# Exploratory Data Analysis

## Mini Project 2: Votes switching from Obama to Trump in 2016 presidential election (S470/670)

Sai Ajay Vutukuri, Himani Anil Deshpande, Murali Krishna Parvataneni, Tushar Kant Samantaray

## Introduction

This project is concerned with understanding, **to what extent do attitudes toward immigration explain the switching of votes of 2012 Obama supporters who became Trump supporters and whether the attitude toward immigrants had influence over different demographic groups to shift**.

The data is made available by Cooperative Congressional Election Study. The CCES is a 50,000+ personal national stratified sample survey administered by YouGov. We will use the 2016 Cooperative Congressional Election Study, a very large survey of a nationally representative sample of 64,600 adults available here.

We divide the analysis into two sections to research the question:

1. Understanding how each of the demographic groups in consideration, namely race, gender, education level, and party identification, sway the attitude towards immigration.
2. Does attitude towards immigration make a substantive difference for the voters to switch from Obama to Trump and whether it matters more for some demographic groups than others.

We use weighted logistic regression models to understand how demographic groups affect attitude towards immigration as we have a binary classification problem where we try to analyze if a person switches their vote from Obama to Trump based on multiple conditions. Logistic regression is a predictive analysis method that is used to describe data and to explain the relationship between one dependent binary variable against one or more independent variables. Since, for various reasons like design and nonresponse bias, modern survey results are rarely a true simple random sample of the population, survey results are **weighted** to adjust for groups being underrepresented or overrepresented in a sample.

For easier analysis, we categorized the 'Strong Democrat', 'Not very Strong Democrat', and 'Lean Democrat' as 'Democrat', and 'Strong Republican, 'Not very Strong Republican', and 'Lean Republican' as 'Republican' for easier analysis. Further, we categorized missing and 'Not sure' values as 'Other'.

## 1. Understanding how each of the demographic groups in consideration, namely race, gender, education level, and party identification, sway the attitude towards immigration.

All the demographic variables, which are race, gender, party identification, and education level, along with the attitude towards immigration are meaningful to us. We are provided with the `commonweight_vv_post` survey weight variable for people who took the post-election survey. However, this does not provide us with information about the individual impact of the concerned variable of voters to switch from Obama to Trump. The weight of evidence (WOE) and information value (IV) provide a great framework for exploratory analysis and variable screening for binary classifiers, which we use here to find the individual weight of each demographic variable which is simple to interpret. We use the following formula to compute the weights of each demographic variable values that voted for Trump in the 2016 election:

$$weightsComputation = \frac{\sum(\text{commonweight\_vv\_post} \times \text{trumpRes})}{\sum(\text{commonweight\_vv\_post})}$$

The weightsComputation() function computes the weight of each individual category in each demographic variable. Here trumpRes contain a value of 1 if a voter voted for Trump in 2016 election, else the value is 0.

Below are the tables showing the weighted proportion by each individual demographic of Obama Voters that have switched to trump voters.

### Weighted proportion by Race

| Race | Race Weight |
|---|---|
| White | 0.138 |
| Black | 0.037 |
| Hispanic | 0.090 |
| Other | 0.081 |

**Table 1**

### Weighted proportion by Party Identification

| Party Identification | Party Identification Weight |
|---|---|
| Democrat | 0.053 |
| Independent | 0.246 |
| Other | 0.312 |
| Republican | 0.573 |

**Table 2**

### Weighted proportion by Education Level

| Education | Education Weight |
|---|---|
| No HS | 0.107 |
| High school graduate | 0.148 |
| Some college | 0.105 |
| 2-year | 0.122 |
| 4-year | 0.084 |
| Post-grad | 0.063 |

**Table 3**

### Weighted proportion by Gender

| Gender | Gender Weight |
|---|---|
| Female | 0.098 |
| Male | 0.122 |

**Table 4**

Table 1 shows the weights of racial demographic groups. Since the sample size of some racial categories is small, groups except "White", "Black", and "Hispanic" are placed under a new group named "Other". From the weights, we observe that the Black group is least likely to switch from Obama to Trump, which is only 3.7%, and the White group is most likely to switch with a percentage weight of 13.8%.

