Mini Project 2

Apurva Gupta, Shailendra Patil, Surbhi Paithankar

March 9, 2018

Question 0: Pre-processing Data

Starting with the pre-processing,we created a subset of respondents who responded to the post election survey and who voted for Barack Obama in 2012.We also created a binary variable that indicates whether the respondent voted for Trump or not.(1 indicates that respondent voted for Trump)

We will need to measure respondent's attitude towards immigration. For this,we created a new variable which counts the number of positive responses towards immigrants across the four variables.

Some of the racial categories have small amount of data. So, we recoded this variable to have four levels:"White","Black","Hispanic","Other".

```
## [1] "integer"
```

Finally we converted the ordinal features:Party and education to numerical values.

```r
nrow(respondants)
```

```
## [1] 23395
```

```r
sum(respondants$trump)
```

```
## [1] 2121
```

After pre-processing, we obtained data consisting of 23,395 individuals who voted for Obama in 2012. Out of these, 2121 said that they voted for Trump in 2016.

Question 1:Weighted proportion of Obama voters for different demographic groups that switched to Trump

Going ahead, We will find the weighted proportion for people belonging to different races who switched to trump in the 2016 elections. Furthermore, we calculate the weighted proportions for people of various educational levels in the same way.

Lets check the table with the weighted proportion of Obama voters for each demographic group.

```r
educ_proportion
```

```
##              Education Proportion
## 1               2-year      0.122
## 2               4-year      0.084
## 3 High school graduate      0.148
## 4                No HS      0.107
## 5            Post-grad      0.063
## 6         Some college      0.105
```

```r
race_proportion
```

```
##       Race Proportion
## 1    Black      0.037
## 2 Hispanic      0.090
## 3    Other      0.081
```

```
## 4    White      0.138
```

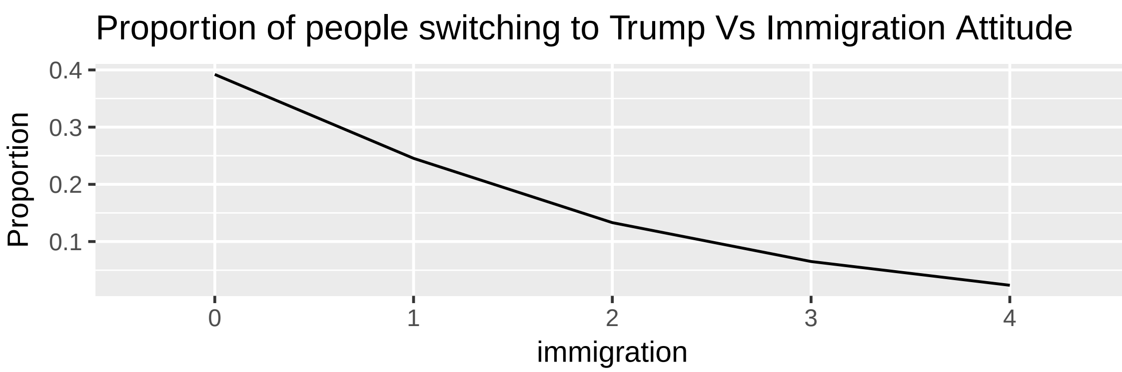
```
gender_proportion
```

```
##   Gender Proportion
## 1 Female      0.098
## 2   Male      0.122
```

```
pid_proportion
```

```
##                        Party Proportion
## 1                Independent      0.246
## 2              Lean Democrat      0.047
## 3            Lean Republican      0.620
## 4                   Not sure      0.312
## 5   Not very strong Democrat      0.115
## 6 Not very strong Republican      0.501
## 7            Strong Democrat      0.029
## 8          Strong Republican      0.683
```

Now we will go ahead and see how the attitude towards immigration changed the tendency to switch from Obama to Trump. To do this, we calculated the weighted proportion for each immigration attitude. Lets look at the graph between these proportions against various immigration attitude now.



Lets look in quantitative terms as well.

```
##   count trump
## 1     0 0.376
## 2     1 0.226
## 3     2 0.123
## 4     3 0.055
## 5     4 0.017
```

We observe that as the attitude towards immigration tends to be more positive(closer to 4), the respondents' tendency of switching to Trump in 2016 decreases. This tells us that about ~38% of the people, who were strongly against immigrants switched to Trump in 2016, whereas the strong supporters of immigrants did not switch.

