Mini-project 2: Obama to Trump

Team Miami
Goutham Arra, Pavan Kumar Madineni, Harika Putti, Jeevan Rachepalli, Priyadarshini Vijjigiri

Problem Statement

To what extent do attitudes toward immigration explain the switching of votes of 2012 Obama supporters who became 2016 Trump supporters?

Introduction

We have been provided with congressional election data and we wish to find out if certain demographic groups were more likely to switch votes from voting for Obama in 2012 to Trump in 2016. We will be fitting logistic regression models using a solo predictor, two predictors and a lot of predictors (about 4!) to analyze the data.

Data Processing

We've chosen only a subset of the main data containing *commonweight_vv_post* depending on if the voters have taken the post-election survey, *tookpost* if the voters have taken the post-election survey, gender, education, party identity, race, who they've voted for in 2012(we limited the data to Obama voters), who they voted for in 2016, and survey questions to judge the individuals attitude towards immigration. After doing this, we obtained a data frame consisting of data for 23,395 individuals who voted for Obama in 2012, of whom 2,121 said they voted for Trump in 2016.

Single Predictor

We found the weighted proportion of Obama voters in each demographic group that switched to Trump. Here are the percentages –

Race	Percentage that Switched
White	13.84%
Black	3.74%
Hispanic	8.96%
Other	8.14%

Gender	Percentage that Switched
Male	12.25%
Female	9.78%

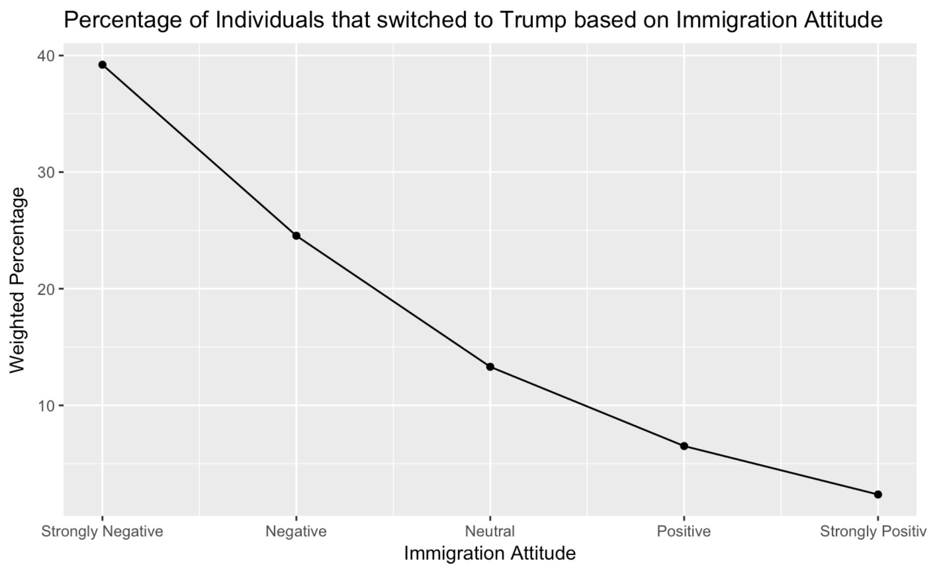
Immigration Attitude	Percentage that Switched
Strongly Negative	39.21%
Negative	24.55%
Neutral	13.31%
Positive	6.51%
Strongly Positive	2.36%

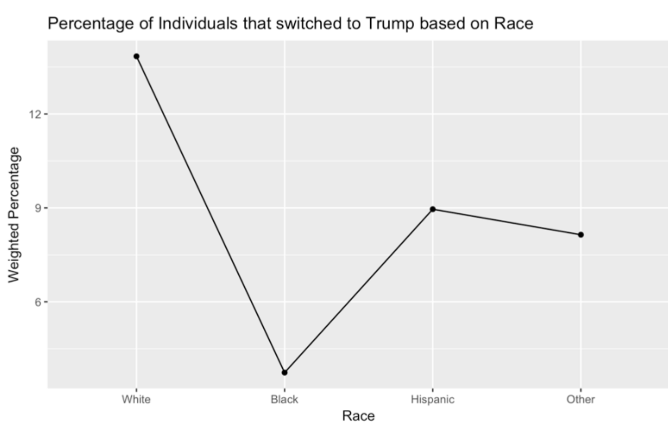
Education	Percentage that Switched
No HS	10.74%
High School graduate	14.78%
Some College	10.48%
2-year	12.15%
4-year	8.39%
Post grad	6.29%

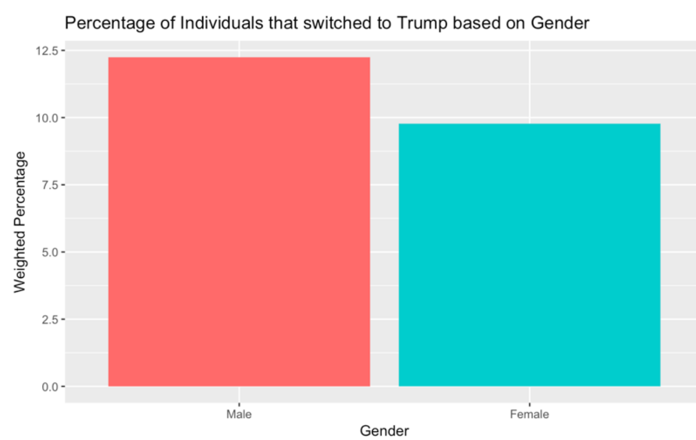
Party Identity	Percentage that Switched
Strong Democrat	2.86%
Not very strong Democrat	11.49%
Lean Democrat	4.71%
Independent	24.62%
Lean Republican	62.04%
Not very strong Republican	50.09%
Strong Republican	68.31%
Not Sure	30.11%

