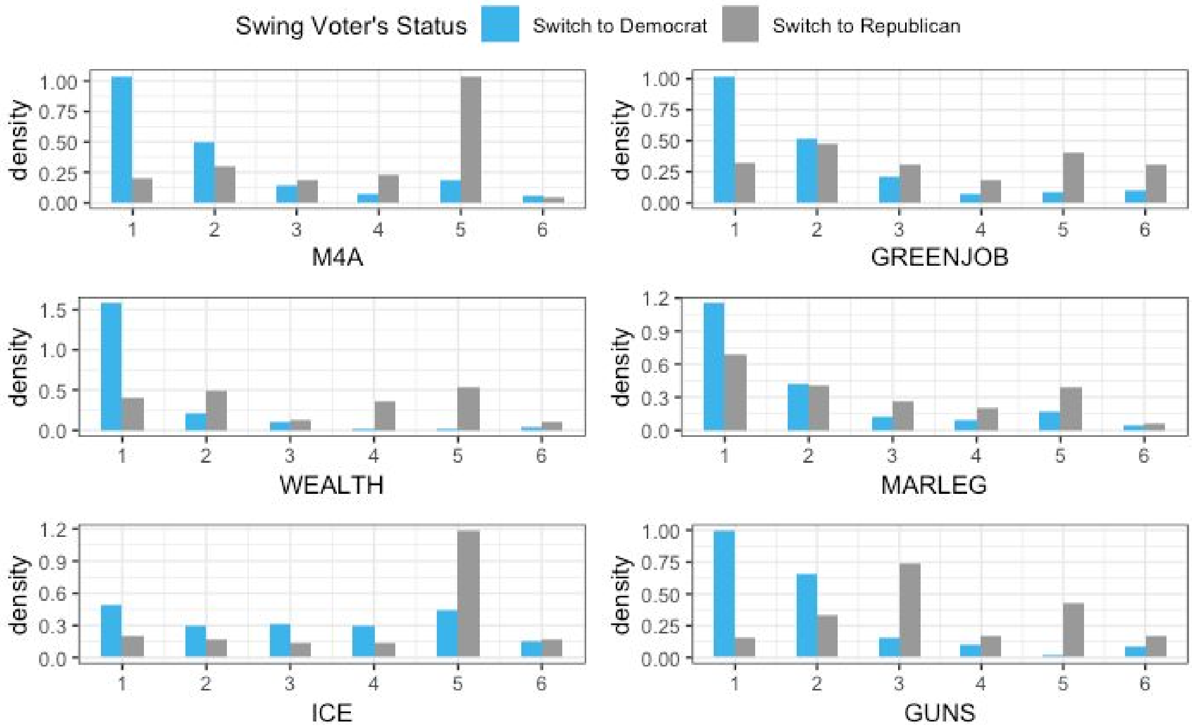
Mini-project 2: Swing voters
S681

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



We can conclude the following from the above plot:

I. A large portion of the switch to Republican voters are strongly against Medicare for all (M4A), and most of the switch to Democrat voters are supporters for Medicare for all.

II. On the issue “A Green Jobs Program(GREENJOB)”: switch to Democrat voters strongly support or somewhat support, and for the switch to Republican voters they don’t have a strong opinion on this issue since their stand on this issue is divided.