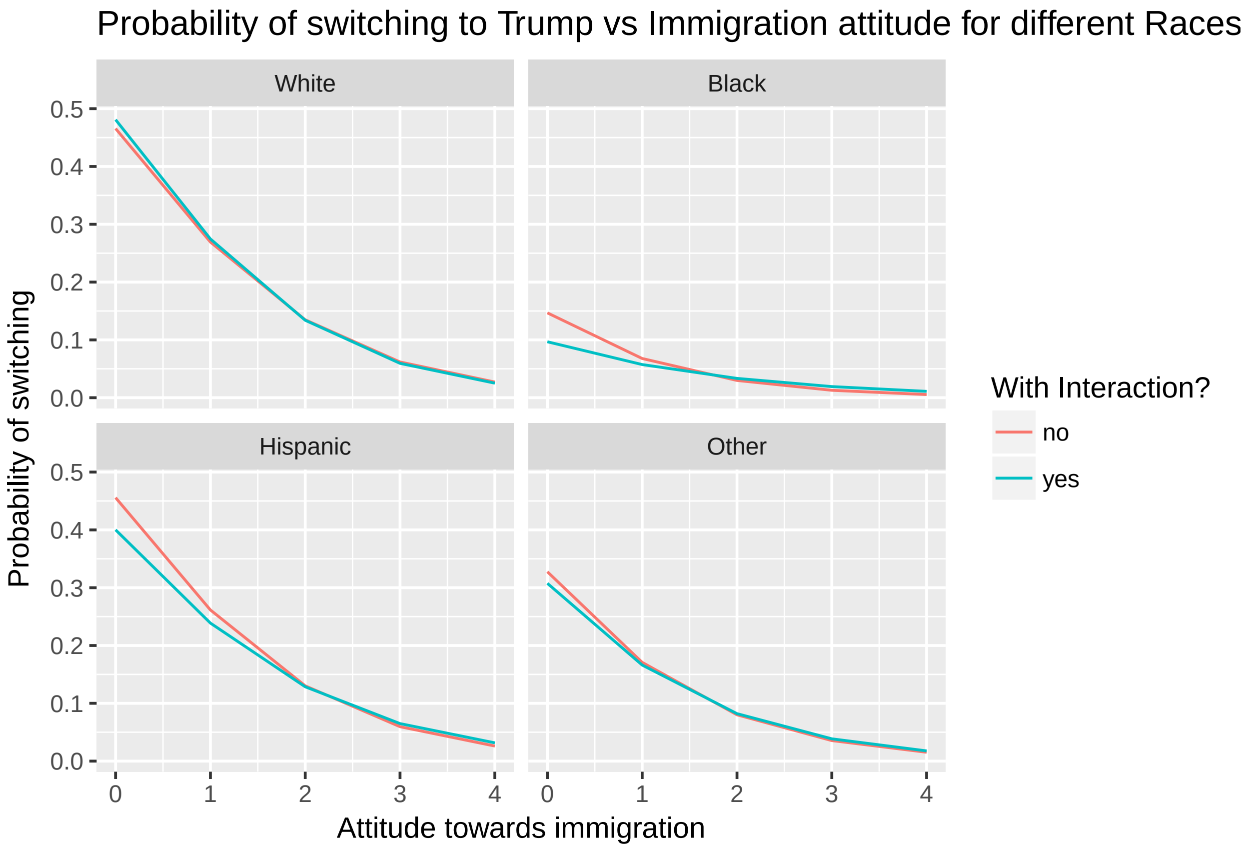
Question 2:Effect of immigration attitude on various demographic categories.

Immigration attitude vs Switching probability for different Races

We will fit a logistic regression using immigration attitude and race.We first fit an additive model with no interaction.

Now we will model with interaction between immigration attitude and Race.

We visualize the fit by drawing curves representing different values of the predictors. Lets display the switching probability as a function of immigration attitude with respect to race. Plotting the graphs using both the models with and without interaction.



