Mini Project 2

Aniket| Ruchik | Shashank | Sumanth

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

Introduction :

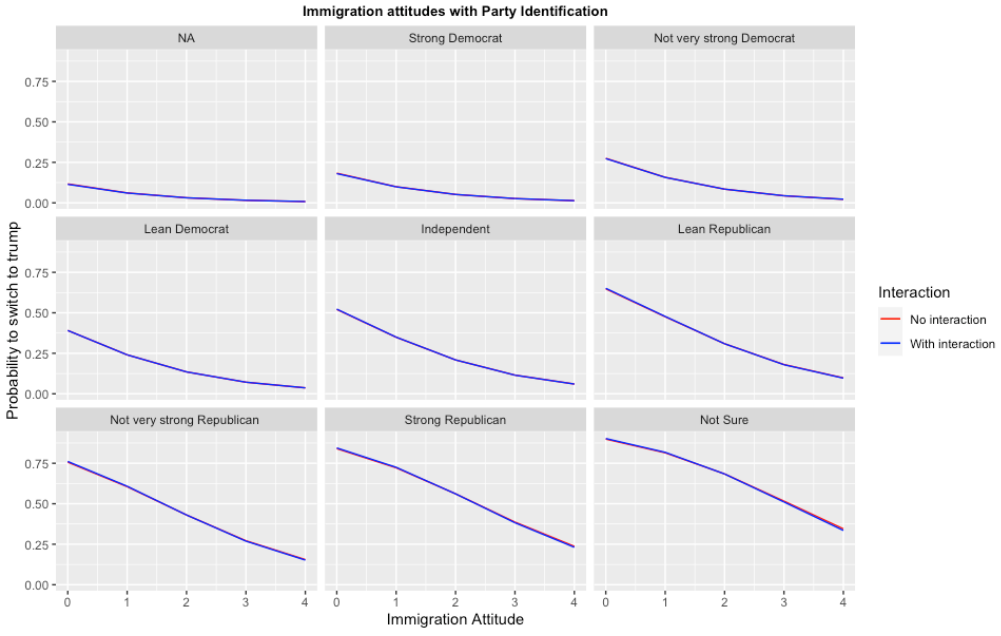
The goal of the mini project is to understand if the immigration attitude had an effect in change of voting behavior from 2012 to 2016 election. In the 2nd question, we try to see if the interaction term with demographic and immigration attitude creates any difference. In the 3rd question

Question 1: Preprocessing the data
We selected the columns which are required in the analysis. Filtered those who responded to the post-election survey and voted for Obama in 2012. We then created a quantitative variable based on pro-immigration response from the four-immigration variable in the data. We grouped the race variable in four levels (White, Black, Hispanic, and Other). Some of the categorical variable like education and party has an order in the data, so we encoded it with numerical variables.

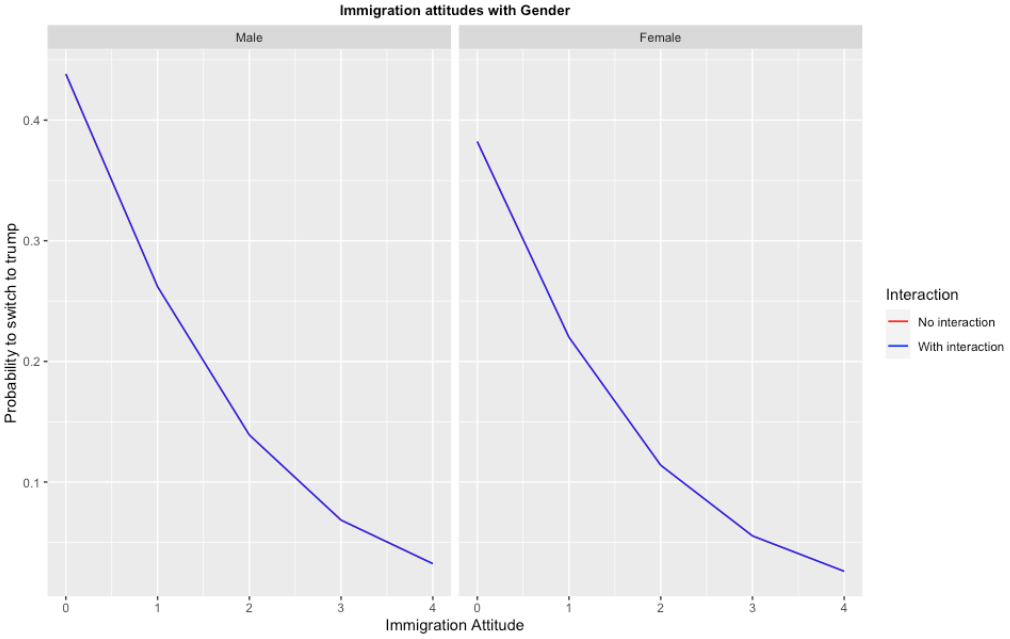
Question 2: Does interaction between demographic variable and immigration attitude has an effect in the model?

