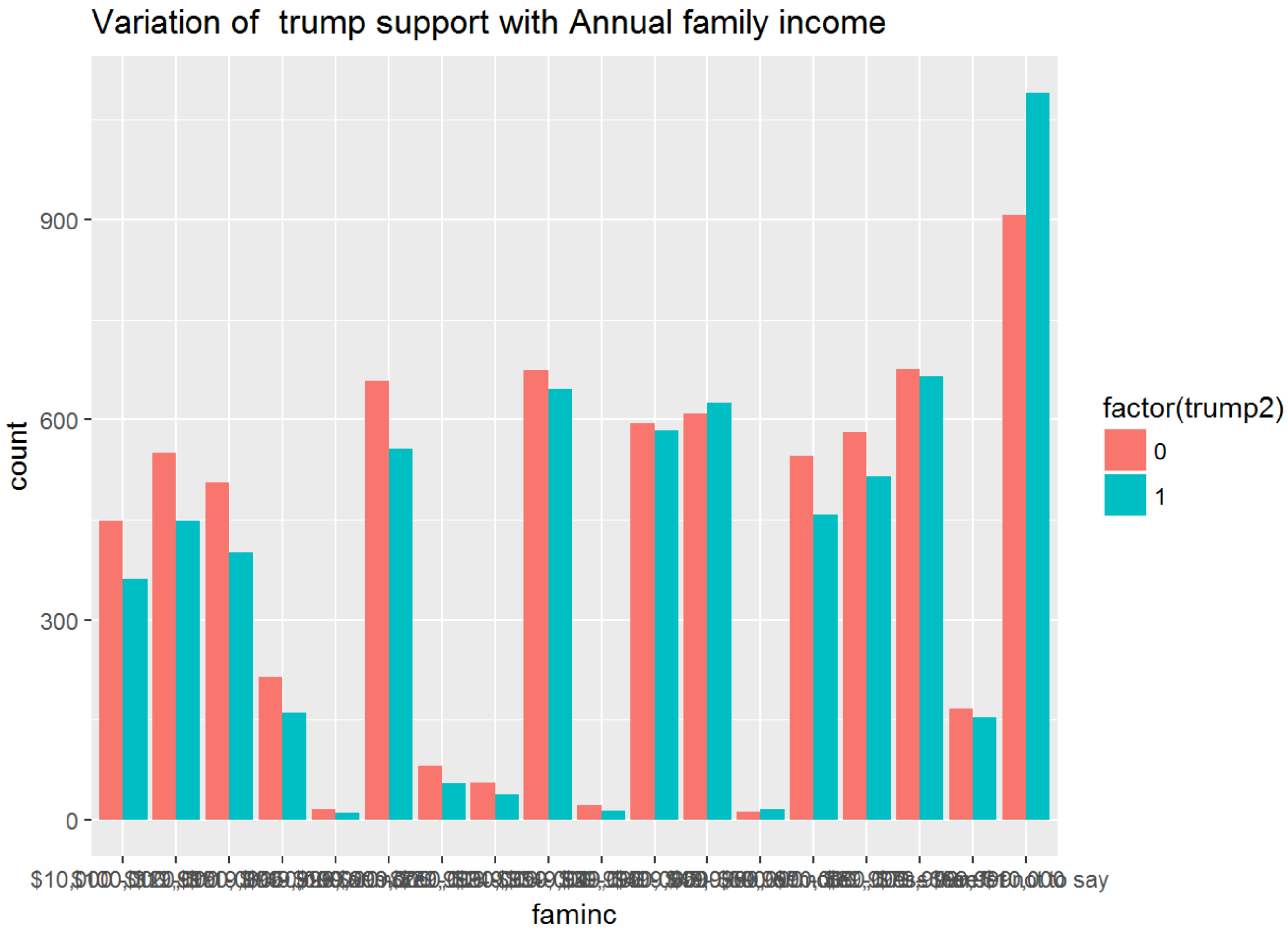


- We see that the trump support is mostly concentrated in white population equally in both genders. However there is very less support for trump in black, Hispanic and Asian population given his policies against them.

```
ggplot(CCES_sub, aes(x=gender,..count..))+geom_bar(aes(fill=factor(trump2)),positi
on = "dodge") + labs(title = "Variation of  trump support with gender")
```




```
ggplot(CCES_sub, aes(x=faminc,..count..))+geom_bar(aes(fill=factor(trump2)),positi
on = "dodge") + labs(title = "Variation of  trump support with Annual family incom
e")
```



- Considering only the gender, we can see that relatively female voters are against voting trump. This might indicate that as a support to Hillary Clinton or a retaliation to Trump's misogynic comments.
- There is no clear trend of support or opposition to trump across the income ranges but, it is an important variable that might provide more information to the model.

```
##convert all the factor variables to unordered variables
CCES_sub$ideo5 = factor(CCES_sub$ideo5, ordered = F)
CCES_sub$educ = factor(CCES_sub$educ, ordered = F)
CCES_sub$inputstate = factor(CCES_sub$inputstate, ordered = F)
CCES_sub$gender = factor(CCES_sub$gender, ordered = F)
CCES_sub$race = factor(CCES_sub$race, ordered = F)
CCES_sub$union = factor(CCES_sub$union, ordered = F)
CCES_sub$immstat = factor(CCES_sub$immstat, ordered = F)
CCES_sub$faminc = factor(CCES_sub$faminc, ordered = F)
CCES_sub$edloan = factor(CCES_sub$edloan, ordered = F)
```