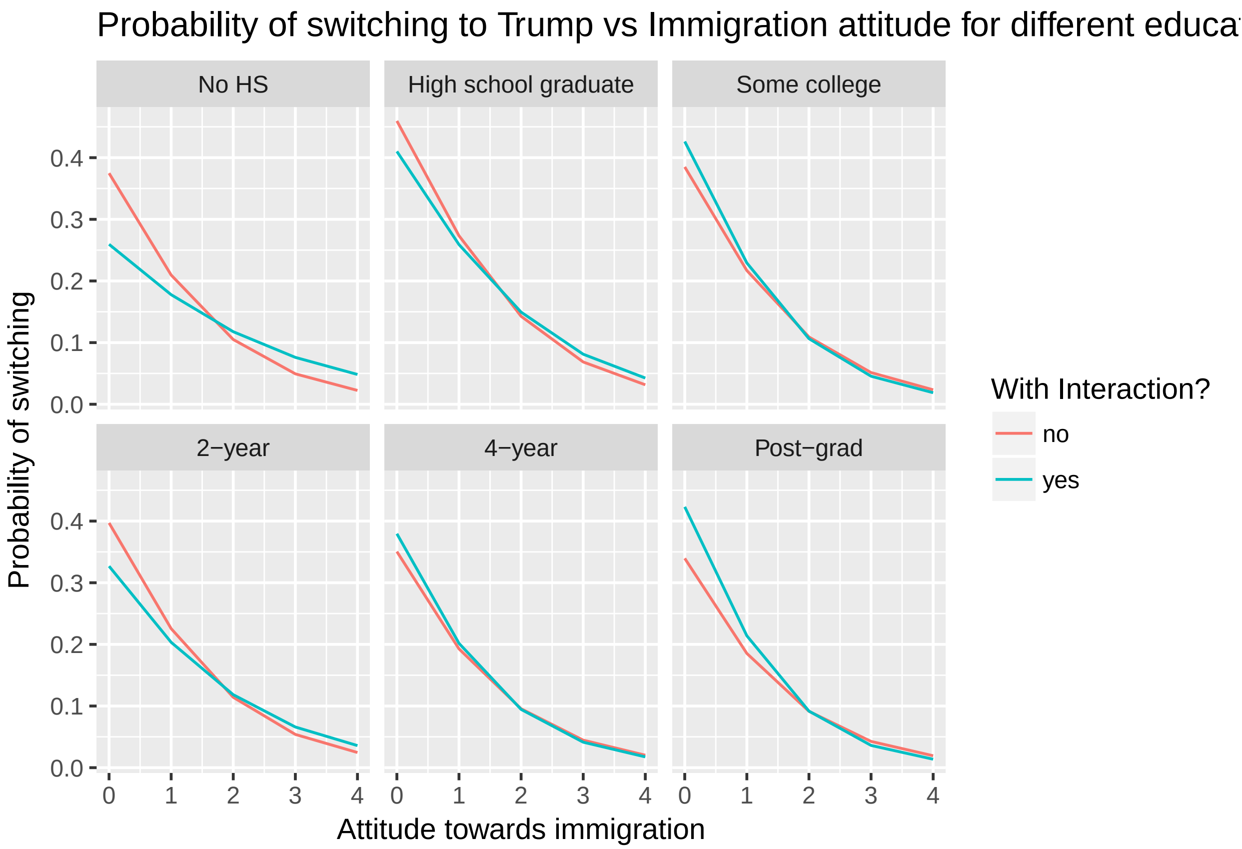
Looking at the graphs we observe that there is an interaction between race and the immigration attitude. The slopes of the curves for hispanic and blacks show a shift of 5% which is significant.

Immigration attitude vs Switching probability for different educational levels.

We will fit a logistic regression using immigration attitude and education.We first fit an additive model with no interaction.

Additionally,we will model with interaction between immigration attitude and education.

Lets visualize by plotting the graphs using both the models with and without interaction.



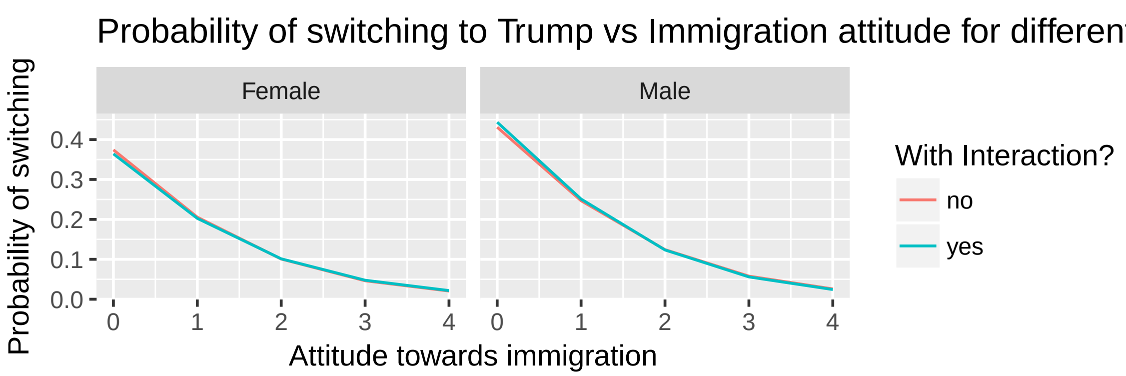
In the above graphs, we can see that there is a difference of almost 5-10% with & without interaction. Hence we can call the interaction between education and immigration attitude to be significant.

Immigration attitude vs Switching probability for different Genders.

We will fit a logistic regression using immigration attitude and genders.We first fit an additive model with no interaction.

Additionally,we will model with interaction between immigration attitude and gender.

Lets visualize by plotting the graphs using both the models with and without interaction.