Table 2 shows how party identification impacts voters to switch from Obama to Trump. Republicans are most likely to switch their votes with a 57.3% weight. Independents are more likely to switch than lean and strong Democrats with a weight of 24.6%, and Democrats seem not very likely to switch their votes with the weight of only 5.3%.

Table 3 shows how education level impacts voters to switch from Obama to Trump in 2016. High school graduates are most likely to switch their votes with 14.8% weight, followed by 2-year college students with a weight of 12.2%.

Post-grads are the least likely to switch their votes with only 6.3% weight, and 4-year college students are also less likely to switch their votes with 8.4% weight. People with no High School are in the middle with a 10.7% weight of switching their votes.

Table 4 shows the weights of gender demographic group, from which we can observe that Males, with 12.2% weight, comparatively are more likely to switch to Trump than Females, with 9.8% by weight.

## Weighted proportion by Attitude Towards Immigration

| Immigration Value | Attitude Towards Immigration Weight | Attitude Towards Immigration |
|---|---|---|
| 0 | 0.392 | Anti-Immigration |
| 1 | 0.245 | Slightly Anti-Immigration |
| 2 | 0.133 | Neutral |
| 3 | 0.065 | Slightly Pro-Immigration |
| 4 | 0.024 | Pro-Immigration |

**Table 5**

Table 5 shows how the attitude towards immigration impacts voters to switch. Voters who are pro-immigrant are least likely to switch with only 2.4% weight while voters who are anti-immigrant are most likely to switch their votes to Trump with a weight of 39.2%.