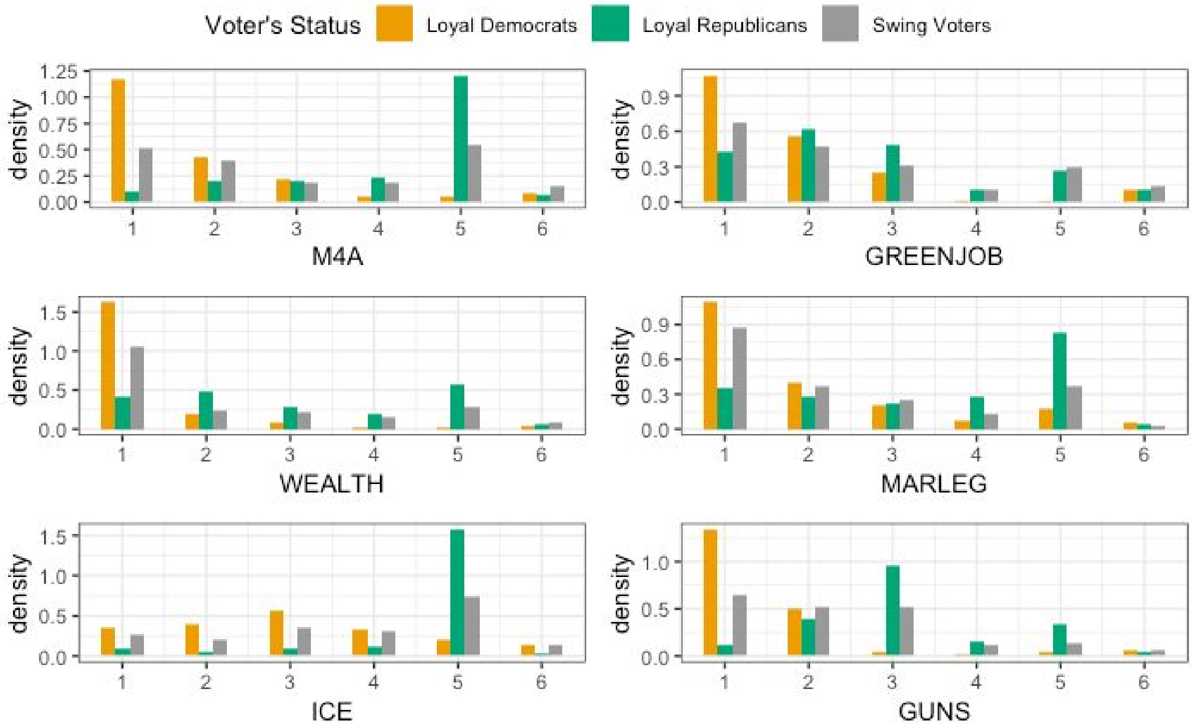
III. On the issue, whether to tax on wealth over $100 million, switch to Democrat voters are a strong supporter of tax on wealth. However, switch to Republican voters are inconsistent, they neither strongly support or against on this issue.

IV. Most of the switch to Democrat voters strongly support to legalizing marijuana. For the switch to Republican voters, they either strongly agree or strongly oppose legalizing marijuana.

V. For the issue on Defunding Immigration and Customs Enforcement, switch to Democrat voters are inconsistent on this issue. They are not strongly opposed or in support of it. For the switch to Republican voters, they are strongly opposed to defunding immigration and customs enforcement.

VI. On the last issue gun control, most the switch to Democrat voters are strongly support to control guns. Surprisingly, switch to Republican voters are neither in support or opposed to it.

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



Generally speaking, compare to loyal Democrats and loyal Republicans voters, swing voters think more like Democrats on some issues and more like Republicans on some other issues, also incoherent on few issues.

I. Loyal Republican and loyal Democrats have an opposite opinion on Medicare for all (M4A), where loyal Republican voters are strongly against, and loyal Democrats strongly support Medicare for all (M4A). However, swing voters are somewhat split on this issue, most of them are strongly support or opposed to this issue.

II. On whether to give every unemployed American who wants one a job building energy-efficient infrastructure (GREENJOB); loyal Democrat voters most strongly support to give a green job to unemployed American. However, for loyal Republican voters, they don't have a strong opinion on this issue but toward to support on this issue, because their score on this issue is somewhat spread. Swing voters are more like loyal Democrat voters, but with a less strong opinion.

III. On the issue of whether to tax on wealth over $100 million (WEALTH), a large amount of loyal Democrat voters are the strong supporters, also same with the swing voters. However, royal Republicans do not have a united opinion on this issue.

IV. On whether to legalize marijuana, most loyal Democratic voters strongly support to legalize it. On the other side, loyal Republican voters are mostly opposed to legalizing it. For swing voters, most of them are more like loyal Democrats, there are also a group of swing voters who opposed legalizing marijuana.

V. For the issue of defunding immigration and customs enforcement, loyal Democrat voters do not have a united opinion to support or oppose this issue. However, for royal Republicans, they are strongly opposed to defunding immigration and customs enforcement. The swing voters are somewhat like loyal Republican voters, most of them are strongly opposed, others are split.

VI. On the gun regulation issue, most loyal Democrats voters strongly support to control guns, and loyal Republican voters are moderates on this issue. Swing voters are like the combination of loyal Democrats and loyal Republicans, they are toward the supporting side on bringing regulation on guns, some are moderate on this issue.

3. What predicts being a swing voter?

predicting swing voters with policy issue variables

We can read from the graph 1 in Appendix that the attitude of swing voter and non-swing voter to the policy issues are very similar, which implies that we may not build a good model by simply adding issue variables to the model.

However, when we examining the (kendall) correlation between issue variables, we found that except GREENJOB-WEALTH issue pair, the correlation of issue pairs are always higher in the non-swing voter samples than in the swing voter samples. We can show this by the appendix table 1, which produced by subtracting correlation matrix of swing voter to the correlation matrix of the non-swing voter.

The correlation difference of issue pairs implies that the interaction effect on these variable pairs work differently for swing voter and non-swing voter, which in return implies we can use the variables pairs which have large correlation differences between swing voter and non-swing voter to build a model which aims to differentiate swing voter from non-swing voters.