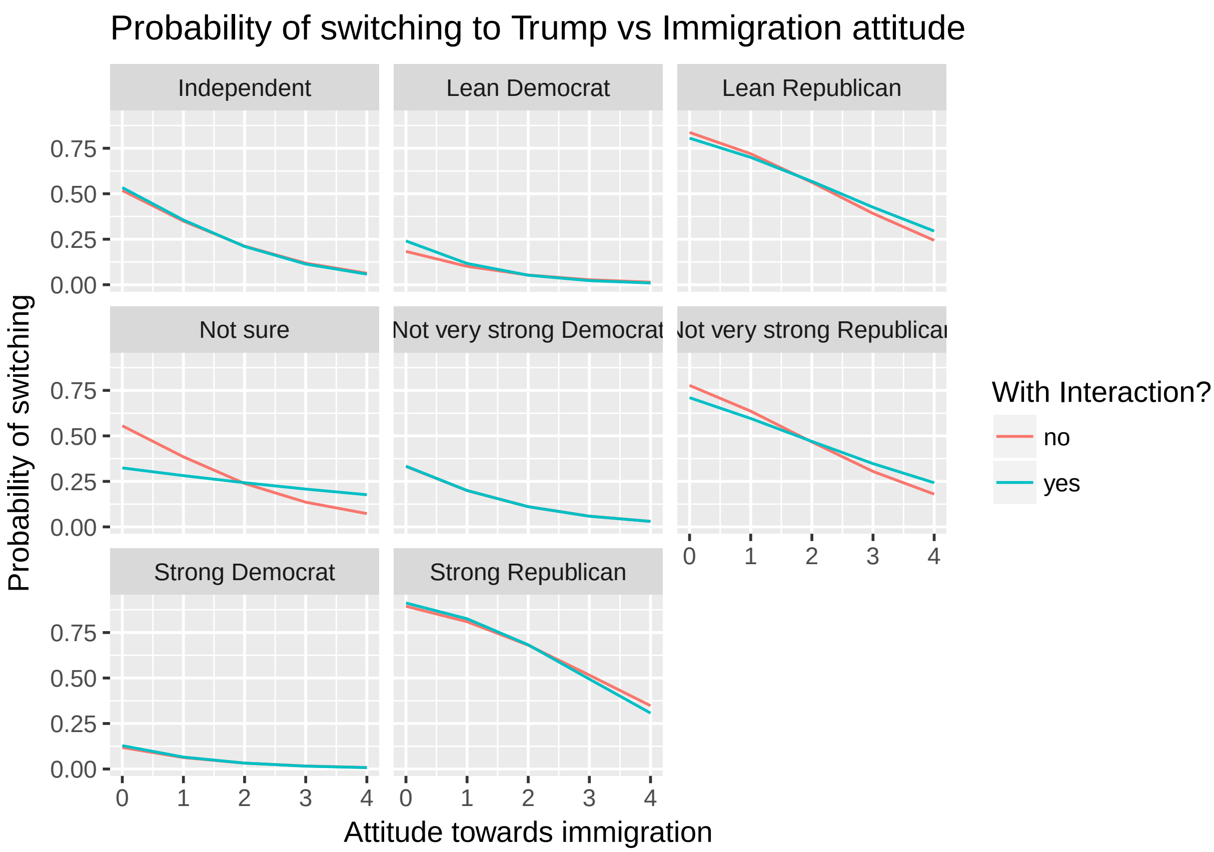
Looking at the graphs we observe that the graphs with/without interaction are almost similar.Hence we can say that, there is no interaction between gender and the immigration attitude.

Immigration attitude vs Switching probability for different Parties.

We will fit a logistic regression using immigration attitude and political parties.We first fit an additive model with no interaction.

Additionally,we will model with interaction between immigration attitude and political parties.

Lets visualize by plotting the graphs using both the models with and without interaction.



The curves for predictions using models with and without interaction are different for most of the political parties. Hence we will keep this interaction in our model.

Question 3:Effect of immigration attitude on various demographic categories using lots of predictors.

We have following demographic predictors: 1. Race 2. Education 3. Gender 4. Party ID 5. Immigration attitude

Firstly lets fit a logistic regression model without considering the immigration attitude. We will begin with fitting a model with interaction between education and party ID.

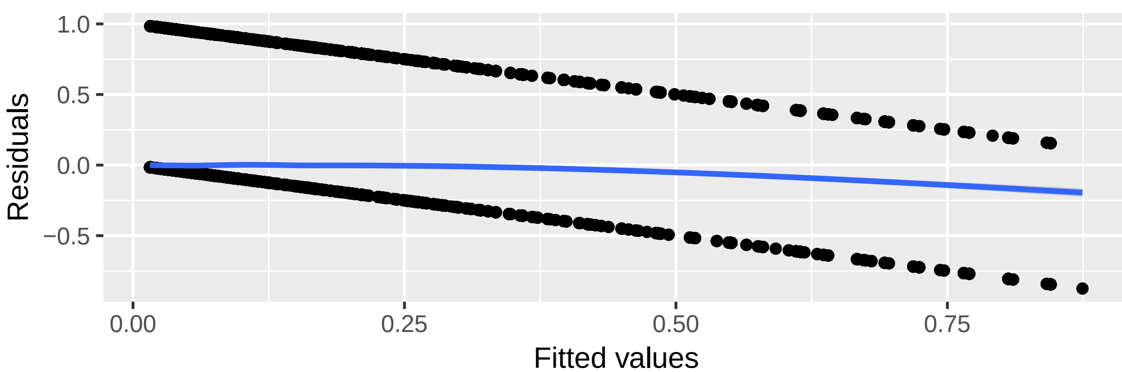
```
##
## Call:
## glm(formula = trump ~ educ_numeric * pid7_numeric + race + gender,
##     family = "binomial", data = respondants)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -1.9072  -0.4151  -0.2926  -0.2061   3.1580
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -2.732162   0.140497 -19.446  < 2e-16 ***
## educ_numeric               -0.242428   0.034875  -6.951 3.62e-12 ***
## pid7_numeric                0.571814   0.035918  15.920  < 2e-16 ***
## raceBlack                  -1.138149   0.105339 -10.805  < 2e-16 ***
## raceHispanic               -0.239253   0.096677  -2.475   0.0133 *
## raceOther                  -0.448771   0.105776  -4.243 2.21e-05 ***
## genderFemale               -0.257702   0.050349  -5.118 3.08e-07 ***
## educ_numeric:pid7_numeric   0.005213   0.009090   0.574   0.5663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 14227  on 23394  degrees of freedom
## Residual deviance: 11500  on 23387  degrees of freedom
## AIC: 11516
##
## Number of Fisher Scoring iterations: 6
```