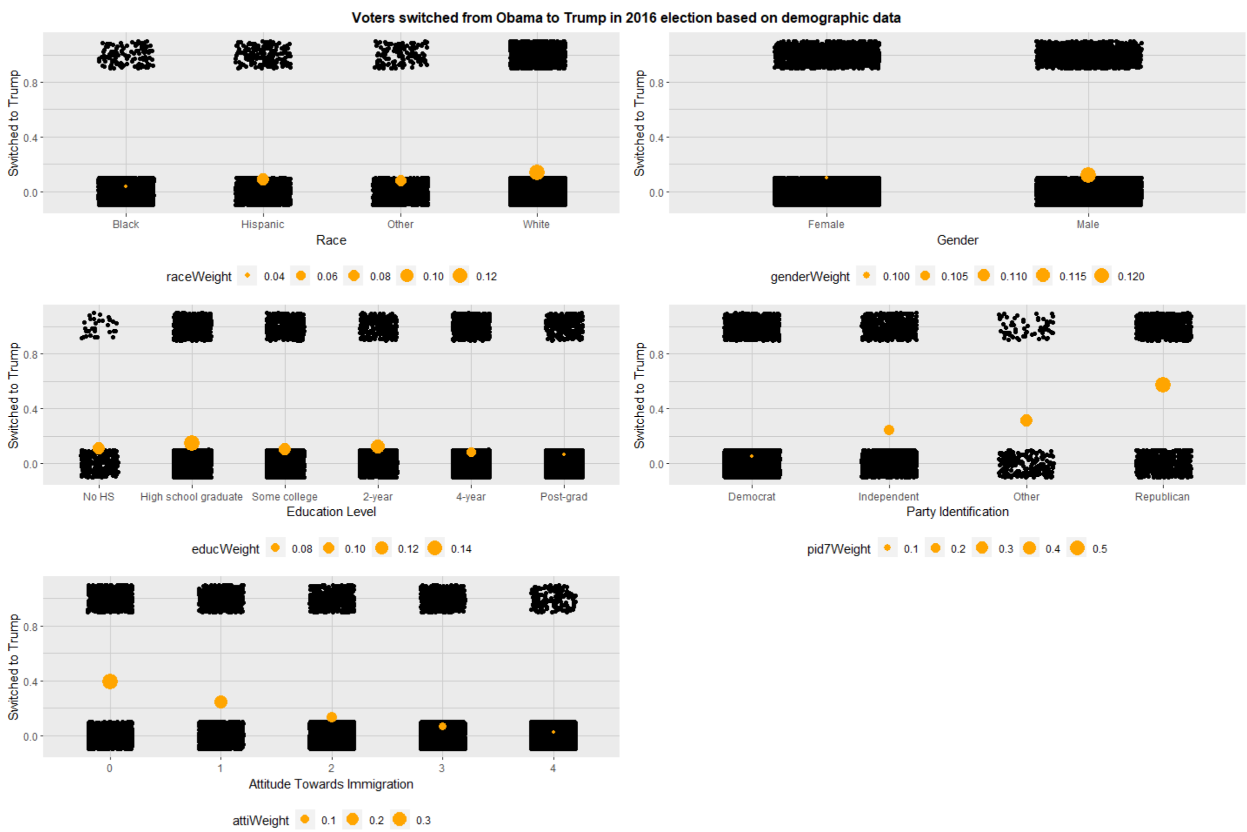


**Figure 1**

Visualizing the impact of individual weights for each demographic variable is helpful. In the above plot in Figure 1, we use the individual categories of each demographic variable in the x-axis, and we use the y-axis to plot the weights of each of the individual categories. The categories who didn't vote for Trump are jittered about y = 0 and the categories who voted for Trump are jittered about y = 1. The orange circle represents the weight of each category. The radius of the circle and the height along the y-axis represent the weight value. The more the weight of a category voting for Trump, the higher and larger the circle gets.

Therefore, we can interpret from the above plot that strong and lean Republicans mostly switched to Trump in 2016. The male voters have a higher chance of switching to Trump. The Black voters are least likely to switch while the Whites are most likely to switch. We could notice that the switch of votes for attitude towards immigration decreases as we go from attitude towards immigration 0(anti-immigration) to attitude towards immigration 4(pro-immigration).