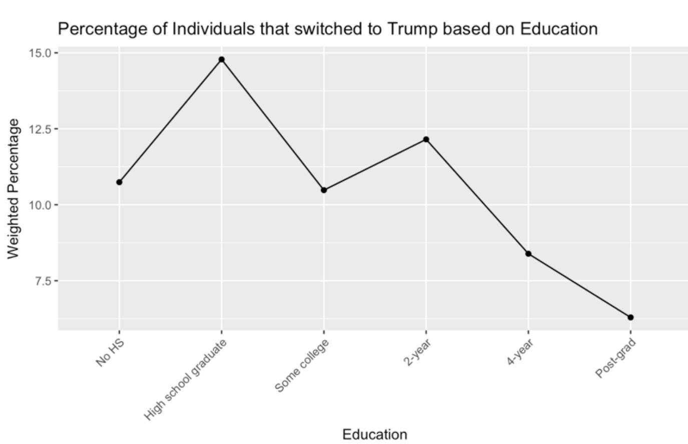
We had constructed a numeric variable that captured the level of positive-ness to negative-ness on the debate of immigration. We can see that as the immigration attitude becomes more positive, the number of individuals that switched to Trump decreases. This means that people who were against the immigration policies switched votes (probably because of the wall). The other demographic variables also showed familiar trends like –

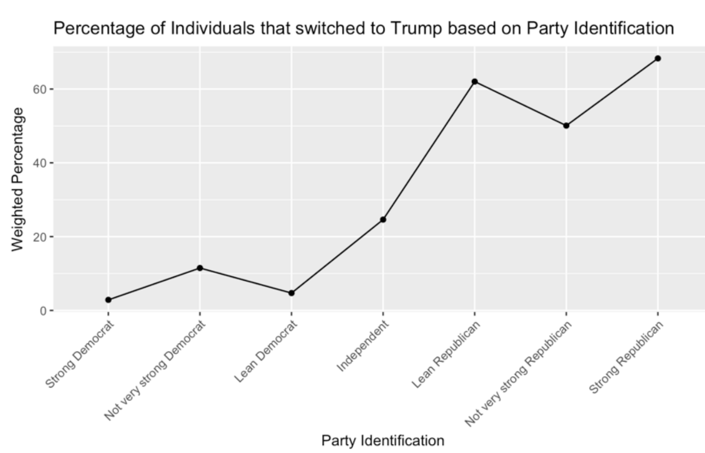
- Individuals with higher education switched their votes less
- Male population seemed to have switched more
- Caucasian population switched the most among other races, and the African Americans switched the least.
- As individuals identified more as republicans, they switched more (this one doesn't come as a surprise).











Two Predictors

We first converted the categorical variables to numerical ones in-order to perform proper logistic regression. Throughout the report the values for immigration attitude are –

- -2 and -1 are "Strongly Negative" and "Slightly Negative" respectively
- 0 id "Neutral"
- 1 and 2 are "Slightly Positive" and "Strongly Positive" respectively

We fit logistic models using immigration attitude and each of the demographic variables.

1. Immigration + Race

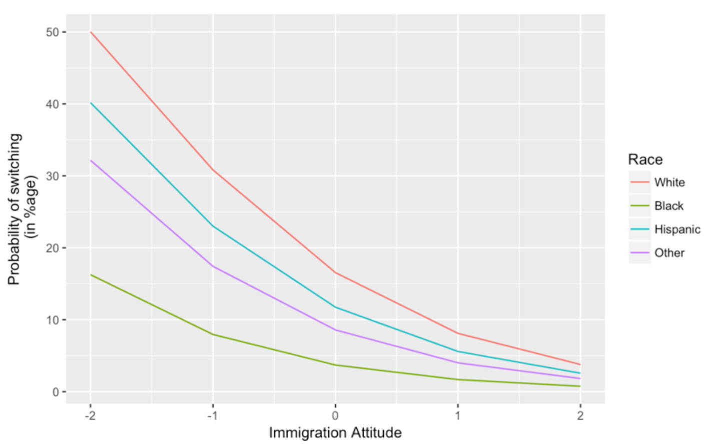
   The log odds for these variables before adding the interaction is –

   - For one-unit change in immigration attitude (it becomes more positive), the log odds of switching votes decreases by 0.81.
   - For race, compared to the white population the log odds decrease by 0.4 for Hispanic population and decreases for rest by 0.75 and decrease by 1.6 for the black population.

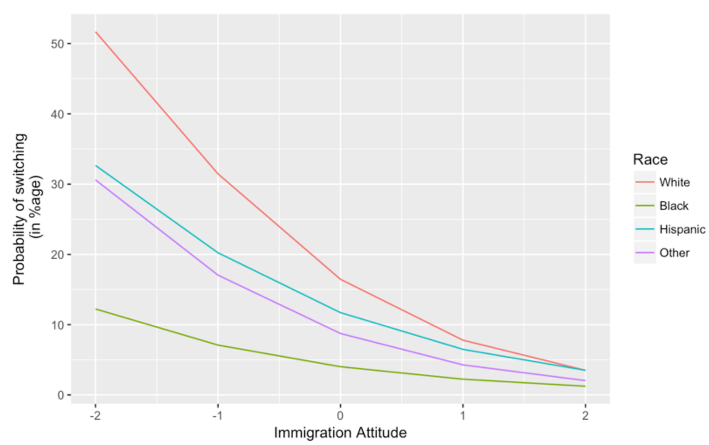
   The log odds after adding the interaction between the two variables is –

   - For one-unit change in immigration attitude (it becomes more positive), the log odds of switching votes decreases by 0.84
   - For race, compared to the white population the log odds decrease by 0.39 for Hispanic population and decreases for rest by 0.72 and decrease by 1.55 for the black population.

