Mini Project 2

Evelyn Delph, Arthur Janes, Nick Mobley

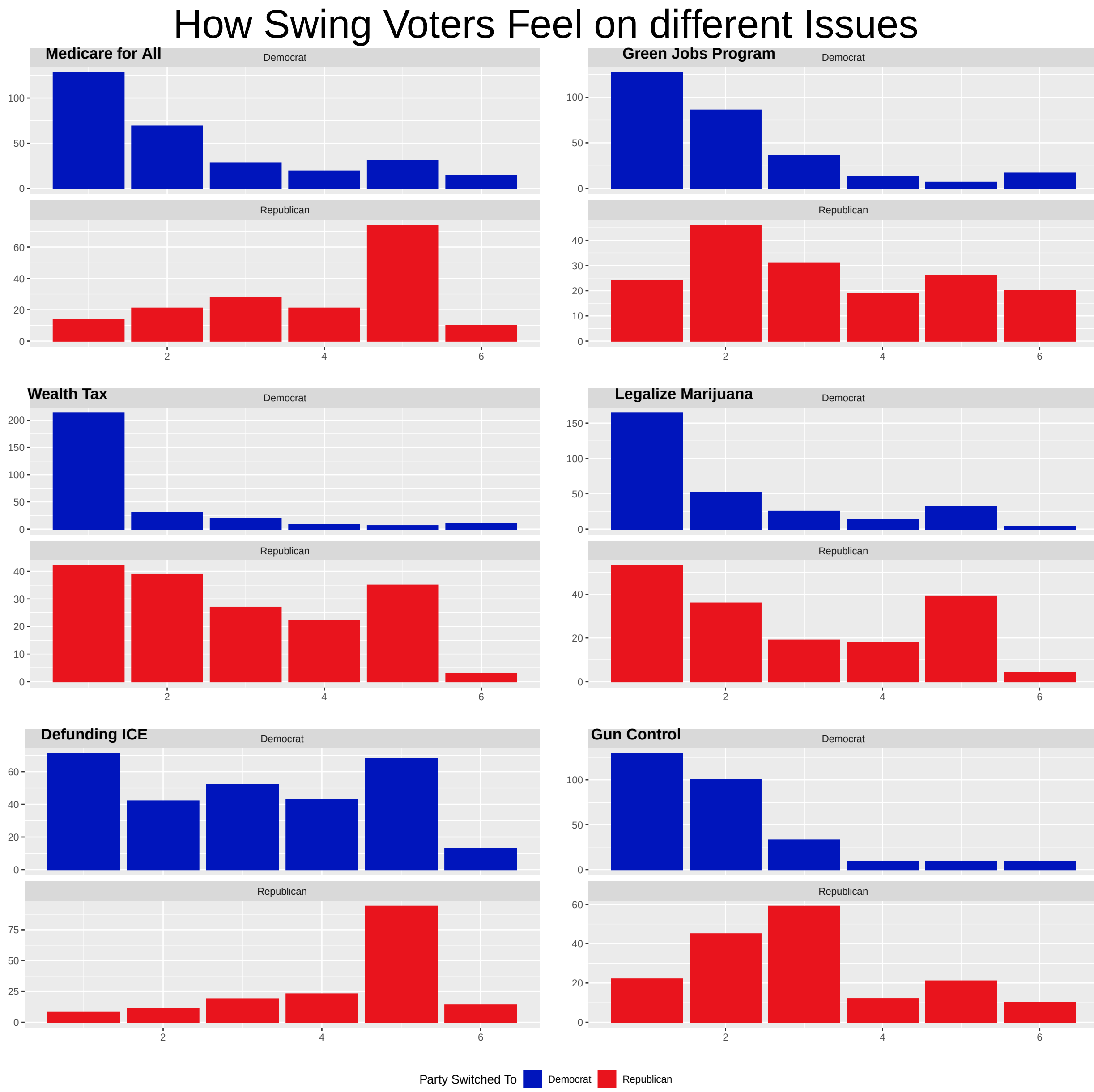
3/26/2020

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

The graph below shows survey results of 'swing voters' on different political issues. Amongst the different issues there are similarities and differences between the two groups:

