

Exploratory Data Analysis - Mini Project 2

Team Colorado

Prashanth Thirukkurungudi Sekar, Pravin Sundar,
Barathwaaj Parthasarathy, Bhavna Sinha

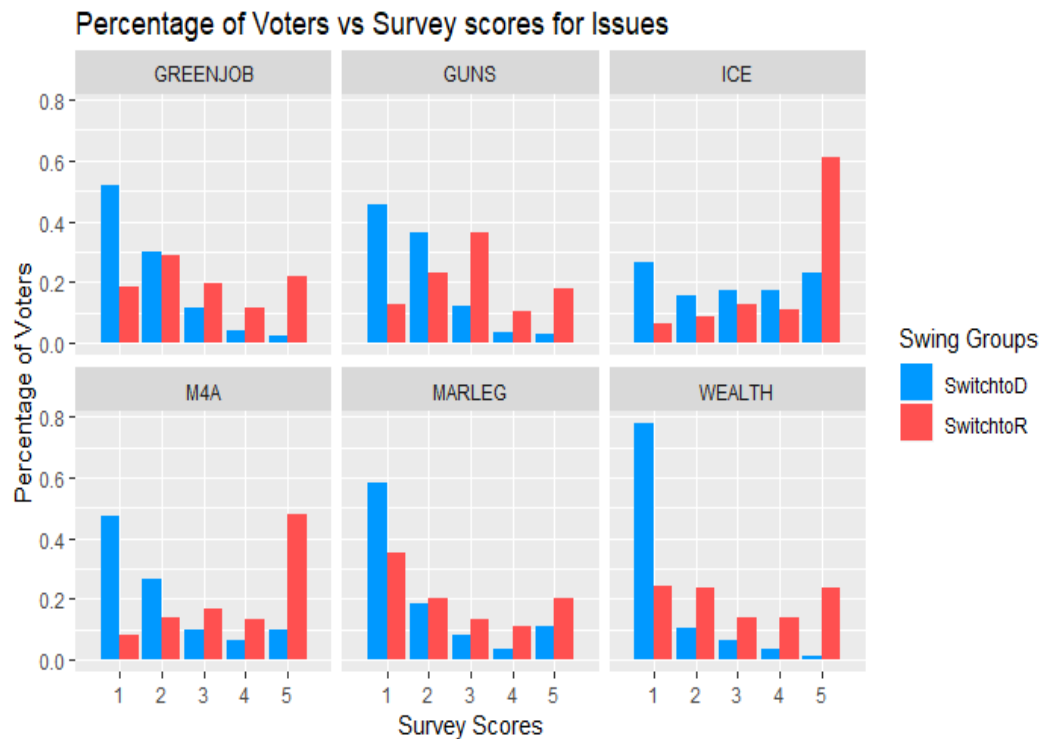
April 8, 2019

1 Introduction

This report explores the distribution of defined voter groups who registered to vote in the 2018 mid-term elections. Specifically, the swing voters, people who voted in 2018 who were neither a Loyal Democrat nor a Loyal Republican. By looking at the differences and similarity of the swing voters with the other groups across different subjective variables (issue variables ; populism variables) models are built to explore what predicts a swing voter.

2 Distribution of Switch to D and Switch to R on Issue Variables

To understand how the two groups vary on the issue variables, percentage of voters (weighted) on each level grouped by the variable is considered.



From the graph above, similarity in opinion between the voters can be clearly observed towards legalization of marijuana (MARLEG). With 60% of switch to D and 35% of switch to R strongly supporting the issue, the percentage decreases in strength of the population opposing the issue with only 10% and 20% of switch to D and switch to R respectively.