To understand if the interaction between demographical variable and immigration attitude plays any role in the model, we built 4 logistic regressions, with each of demographic variable (race, education, party, gender) and immigration attitude as a predictor in the model. We compared these models with the corresponding interaction term of demographics and immigration to check if the probability lines are overlapping or its different.

Immigration attitudes with Party Identification

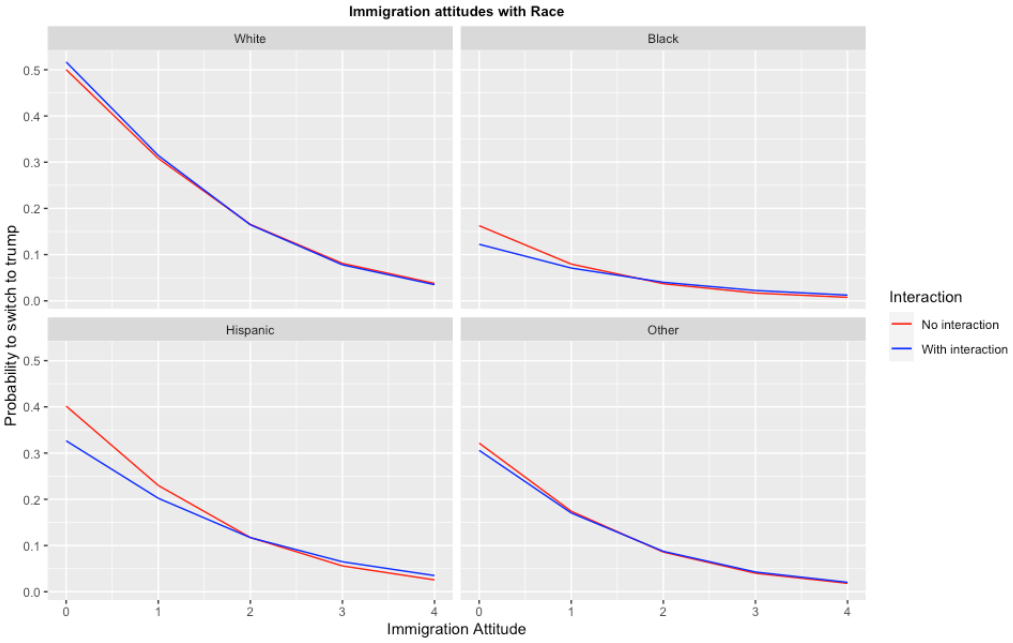


Immigration attitudes with Gender

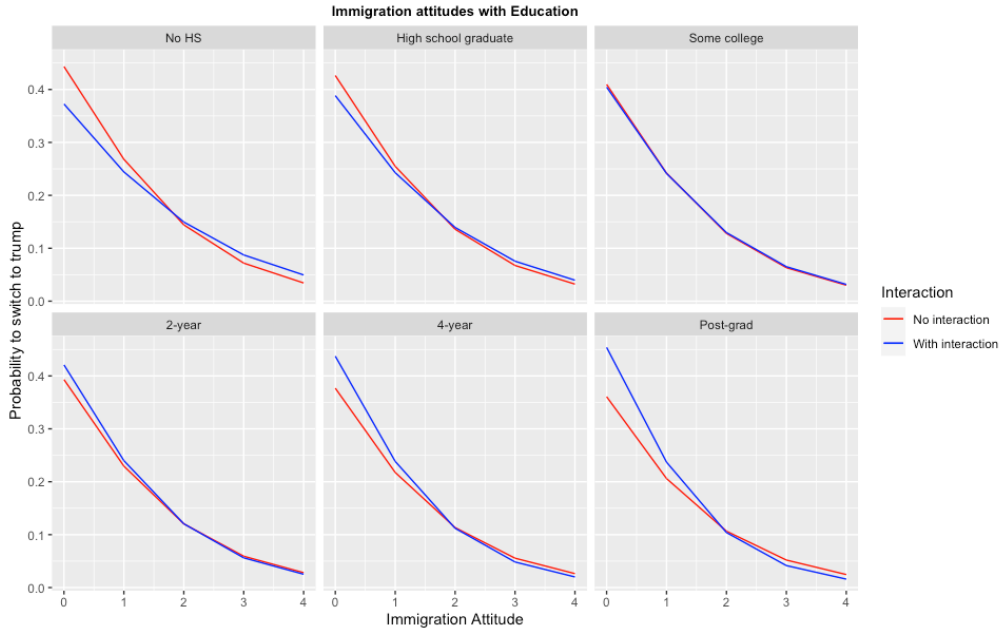


From the above charts we notice that the probability line for gender and party are overlapping suggesting that we don't need any interaction term of gender and party with immigration. This observation can also be confirmed when we see the p-value of the predictors. The interaction term does not show any statistical significance (p-value > significance level)

Immigration attitudes with Race



Immigration attitudes with Eduaction



When we add interaction term in race, we see that there is a deviation in probability line for each of the levels. Hispanic and Black race group has a greater deviation in probability line compared to Whites and Others

Each of the subgroups in Education has a different probability line apart from “Some college” suggesting the interaction plays a role in the model.

