

Mini Project 2: Thinktank Dataset

Chhavi Sharma, Gopal Seshadri, Prahasan Gadugu, Supriya Ayalur Balasubramanian

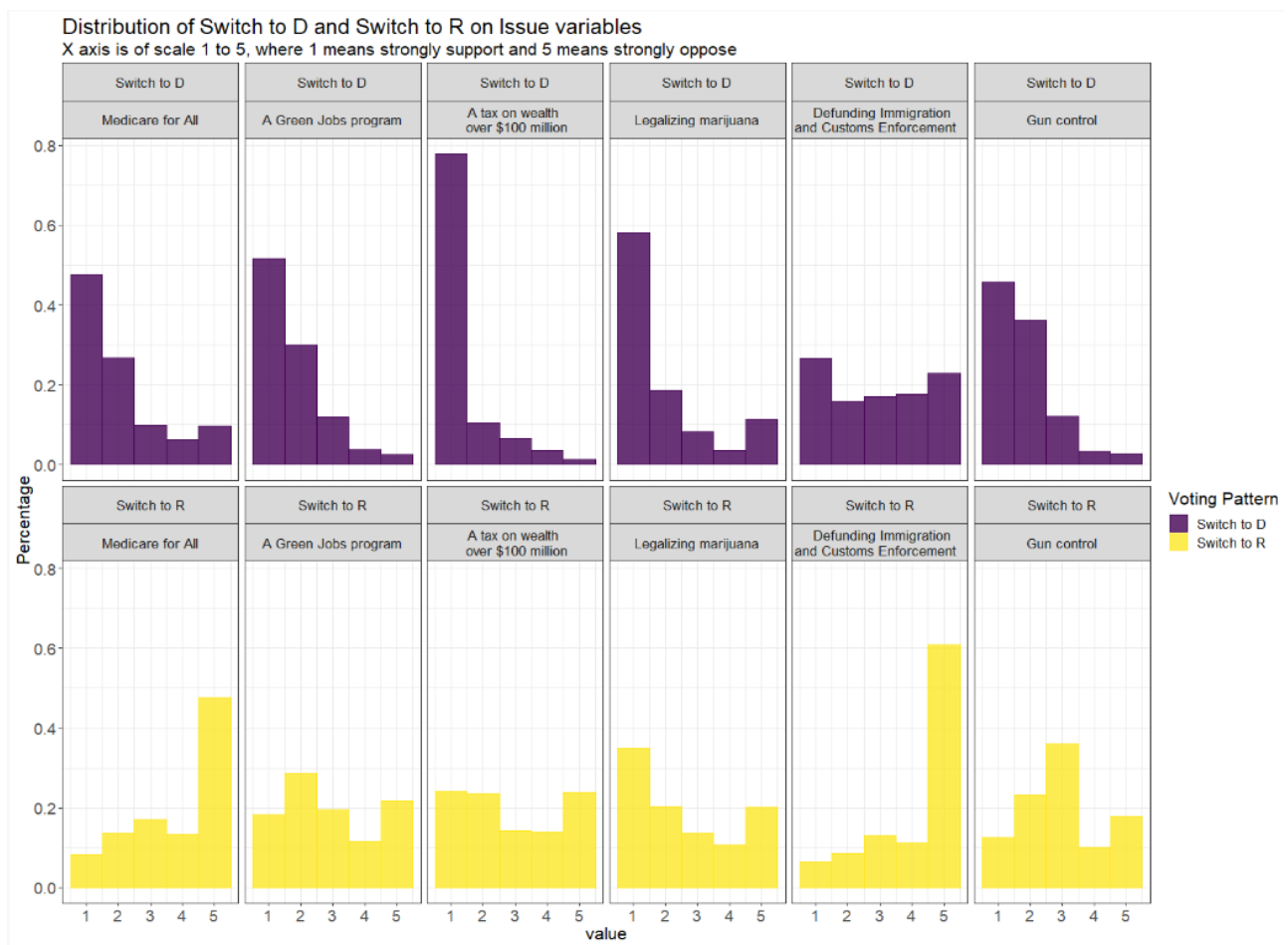
(Team Kansas)

Introduction:

The Thinktank Data for Progress collected survey data (in DFP WTHH release.csv) that represents the population of people registered to vote in the 2018 midterm elections. As per the mini-project, we need to address the following questions:

Question 1:

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



As per the above histogram graph, Switch to D and Switch to R voters differ on the issue variables in the following manner:

Medicare for All: People switching to Democratic Party and people switching to Republican Party are strongly divided over this issue variable. We can see a clear divide wherein people switching to Democratic

party are in strong support of the reform indicated by the much higher percentage of voters near the 1 (about 48%) and 2 (about 28%) sections on the scale of support and the contrasting decrease in percentage distribution of voters as we move towards sections 4 and 5 (about 17% combined) that signify opposition towards the reform.