Following the reasoning above, we pick three variable pairs which have largest correlation differences, i.e., WEALTH*MARLEG, WEALTH*ICE, and MARLEG*GUNS, and use these three variable pairs to build a logistic regression model to predict the probability of a voter being a swing voter. The following graph shows the model summary.

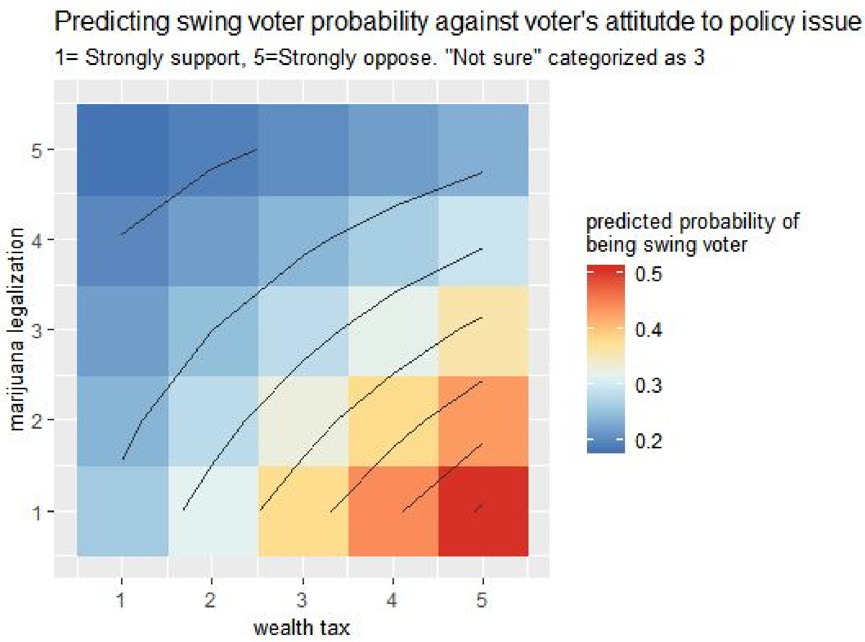
```
Call:
glm(formula = swing ~ WEALTH * MARLEG + WEALTH * ICE + MARLEG *
    GUNS, family = quasibinomial, data = mdl_data, weights = weight_DFP)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0638  -0.6942  -0.5648  -0.4003   3.8406

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.37478    0.23272 -10.204  < 2e-16 ***
WEALTH         0.62064    0.11791   5.264 1.51e-07 ***
MARLEG         0.15031    0.06988   2.151 0.031559 *
ICE            0.08525    0.05760   1.480 0.138977
GUNS           0.28316    0.07776   3.641 0.000276 ***
WEALTH:MARLEG -0.05370    0.02391  -2.246 0.024756 *
WEALTH:ICE    -0.09898    0.02425  -4.081 4.59e-05 ***
MARLEG:GUNS   -0.06840    0.02757  -2.481 0.013153 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 1.000888)

    Null deviance: 3228.6  on 3101  degrees of freedom
Residual deviance: 3140.7  on 3094  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
```

For most of the performance metrics of a binary classifier, the value of metrics depends on the cut-off threshold you choose to split different class. If we choose 0.23 as our cut point to split swing voter from the non-swing voter, then we can get the performance metrics of our logistic classifier as the following table:

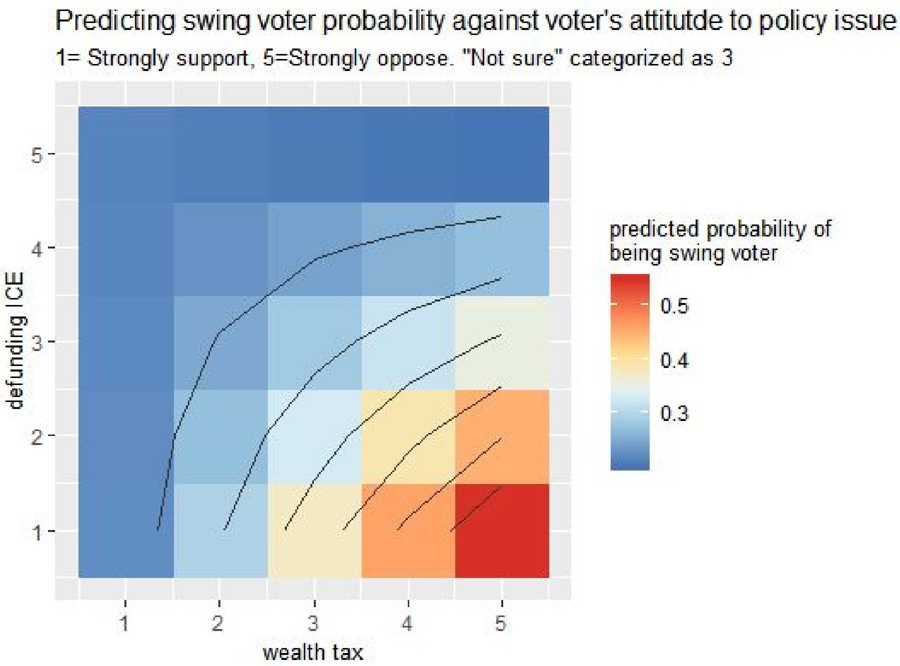
Probability Threshold	0.23	Classification Error	0.2847
Precision	0.2882	Accuracy	0.7153
Recall	0.3396	AUC	0.6131