Now, we create two models, for each demographic variable, one with interaction with attitude towards immigration and another without interaction.

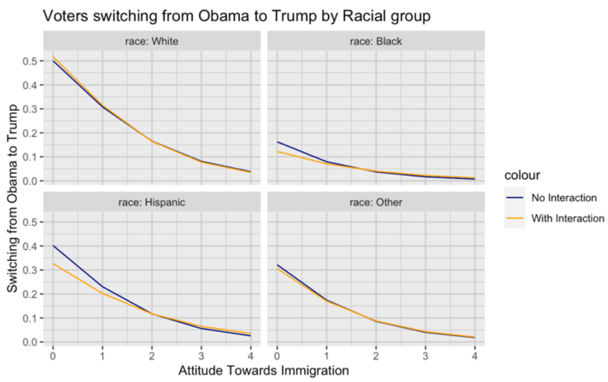


**Figure 2**

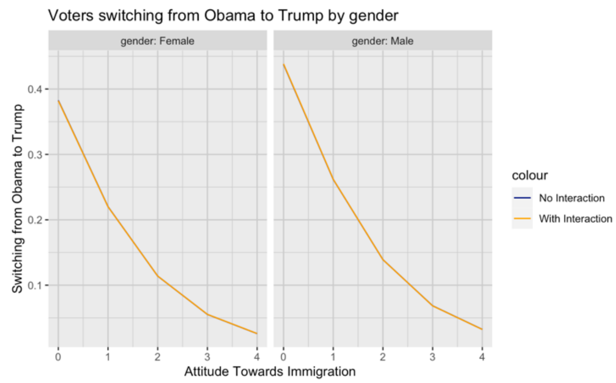


**Figure 3**

In the above plot in Figure 2, we notice the plots for models with and without interaction are different, and hence the race variable does interact with the attitude towards immigration. Further, all the slopes are decreasing and have a negative slope going from left to right, that is the probability of voters switching to Trump decreases across all racial groups when the attitude changes from anti-immigration to pro-immigration.

From the above plot in Figure 3, we notice that there is almost no significant difference in the plots of models with and without interaction, suggesting there is no interaction between gender and attitude towards immigration. However, the probability of voters switching to Trump decreases as the attitude towards immigration changes from anti-immigration to pro-immigration for both the genders.