Green Jobs program: People switching to Democratic party strongly support this reform as can be seen clearly by the huge percentage of voters at 1 (about 51% of voters) and 2 (30% voters) on the scale of support. Moreover, we observe a gradual decrease as we move towards 5 on the support scale with only around 1-2% voters strongly opposing the reform. In contrast, people switching to Republican party have a minor peak around 2 (about 29%) and a minor low at 4 (around 12%). However, the percentages of voters across 1-5 on the scale of support is equal across all sections. This indicates a near neutral approach of 'Switch to Republican' voters towards this reform.

Tax on wealth over \$100 million: People switching to Democratic party provide a very strong backing to this reform as can be seen in the plot where the percentage of voters around area 1 (scale of support) towers across all other sections with a huge majority of about 78% whereas the percentage distribution near opposing scales is almost minimal with just about 1% people strongly opposing the reform. However, people switching to the Republican party have a neutral attitude towards the plan owing to almost equal percentage of voters across all sections of the scale.

Legalizing Marijuana: Similar to 'Green Jobs program' issue, people switching to the Democratic party strongly support legalizing Marijuana owing to higher percentage of people near 1 on the support scale (about 58-59%) and in contrast, a sharp decrease in percentages as we move from 1 to 4 with 4 having as low as 3% voters opposing the reform. However, we observe another minor peak at 5 comprising around 11-12% voters which suggests a much smaller section of people switching from Republican to Democratic strongly opposing the reform. For people switching to Republican, the percentages of voters are again almost equal across all sections of the scale of support.

Defunding Immigration and Customs Enforcements: This issue demonstrates an opinion near-opposite to the opinion on Legalizing Marijuana. In this case, people switching to Democratic Party have almost equal percentages of voters across all sections on the support scale which suggests no strong opinion for these kind of swing voters towards the reform. However, voters switching to the Republican party have a minor peak around 3 (about 12%) which suggests a small section of people with a neutral approach towards the issue, and a sharp peak at 5 (about 61%) which indicates a strong opposition towards the reform.

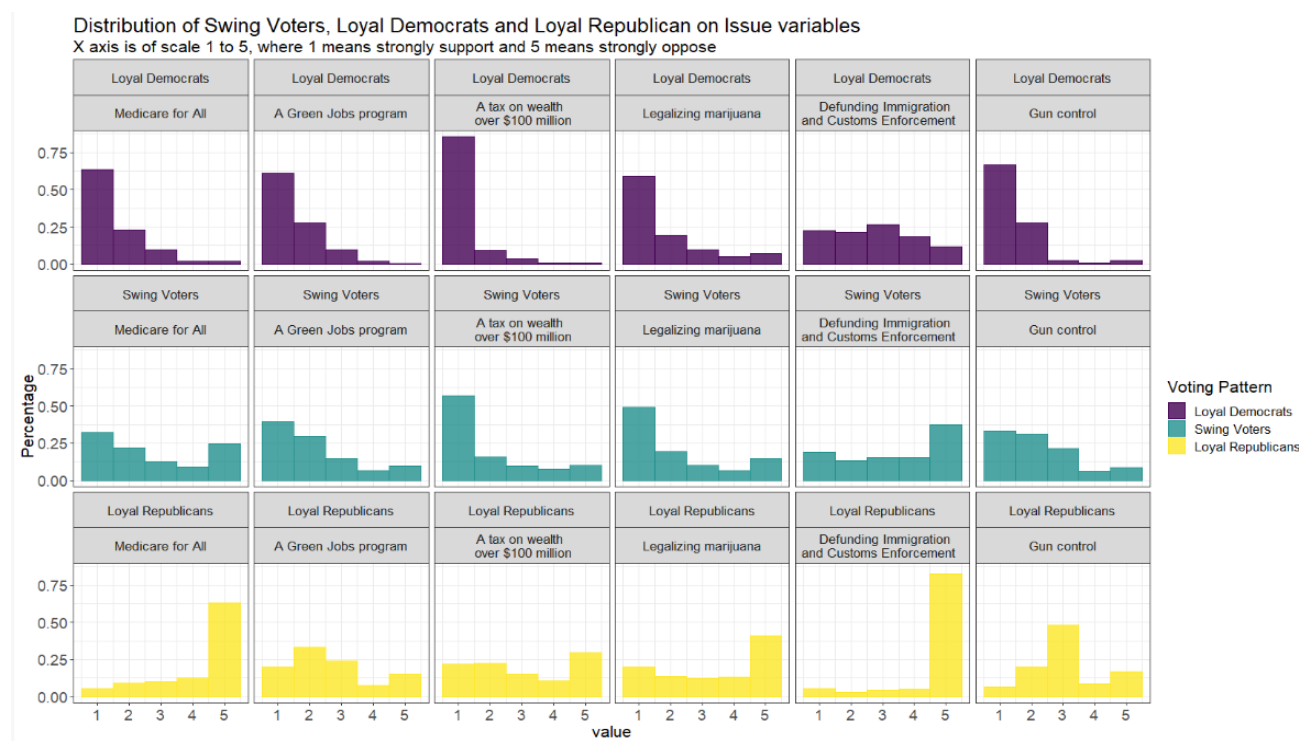
Gun Control: For this issue, people switching to the democratic party have major percentage peaks at 1 and 2 sections on the support scale which exhibits a strong support towards the reform comprising a combined percentage of about 82% voters. Thereafter, the percentages gradually decrease as we move towards 5 with a mere 2% voters strongly opposing the issue. For voters switching to the Republican party, the opinion is neutral for most voters since the highest peak is at 3 on the support scale. The percentage of voters increases until 3 and decreases from there as we move until 4. There is another small peak at 5 which indicates a small section of these swing voters opposing the reform. Therefore, it indicates a divided house as far as voters switching to the Republican party are concerned with a slight bend towards a neutral opinion.