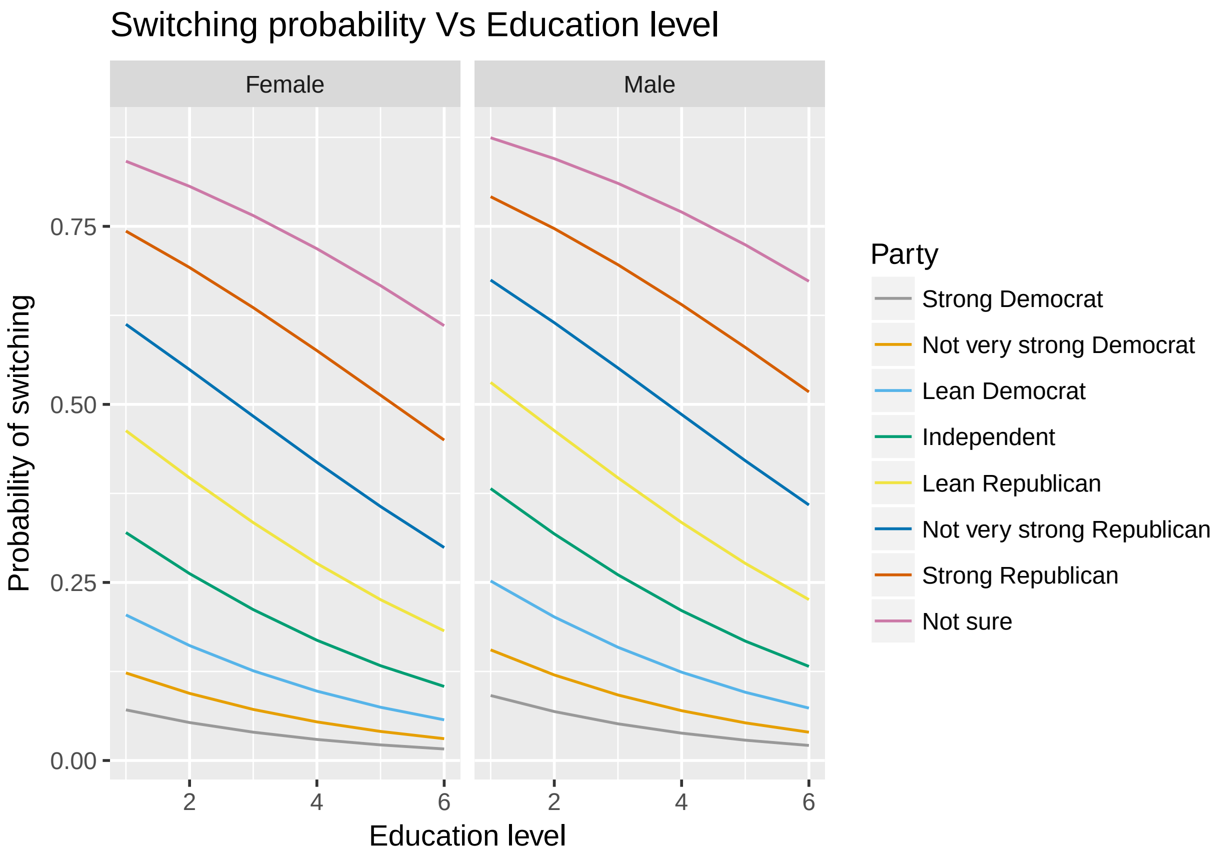
```
## [1] 11499.57
```

Looking at the coefficients, a positive value for party Id indicates that person who is more inclined towards Republicans are more likely to switch. Furthermore, as we go from whites to blacks, the chances to switch decreases. Finally females are less likely to switch compared to males. We have a residual deviance of 11500. Lets incorporate some more interactions in our model, just to make it better. Additionally, the interactions between race & party, race & education looks to be sensible. Lets fit a more complex model and check the deviance.

```
## [1] 11413.36
```