- All the predictors are converted to unordered factor variables, but still the order is preserved. This is done because, packages in R are not able to give interpretations to coefficient levels when ordinal factor variables with multiple levels are used. ##7.Model Fitting

```
m1.US = glm(trump2~gender+educ+ideo5+race+edloan+faminc+immstat,weights = CCES_sub
$commonweight_post,
            family = quasibinomial, data = CCES_sub)
Anova(m1.US)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: trump2
##          LR Chisq Df Pr(>Chisq)
## gender      30.9  1  2.762e-08 ***
## educ        97.1  5  < 2.2e-16 ***
## ideo5     6598.3  5  < 2.2e-16 ***
## race       818.0  7  < 2.2e-16 ***
## edloan      16.0  1  6.299e-05 ***
## faminc      35.4 17  0.0055554 **
## immstat     20.6  4  0.0003744 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m2.US = update(m1.US, .~. - immstat)
Anova(m2.US)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: trump2
##        LR Chisq Df Pr(>Chisq)
## gender     33.1  1  8.894e-09 ***
## educ      101.1  5  < 2.2e-16 ***
## ideo5    6774.9  5  < 2.2e-16 ***
## race      856.3  7  < 2.2e-16 ***
## edloan     17.8  1  2.441e-05 ***
## faminc     37.0 17   0.003389 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m3.US = glm(trump2~gender+educ+ideo5+race+edloan+faminc+educ:edloan+edloan:faminc+
              gender:race+race:ideo5,weights = CCES_sub$commonweight_post,
            family = quasibinomial, data = CCES_sub)
Anova(m3.US)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: trump2
##               LR Chisq Df Pr(>Chisq)
## gender            35.9  1  2.114e-09 ***
## educ             117.7  5  < 2.2e-16 ***
## ideo5           7739.4  5  < 2.2e-16 ***
## race             979.2  7  < 2.2e-16 ***
## edloan            17.4  1  2.965e-05 ***
## faminc            41.0 17  0.0009307 ***
## educ:edloan       10.6  5  0.0595921 .
## edloan:faminc     27.2 17  0.0550045 .
## gender:race       29.1  7  0.0001394 ***
## ideo5:race        87.9 35  1.957e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m4.US = update(m3.US, .~. - educ:edloan- edloan:faminc)
summary(m4.US)
```

```
##
## Call:
## glm(formula = trump2 ~ gender + educ + ideo5 + race + edloan +
##     faminc + gender:race + ideo5:race, family = quasibinomial,
##     data = CCES_sub, weights = CCES_sub$commonweight_post)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -7.8073  -0.2517  -0.0398   0.3477   5.8127
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                 9.803e-01  4.097e-01   2.393
## genderMale                                 -3.521e-01  4.360e-01  -0.808
## educ4-year                                 -4.347e-01  9.882e-02  -4.399
## educHigh school graduate                    2.125e-01  1.013e-01   2.098
## educNo HS                                   4.494e-01  1.463e-01   3.072
## educPost-grad                              -4.287e-01  1.083e-01  -3.958
## educSome college                            1.577e-01  9.902e-02   1.593
## ideo5Liberal                               -1.753e+01  3.478e+02  -0.050
## ideo5Moderate                              -1.192e+00  4.682e-01  -2.546
## ideo5Not sure                              -1.494e+00  2.015e+00  -0.741
## ideo5Very conservative                      1.271e+00  1.563e+00   0.813
## ideo5Very liberal                          -1.725e+01  7.166e+02  -0.024
## raceBlack                                  -2.605e+00  4.744e-01  -5.492
## raceHispanic                               -7.127e-01  4.810e-01  -1.482
## raceMiddle Eastern                          1.899e+00  2.105e+00   0.902
## raceMixed                                   1.338e+00  6.661e-01   2.009
## raceNative American                         9.695e-01  7.362e-01   1.317
## raceOther                                   2.618e+00  1.173e+00   2.232
## raceWhite                                   1.623e+00  3.945e-01   4.114
## edloanYes                                  -3.112e-01  7.461e-02  -4.171
## faminc$100,000 - $119,999                   9.083e-02  1.487e-01   0.611
## faminc$120,000 - $149,999                   1.477e-01  1.479e-01   0.999
## faminc$150,000 - $199,999                  -1.487e-01  2.045e-01  -0.727
## faminc$150,000 or more                     -2.101e-01  5.661e-01  -0.371
## faminc$20,000 - $29,999                     7.400e-02  1.315e-01   0.563
## faminc$200,000 - $249,999                   1.845e-01  2.932e-01   0.629
## faminc$250,000 - $349,999                   1.193e-01  3.533e-01   0.338
## faminc$30,000 - $39,999                    -7.402e-03  1.354e-01  -0.055
## faminc$350,000 - $499,999                   1.137e-01  5.673e-01   0.200
## faminc$40,000 - $49,999                     1.423e-01  1.374e-01   1.036
## faminc$50,000 - $59,999                     4.544e-01  1.349e-01   3.368
## faminc$500,000 or more                      8.581e-01  7.043e-01   1.218
## faminc$60,000 - $69,999                    -3.737e-02  1.457e-01  -0.256
## faminc$70,000 - $79,999                     2.078e-01  1.434e-01   1.449
## faminc$80,000 - $99,999                     3.107e-01  1.381e-01   2.250
## famincLess than $10,000                     1.696e-01  1.868e-01   0.908
## famincPrefer not to say                     3.781e-01  1.256e-01   3.010
## genderMale:raceBlack                        1.337e+00  5.129e-01   2.607
## genderMale:raceHispanic                     1.517e+00  5.056e-01   3.000
## genderMale:raceMiddle Eastern              -2.813e+00  2.105e+00  -1.336
## genderMale:raceMixed                        1.139e+00  5.838e-01   1.951
```

```
## genderMale:raceNative American                  4.808e-01  7.343e-01   0.655
## genderMale:raceOther                             1.601e+00  7.031e-01   2.278
## genderMale:raceWhite                             5.915e-01  4.394e-01   1.346
## ideo5Liberal:raceBlack                           1.323e+01  3.478e+02   0.038
## ideo5Moderate:raceBlack                         -1.003e+00  5.696e-01  -1.762
## ideo5Not sure:raceBlack                          2.409e-01  2.055e+00   0.117
## ideo5Very conservative:raceBlack                -1.033e+00  1.614e+00  -0.640
## ideo5Very liberal:raceBlack                      1.350e+01  7.166e+02   0.019
## ideo5Liberal:raceHispanic                        1.292e+01  3.478e+02   0.037
## ideo5Moderate:raceHispanic                      -6.696e-01  5.534e-01  -1.210
## ideo5Not sure:raceHispanic                      -2.429e+00  2.416e+00  -1.005
## ideo5Very conservative:raceHispanic              1.096e-01  1.731e+00   0.063
## ideo5Very liberal:raceHispanic                   1.134e+01  7.166e+02   0.016
## ideo5Liberal:raceMiddle Eastern                  1.345e+01  