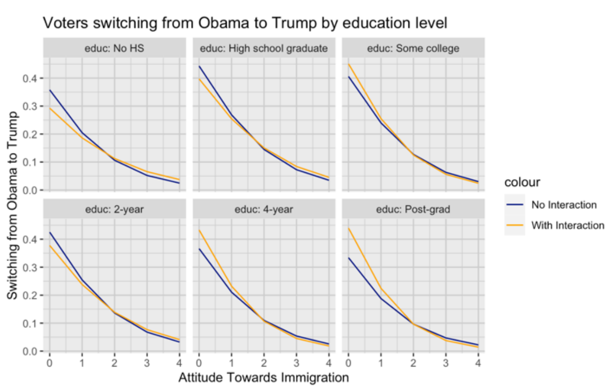


**Figure 4**

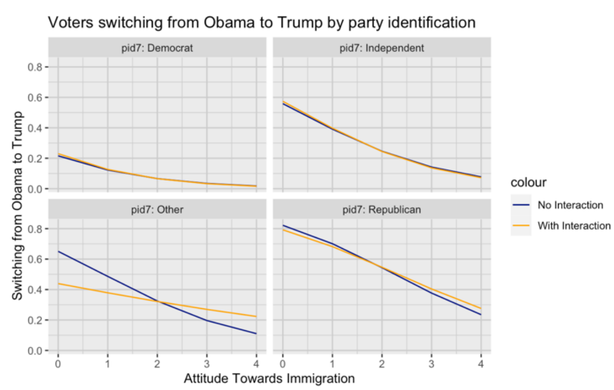


**Figure 5**

The above Figure 4, the education variable interacts with attitude towards immigration. All the curves have a negative slope meaning voters who are anti-immigration are more likely to switch votes to Trump than voters who are pro-immigration across every education level. Since the curves for 'post-grad', '4-year' and 'Some college' are steeper than

'No HS' and 'High School graduate' suggesting that highly education voters are less likely to switch their voter to Trump in 2016.

In the plot shown in Figure 5, we notice that there is an interaction between attitude towards immigration and party identification since the slope between the two curves changes. For all the categories, the probability of switching from Trump decreases as the attitude towards from anti-immigration changes to pro-immigration. The probability to switch changes the most for Republicans, while the probability of switching to Trump is generally low for 'Democrat'. The Independent category is more likely to switch than the Democrats, but less likely to switch than the Republicans. Also, the voters in the 'Other' category are less likely to switch as they go from anti-immigration to pro-immigration.

## 2. Does attitude towards immigration make a substantive difference for the voters to switch from Obama to Trump and whether it matters more for some demographic groups than others.

From the above plots from Figure 2 to Figure 5, we found that attitude towards immigration does interact with race, education level, and party identification demographic variables but doesn't interact with gender. To understand if the attitude towards immigration makes a substantive difference, we create two-weight logistic models, one with attitude immigration as a predictor and another without it and compare. We added interactions of party identification with gender and race, the interaction of gender with race, and race with education, but left out interaction with education, and party identification with education since Pr(>|t|) was about 0.9, like attitude towards immigration and gender.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -5.973 | 0.417 | -14.320 | 0.000 |
| attiTwrdsImmi | -0.425 | 0.077 | -5.515 | 0.000 |
| pid7Num | 0.933 | 0.108 | 8.674 | 0.000 |
| genderNum | 1.496 | 0.182 | 8.241 | 0.000 |
| raceHispanic | 1.926 | 0.598 | 3.222 | 0.001 |
| raceOther | 1.820 | 0.645 | 2.823 | 0.005 |
| raceWhite | 3.836 | 0.412 | 9.320 | 0.000 |
| educNum | 0.216 | 0.060 | 3.632 | 0.000 |
| attiTwrdsImmi:pid7Num | 0.061 | 0.016 | 3.687 | 0.000 |
| attiTwrdsImmi:raceHispanic | -0.007 | 0.093 | -0.074 | 0.941 |
| attiTwrdsImmi:raceOther | -0.020 | 0.100 | -0.201 | 0.841 |
| attiTwrdsImmi:raceWhite | -0.155 | 0.066 | -2.335 | 0.020 |
| attiTwrdsImmi:educNum | -0.076 | 0.013 | -5.903 | 0.000 |
| pid7Num:genderNum | -0.152 | 0.042 | -3.636 | 0.000 |
| pid7Num:raceHispanic | -0.005 | 0.108 | -0.042 | 0.966 |
| pid7Num:raceOther | 0.072 | 0.110 | 0.649 | 0.516 |
| pid7Num:raceWhite | 0.113 | 0.080 | 1.412 | 0.158 |
| genderNum:raceHispanic | -0.759 | 0.250 | -3.034 | 0.002 |
| genderNum:raceOther | -1.306 | 0.265 | -4.924 | 0.000 |
| genderNum:raceWhite | -1.038 | 0.179 | -5.804 | 0.000 |
| raceHispanic:educNum | 0.091 | 0.085 | 1.068 | 0.285 |
| raceOther:educNum | 0.129 | 0.088 | 1.466 | 0.143 |
| raceWhite:educNum | -0.241 | 0.058 | -4.133 | 0.000 |