   Displaying the voter_Trump probability as a function of immigration attitude for a few values of race –



Immigration Attitude + Race 1: Without Interaction



Immigration Attitude vs Race 2: With Interaction

It's not very different but there seems to be some interaction between few of the values of race. We can consider this interaction in our model to have a better understanding. But at a glance, we can see that as the immigration attitude becomes more negative, the White population seemed to switch more (look how steep the curve gets), and the opposite happens to the Hispanic and the Black population, upon adding an interaction their slopes decrease – this means that no matter how negative their immigration attitude is, they're less likely to switch.

2. Immigration + Party Identity

   The log odds for these variables before adding the interaction is –

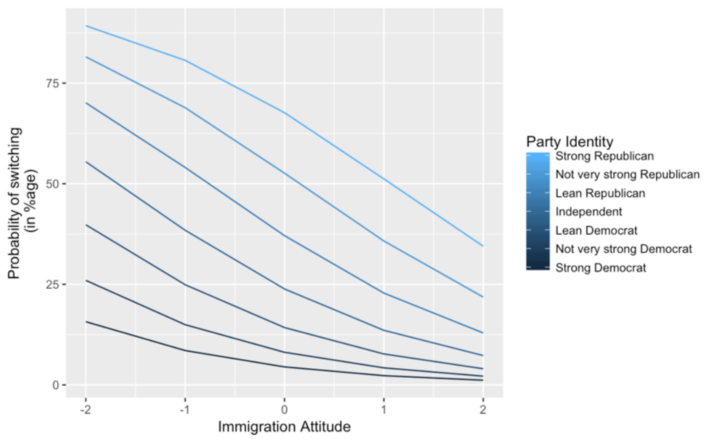
   - For one-unit change in immigration attitude (it becomes more positive), the log odds of switching votes decreases by 0.69.

- For one-unit change in party identification (i.e., they become increasingly republican), the log odds of switching votes increases by 0.63

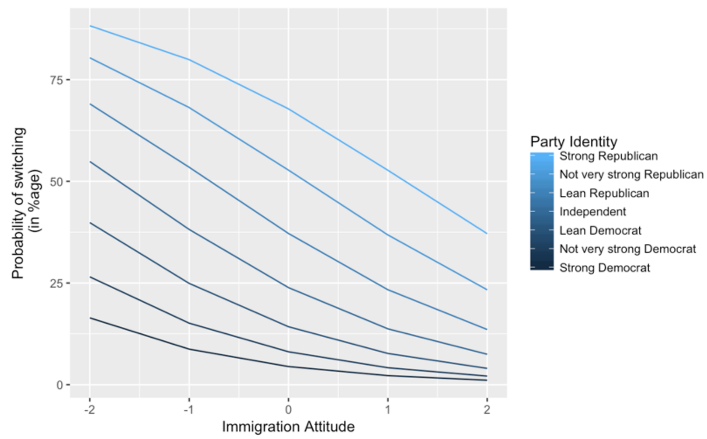
The log odds after adding the interaction between the two variables is –

- For one-unit change in immigration attitude (it becomes more positive), the log odds of switching votes decreases by 0.73
- For one-unit change in party identification (i.e., they become increasingly republican), the log odds of switching votes increases by 0.63
- The interaction coefficient is imm_att : pid7.f 0.01382

The voter_Trump probability as a function of immigration attitude for a few values of party identity –



Immigration Attitude + Party Identity 1: Without Interaction



Immigration Attitude + Party Identity 2: With Interaction

We can see that, as the individuals' party identity changed from democrat to republican, the individuals tended to switch more as the immigration attitude became more negative. But, there is no interaction here either.

3. Immigration + Gender

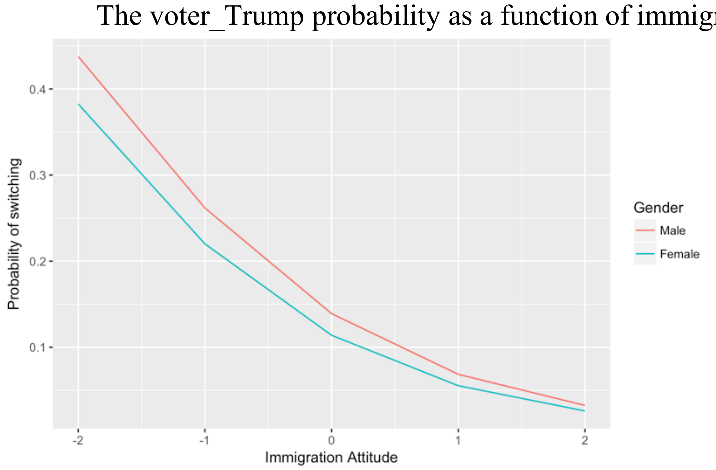
The log odds for these variables before adding the interaction is –

- For one-unit change in immigration attitude, the log odds of switching votes decreases by 0.79.
- For gender, compared to the male population the log odds decrease by 0.23 for female population