3.478e+02   0.039
## ideo5Moderate:raceMiddle Eastern                -3.210e+00  2.157e+00  -1.488
## ideo5Not sure:raceMiddle Eastern                -1.769e+01  4.684e+03  -0.004
## ideo5Very conservative:raceMiddle Eastern        1.565e+01  2.965e+03   0.005
## ideo5Very liberal:raceMiddle Eastern            -1.039e+00  1.356e+03  -0.001
## ideo5Liberal:raceMixed                           1.219e+01  3.478e+02   0.035
## ideo5Moderate:raceMixed                         -2.116e+00  7.474e-01  -2.831
## ideo5Not sure:raceMixed                         -1.435e+00  2.160e+00  -0.664
## ideo5Very conservative:raceMixed                -1.993e+00  1.842e+00  -1.082
## ideo5Very liberal:raceMixed                     -1.882e+00  8.349e+02  -0.002
## ideo5Liberal:raceNative American                 1.356e+01  3.478e+02   0.039
## ideo5Moderate:raceNative American                9.147e-02  7.927e-01   0.115
## ideo5Not sure:raceNative American                7.774e-01  2.766e+00   0.281
## ideo5Very conservative:raceNative American  1.312e+01  8.271e+02   0.016
## ideo5Very liberal:raceNative American           -1.208e+00  1.172e+03  -0.001
## ideo5Liberal:raceOther                          -3.026e+00  7.422e+02  -0.004
## ideo5Moderate:raceOther                         -3.029e+00  1.201e+00  -2.522
## ideo5Not sure:raceOther                         -4.306e+00  2.487e+00  -1.732
## ideo5Very conservative:raceOther                 1.063e+01  4.077e+02   0.026
## ideo5Very liberal:raceOther                      9.748e+00  7.166e+02   0.014
## ideo5Liberal:raceWhite                           1.163e+01  3.478e+02   0.033
## ideo5Moderate:raceWhite                         -1.822e+00  4.742e-01  -3.843
## ideo5Not sure:raceWhite                         -2.623e-01  2.024e+00  -0.130
## ideo5Very conservative:raceWhite                -2.577e-01  1.579e+00  -0.163
## ideo5Very liberal:raceWhite                      1.065e+01  7.166e+02   0.015
##                                                 Pr(>|t|)
## (Intercept)                                     0.016739 *
## genderMale                                      0.419369
## educ4-year                                      1.10e-05 ***
## educHigh school graduate                        0.035932 *
## educNo HS                                       0.002133 **
## educPost-grad                                   7.59e-05 ***
## educSome college                                0.111185
## ideo5Liberal                                    0.959801
## ideo5Moderate                                   0.010921 *
## ideo5Not sure                                   0.458566
## ideo5Very conservative                          0.416169
## ideo5Very liberal                               0.980791
## raceBlack                                       4.04e-08 ***
## raceHispanic                                    0.138443
## raceMiddle Eastern                              0.367132
## raceMixed                                       0.044516 *
## raceNative American                             0.187907
```

```
## raceOther                                 0.025639 *
## raceWhite                                 3.90e-05 ***
## edloanYes                                 3.05e-05 ***
## faminc$100,000 - $119,999                 0.541322
## faminc$120,000 - $149,999                 0.317933
## faminc$150,000 - $199,999                 0.467083
## faminc$150,000 or more                    0.710572
## faminc$20,000 - $29,999                   0.573534
## faminc$200,000 - $249,999                 0.529204
## faminc$250,000 - $349,999                 0.735564
## faminc$30,000 - $39,999                   0.956409
## faminc$350,000 - $499,999                 0.841184
## faminc$40,000 - $49,999                   0.300265
## faminc$50,000 - $59,999                   0.000758 ***
## faminc$500,000 or more                    0.223074
## faminc$60,000 - $69,999                   0.797571
## faminc$70,000 - $79,999                   0.147396
## faminc$80,000 - $99,999                   0.024446 *
## fammincLess than $10,000                  0.363756
## famincPrefer not to say                   0.002613 **
## genderMale:raceBlack                      0.009149 **
## genderMale:raceHispanic                   0.002703 **
## genderMale:raceMiddle Eastern             0.181511
## genderMale:raceMixed                      0.051110 .
## genderMale:raceNative American            0.512639
## genderMale:raceOther                      0.022767 *
## genderMale:raceWhite                      0.178298
## ideo5Liberal:raceBlack                    0.969650
## ideo5Moderate:raceBlack                   0.078172 .
## ideo5Not sure:raceBlack                   0.906692
## ideo5Very conservative:raceBlack          0.522465
## ideo5Very liberal:raceBlack               0.984974
## ideo5Liberal:raceHispanic                 0.970360
## ideo5Moderate:raceHispanic                0.226360
## ideo5Not sure:raceHispanic                0.314700
## ideo5Very conservative:raceHispanic       0.949538
## ideo5Very liberal:raceHispanic            0.987375
## ideo5Liberal:raceMiddle Eastern           0.969158
## ideo5Moderate:raceMiddle Eastern          0.136648
## ideo5Not sure:raceMiddle Eastern          0.996986
## ideo5Very conservative:raceMiddle Eastern  0.995790
## ideo5Very liberal:raceMiddle Eastern      0.999388
## ideo5Liberal:raceMixed                    0.972033
## ideo5Moderate:raceMixed                   0.004641 **
## ideo5Not sure:raceMixed                   0.506412
## ideo5Very conservative:raceMixed          0.279320
## ideo5Very liberal:raceMixed               0.998201
## ideo5Liberal:raceNative American          0.968906
## ideo5Moderate:raceNative American         0.908142
## ideo5Not sure:raceNative American         0.778681
## ideo5Very conservative:raceNative American 0.987348
## ideo5Very liberal:raceNative American     0.999178
## ideo5Liberal:raceOther                    0.996747
## ideo5Moderate:raceOther                   0.011670 *
## ideo5Not sure:raceOther                   0.083339 .
## ideo5Very conservative:raceOther          0.979197
## ideo5Very liberal:raceOther               0.989148
```

```
## ideo5Liberal:raceWhite                           0.973331
## ideo5Moderate:raceWhite                          0.000122 ***
## ideo5Not sure:raceWhite                          0.896882
## ideo5Very conservative:raceWhite                 0.870334
## ideo5Very liberal:raceWhite                      0.988143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.8695947)
##
##     Null deviance: 16783.2  on 14123  degrees of freedom
## Residual deviance:  8187.4  on 14045  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 15
```