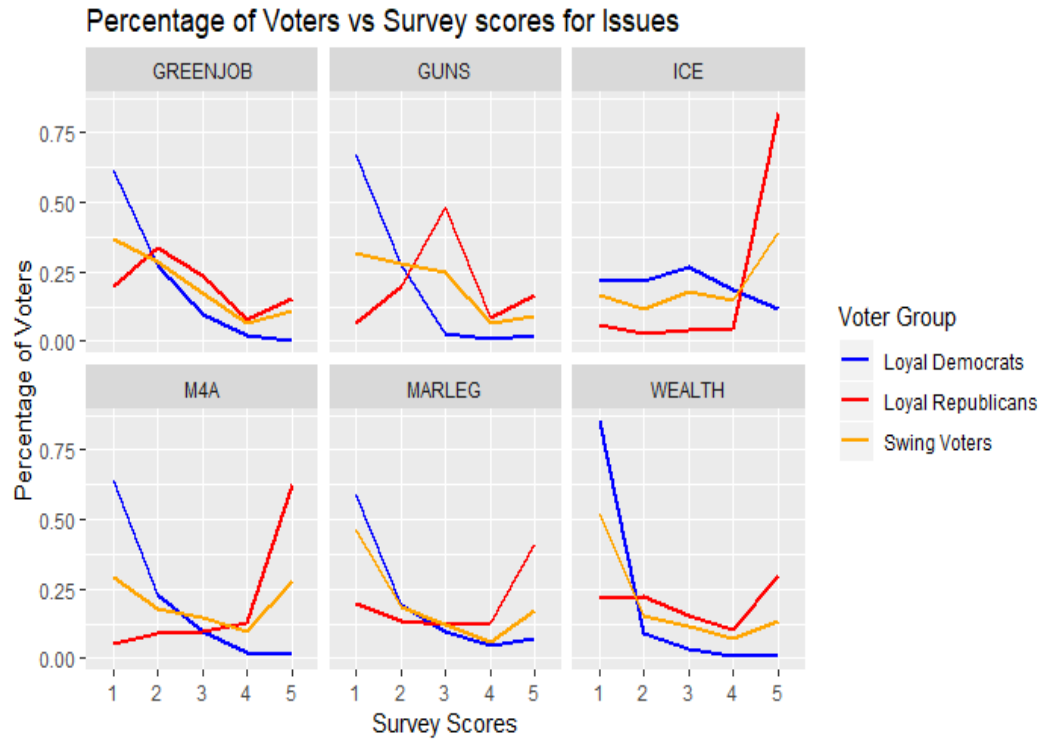
A clear difference in opinion can be observed with respect to Medicare for All(M4A) issue. The majority of switch to D voters strongly support it (about 50%) and the majority of switch to R voters strongly oppose it (about 50%).

With respect to the issues: Defunding Immigration and Customs Enforcement(ICE), tax on wealth over \$100 million(WEALTH) and the Green Jobs program(GREENJOB) a majority of a single group of voter either supporting or opposing the issue with the other side evenly spread across opinions can be seen. Specifically, with Defunding immigration(ICE) we see 60% of switch to R voters opposing it and switch to D evenly spread. In the case of tax on wealth(WEALTH) switch to D voters shows complete support with 80% of the people in the voter group. Finally similar trend can be observed for Green Jobs program(GREENJOB) with about 50% of switch to D voters showing strong support and an evenly spread population of switch to R.

What can be said about Gun control(GUNS) is that majority of switch to D voters are in support with 45% and 35% of people with a score of 1 and 2 respectively. But majority of switch to R voters have a neutral opinion towards gun control(35%).

3 Distribution of Swing Voters and Loyalists on Issue Variables

Swing voters are those who are not loyal to a specific party. If they feel that a party's views on issues are not beneficial to them, they will switch to the party that favours their opinion. In this part, we will see how these group of voters, the swing voters, similar to or vary from those who are loyal to a single party.



The above graph shows the survey scores and percentage of people (weighted) of each voter group who selected the score.

- For the Green jobs Program(GREENJOB) issue, It can be seen that around 60% of the loyal Democrats strongly support the issue. The majority of the swing voters seem to support but not to the level of the Democrats. The loyal Republicans are also in favour of the issue but their scores are not skewed towards a score. But, the distribution of the swing voters is similar to that of the loyal republicans
- In the case of Gun control(GUNS) issue. The majority of the swing voters strongly support the issue which is similar to the loyal Democrats. The loyal Republicans seem to be neutral in this case
- For the Defunding Immigration and Customs enforcement issue(ICE), the swing voters strongly oppose the policy similar to that of the loyal Republicans. It can be seen that even though the orange line is between the red and blue, the distribution of swing voters is very similar to that of the Loyal Republicans.
- Looking at the Medicare for all issue(M4A), it can be observed that some part of the swing voters behave like loyal democrats(1,2,3) and the others