**Table 6**

Looking at the coefficients we see that pid7Num has a positive coefficient, which shows that people in Republic party are more likely to switch.

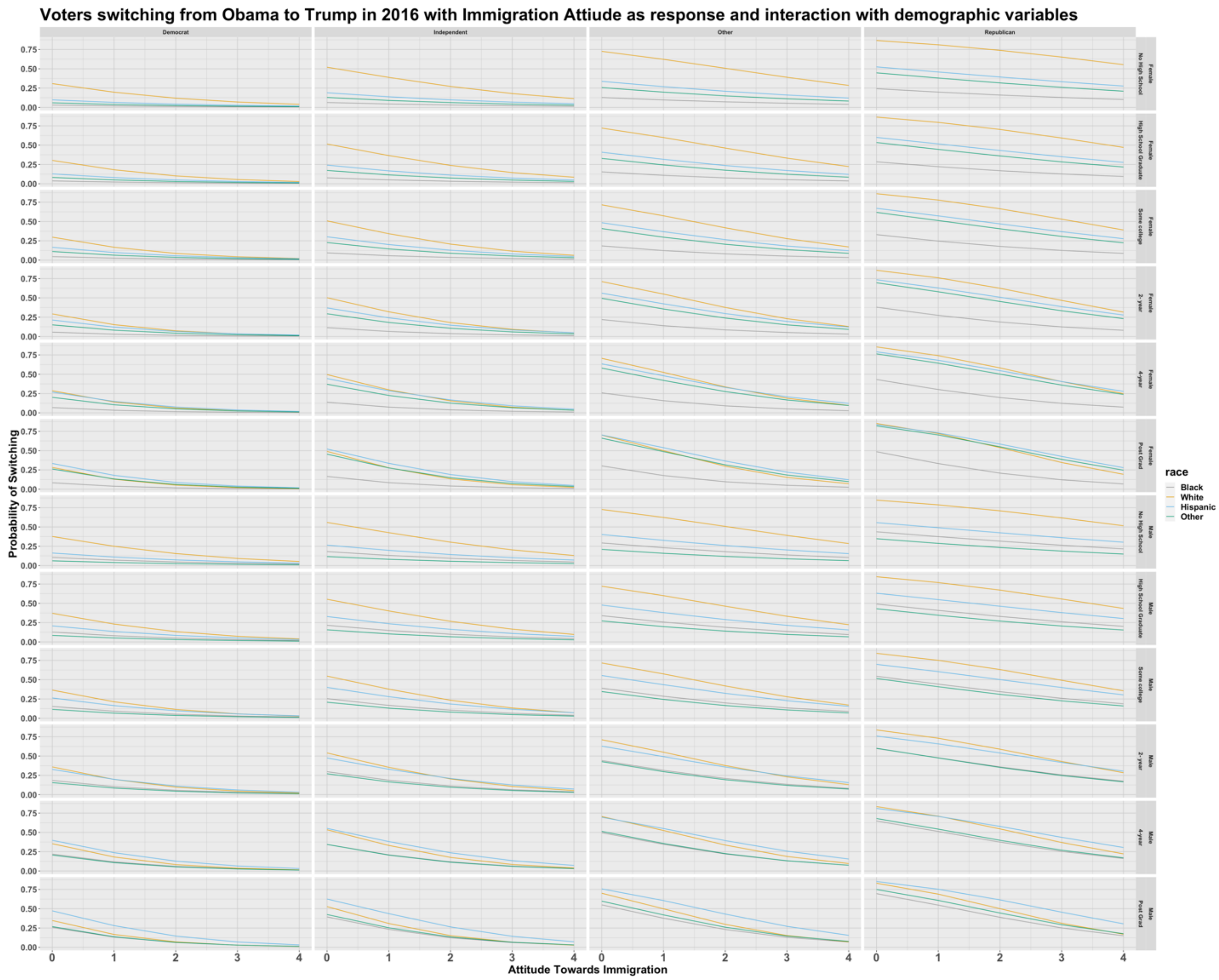


**Figure 6**

From the above plot in Figure 6, we notice that the probability of voters to switch decreases for all the cases as all the plots have a negative slope. In almost all the cases, the White voters are more likely to switch than the Black voters and Hispanic voters, except for 'post-grad' education level shown in the last row, where highly educated Hispanic voters are more likely to switch. Democrat voters shown in the first column are less likely to switch their votes than the rest of the three voters when party identification is concerned, while Republicans are most likely to switch, as seen in the 4th column.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -6.792 | 0.394 | -17.234 | 0.000 |
| pid7Num | 1.085 | 0.102 | 10.595 | 0.000 |
| genderNum | 1.426 | 0.181 | 7.874 | 0.000 |
| raceHispanic | 1.571 | 0.563 | 2.792 | 0.005 |
| raceOther | 1.161 | 0.610 | 1.904 | 0.057 |
| raceWhite | 3.444 | 0.395 | 8.713 | 0.000 |
| educNum | 0.037 | 0.055 | 0.668 | 0.504 |
| pid7Num:genderNum | -0.131 | 0.040 | -3.240 | 0.001 |
| pid7Num:raceHispanic | 0.053 | 0.106 | 0.496 | 0.620 |
| pid7Num:raceOther | 0.113 | 0.108 | 1.043 | 0.297 |
| pid7Num:raceWhite | 0.175 | 0.079 | 2.197 | 0.028 |
| genderNum:raceHispanic | -0.671 | 0.249 | -2.700 | 0.007 |
| genderNum:raceOther | -1.057 | 0.261 | -4.050 | 0.000 |
| genderNum:raceWhite | -0.985 | 0.178 | -5.520 | 0.000 |
| raceHispanic:educNum | 0.053 | 0.083 | 0.638 | 0.523 |
| raceOther:educNum | 0.171 | 0.088 | 1.946 | 0.052 |
| raceWhite:educNum | -0.302 | 0.058 | -5.218 | 0.000 |