The p-value of interaction term of race and education is lower than the significance level(0.05) suggesting the interaction term plays a role in the model.

Question3:

Feature *educ* was transformed to be an ordinal attribute. But the feature *race* was kept as categorical (non-ordinal) attribute as it would be incorrect to give an ordering to the *race* feature.

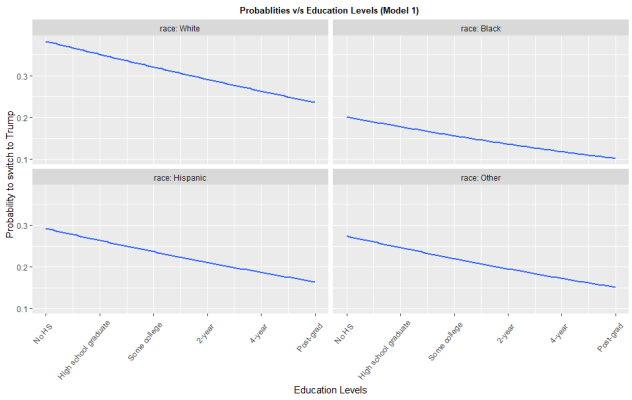
Fitting the two models

1. **Model 1: Fitting a weighted logistic regression model (without *immigration attitude*):**
   - In this model the predictors are: pid7(), race, educ (education), gender.
   - After comparing the plots (for displaying probabilities based on one or two features at a time) for these features separately, *race* and *educ* were selected as the most important features to draw conclusions for comparing model inferences.
   - Interaction between features *race* and *educ*:
     - Without interaction between *race* and *educ,* the probability v/s educ faceted by race plot showed the same downwards trend (shifted intercept).
     - But when the interaction was added between these features, we saw difference in the probability shifts for all the facets.
     - Hence, this interaction was necessary to be included in the model.
     - We made this decision to include the interaction based on the p-value (mentioned in Part 2).