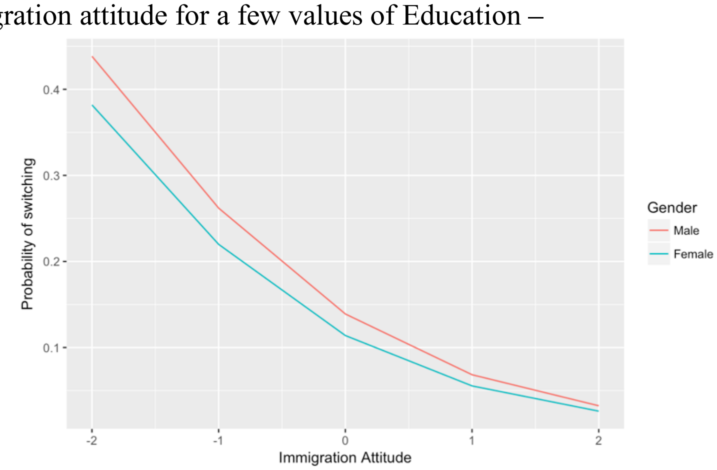
The log odds after adding the interaction between the two variables is –

- For one-unit change in immigration attitude, the log odds of switching votes decreases by 0.79
- For gender, compared to the male population the log odds decrease by 0.23 for female population

The voter_Trump probability as a function of immigration attitude for a few values of Education –



Immigration Attitude + Gender 1: Without Interaction



Immigration Attitude + Gender 2: With Interaction

4. Immigration + Education

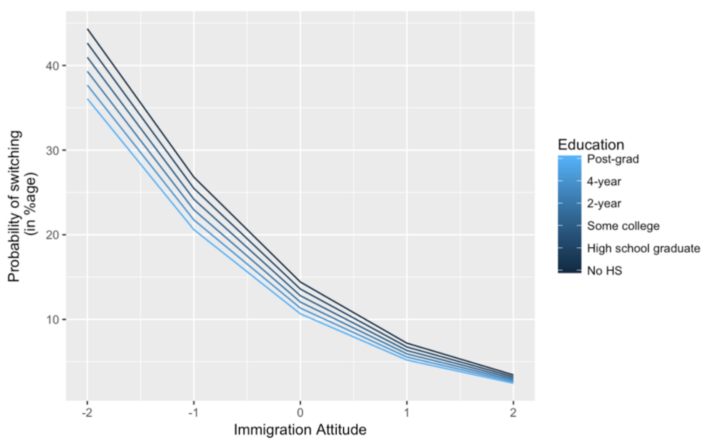
The log odds for these variables before adding the interaction is –

- For one-unit change in immigration attitude (it becomes more positive), the log odds of switching votes decreases by 0.78.
- For education, one-unit change (i.e., as their level of education increases) it leads to a decrease in the log odds by 0.069

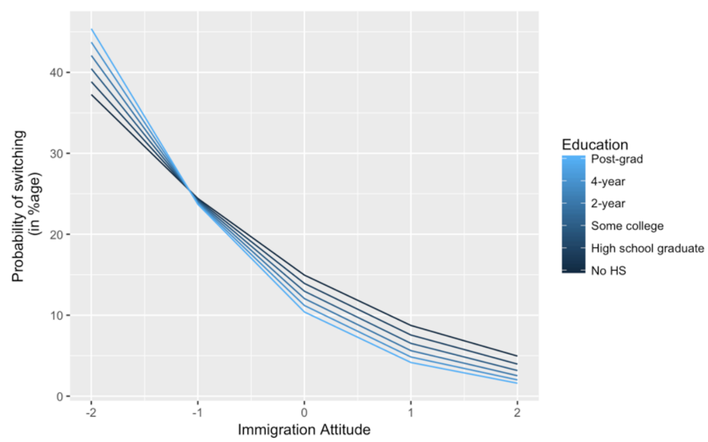
The log odds after adding the interaction between the two variables is –

- For one-unit change in immigration attitude (it becomes more positive), the log odds of switching votes decreases by 0.53
- For education, one-unit change (i.e., as their level of education increases) it leads to a decrease in the log odds by 0.083
- The interaction coefficient imm_att : educ.f -0.07515

The voter_Trump probability as a function of immigration attitude for a few values of Education –



Immigration Attitude + Education 1: Without Interaction



Immigration Attitude + Education 2: With Interaction

Here we can see a clear interaction between the immigration attitude and the education variable. As the immigration attitude becomes more negative, the curves come together. That is, as the attitude towards immigration becomes too negative an individual is more likely to switch. This also tells us that people who are well educated were more likely to switch than those who weren't if their immigration attitude was negative.

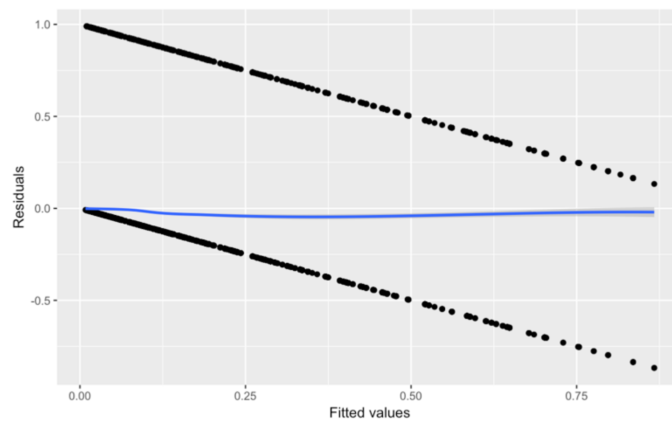
Lots of Predictors

1. Without using Immigration Attitude as a predictor.

   We're going to use a model with all the demographic variables but excluding immigration attitude. We've fit a weighted regression model using Generalized Linear Model on the data and here's how the coefficients and the fit look –