**Table 7**

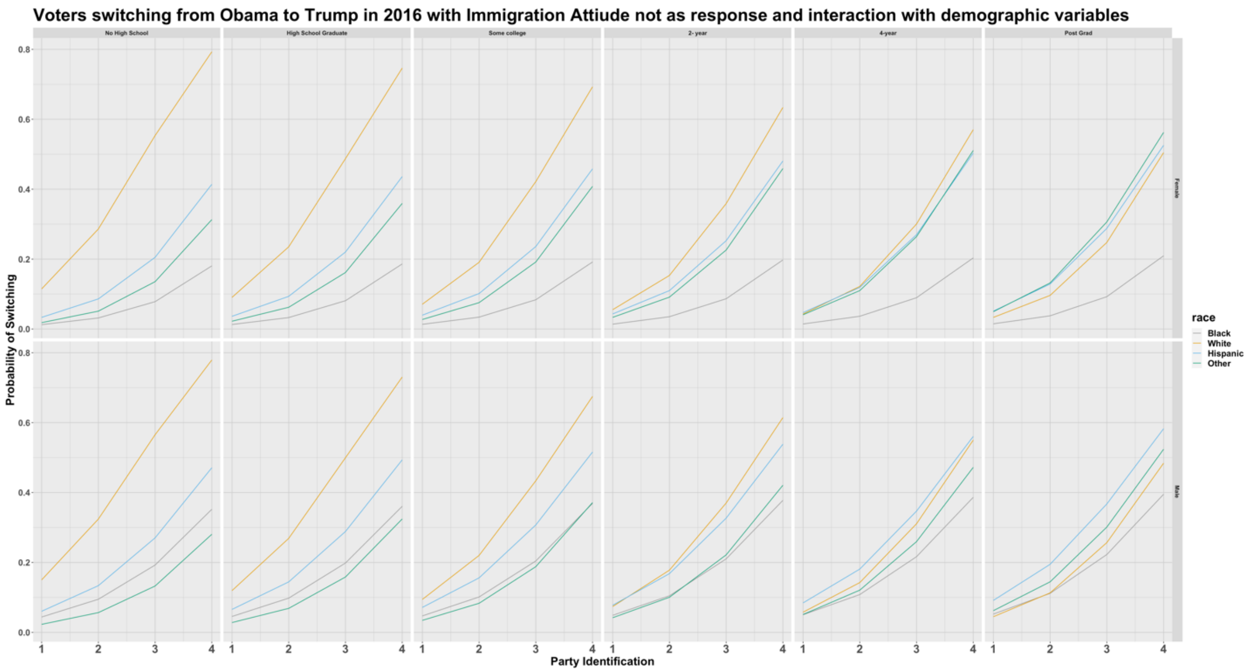


**Figure 7**

Looking at the coefficients we can see that party num has a positive coefficient indicating people who are inclined towards republic party are more likely to switch, similarly education numeric has a negative coefficient saying people with higher education are less likely to switch.

When attitude towards immigration is not used as a predictor, in Figure 7, the voters are less likely to switch their votes to Trump as the education level increases, for both the gender and all 4 race categories. The White voters still seem to have the highest probability of switching and the Black voters are the least. Also, Democrat voters who are highly educated are less likely to switch, and highly educated Hispanic voters are more likely. The Male voters are more likely to switch than the Female voters. However, we notice that highly educated Black, Hispanic, and other voters are more likely to switch for both the gender, which is not the same when the attitude towards immigration is considered. Hence, attitude towards immigration does affect the voters switching for racial groups and education.

## Conclusion

From the study, we found that the Republicans are more likely to switch their votes from Obama to Trump than the Independent and Democrat voters. Further, we notice that other demographic variables such a race, and education level did play a significant role in the voters to switch their votes, however, gender did not. The White voters are more likely to switch to Trump than the Black voters, and highly educated voters are less likely to switch their votes to Trump.

We see that the attitude towards immigration does make a substantive effect on the voters to switch to Trump based on race and education level, but not due to their party identification. Further, fitted values vs the residual plots of the models, shown in Figure 8 and Figure 9 in the appendix section, with using and not using attitude towards immigration as a predictor respectively, shows that the model considering attitude towards immigration, since the plotted line is a straight line close to 0 for the former model and the latter one shows a bending curve, so former model predicts voters switching to Trump better.

# Appendix

Fitted values vs Residual's plot of model with immigration attitude variable as response