‘Medicare for All’ demonstrates a strong disagreement in opinion since voters switching to the Democratic party exhibit a strong support for the issue, however, voters switching to the Republican party exhibit a strong opposition towards the reform.

The opinion on ‘Legalizing marijuana’ seems to be somewhat similar as compared to the evidently different opinions on all other issues. This can be seen from the fact that the swing voters switching to Republican or Democratic party- both exhibit a strong support towards the reform with their major peaks at 1. Thereafter, both kinds of swing voters exhibit a decrease in support as we move from 1 to 4 until 5 where both have a smaller section of voters who strongly oppose the reform as can be seen through their minor peaks at 5 indicating strong opposition.

Question 2:

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



On most issue variables, swing voters are split, with some of them acting more like Democrats and others acting more like Republicans.

The hypothesis ‘Swing voters are moderates and tend to be in the middle of the distribution when Democrats are on one side and Republicans are on the other’ fits the ‘Medicare for All’ issue variable best since as per the plot, the swing voters are moderates since their percentage distribution is spread out evenly across all sections and is relatively uniform. In case of the loyal Democrats (right-skewed), there is a higher percentage of around 63% of loyal Democrats indicating strong support to the reform whereas loyal Republicans (left-skewed) have a higher percentage of voters of around 63% at 5 indicating strong opposition. On a closer look at the above graphs, we can conclude that the swing voters are split in their opinion towards Democrats and Republicans as far as these issues are concerned. They have a significant percentage distribution across all sections unlike loyal Democrats and loyal Republicans.

For issue variables like 'Green Jobs program', 'Tax on wealth over \$100 million', 'Legalizing Marijuana' and 'Gun Control', it is evident that the swing voters have more voters supporting the ideologies of the Democrats rather than the Republicans by maximizing at the same value as the maxima for the Democrats. However, for 'Defunding Immigration and Customs Enforcement', the Swing voters exhibit a higher percentage of voters behaving like loyal Republicans by maximizing at 5 (like Loyal Republicans). This indicates their split opinions wherein they sometimes act like Democrats and at other times, act like the Republicans.

Therefore, the general hypothesis describing the swing voters is:

On most issues, swing voters are split, with some of them acting more like Democrats and others acting more like Republicans. However, it may be noted that more frequently, the swing voters mirror the ideologies of the Democrats rather than Republicans since for most issue variables like 'Green Jobs program', 'Tax on wealth over \$100 million', 'Legalizing Marijuana' and 'Gun Control', their maxima are completely or evidently bent towards the maxima of the loyal Democrats and the curve follows a trend similar to Democrats. Therefore, it appears like there are far more loyal Democrats than any other group.

Question 3:

What predicts being a swing voter?

We developed two logistic regression models, one using the issue variables as predictors (Model 1) and the other using the populism variables as predictors (Model 2), to probabilistically predict whether a registered voter is a swing voter or not.