To show how the predicted probability changes with the variables we use, we can draw a raster-and-contour plot with the predicted probability against the variable pair we choose.

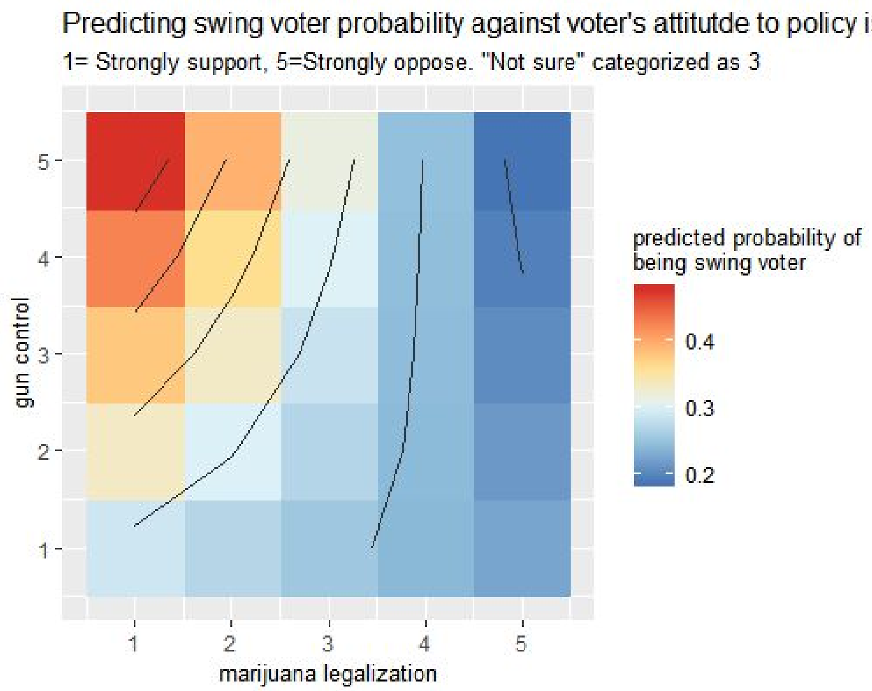
First, for people who have a low level of support on wealth tax and a high level of support on marijuana legalization, the probability of being a swing voter tends to be high.



Second, for people who have a low level of support on wealth tax and high level of support on defunding ICE (Immigration and Customs Enforcement), the probability to being a swing voter tends to be high.



Third, for the people who have high support on marijuana and have high support on gun control, the probability of being a swing voter tends to be high.