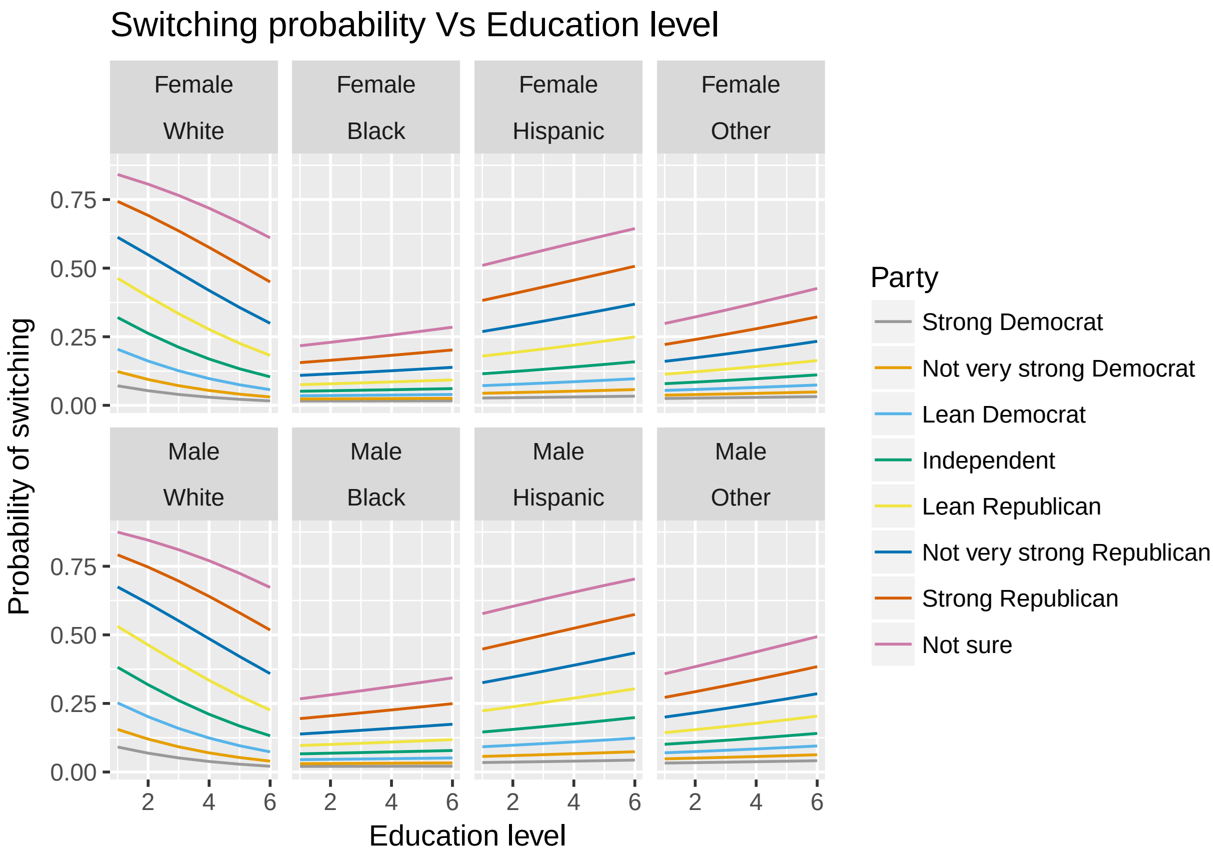


The deviance has reduced but not significantly. The plot shows a bend at the tails. However the model looks to do a fine job. Going ahead, we will plot our predictions using this model. Lets fix race: white.



In the plot above we observe that white males are 5% more likely to switch from Obama to Trump. Moreover, as the education of a person increases, he is less likely to switch. We also notice that the political preference of a person also contributes to shifting of votes. Republicans are more likely to switch compared to democrats. The probability of switching increases by ~10% as the proclivity changes from democrats to republicans.

We will now look for at all the races and see if we can see similar behaviors in them as well.



For Blacks, the effect of education level is low. The lines are almost horizontal & parallel.In case of Hispanics and Other races as the education level increases the probability of switching increases as opposed to the behavior observed in "Whites". However the effect of gender looks to be similar in all the races.