Coefficients for Model 1:

```
Coefficients:
                Estimate
(Intercept) -4.6997276
pid7         0.5717425
gender.L    -0.1900613
race.L      -1.5800084
race.Q       0.6397559
race.C      -0.6914033
educ         0.0097877
race.L:educ  0.3318896
race.Q:educ -0.0903996
race.C:educ  0.0009077
```

Plot for Model 1:



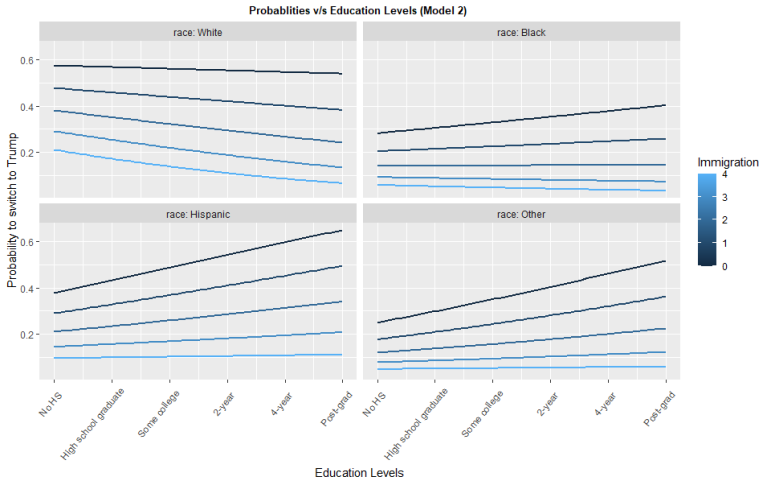
Plot (1) [Model 1]

2. **Model 2: Fitting a weighted logistic regression model (with *immigration attitude*):**
   - In this model the predictors are: pid7(), race, educ (education), gender, imat (*immigration attitude*).
   - For this model we decided to keep the *race* and *educ* interaction as the above model (for consistency). We also found that fitting an interaction between *imat & educ* showed more comprehensive plots, suggesting that this interaction is important as well.

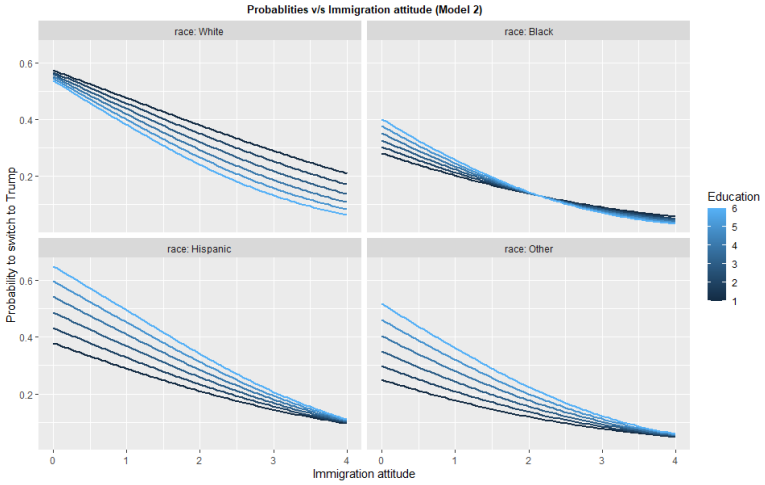
Co-efficient of Model 2:

```
Coefficients:
             Estimate
(Intercept) -3.45943
imat        -0.47957
educ         0.18279
pid7         0.50385
gender.L    -0.18617
race.L      -1.42327
race.Q       0.53347
race.C      -0.79741
imat:educ   -0.06743
educ:race.L  0.27621
educ:race.Q -0.08813
educ:race.C -0.02880
```

Plot for Model 2:



Plot (2) [Model 2]



Plot (3) [Model 2]

Compare the results of the two models. Does including immigration attitudes make a substantive difference?

