MiniProject2

Ojaas, Meet, Vijeet

Dataset Description:
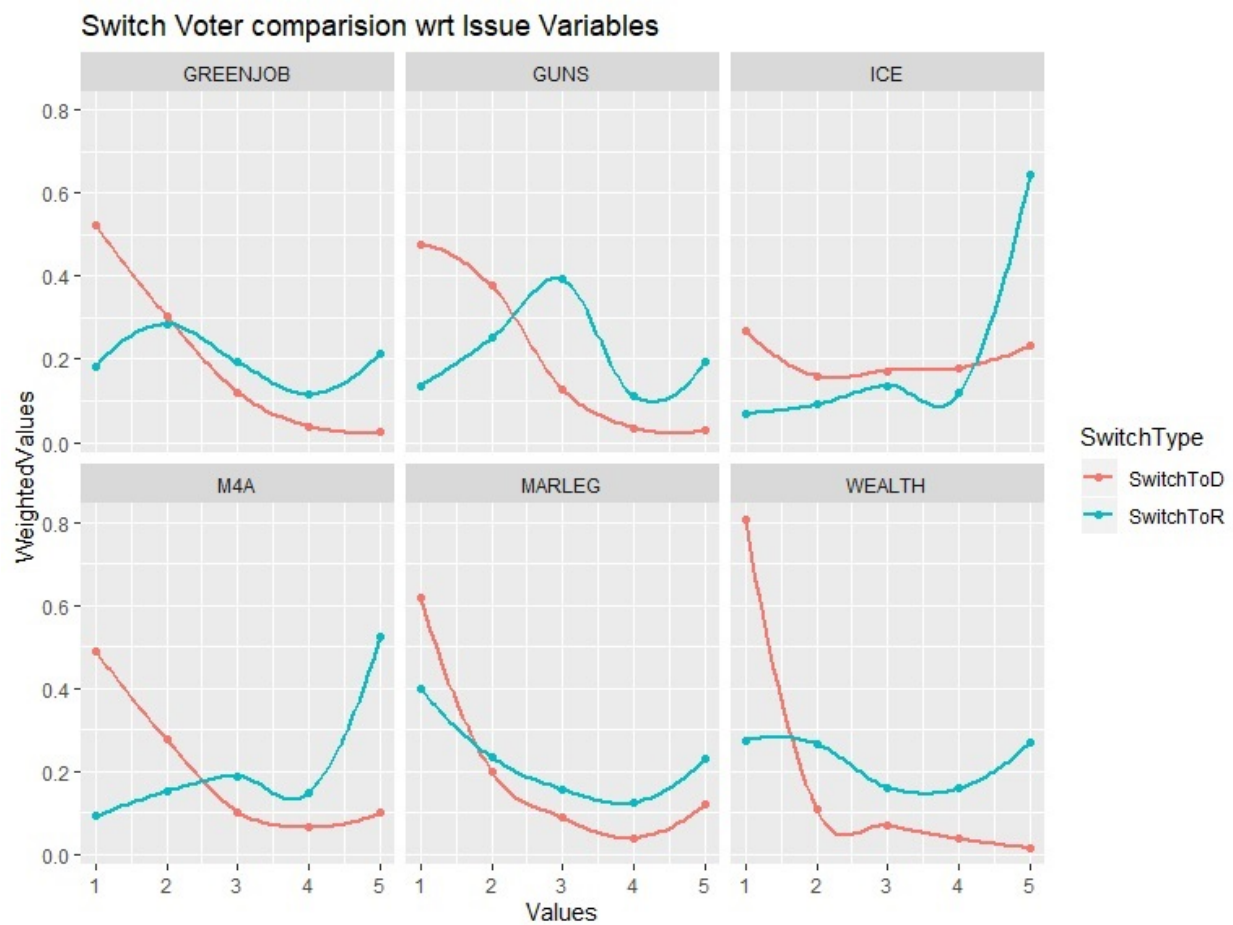
Issue Variables: M4A GREENJOB WEALTH MARLEG ICE GUNS

Populism Variables: POP_1 POP_2 POP_3

Weights: weights_DFP

Question 1:

How do the Switch to R and Switch to D voters differ on the issue variables?