- Medicare for All has an almost mirror distribution between the two different groups. Democrats mostly support it while Republicans generally strongly oppose it.
- Green Jobs Program is less clear. It is pretty obvious that most Democrats strongly agree with the program. However for the Republicans it isn't as clear. The most common response was 2 meaning they agree, but the second most common response was strongly oppose. So overall most swing voters will generally agree with democrat swing voters on the green jobs program, but there is still a decent portion that will strongly disagree.
- Wealth Tax for the democratic party is nearly unamious in favor for it. An overwhelming number agree with it, however the sentiment is ambigious within the Republican swingers. The results almost look uniform, except they respondants were more likely to give a 'strong' opinion.
- Leagalize Marijuana has an almost idnetical distribution between the two groups. Although it seems the Republican swing voters have a little bit more opposition within the party than the Democrats do, a majority of both groups will agree on the legalization.
- Defunding ICE seems to be a split topic within Democrat voters. It seems that they mostly have a uniform distirbution, except similar to wealth tax for republican are more likely to give a 'strong' opinion. For the Republicans it seems obvious that generally they are in consensus to defunding ICE.
- Gun control for the Democrats is in strong favor to put more limits on gun control. While for the first time it seems the republican party doesn't really have a strong opinon. The most common response was neutral, but the most common after that were 2 and 1 which shows that most people either don't have a strong opinion, or they are in favor for increasing gun control.

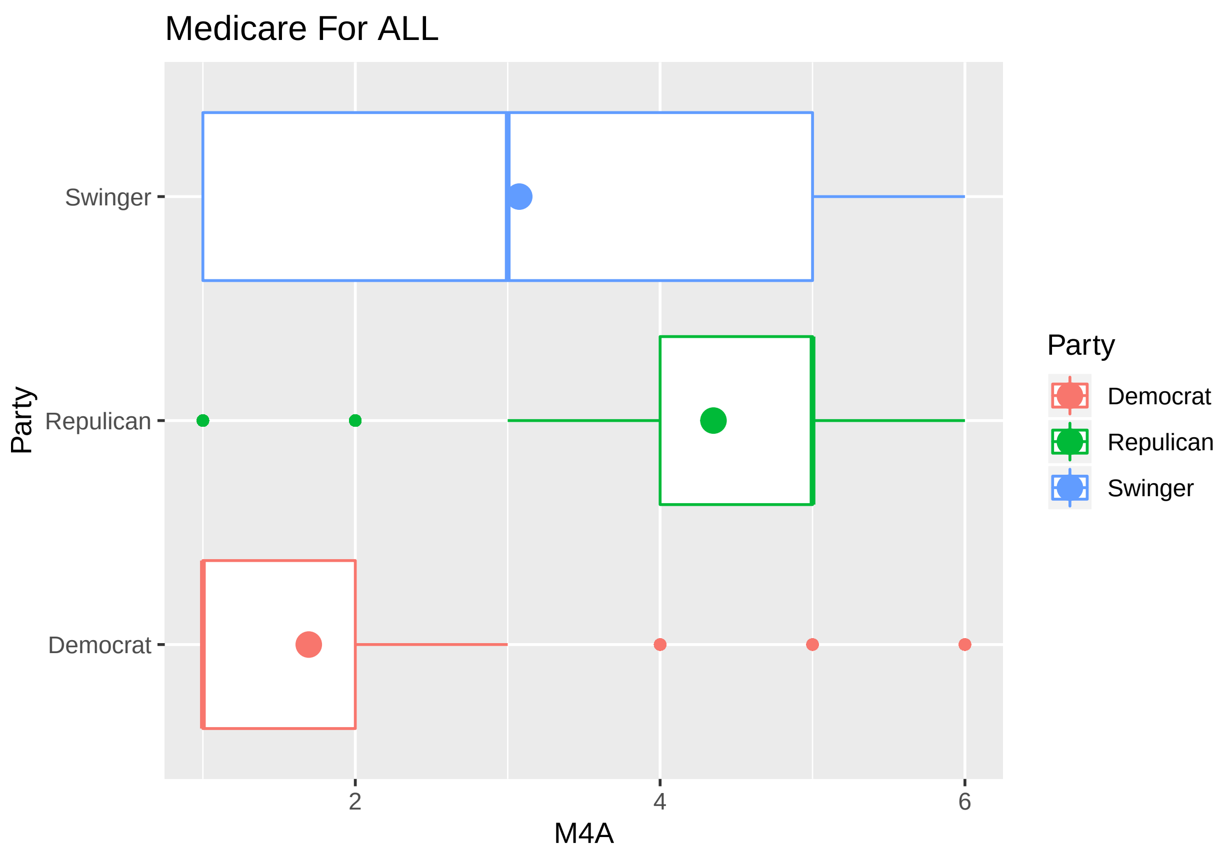
To highlight some key differences here. It seems Democrats have the strongest opinions on Wealth Tax and LEgalize Marijuana, while the republican party is much more split on these issues. On the otherhand republican party has a strong opinion on Medicare for All and Defunding Ice. Because of these strong polarities, these issues would be the most useful in determining which group a swing voter would be a part of.