Note: Interaction effects have been ignored due to complex computations.

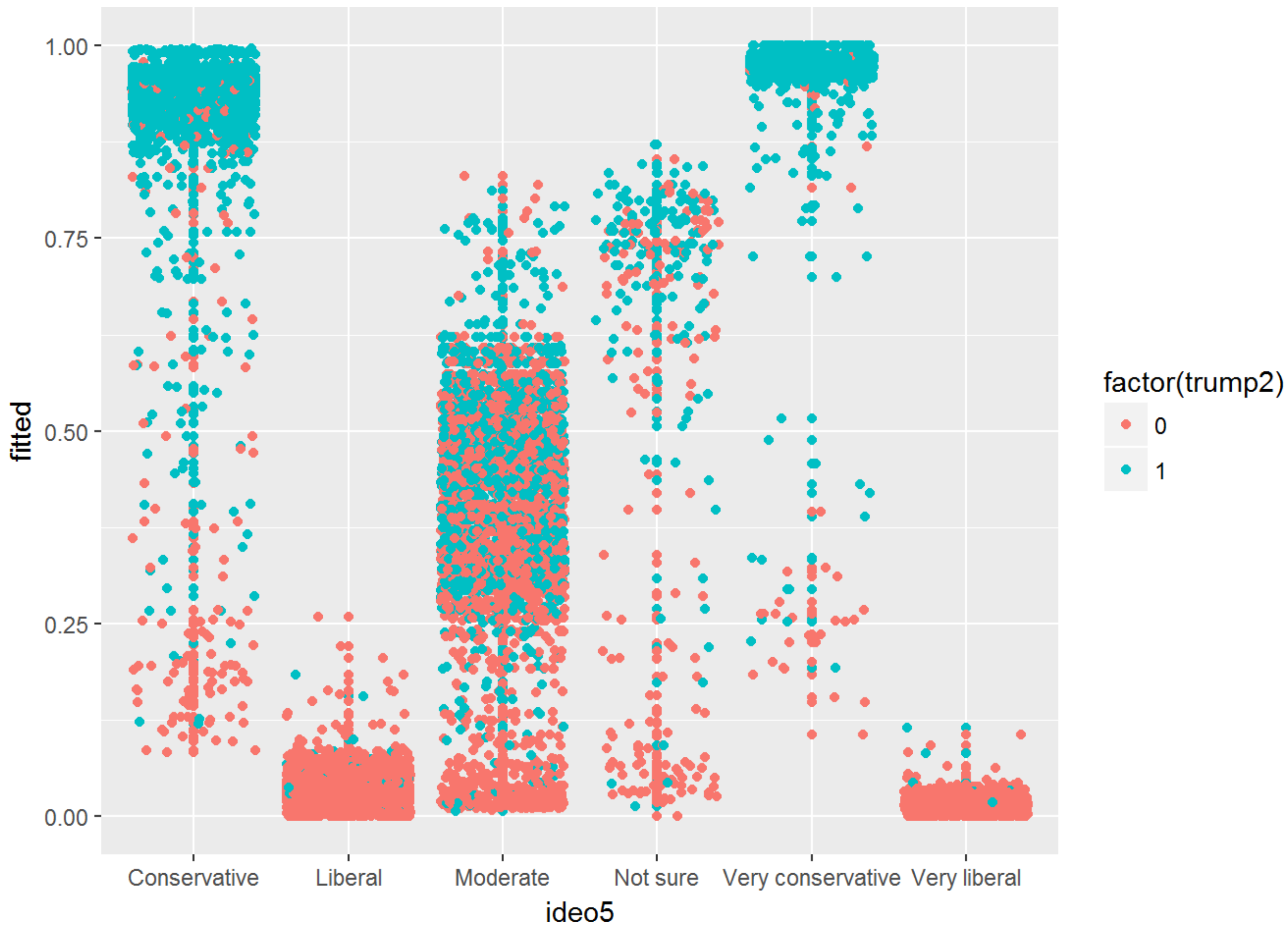
8.Interpretations:

- The baseline level here is the respondent being a white male with no high school education, a very liberal ideology and family income less than $10,000.
- Gender: The coefficient of gender female is -0.24. This says that the odds of a female voting to trump are 21 percent lower than males with the characteristics of baseline level.
- Education: We see that as the education level increases, the chances to vote for trump instead of Hillary reduce. People with an education have 40% lesser odds of voting to trump than people with education less than high school.
- Ideology: The chance of people voting to Trump instead of Hillary have been the lowest for those who termed themselves as very liberals and liberals. As the ideology level moved toward being more conservative, support to trump soared.
- Race: From the coefficients, the odds of Blacks and Hispanics voting to Trump are 50% lesser than Whites. The odds of Native Americans voting to Trump are way lesser than that. However, Trump has a slightly higher support in Asian community.
- Educational Loan: Respondents with educational loan have less chance to vote to Trump. Hillary’s policies towards educational loan debt might explain this behaviour.