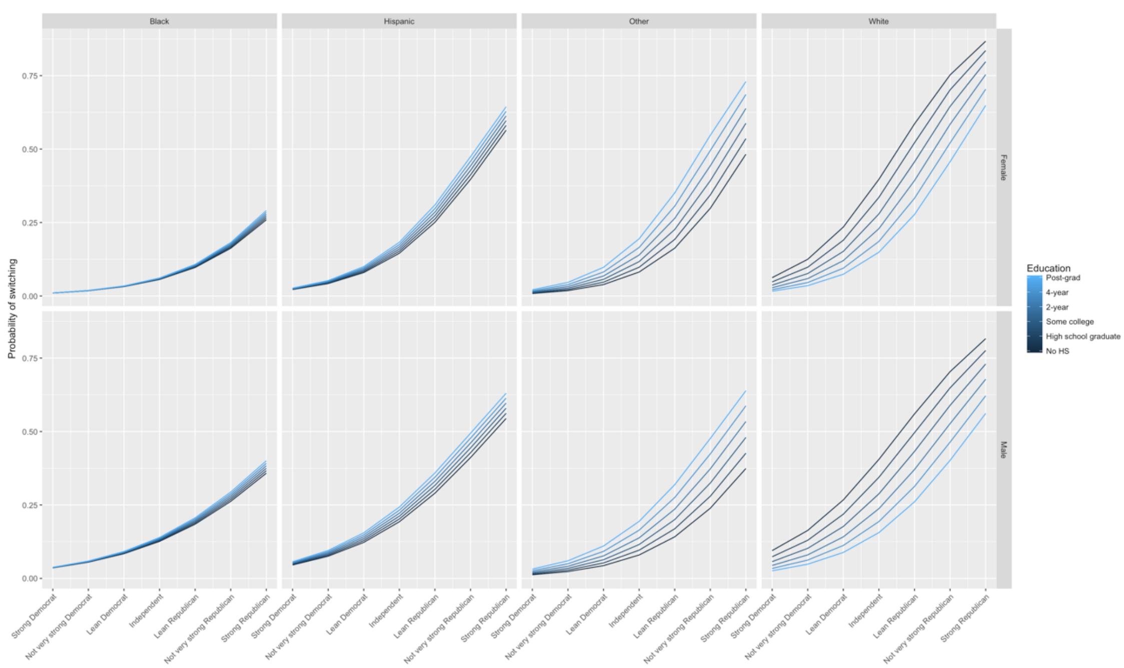
```
Call:
glm(formula = voted_Trump ~ educ.f + pid7.f * race.f + gender.f,
    family = "quasibinomial", data = df2, weights = commonweight_vv_post)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.5879  -0.3700  -0.2326  -0.1546  10.3275

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          -3.03480    0.08184 -37.080  < 2e-16 ***
educ.f               -0.20378    0.01608 -12.670  < 2e-16 ***
pid7.f                0.70510    0.01604  43.948  < 2e-16 ***
race.fBlack          -0.62254    0.15031  -4.142 3.46e-05 ***
race.fHispanic       -0.23458    0.18339  -1.279 0.200874
race.fOther          -0.61999    0.24763  -2.504 0.012297 *
gender.fFemale       -0.22342    0.04719  -4.734 2.21e-06 ***
pid7.f:race.fBlack   -0.16638    0.04823  -3.450 0.000562 ***
pid7.f:race.fHispanic -0.07027   0.05109  -1.376 0.168968
pid7.f:race.fOther   -0.02556    0.06016  -0.425 0.670899
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.9101006)

    Null deviance: 14732  on 23394  degrees of freedom
Residual deviance: 11539  on 23385  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 6
```

Model Coefficients



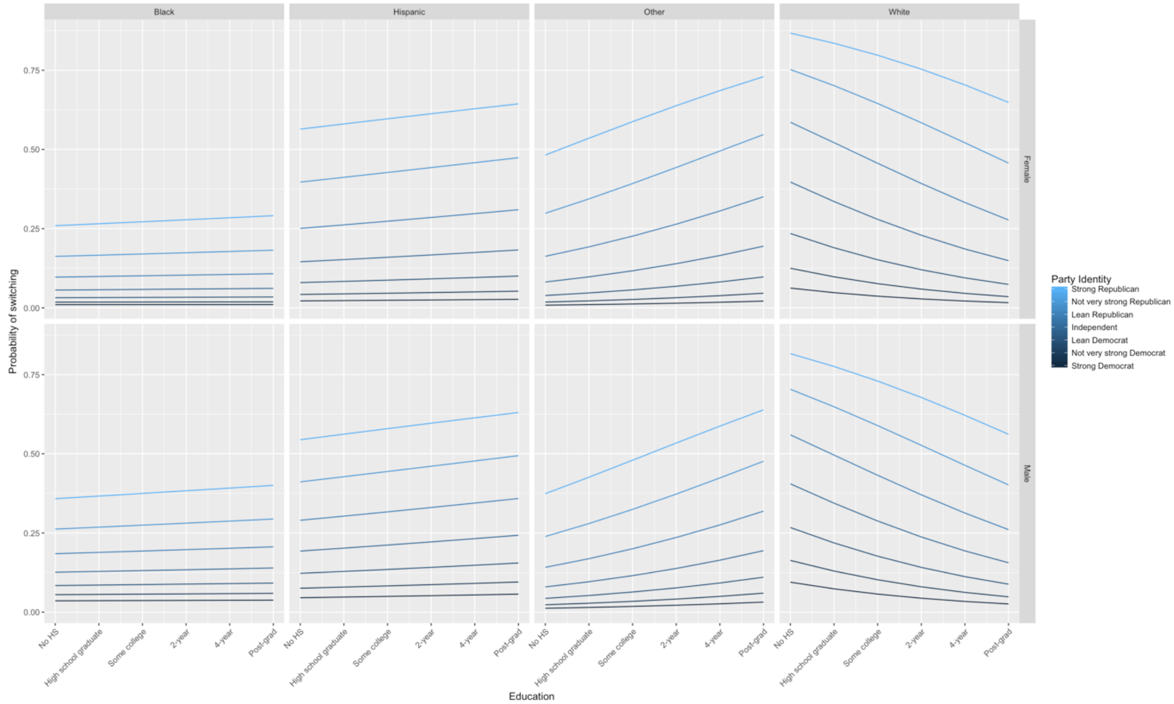
Residual Plot

There are no weird coefficients as far as we can tell but explaining the model from the coefficients alone can be pretty tricky. The Residual deviance is 11539, which is pretty high, but the fit of the model looks good, so we'll proceed with using this model. Let's start simple and use a small portion of the demographic to see if there's any trends –