---

1. By comparing both the models, with and without *immigration attitude,* we are trying to quantity whether *immigration attitude* has any effect or not by looking at the probability-plots using the features: *(1) race (2) educ.*
2. We also found *pid7* can be used as to draw conclusions, but we decided to drop the inferences from *pid7* probability-plots as the results from inferences weren't as comprehensive as *race & educ.* (probability-plots showed the same trend in all the facets for **Model 2**).
3. Comparison of Model 1 and Model 2:

a. For Model 1, all the facets (by race) show a generalized trend of probabilities as per the Education levels. But when we plot the same graph, after adding *immigration attitude* as an additional feature, according to Model 2, we can see more comprehensive differences in the trends if we look at the results from the perspective of *immigration attitude.*
b. In other words, in Model 2 plot, for all races, using *immigration attitude* as an additional feature shows varying trends for different *immigration attitude* levels (which is being generalized in plot for Model 1).
c. Hence, using *immigration attitude* as a predictor gives us more information about the voting patterns accounted by important features *race & educ. (As discussed below*)*

4. Comparing Plots (2) and (3) [Model 2]:
[Considering key demographics: Education levels & Race]

Using Plot (2) as the reference:

- For race: White,
    - The general trend is that with increasing education levels, the prob. of switching to Trump decreases.
    - But for higher *immigration attitude* levels (i.e. levels 2, 3 & 4) the prob. of switching to Trump decreases slightly more as compared to the lower levels (i.e 0, 1).
- For race: Hispanic/Others,
    - As the level of education increases, for the lower *immigration attitude* levels (i.e. levels 0, 1 & 2) we can see that the probability of switching to Trump increases.
    - But for the higher *immigration attitude* levels (i.e. levels 3 & 4) we can see that there is no change in the prob. to switch to Trump over the different education levels.
- For race: Black,
    - As the level of education increases, for the *immigration attitude* level 0, there is a slight increase in the prob. to switch to Trump. Inversely there is a slight decrease in the same, for *immigration attitude* level 4.
    - For the rest of the *immigration attitude* levels, there is no change in the prob. to switch to Trump, as the level of education increases.

Using Plot (2) as the reference:

- The general trend is that as the *immigration attitude* level is increasing the prob. of voting for Trump is decreasing, which means as we higher on the *immigration attitude* level scale the prob. of switching decreases.
- For race: White,
    - As the *immigration attitude* level increases, and as the education level increases there is a steeper decline in the prob. of switching to Trump.
    - For lower *immigration attitude* levels, there isn't much divergence in the prob. of voting Trump. But for the higher *immigration attitude* levels, the prob. of voting Trump diverges according to the level of Education (as mentioned above).
- For race: Hispanic/Others,
    - The interaction brings the curve together as the *immigration attitude* levels increase.
    - For lower *immigration attitude* levels, the level of education (low or high) makes a big difference on the prob. to switch to Trump.
    - In contrast to the impact of *Education* level in race: White, lower the education corresponds to lower the prob. to vote to Trump.

    - However, as for the higher (or highest) *immigration attitude* level, the impact of level of *Education* diminishes. At this point, people are unlikely to vote Trump irrespective of *Education*.
- For race: Black,
    - The level of *Education* makes no difference i.e. for all Education levels, as the *immigration attitude* level increases, the general trend to switch to Trump also decreases.

Does *immigration attitude* matter more for some demographic groups than others?
YES. As mentioned in the above points, different people belonging to different demographic groups have different prob. to switch to Trump after accounting for *immigration attitude,* according to their own trend. for eg. race: Black has disparaging trends as compared to race: White. These trends are different for race: Hispanic & Other (but Hispanic & Other they have the same trends).

CONCLUSION:

As discussed in the second question, immigration has an interaction with race and education. After trying multiple combination of interactions of demographic variables with immigrations, we finalized our model with the interaction of *immigration* and *education* along with *race* and *education*.
Immigration attitude has a substantial effect in switching of the voting behavior. In the 3rd question, we compared the models where one had all the demographic variable except *immigration* and other had *immigration* and all other demographic variables, in the latter one we could see that different immigration levels had an impact on probability of switching to Trump based on the different racial and educational demographic groups.