- The interaction effects shows that in females, black and Middle eastern females especially have lesser odds to support trump.
- Also, Blacks and Hispanics tend to oppose Trump irrespective of their ideologies. However, Native americans supported Trump at all ideology levels.

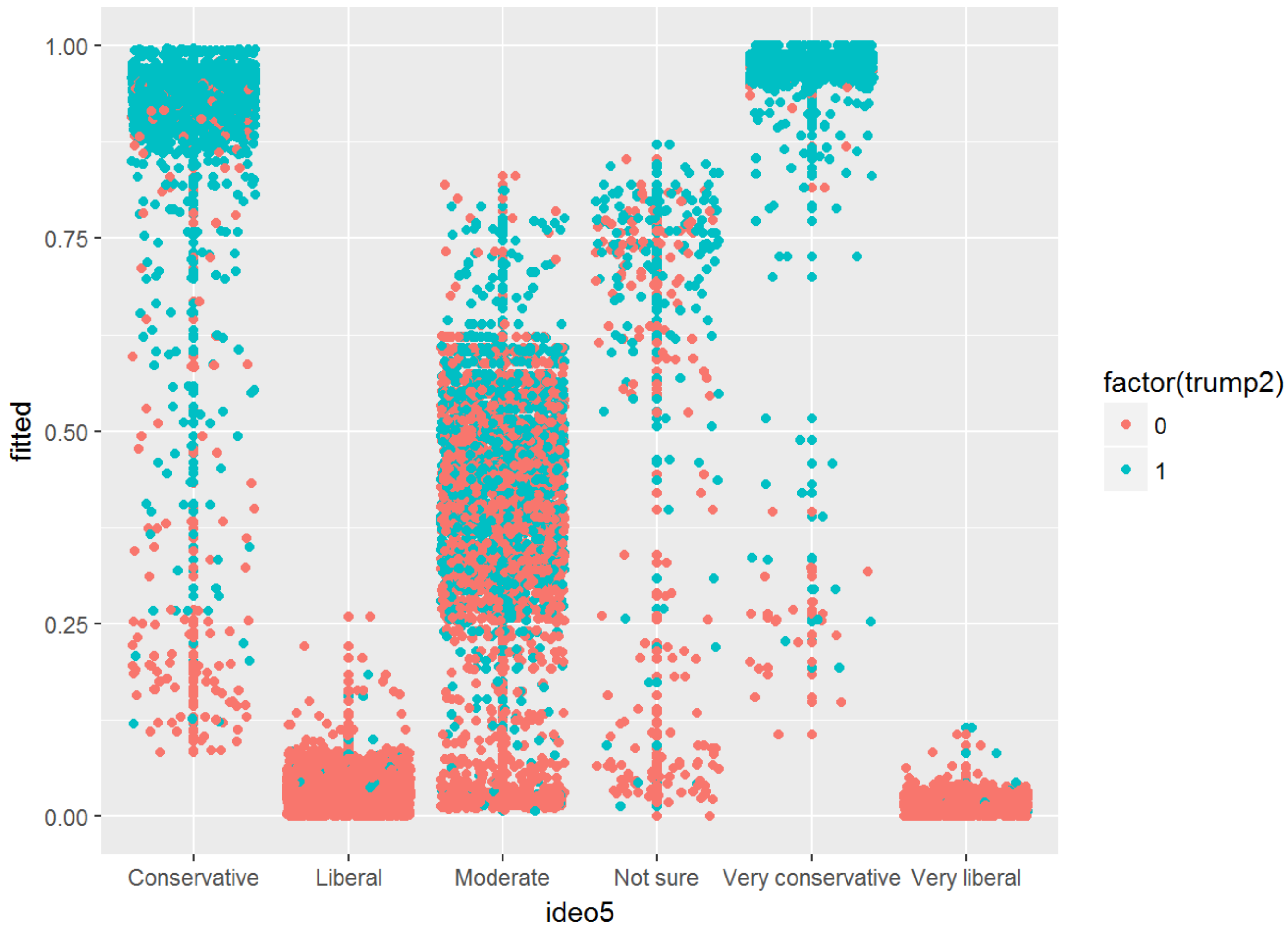
9. Some visualizations with the model fitted values.

```
CCES_sub$fitted = fitted.values(m4.US)
ggplot(CCES_sub,aes(x= ideo5, y = fitted, color = factor(trump2)))+geom_point()+ge
om_jitter()
```




```
ggplot(CCES_sub,aes(x= gender, y = fitted, color = factor(trump2)))+geom_point()+g
eom_jitter()
```


```
ggplot(CCES_sub,aes(x= ideo5, y = fitted, color = factor(trump2)))+geom_point()+ge
om_jitter()
```



10.Model Testing

```
with(m4.US, pchisq(null.deviance - deviance, df.null - df.residual, lower.tail = F
ALSE))
```

```
## [1] 0
```

```
Anova(m4.US, test = "Wald")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: trump2
##              Df    Chisq Pr(>Chisq)
## gender        1   35.051  3.211e-09 ***
## educ          5  114.622  < 2.2e-16 ***
## ideo5         5 3145.649  < 2.2e-16 ***
## race          7  639.904  < 2.2e-16 ***
## edloan        1   17.398  3.031e-05 ***
## faminc       17   40.328  0.0011633 **
## gender:race   7   28.616  0.0001699 ***
## ideo5:race   35   80.166  2.125e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
Anova(m4.US, test = "LR")
```