Here we can tell that, in general, as the individuals become more republican, they tend to switch more. In case of white individuals, those who have less education seemed to switch more as their party identity changed and in case of rest of the races, the individuals with higher education switched more. Also, there is no difference between male and female trends. There is not much else we can tell from the graph.

Let’s do another graph –



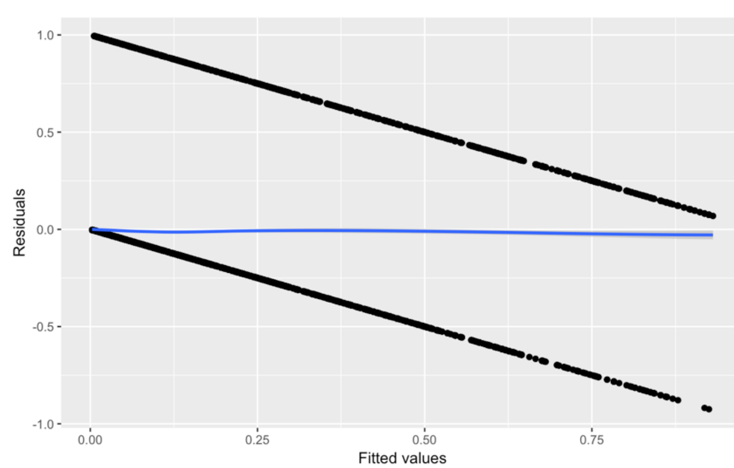
From this we can tell that party identity and education are not that interactive which is a little misleading. So this one is a bad representation of the data since it’s hard to explaining anything. Let do a different model.

2. Using Immigration Attitude as a predictor

We have seen earlier that we have an interaction between Immigration Attitude and Education and Immigration attitude and Race. We are going to use those for our model and also put in all the other demographic variables and check if we have anything meaningful. We’ve fit a weighted regression model using Generalized Linear Model on the data and here’s how the coefficients and the fit look –

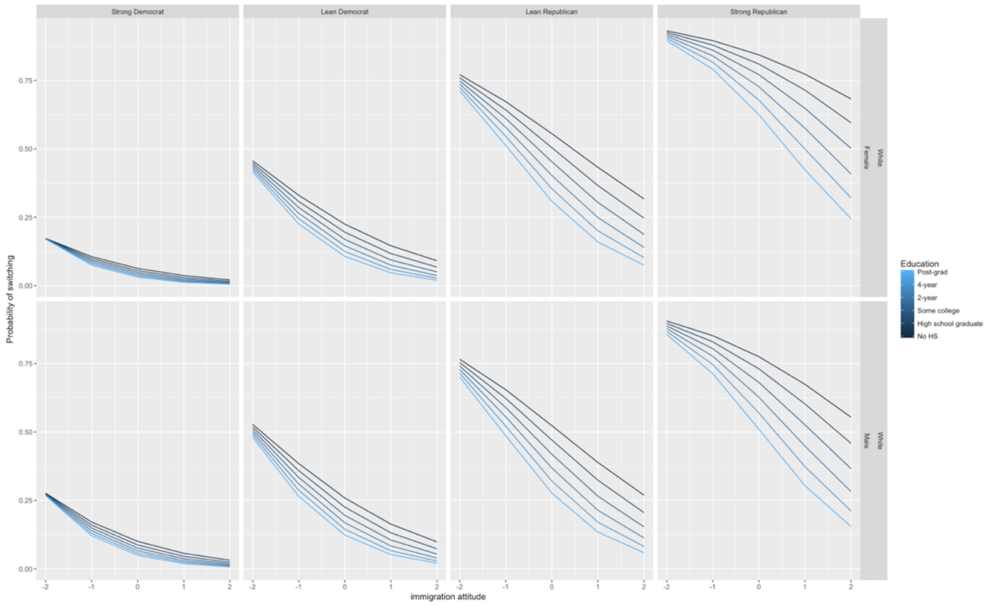
```
Call:
glm(formula = voted_Trump ~ imm_att * educ.f * race.f + pid7.f +
    gender.f, family = "quasibinomial", data = df2, weights = commonweight_vv_post)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.0091  -0.3277  -0.1838  -0.1060  11.5352

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -2.63092    0.08457 -31.110  < 2e-16 ***
imm_att                       -0.46302    0.05211  -8.886  < 2e-16 ***
educ.f                        -0.18371    0.01933  -9.506  < 2e-16 ***
race.fBlack                   -1.81198    0.20900  -8.670  < 2e-16 ***
race.fHispanic                -1.14745    0.23426  -4.898 9.73e-07 ***
race.fOther                   -2.12120    0.31761  -6.679 2.47e-11 ***
pid7.f                         0.60081    0.01459  41.189  < 2e-16 ***
gender.fFemale                -0.23105    0.04945  -4.672 3.00e-06 ***
imm_att:educ.f                -0.07786    0.01498  -5.197 2.04e-07 ***
imm_att:race.fBlack           -0.27621    0.17105  -1.615 0.106365
imm_att:race.fHispanic         0.12523    0.18843   0.665 0.506318
imm_att:race.fOther            0.28041    0.21564   1.300 0.193484
educ.f:race.fBlack             0.21083    0.06054   3.483 0.000497 ***
educ.f:race.fHispanic          0.27468    0.07138   3.848 0.000119 ***
educ.f:race.fOther             0.33584    0.07389   4.545 5.51e-06 ***
imm_att:educ.f:race.fBlack     0.12972    0.04919   2.637 0.008367 **
imm_att:educ.f:race.fHispanic  0.01375    0.05679   0.242 0.808761
imm_att:educ.f:race.fOther    -0.02904    0.05147  -0.564 0.572690
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 0.8906477)

    Null deviance: 14732  on 23394  degrees of freedom
Residual deviance: 10216  on 23377  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 6
```