Brief Desription of Graph Parameters:
Y-Axis (WeightedValues) – Mean(Survey Weight * Opinion)

X-Axis (Values) – Opinion (1-Strongly Support, 5-Strongly Oppose)

SwitchToD – Switched Vote from Republican to Democrats in 2019

SwitchToR – Switched Vote from Democrats to Republican in 2019

Analysis:

While accounting for the trends between the two category of voters namely Switching to Democrats and Switching to Republicans, we get a scale of 1 to 5 with 1 being strong support and 5 means strong opposition, and there is one more value 6 which corresponds to the 'Not Sure' class of voters. While observing the difference between voters of the two swinging categories, we drop the 6 value and consider all other classes i.e. 1 to 5. 1. In the case of GreenJob variable, we can observe that the voters only differ on the strong support parameter (1), otherwise the graph for both the category of switch voter follows almost the same characteristics. 2. Strong supporters of Gun Control have a high chance to Switch votes to Democrats whereas as for the voters who neither strongly support nor strongly oppose to Gun Control have a high chance to switch their votes to Republicans. 3. Voters with strong opposition towards Defunding Immigration and Custom enforcements have a very high probability to switch votes to Republicans. 4. Strong supporters of Medicare For All might switch votes to Democrats while strong opposition have a high probability to switch votes to Repiblicans 5. Legalizing Marijuana has a similar trend with very menial deviations for either category of switch voters. 6. Strong supporters of Tax on Wealth over $100 million will most probably switch their votes to Democrats, while the people who are bound to Switch votes to Republicans follow a neutral trend throughout the graph.

Question 2:

How do swing voters differ from Loyal Democrats and Loyal Republicans on the Issue Variables?
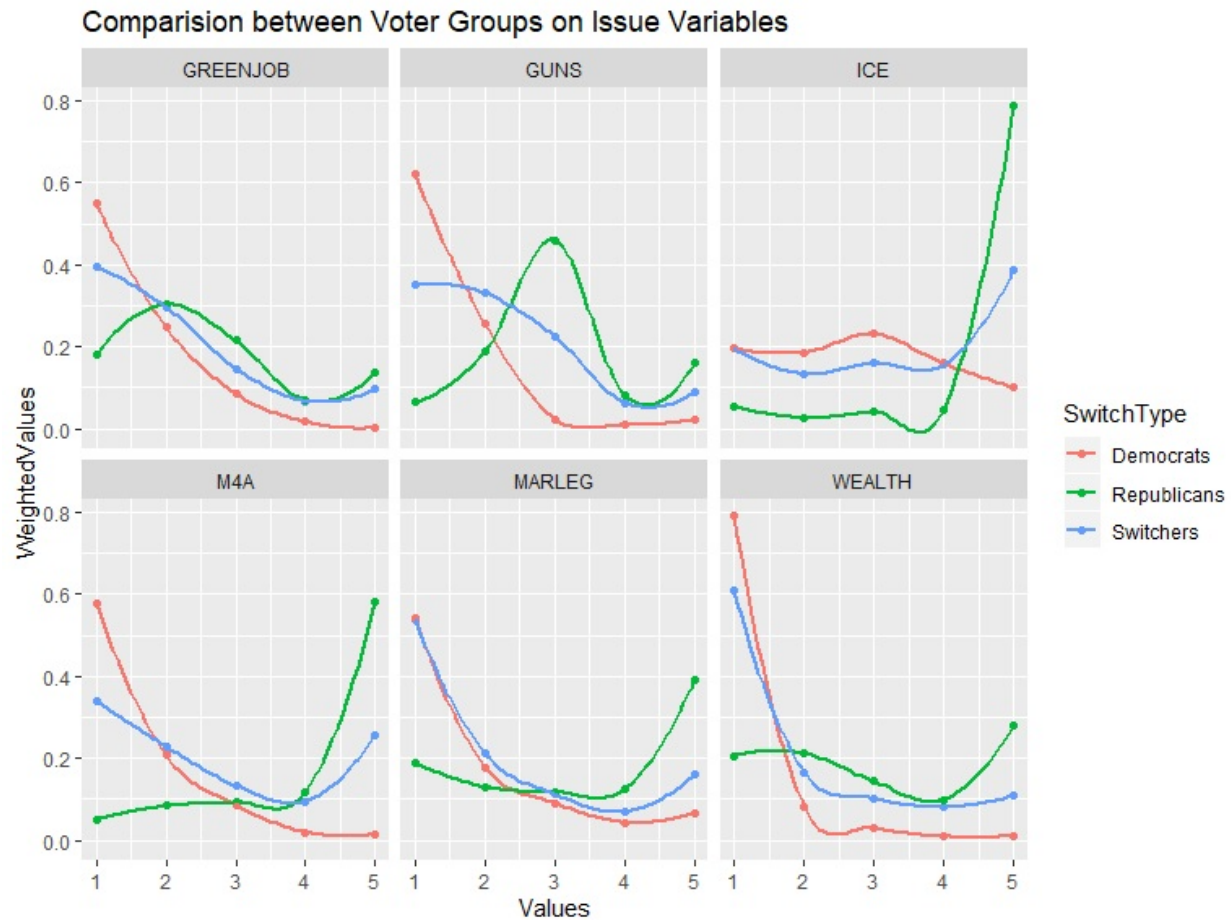
Comparision between Voter Groups on Issue Variables