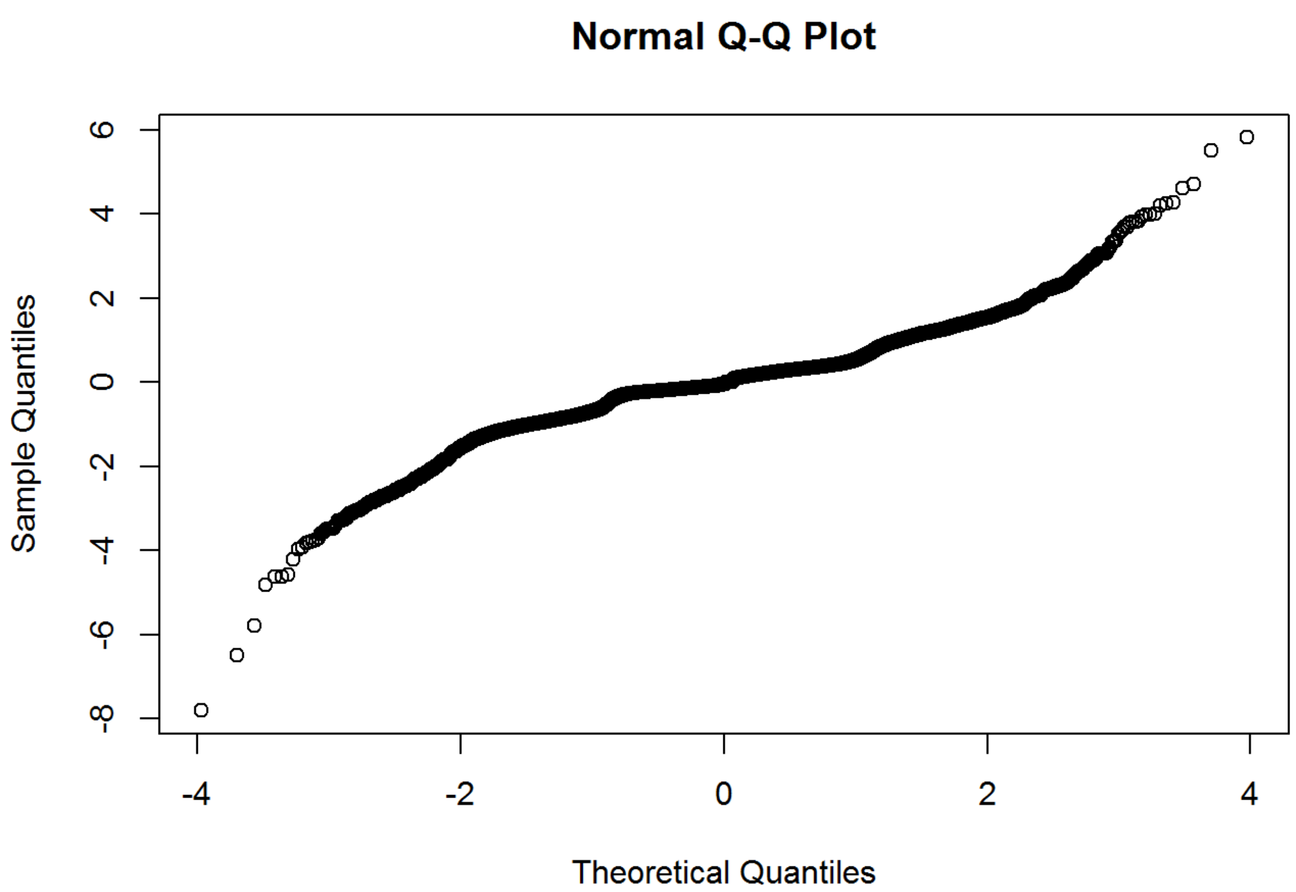
```
## Analysis of Deviance Table (Type II tests)
##
## Response: trump2
##              LR Chisq Df Pr(>Chisq)
## gender           35.6  1  2.431e-09 ***
## educ            116.3  5  < 2.2e-16 ***
## ideo5          7781.1  5  < 2.2e-16 ***
## race            984.8  7  < 2.2e-16 ***
## edloan           17.5  1  2.875e-05 ***
## faminc           40.6 17   0.001059 **
## gender:race      31.5  7  5.078e-05 ***
## ideo5:race       88.4 35  1.667e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The model is significant compared to the null model.Both the Wald and Likelihood Ratio tests indicate that the coefficients can significantly explain the response variable.

11.Diagnostics

```r
##Normality check- qqplot

qqnorm(residuals(m4.US))
```



- The qq norm plot is mostly linear at the middle, with curvature at the ends which says that the assumoption of normality is satisfied.

Residuals vs fitted plot with bins

```
head(predict(m4.US)) ## predicted linear responses
```

```
##          1          2          3          4          5          6
##  0.1798674  0.2561483 -3.7713977  4.1422545  3.0478842 -0.2356143
```