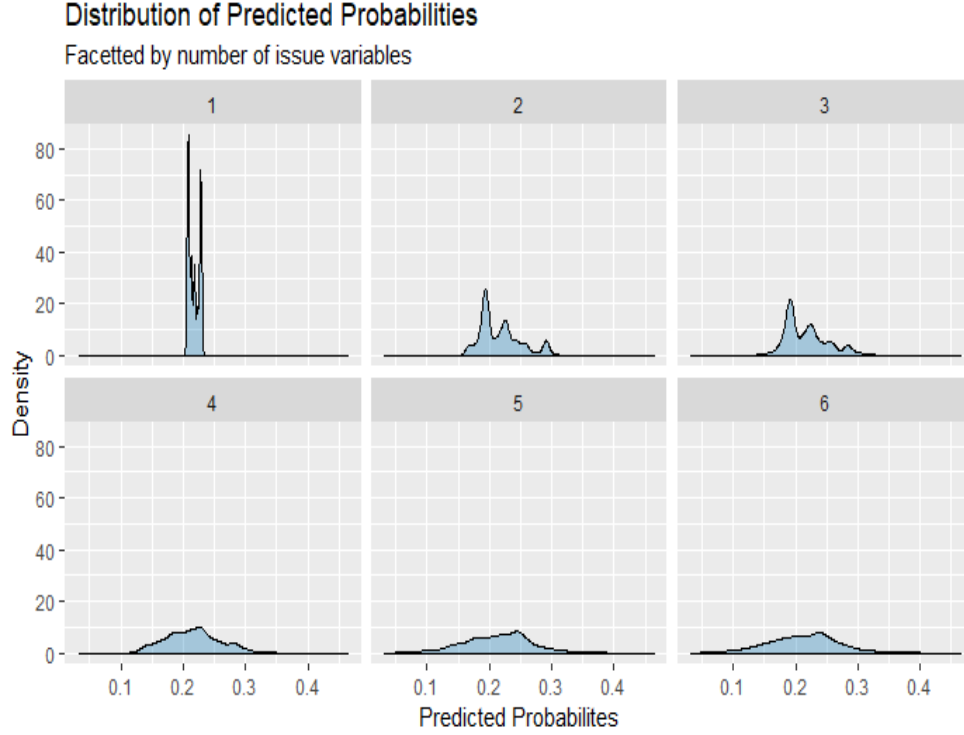
behave like loyal republicans(4,5). The number of people who strongly support and oppose is almost the same

- In the case of the Marijuana Legalization(MARLEG) issue, the swing voters behave similar to the loyal democrats(strongly support) as seen by the similar orange and blue lines
- For the Tax on Wealth(WEALTH) over \$100 million issue, the behaviour of swing voters is very similar to that of the democrats. They seem to strongly support the law.

4 Probabilistic Model on Swing Voters

To determine what predicts being a swing voter, two models are built, one considering only the issue variables and another considering only the populism variables. From the data, survey score of 6 meant that people were not sure about their opinion, for the ease of exploration we treat this score as an equivalent of a neutral score (Score 3). The models were built in a sequential order with increasing number of variables selecting the variable combination with lowest error. The predictions were made with a threshold value of 0.25 for individual variables and 0.45 for additional variable combinations. The error distribution can be observed from Table 1, in the appendix. There is little to no difference in the error term with single variables, with Green jobs program (GREENJOB) and Gun control (GUNS) being an exception. The combinations of additional variables after selecting Medicare for all (M4A) as the first variable only has difference in error after the 4th decimal. Similar trend is followed with every subsequent addition of the variable with very minimal change in error. The lowest error of 19.13% can be observed when all the 6 variables were included. The distribution of predicted probabilities for the chosen combinations can be seen below.

4.1 Issue Variables



1:M4A, 2:M4A+GREENJOB, 3:M4A+GREENJOB+WEALTH,
4:M4A+GREENJOB+WEALTH+ICE,
5:M4A+GREENJOB+WEALTH+ICE+GUNS,
6:M4A+GREENJOB+WEALTH+ICE+GUNS+MARLEG

It can be observed that the probabilities tend to spread out more as the variables increase. With most frequent probability around 0.15 to 0.25.

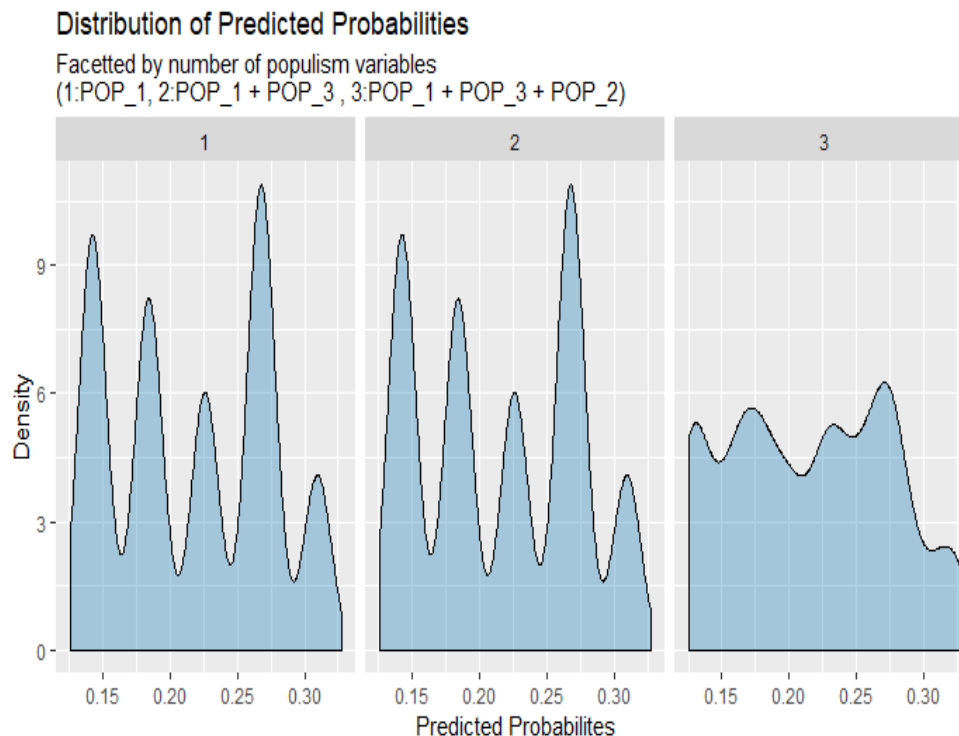
Comparing all the models together, as more and more predictors are added the prediction of swing voters become better, as can be seen from the marginal decrease in the prediction error.