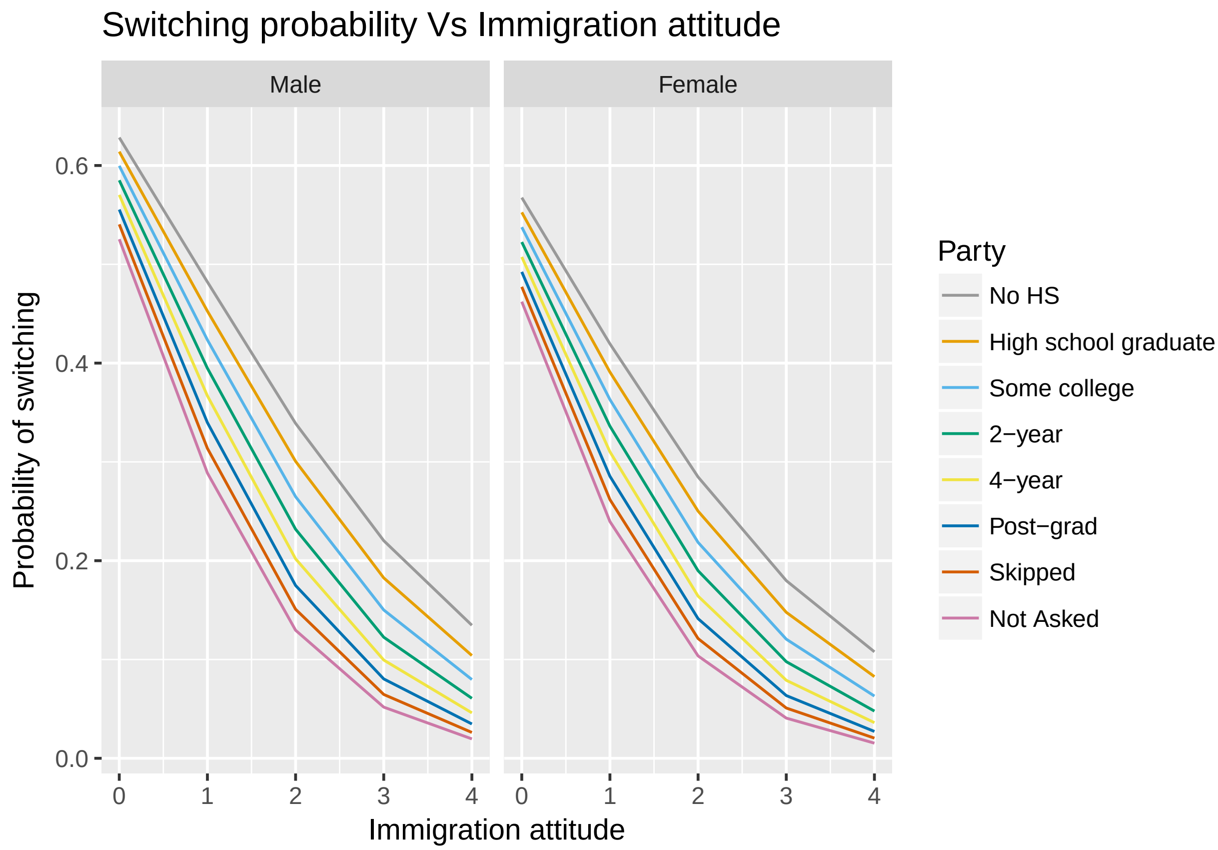
Now we will go ahead and fit another logistic model with immigration attitude as a predictor as well. From the question 2, we observed a significant interaction going on between immigration attitude: education,immigration attitude:race and immigration attitude:party. Therefore, we will keep them in our model along with the interactions in the previous model. Lets look at the summary of the new model.

```
##
## Call:
## glm(formula = trump ~ educ_numeric:count + pid7_numeric:count +
##     race:educ_numeric + race:pid7_numeric + educ_numeric:pid7_numeric +
##     race:count + race + pid7_numeric + gender + educ_numeric +
##     count, family = "binomial", data = respondants)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3017  -0.3452  -0.2163  -0.1450   3.1924
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -1.502110   0.189850  -7.912 2.53e-15 ***
## raceBlack                   -2.158564   0.381504  -5.658 1.53e-08 ***
## raceHispanic                -1.159667   0.389561  -2.977 0.002912 **
## raceOther                   -1.097181   0.465480  -2.357 0.018418 *
## pid7_numeric                 0.521878   0.042952  12.150  < 2e-16 ***
## genderFemale                -0.253279   0.053762  -4.711 2.46e-06 ***
```

```
## educ_numeric              -0.023512   0.047433  -0.496 0.620117    
## count                     -0.659997   0.066474  -9.929  < 2e-16 ***
## educ_numeric:count        -0.058102   0.014624  -3.973 7.09e-05 ***
## count:pid7_numeric         0.030424   0.010940   2.781 0.005418 ** 
## educ_numeric:raceBlack     0.221365   0.078710   2.812 0.004917 ** 
## educ_numeric:raceHispanic  0.288423   0.075815   3.804 0.000142 ***
## educ_numeric:raceOther     0.232469   0.080852   2.875 0.004037 ** 
## pid7_numeric:raceBlack    -0.136523   0.047826  -2.855 0.004309 ** 
## pid7_numeric:raceHispanic -0.069869   0.048713  -1.434 0.151487    
## pid7_numeric:raceOther    -0.208092   0.057660  -3.609 0.000307 ***
## educ_numeric:pid7_numeric -0.009241   0.009860  -0.937 0.348647    
## count:raceBlack            0.273980   0.082274   3.330 0.000868 ***
## count:raceHispanic         0.125036   0.078548   1.592 0.111422    
## count:raceOther            0.123400   0.080959   1.524 0.127453    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 14227  on 23394  degrees of freedom
## Residual deviance:  9967  on 23375  degrees of freedom
## AIC: 10007
## 
## Number of Fisher Scoring iterations: 7
```