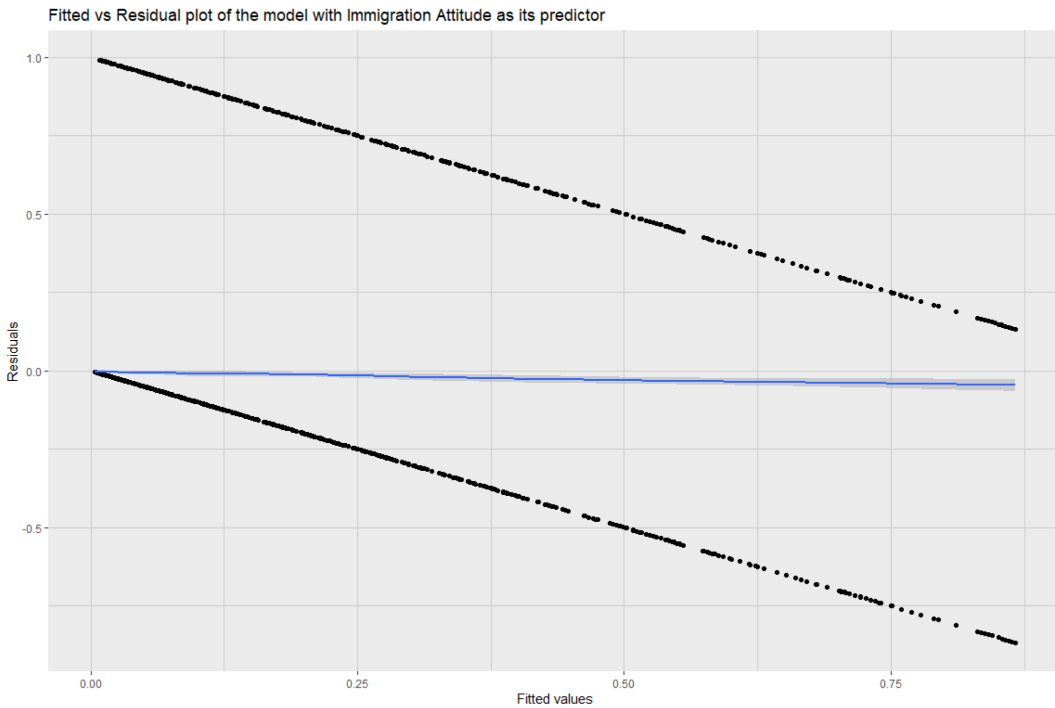


**Figure 8**

Fitted values vs Residual's plot of model without immigration attitude variable as response

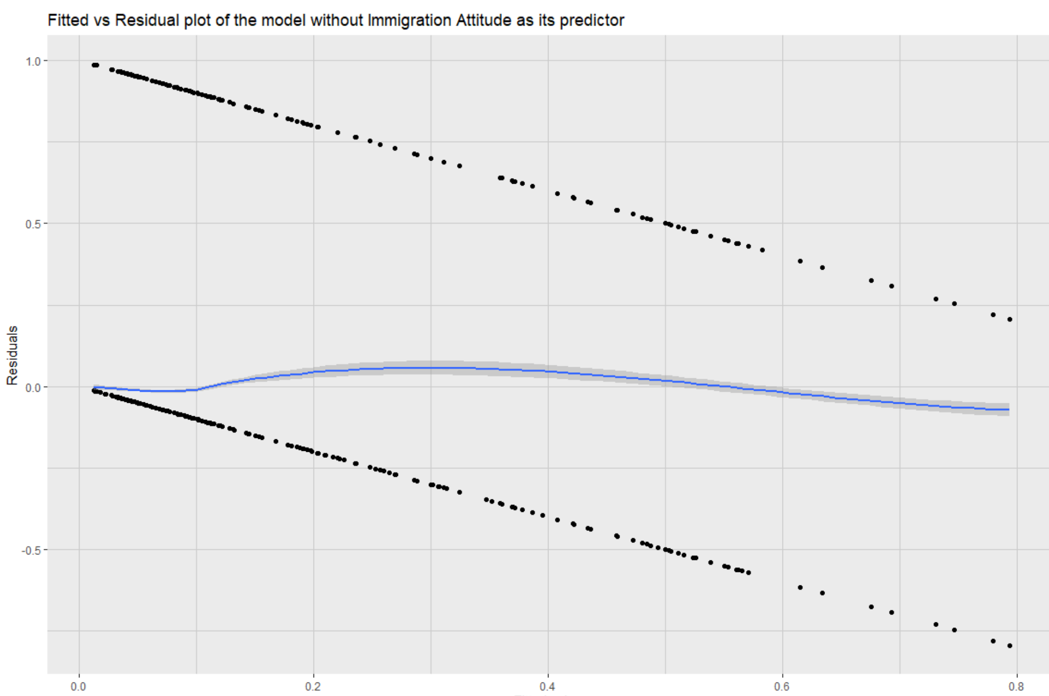


**Figure 9**

## References


1. Weight of evidence (WoE) and Information value (IV) - https://multithreaded.stitchfix.com/blog/2015/08/13/weight-of-evidence/
2. Using rm() function - https://stackoverflow.com/questions/32811527/error-in-evalexpr-envir-enclos-during-knit-in-r-markdown
3. Using ggpubr() - https://github.com/kassambara/ggpubr/issues/97
4. Using ggarrange() - https://stackoverflow.com/questions/65694397/add-main-title-multiple-plots-ggarange
5. Using Labeller - https://stackoverflow.com/questions/3472980/how-to-change-facet-labels