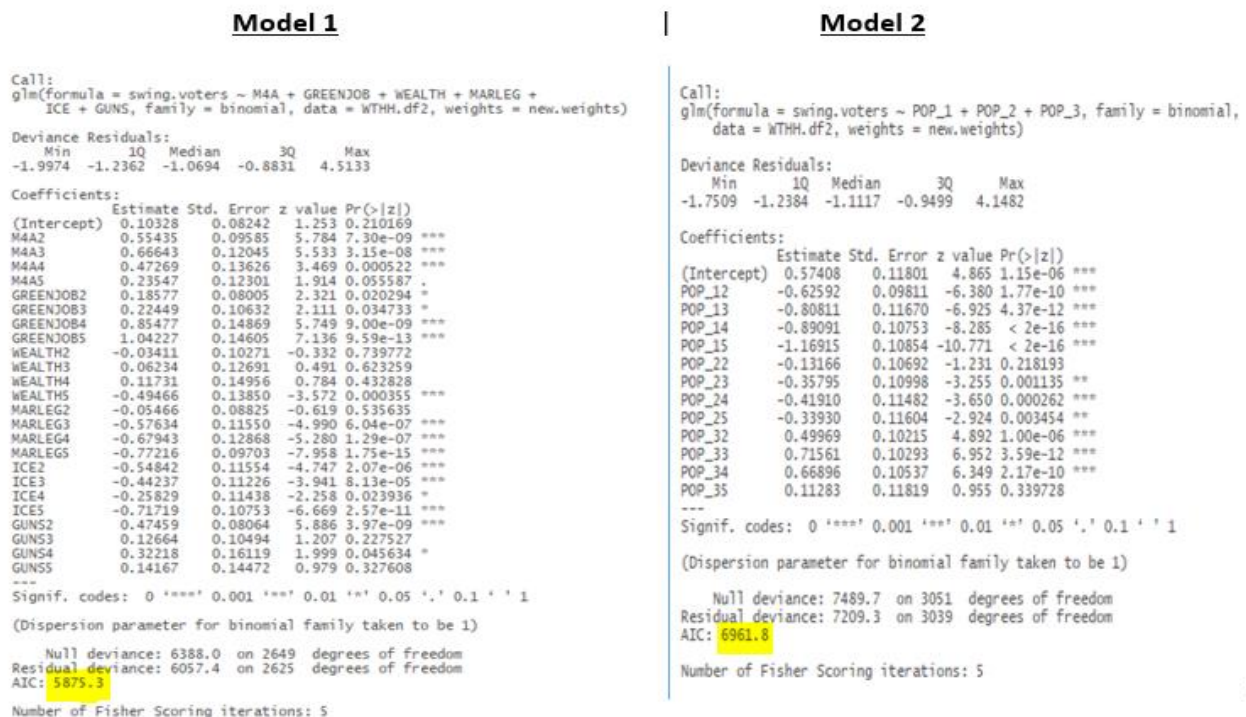
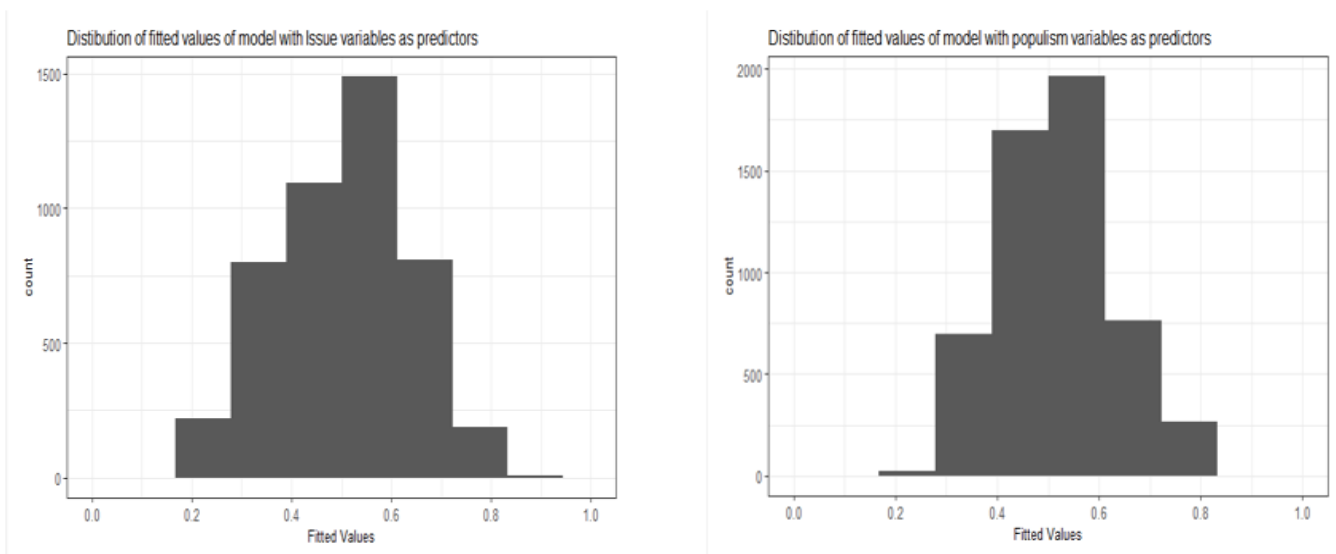