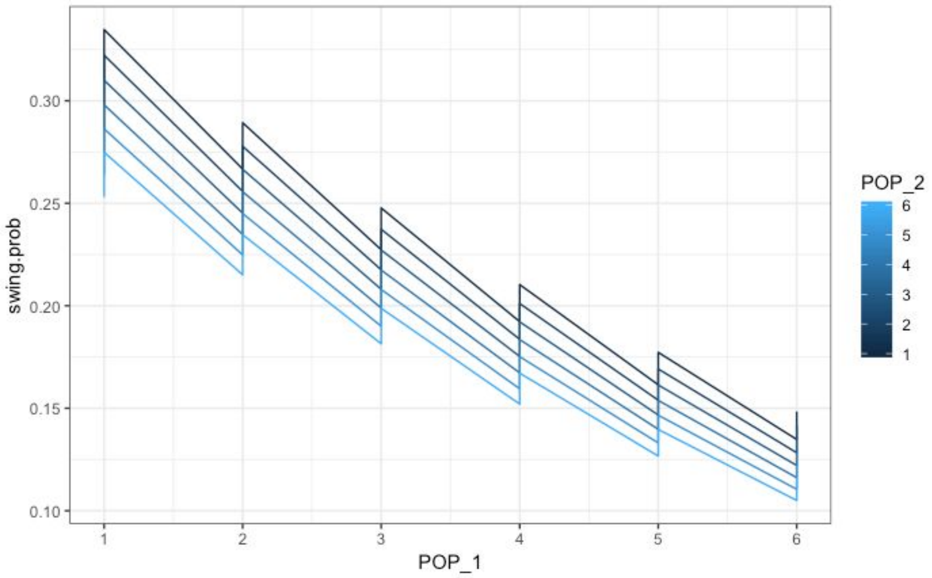


predicting swing voters with populism variables

```
Call:
glm(formula = swing ~ POP_1 + POP_2 + POP_3, family = quasibinomial,
    data = mdl_data, weights = weight_DFP)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8924  -0.7231  -0.5689  -0.4130   3.6658

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.360949   0.159278  -2.266   0.0235 *
POP_1       -0.222736   0.035690  -6.241 4.95e-10 ***
POP_2       -0.082714   0.038286  -2.160   0.0308 *
POP_3       -0.003266   0.036424  -0.090   0.9286
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 1.004001)

    Null deviance: 3228.6  on 3101  degrees of freedom
Residual deviance: 3163.8  on 3098  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
```

For populism model**,** If we choose the probability threshold which generate similar precision/recall performance as the issue model, then we can specified the performance of populism as following table:

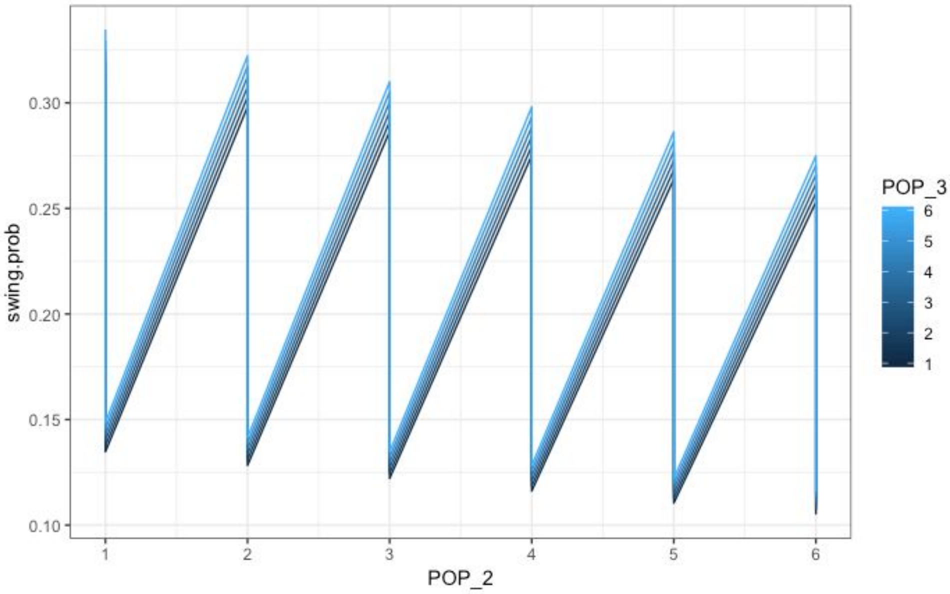
Probability Threshold	0.26	Classification Error	0.2847
Precision	0.2876	Accuracy	0.7153
Recall	0.3379	AUC	0.6052



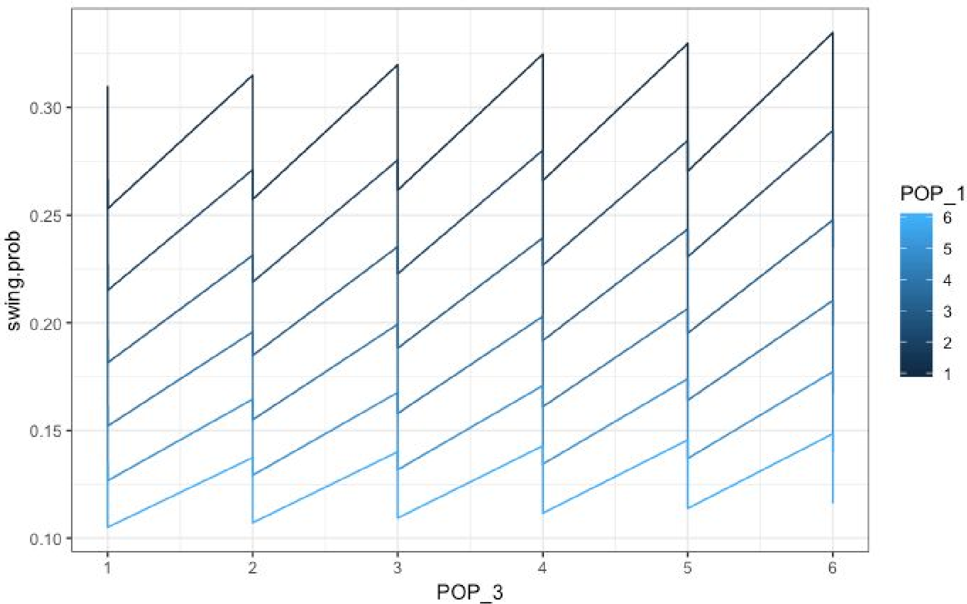
Here we have created an additive model taking all 3 populism variables into consideration to train our model.