Brief Desription of Graph Parameters:

Y-Axis (WeightedValues) – Mean(Survey Weight * Opinion)

X-Axis (Values) – Opinion (1-Strongly Support, 5-Strongly Oppose)

Democrats –Loyal Democrats

Republicans –Loyal Republican

Switchers – Switched Votes from Republicans to Democrats or vice versa

Analysis:

Based on the comparison between the Switch voters and the loyal republican as well as democrat voters we can see that two of the given hypotheses can be proven with ease.

Hypothesis 1 - "On most issues, swing voters are split, with some of them acting more like Democrats and others acting more like Republicans." This hypothesis can be seen very clearly in the Medicine for All variable. The strong supporters if the M4A program tend to have a high probability of switching to Democrats, while the strong opponents of the M4A program tend to have high chances of switching their votes to Republicans.

Hypothesis 2 - "Swing voters think more like Democrats on some issues and more like Republicans on other issues." This hypothesis is proven in all other Issue variables apart from the M4A variable. The switch voters follow the general trend followed by the loyal Republican voters in the ICE (Defunding Immigration and Customs Enforcement) Issue Variable. While for all other issue variables (i.e. Greenjobs program, Gun Control, Marijuana Legalization, Tax on Wealth over $100 Mn), we can see from the graph that the switch voters tend to follow the characteristics of Loyal Democrat voters.

Question 3:

What predicts being a swing voter?

Build two models to probabilistically predict whether a registered voter is a swing voter:

- One model should use ONLY the issue variables as predictors.
- The other model should use ONLY the populism variables as predictors.
-

How well do your models do? Which of your models does better? If you had to guess, what factors are most important in determining what makes a voter a swing voter?

We have created a column called SwitchType that states the probability of a Voter to switch his/her Vote from Republican to Democrat and vice versa.

Question 3 Part A: Model based on Issue Variables

The model gives an increase in accuracy values as we add various predictor variables to the model in order to predict the SwitchType.

 "Train accuracy for model with 6 predictors is: 85.1 %"