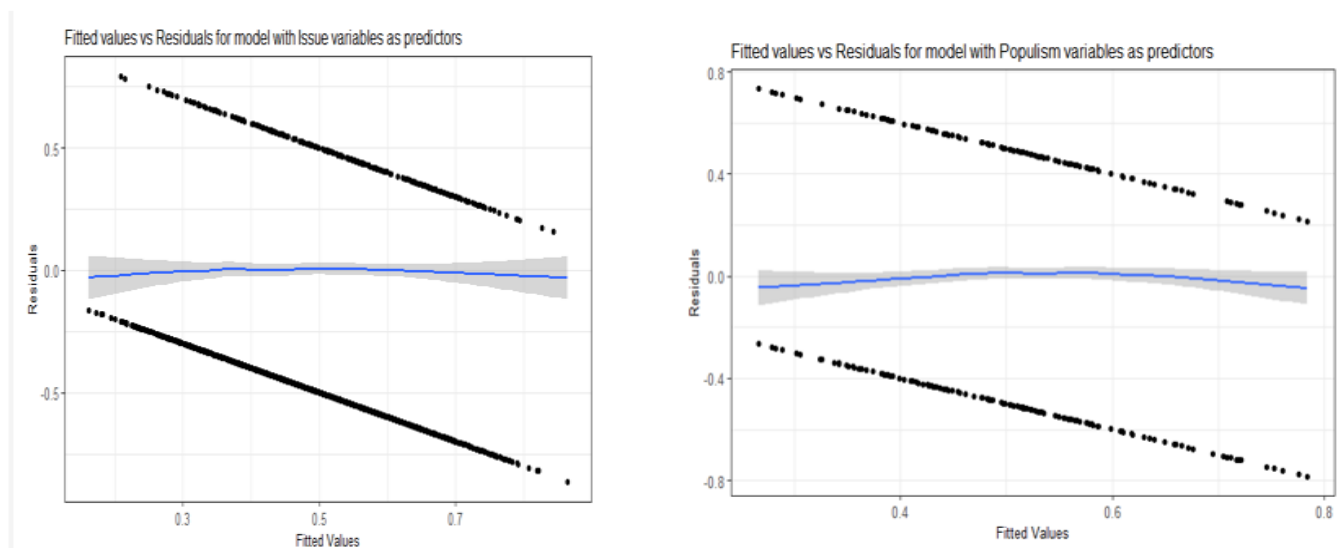
Figure 3.1

The above figure (Figure 3.1) compares the summary statistics of both the models against each other. The left-hand side summary shows the model fitted with issue variables and the right-hand side summary shows the model fitted with populism variables. After fitting the models, we compared their AIC (Akaike information criterion) scores which is an estimator of the relative quality of statistical models. As seen in the image, the AIC score for Model 1 (5875.3) is less than that of Model 2 (6961.8). The second model is more probable than the first model to minimize the information loss. Hence, we choose the first model (Model 1) which is fitted with issue variables.

We further analyzed our models by plotting the distribution of predicted values across different probabilities. We can observe from the following graphs that the distribution of the probabilities is spread out more for Model 1, where the model is able to predict the fitted values with a probability approaching 1 (Around 0.95) as compared to Model 2.



Further, we analyzed by plotting the residual values against the fitted values for both the models, where we can clearly infer that the blue line is nearer to zero for the Model 1 compared to Model 2.



Conclusion:

Based on the above analysis, we can conclude that the model fitted with the issue variables (Model 1) performs better when compared with the model fitted with populism variables (Model 2).

As shown in the summary statistics of our first model (Figure 3.1), based on the significance codes (number of asterisk marks associated with each variable), we can guess that the “Medical for All” and the “Green Jobs” are most important variables while deciding what makes a registered voter a swing voter.