We observed from the above graph that-

I. People who strongly agree on POP_1 that is who believe in the statement that “It doesn’t really matter who you vote for because the rich control both political parties.” are most probably swing voters. People who strongly opposed to that statements are less likely to be swing voters.
II. We have grouped the above values and used color on POP_2 variable. People who believe in the statement that “The system is stacked against people like me.” are more likely be swing voters and vice versa.
III. The graph shows when people strongly agree on both POP_1 and POP_2, they are most probably be swing voters.
IV. People who strongly agree on POP_2 are most probably swing voters. People who strongly opposed to that statements are less likely to be swing voters.
V. We have grouped the above values and used color on POP_3 variable. People who believe in the statement that “I’d rather put my trust in the wisdom of ordinary people than in the opinions of experts and intellectuals.” are less likely be swing voters and vice versa.



VI. The graph shows when people strongly agree on both POP_2 and strongly oppose POP_3, they are most probably be swing voters.



As seen from the above graph-

VII. When people disagree on POP_3 and strongly agree on POP_1, they are most probably be swing voters.

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?

By looking at both models and AUC scores we can conclude that our model with issue variables is better than our model with populism variables. For issue variable model, the interaction Wealth: ICE are the major predictors and for populism variable model POP_1 is the major predictor of swing voters.