Here we observed that the model gives the highest accuracy of 85.1% when all the variables are added in it.
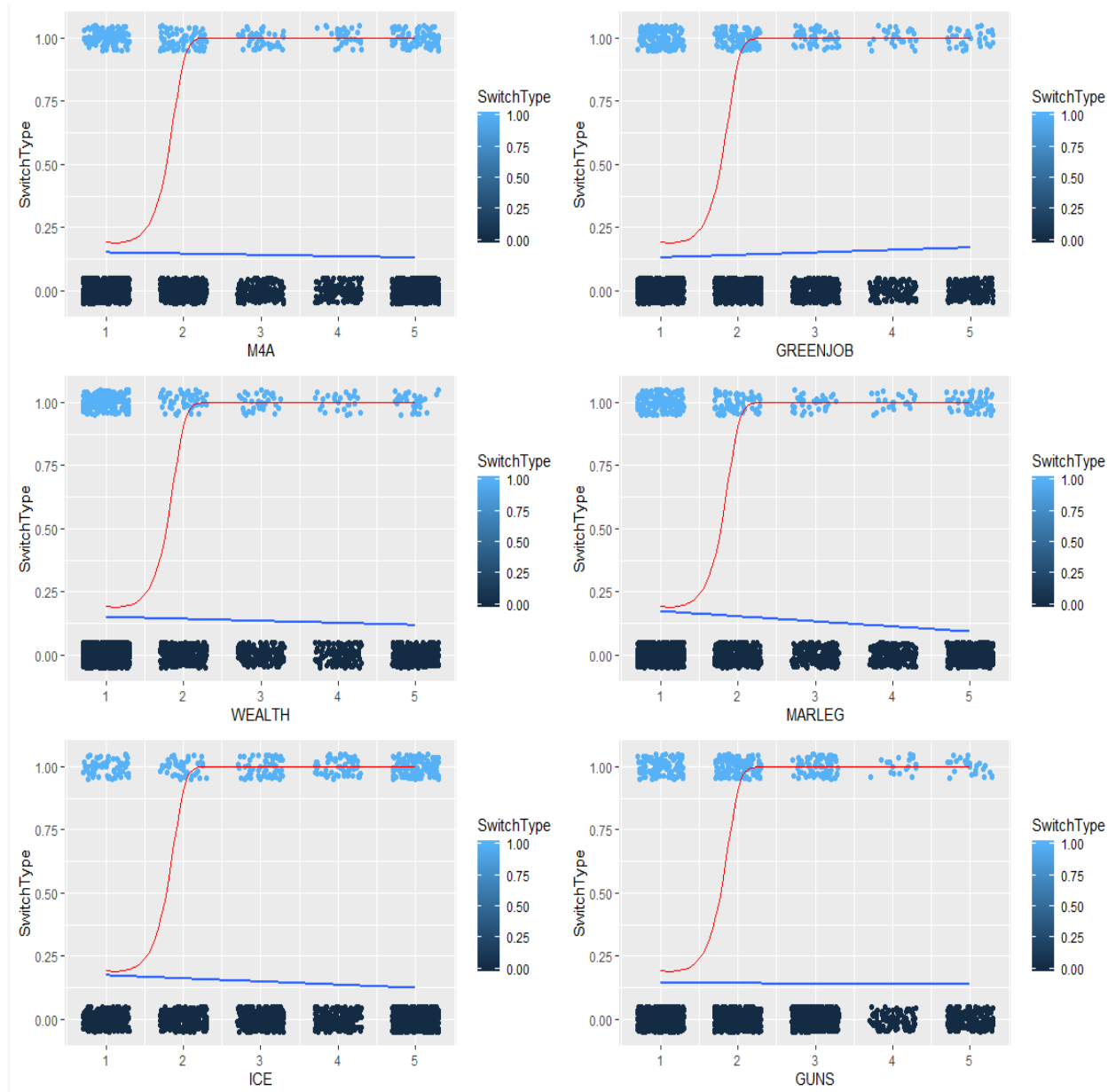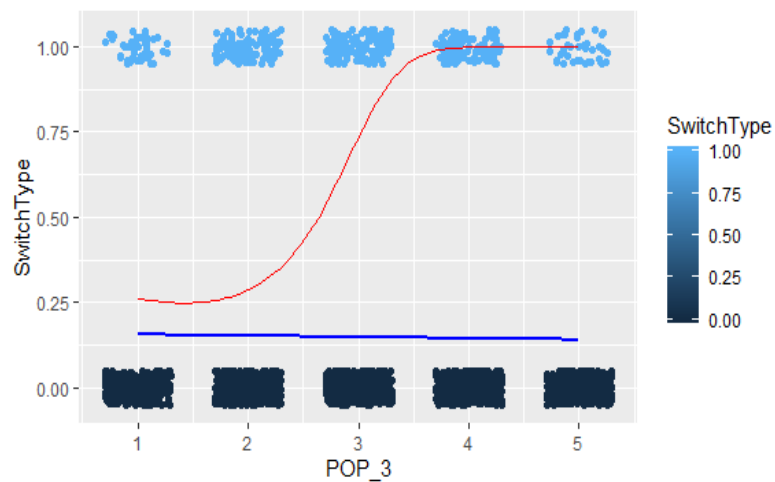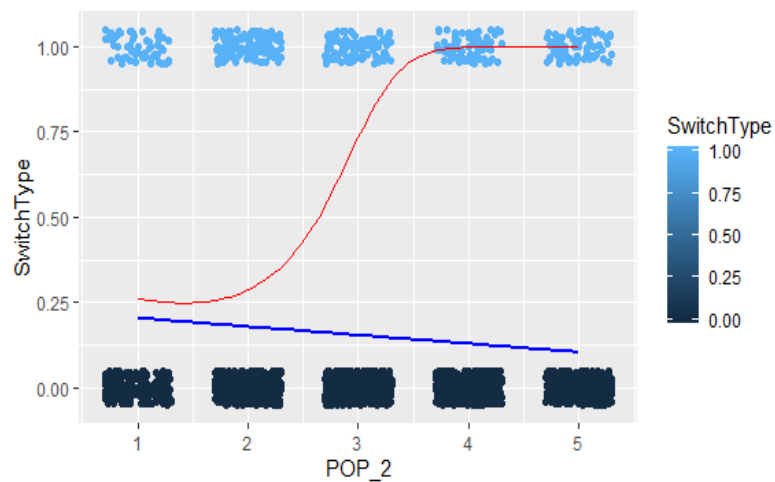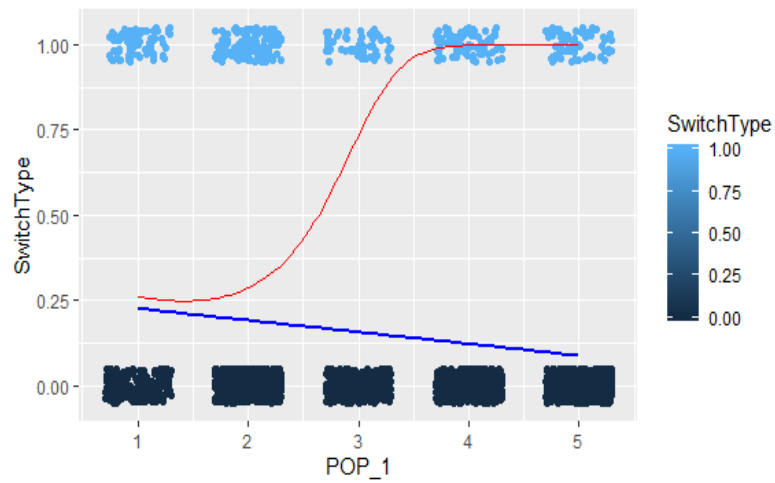
From observing the graph above, we can see that the red line represents the overall trend that the model follows while taking all the variables into account. We can see from the red line that the probability of being a swing voter is high only and only when the person is a strong supporter of the policies, whereas the probability of the person swinging his vote increases as his/her opinions become neutral or opposing. The individual graphs for the variables with respect to swinging probability is shown by the blue lines in the graph. From the graph for the blue lines, we can conclude that the issue variables like Greenjob Issues (GREENJOB), Marijuana Legalisation (MARLEG) and defunding Immigration and Custom Enforcements (ICE) show a stronger correlation to changes in Y wrt changes in X rather than the other three issue variables.

Question 3 Part B : Model based on Pop Variables

Train accuracy for model with 3 predictors is: 84.9 %

In case of the populism variables, the model gives a highest accuracy value of 84.9% when all the three pop variables are added to it.

From observing the graph above, we can see that the red line represents the overall trend that the model follows while taking all the variables into account. We can see from the red line that the probability of being a swing voter is high when the person is a supporter of the policies (opinion rating of 1 or 2), whereas the probability of the person swinging his vote increases as his/her opinions become neutral or opposing (opinion rating between 3 and 5). The individual graphs for the variables with respect to swinging probability is shown by the blue lines in the graph. From the graph for the blue lines, we can conclude that the populism variables like Pop1 and Pop2 show a stronger correlation to changes in Y wrt changes in X rather than the Pop3 variables.

Model	Accuracy
Model using all the 6 issue variables	85.1%
Model using all the 3 populism variables	84.9%

When we compare the two models, we can see that the model accuracy for Pop variables is 84.9% while that for the issue variables is 85.1%. Hence, we can conclude that the model based on issue variables works better than the one based on pop variables.