```
head(fitted(m4.US)) ## predicted probabilities
```

```
##          1          2          3          4          5          6
## 0.54484602 0.56368922 0.02250188 0.98436145 0.95469109 0.44136743
```

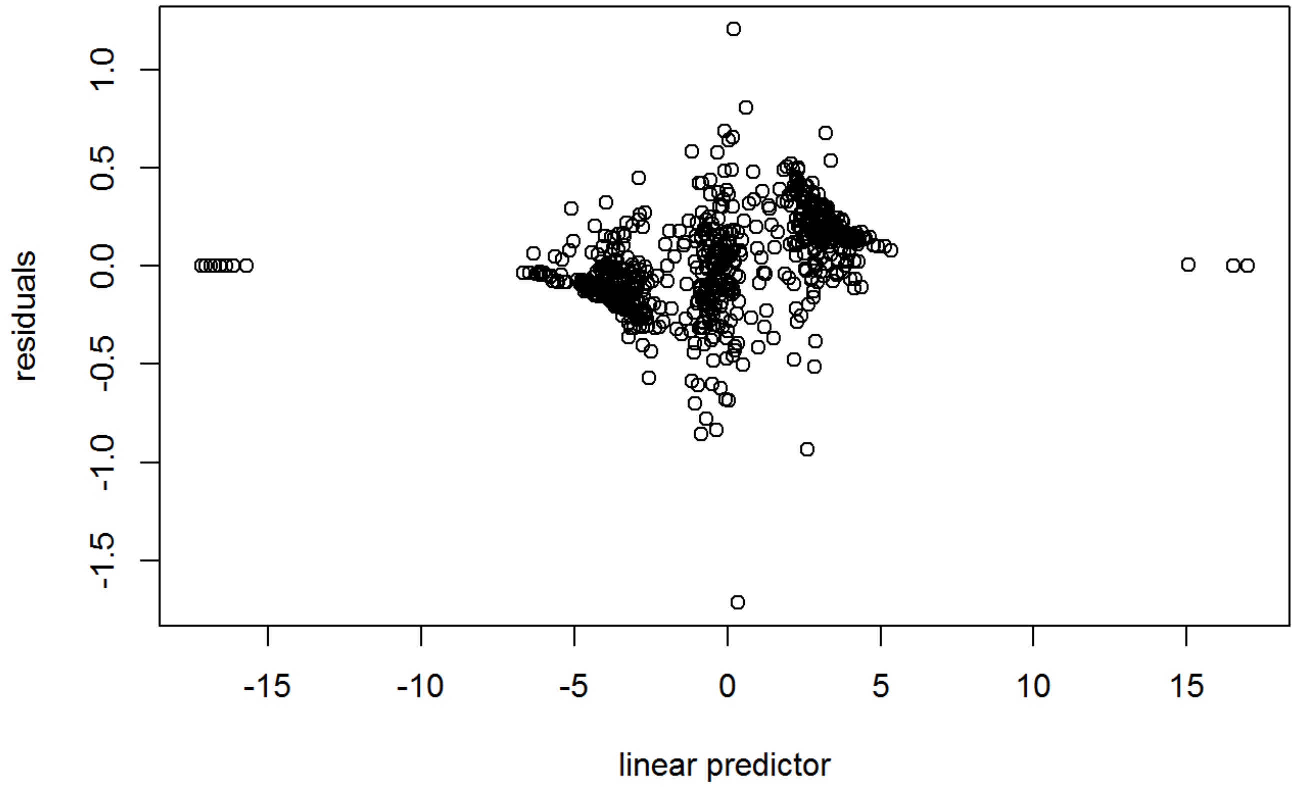
```
CCES_sub = mutate(CCES_sub, residuals = residuals(m4.US), eta = predict(m4.US))
gdf = group_by(CCES_sub, cut(eta, breaks=unique(quantile(eta, (1:1000)/1001))))
dim(gdf)
```

```
## [1] 14124    15
```

```
gdf[,14]
```

```
## # A tibble: 14,124 × 1
##           eta
##         <dbl>
## 1   0.1798674
## 2   0.2561483
## 3  -3.7713977
## 4   4.1422545
## 5   3.0478842
## 6  -0.2356143
## 7  -0.5554013
## 8  -1.6504335
## 9  -3.1415747
## 10  0.2831071
## # ... with 14,114 more rows
```

```
diagdf = summarise(gdf, residuals=mean(residuals), eta=mean(eta))
plot(residuals ~ eta, diagdf, xlab="linear predictor")
```



- The residuals plot indicate there might be hints of some heteroscedasticity in the data. - Further analysis should be done to rectify this issue which I could not carry out as a part of this analysis.

Accuracy and ROC plot

```
CCES_sub = mutate(CCES_sub, predprob = predict(m4.US, type = "response"))
CCES_sub = mutate(CCES_sub, predout=ifelse(predprob < 0.5, "no", "yes"))
xtabs( ~ trump2 + predout, CCES_sub)
```

```
##        predout
## trump2   no  yes
##      0 6517  804
##      1 1471 5332
```

```
##The correct classification rate
(19168+14117)/(2828+4545+19168+14117)
```

```
## [1] 0.8186581
```

```
##The misclassification rate
1-(19168+14117)/(2828+4545+19168+14117)
```

```
## [1] 0.1813419
```

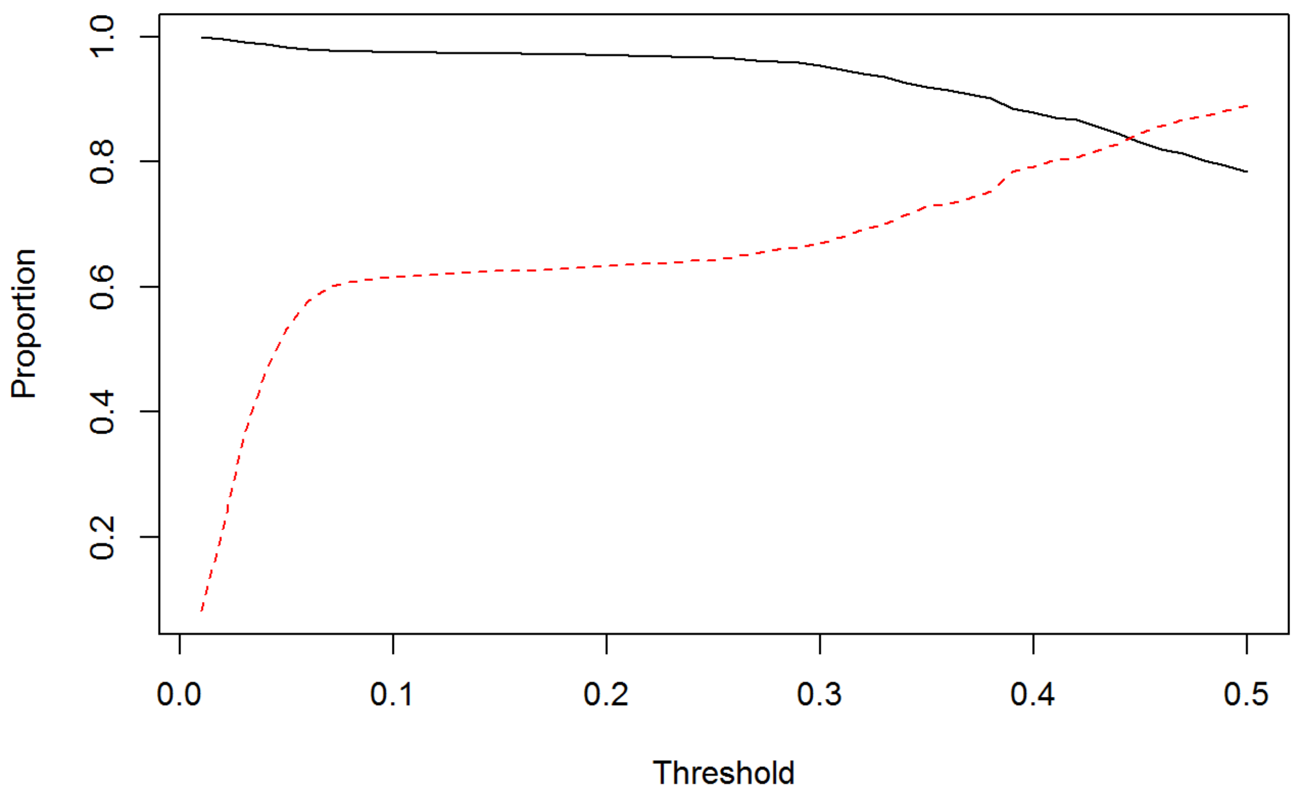
```
##Sensitivity
14117/(14117+4545)
```