========================Finished=============================

Appendix

graph 1

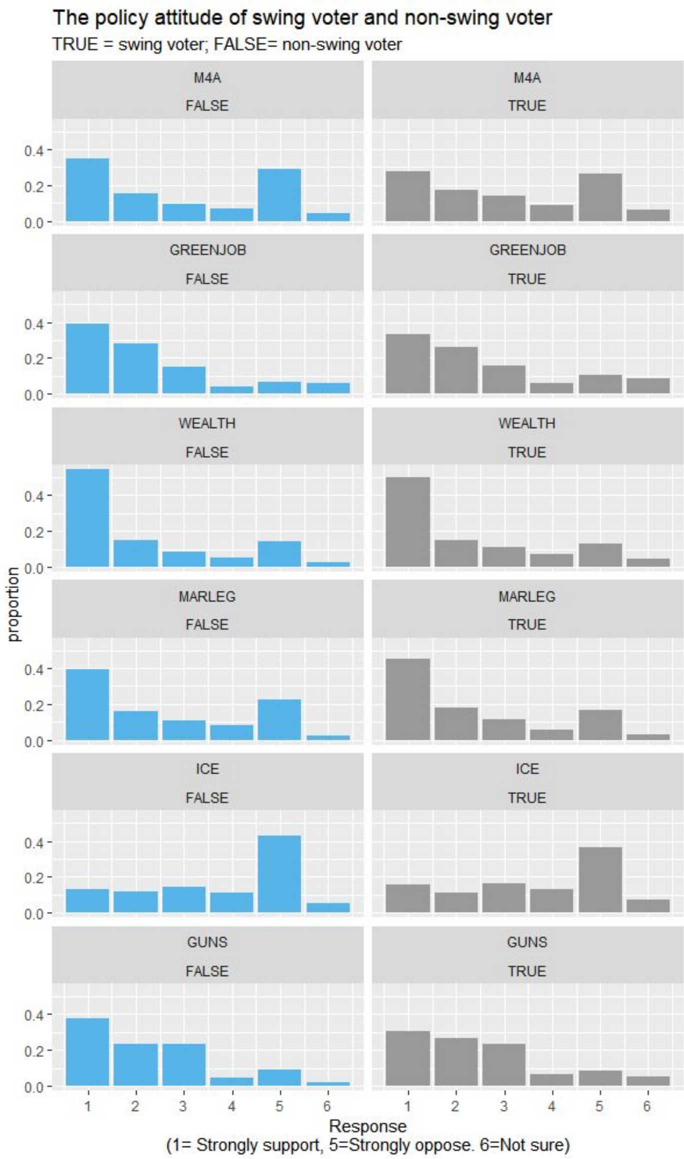
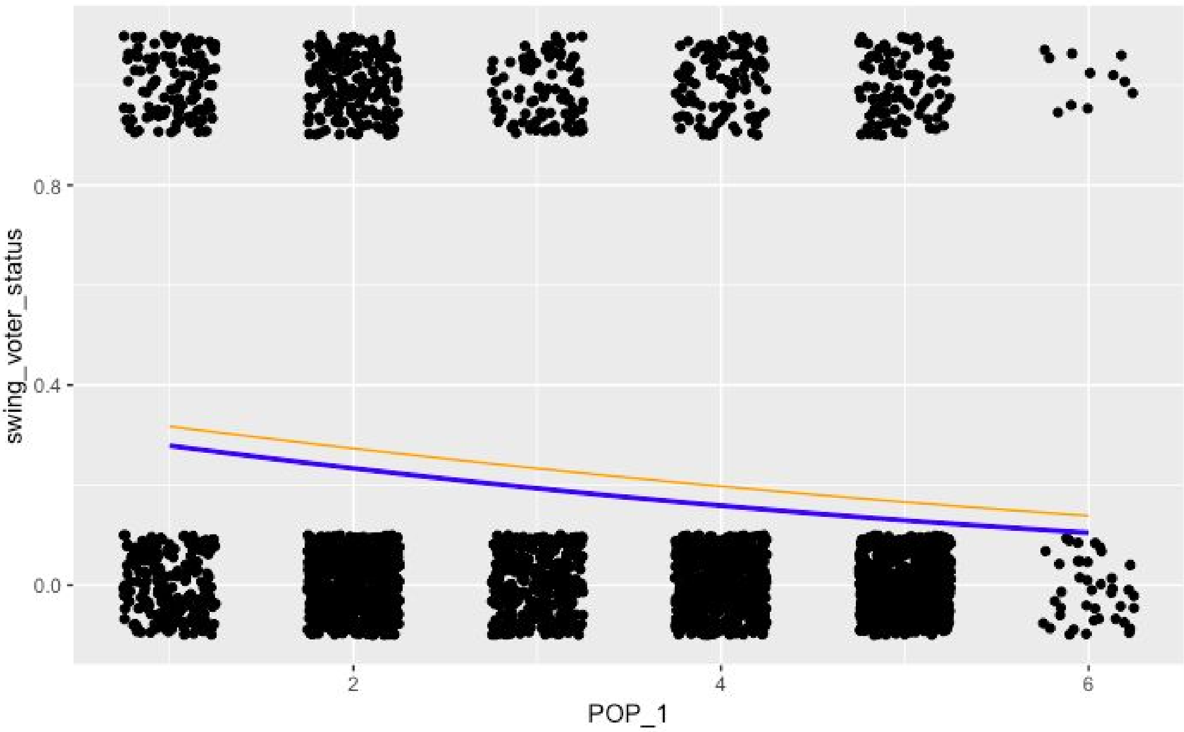
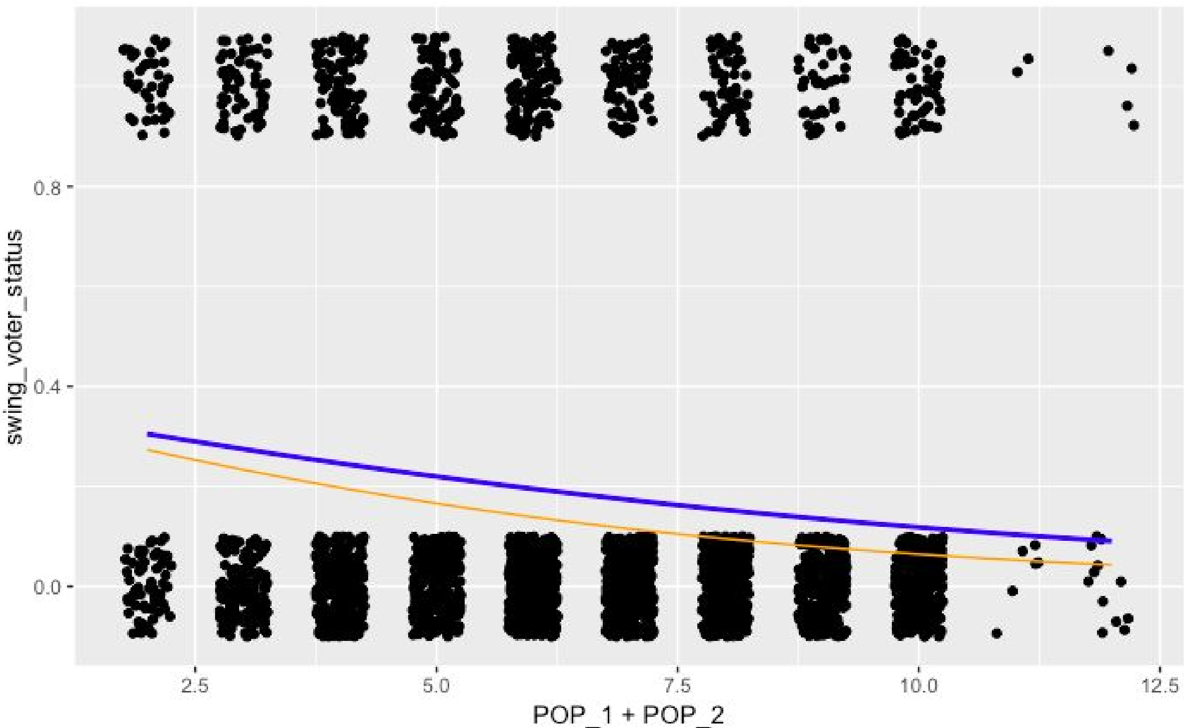