Model Coefficients



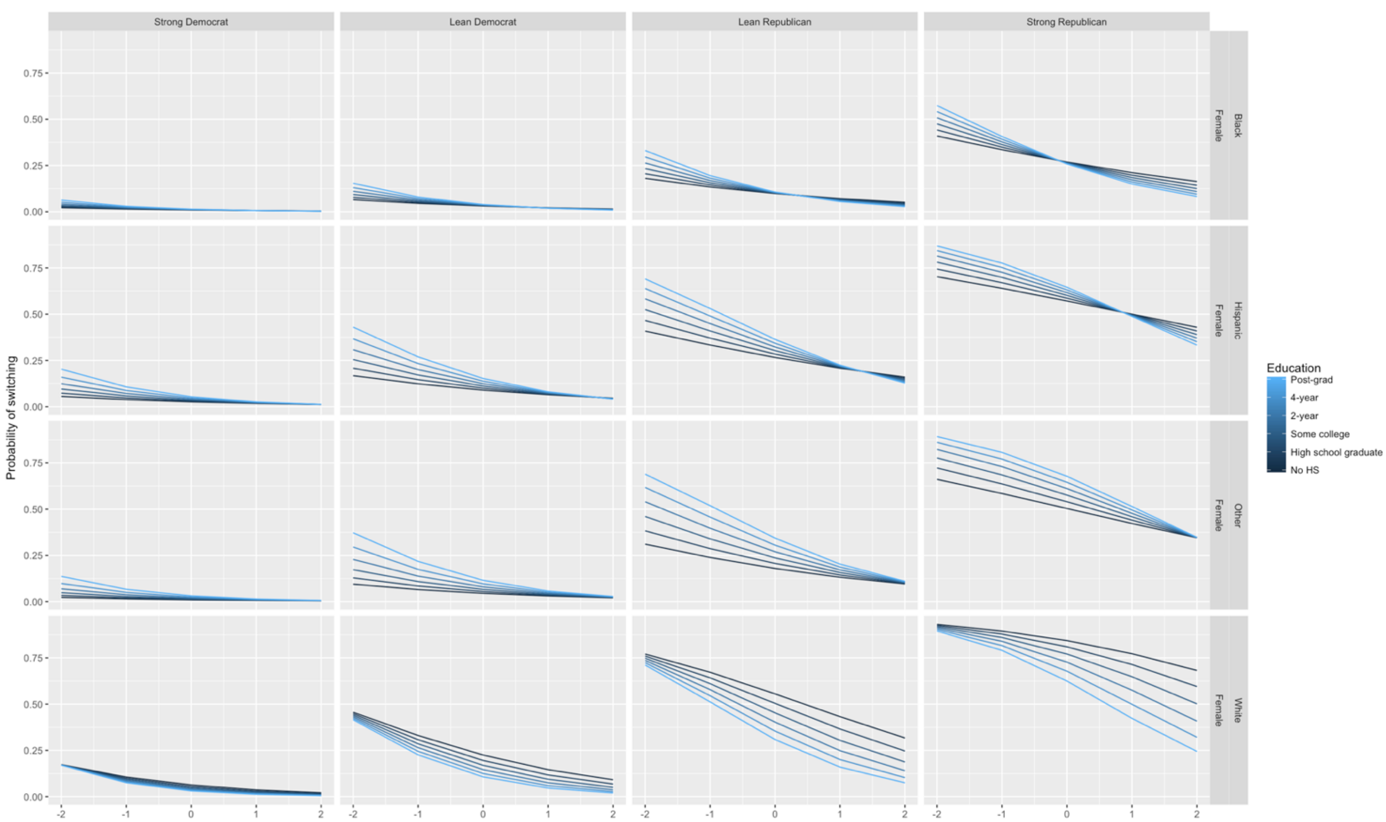
Residual Plot

There are no weird coefficients here either. The residual deviance is 10216 which is significantly lower than when we didn't use immigration attitude. This shows that immigration attitude is an important predictor. The fit of the model looks good, so we'll proceed with using this model. Let's start simple and use a small portion of the demographic to see if there's any trends –



Here we have taken only a few of the values of party identification and only one value from the race demographic. This is to check if we have any differences when it comes to the gender demographic. We can see that the gender has no impact on how the trends are varying, implying that gender might not tell us any new information, so we can exclude that and make a new model, or just use one of the values of gender. We did the latter.