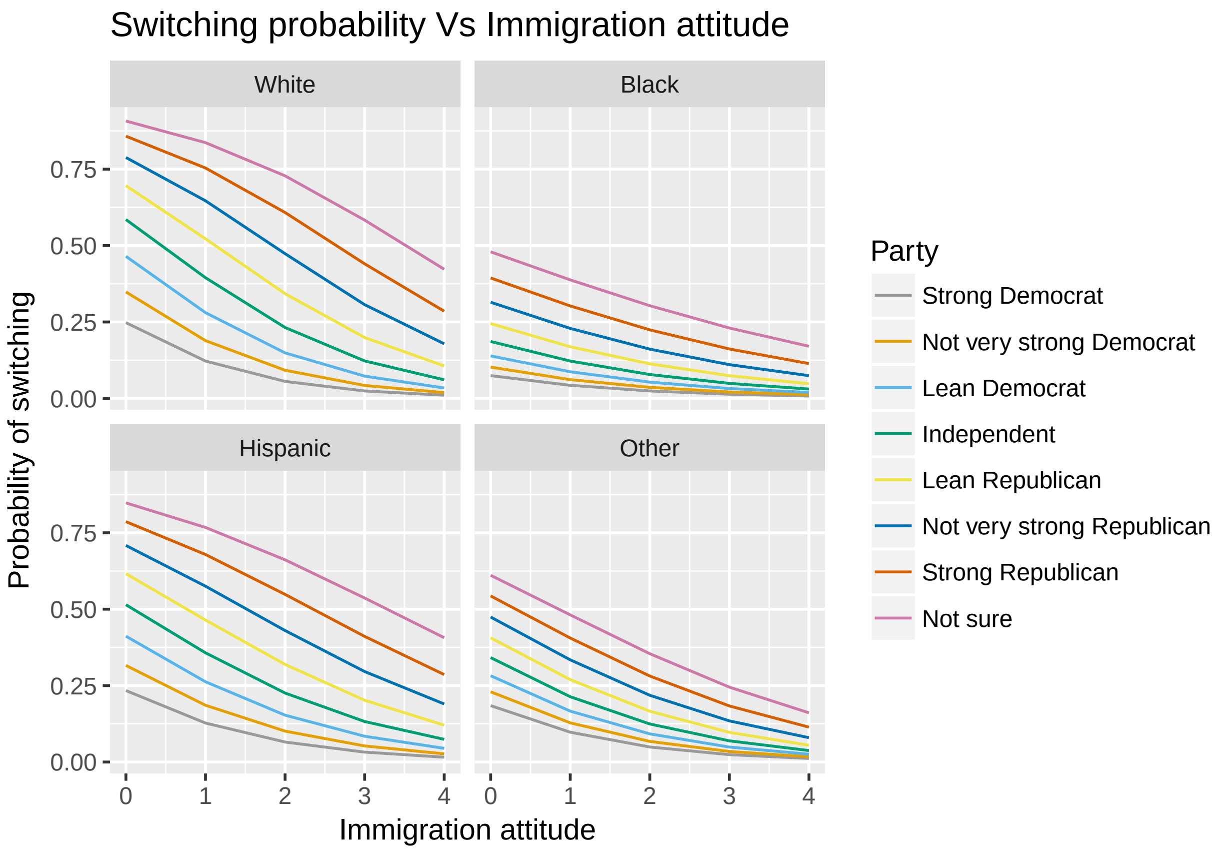
We obtain a residual deviance of 9980 which is a significant reduction when compared to the previous model. This tells us that the immigration attitudes plays a major role in predicting the probability of switching votes.

Lets try to visualize this graphically. To do this, we will fix the race as White belonging to independent party for various education levels.



We can clearly see that as the immigration attitude gets more positive, the probability of switching decreases. For every unit change of immigration attitude, the switching probability decreases by roughly 20%.

Lets again make the predictions. This time we will consider the males having median level of education. We will take all the races into consideration.



The above plots also indicate that for all the races, the immigration attitude affects the switching probability for all the political parties.

Conclusion

The fitted logistic models show that the immigration attitude makes a substantive difference to the probability of switching from Obama to trump. We can say that republicans were more likely to switch compared to democrats.With increase in education level, the switching probability decreases. Furthermore, females within same demographic group were 5% less probable to change their votes.

Finally, keeping all the demographic variables constant, had the immigration attitude was insignificant, we would expect the curves to be horizontal. However we see a systematic pattern in the curves. Hence we can conclude that immigration attitude played an important role for shifting their votes from Obama to Trump.