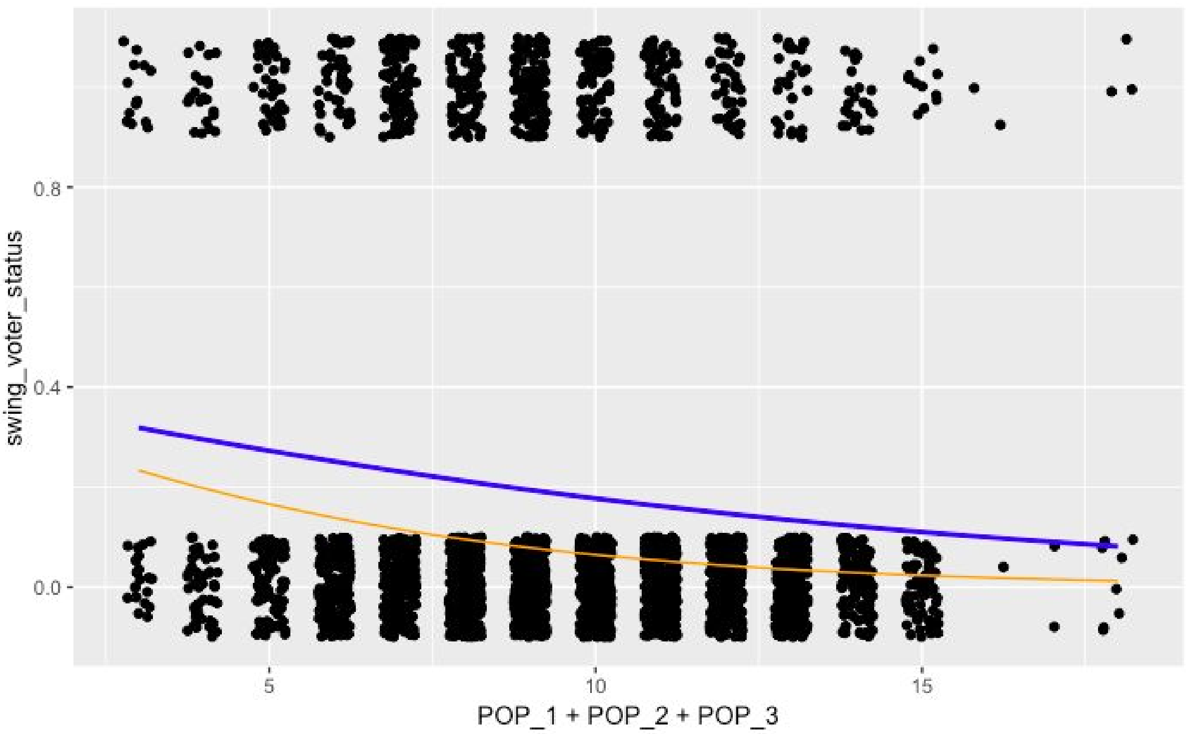
table 1

	M4A	GREENJOB	WEALTH	MARLEG	ICE	GUNS
M4A	0.0000	-0.0673	-0.1166	-0.1351	-0.1057	-0.1255
GREENJOB	-0.0673	0.0000	-0.0184	-0.1087	-0.0572	-0.0516
WEALTH	-0.1166	-0.0184	0.0000	-0.1768	-0.1789	-0.1094
MARLEG	-0.1351	-0.1087	-0.1768	0.0000	-0.0951	-0.1964
ICE	-0.1057	-0.0572	-0.1789	-0.0951	0.0000	-0.1401
GUNS	-0.1255	-0.0516	-0.1094	-0.1964	-0.1401	0.0000

3.


swing_voter_status
0.8
0.4
0.0
2
4
6
POP_1


swing_voter_status
0.8
0.4
0.0
2.5
5.0
7.5
10.0
12.5
POP_1 + POP_2


0.8
0.4
0.0
swing_voter_status
5
10
15
POP_1 + POP_2 + POP_3