Now let's take more values of race and keep only one value of gender (I'm using female, because there's more records) –

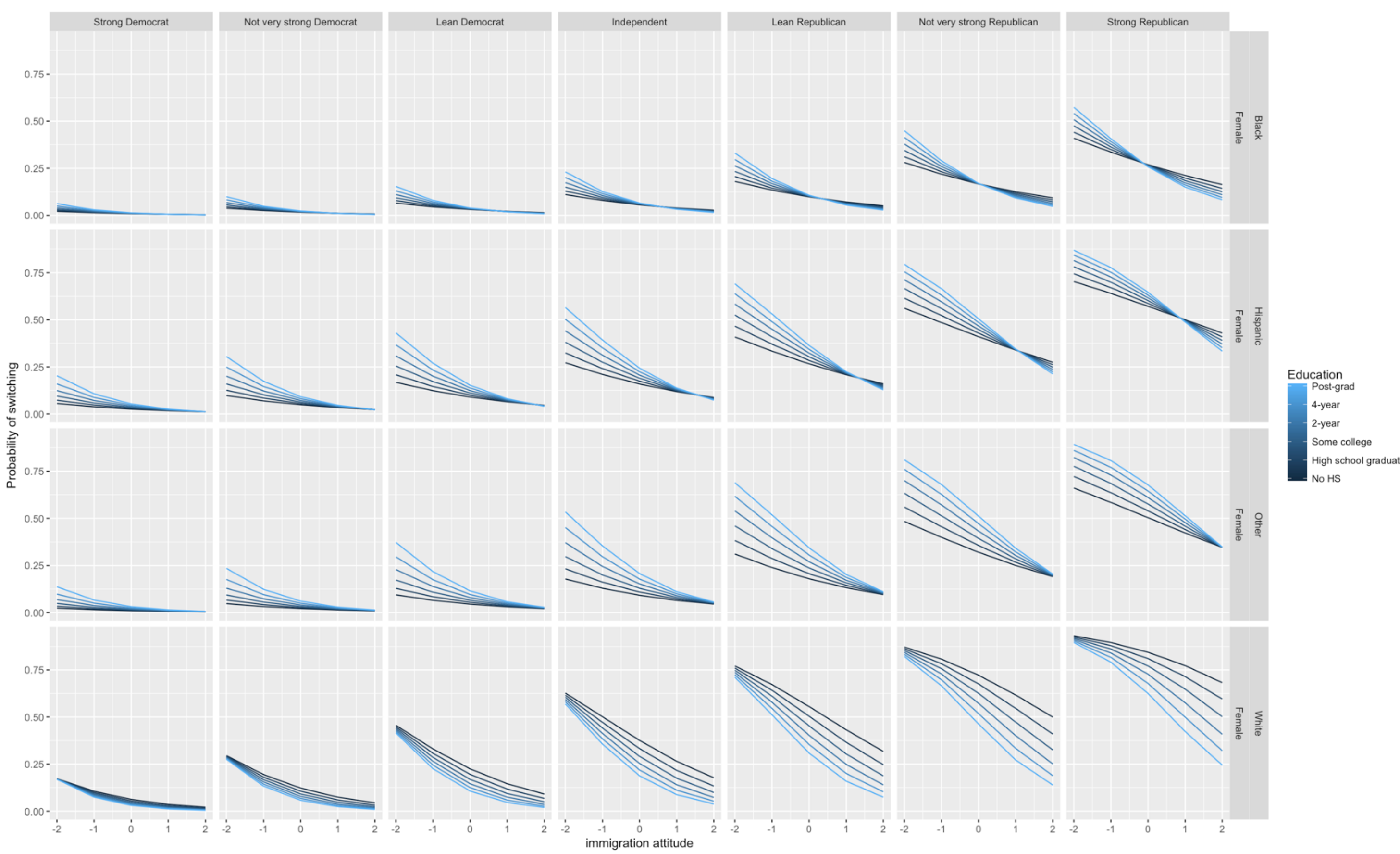


Let's unpack this by individual facet and then a couple of trends that we observe –

As the party identity changes, as observed earlier, there is a lot of shift in the probability of switching votes. Let's look at the race wise trends –

- For the "black-female" population, in all cases of party identification, education and immigration attitude seemed to depend on how much they've switched. For example, a well-educated republican's trend in switching increased rather quickly, compared to a non-educated person, as her immigration attitude became negative. It can also be said that the rate of switching increases a bit exponentially as their immigration attitude becomes more negative and it becomes increasingly more exponential as their education increases.
- For a "Hispanic-female", the trend is linear when their party identity is somewhere between lean democrat and lean republican but as they become more republican, their trend in switching increases logarithmically (which again depends on their education as well).
- For the "white-female" population, it's a whole other story. While for all other races, the amount of education didn't matter when their immigration attitudes were positive (their probability of switching was almost similar) but for the white population, lesser the education is the probability of switching is more (even with positive immigration attitudes). But no matter what education they had, they had similar probabilities of switching when their immigration attitude is the most negative (because the rate of switching for highly educated people is more when their attitude is more negative).

Here we have an extensive look of how the probabilities change for black, white and Hispanic-male populations for all party identities. We can see how they gradually change.



Let's look at the party identity wise trends for each race at a time –

- For a "white-female", who is a "strong democrat", the probability of switching is small no matter what education they've had. But they tend to switch a bit more if their immigration attitude is negative.

- For a “black-female”, who is a “strong democrat”, no matter what education they’ve had or their immigration attitude, they did not switch votes.
- For a “strong democrat”, the percentage of switching is very low in general. But when the individual is either Hispanic or other race, if they were more educated they seemed to have switched more if their immigration attitude was negative.
- Similar trends are followed by the other two democratic party identities.
- For the republican party identities, the black population switched more only when their immigration attitude was around “slightly negative”, so their trend increases slower than the rest. The Hispanic population switched more quickly (slopes are steeper) with increasing education. The other races also have switched more with increasing education.

To summarize, overall the probability of switching based on immigration, was more when a person was well educated. At the same time a black individual was less likely to switch than a white individual even when they were well educated and identified as a republican. So for individuals who identified more as republican, education and immigration attitude mattered the most in order to switch votes.