If it had to be guessed, among the issue variables, those variables that have a difference in the policy between the political party might sway the voters depending on their beliefs. As can be seen from Table 1. in appendix that all issue variables as a combination gives the best prediction.

4.2 Populism Variables

Following the previous exploration, similar approach was adapted to predict the swing voters using the populism variables. From Table 2. in appendix, the distribution of error with inclusion of every variable can be observed. There is

little to no difference in the prediction error as each variable is included. Below is the distribution of probabilities across the three models chosen.



The distribution does not change much moving from a one variable plot to two variable to three variables. Given the subjective nature of the variables and the model performances, there is no clear distinguishing variable or combination of variable that can be used to predict a swing voter. This error percentage did not improve even with the inclusion of interaction variables validating the above conclusion. Hence, it is better to keep all combination of variables to include the variation in opinions.

5 Conclusion

How do Switch to D and Switch to R voters differ on the issue variables? From the distribution, the two groups share similar opinion towards legalization of marijuana with both groups in support of the issue and difference of opinion in Medicare for All where the two groups stand in opposite sides. A strong support from switch to R voters for Defunding Immigration and Customs Enforcement and a strong support from Switch to D for Green Jobs program and tax on wealth over \$ 100 million can be observed.

Both parties are somewhat in support of gun control but with majority of switch to R taking a neutral stance.

How do swing voters differ from loyal Democrats and loyal Republicans on the issue variables?

- Swing voters think more like Democrats on Tax on Wealth(WEALTH) and Marijuana Legalization(MARLEG) and Gun control(GUNS) issues and more like Republicans on Green jobs Program(GREENJOB), Defunding Immigration and Customs enforcement(ICE) issues.
- For the Medicare for All(M4A) issue, swing voters are split, with some of them acting more like Democrats and others acting more like Republicans.

What predicts being a swing voter? Among the issue variables, an additive model consisting of all the 6 variables seemed to be the best combination to predict a swing voter. Among the populism variables, all combination of the variables gave similar probability distribution, but keeping all variables to predict a swing voter includes the variation in opinions.

6 Appendix

Variable Combination	Error
M4A	0.1919
GREENJOB	0.2522
WEALTH	0.1919
MARLEG	0.1919
ICE	0.1919
GUNS	0.2423
M4A + GREENJOB	0.1919
M4A + WEALTH	0.1919
M4A + MARLEG	0.1919
M4A + ICE	0.1919
M4A + GUNS	0.1919
GREENJOB + M4A + WEALTH	0.1919
GREENJOB + M4A + MARLEG	0.1919
GREENJOB + M4A + ICE	0.1919
GREENJOB + M4A + GUNS	0.1919
M4A + WEALTH + GREENJOB + ICE	0.1919
M4A + WEALTH + GREENJOB + MARLEG	0.1919
M4A + WEALTH + GREENJOB + GUNS	0.1919
M4A + WEALTH + ICE + GREENJOB + GUNS	0.1925
M4A + WEALTH + ICE + GREENJOB + MARLEG	0.1922
M4A + WEALTH + ICE + GREENJOB + MARLEG + GUNS	0.1913

Table 1: Model Errors for Issue Variables

Variable Combination	Error
POP_1	0.1941
POP_2	0.1941
POP_3	0.1941
POP_1 + POP_3	0.1941
POP_1 + POP_2	0.1941
POP_3 + POP_1 + POP_2	0.1941

Table 2: Model Errors for Populism Variables