```
## [1] 0.756457
```

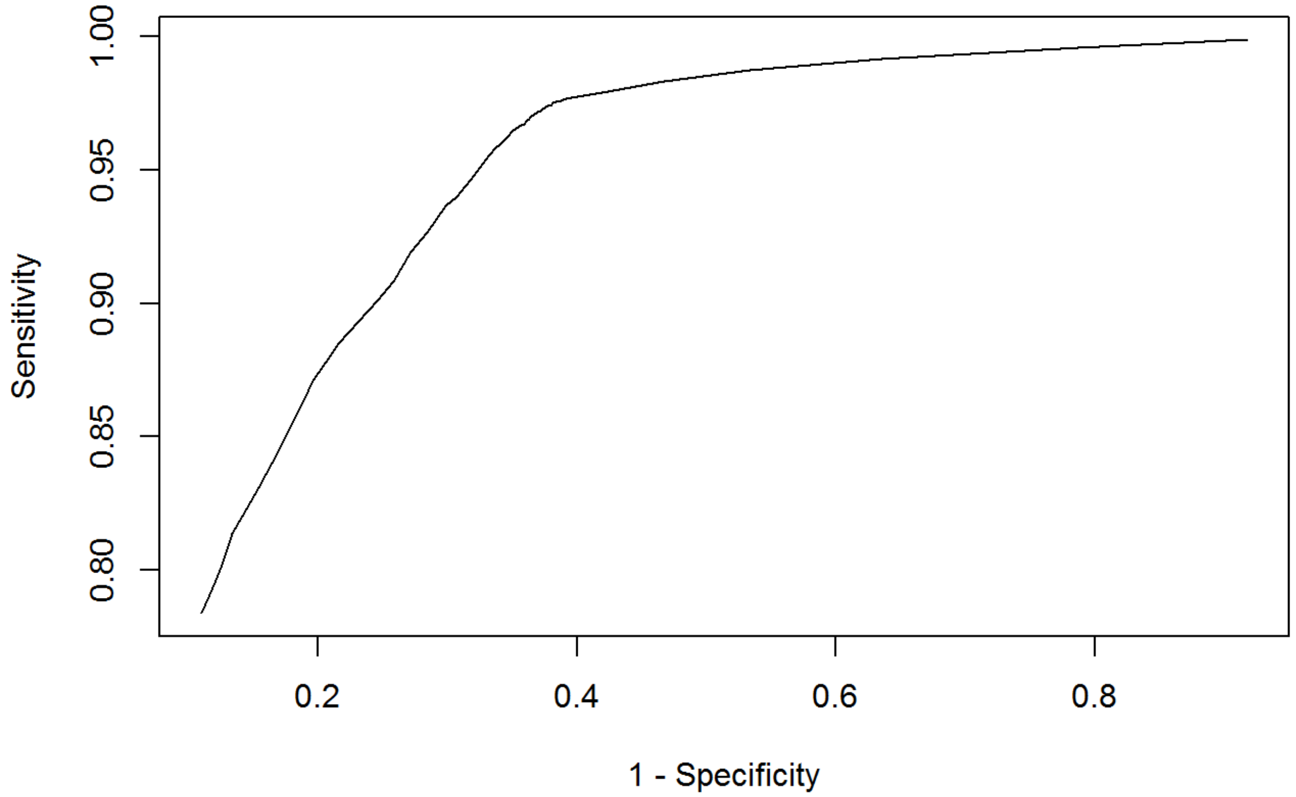
```
##Specificity
19168/(19168+2828)
```

```
## [1] 0.8714312
```

```
thresh = seq(0.01,0.5,0.01)
Sensitivity = numeric(length(thresh))
Specificity = numeric(length(thresh))
for(j in seq(along=thresh)){
  pp = ifelse(CCES_sub$predprob < thresh[j],"no","yes")
  xx = xtabs( ~ trump2 + pp, CCES_sub)
  Specificity[j] = xx[1,1]/(xx[1,1]+xx[1,2])
  Sensitivity[j] = xx[2,2]/(xx[2,1]+xx[2,2])
}
matplot(thresh,cbind(Sensitivity,Specificity),type="l",xlab="Threshold",ylab="Propo
rtion",lty=1:2)
```




```
plot(1-Specificity,Sensitivity,type="l")
```



- The model has good sensitivity of 0.75 and specificity of 0.87 at the given threshold of 0.5.
- However, at a threshold of 0.45, we have the ideal match of sensitivity and specificity.

12.Conclusion

- From the analysis performed on the CCES survey data, it can be understood that Trump's victory can be attributed predominantly to white conservative population with income levels less than 30000 dollars. Education level of the voters also played a predominant role in Trump's victory.
- Women, blacks, Hispanics have extended their support to Hillary, however there is a reduction in support to Hillary compared to Obama, particularly in Black Male population.