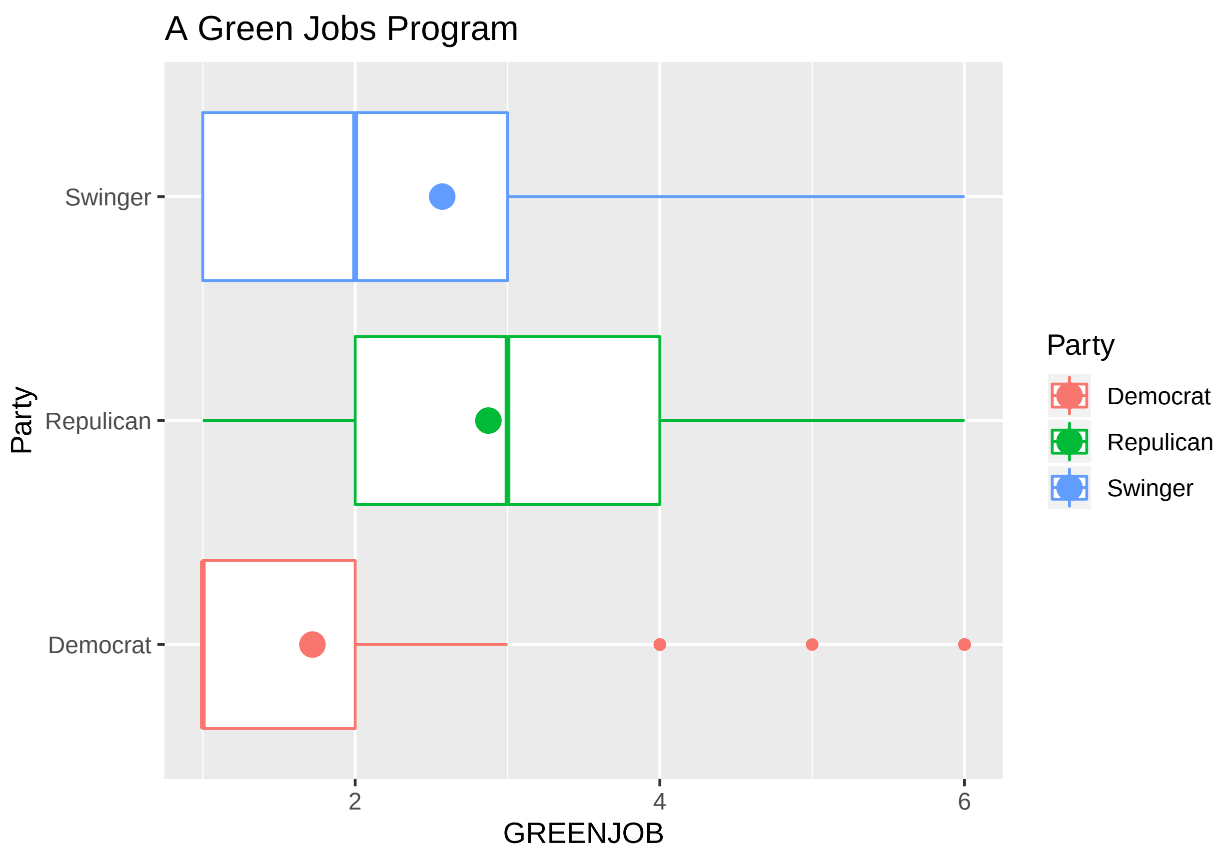
Survey Results on Scale 1−5, 1 = Strongly Support, 5 = Strongly Oppose (6 = "Not Sure")

Section 2: How do Swing voters differ from loyal Republicans and loyal Democrats on the issue variables?

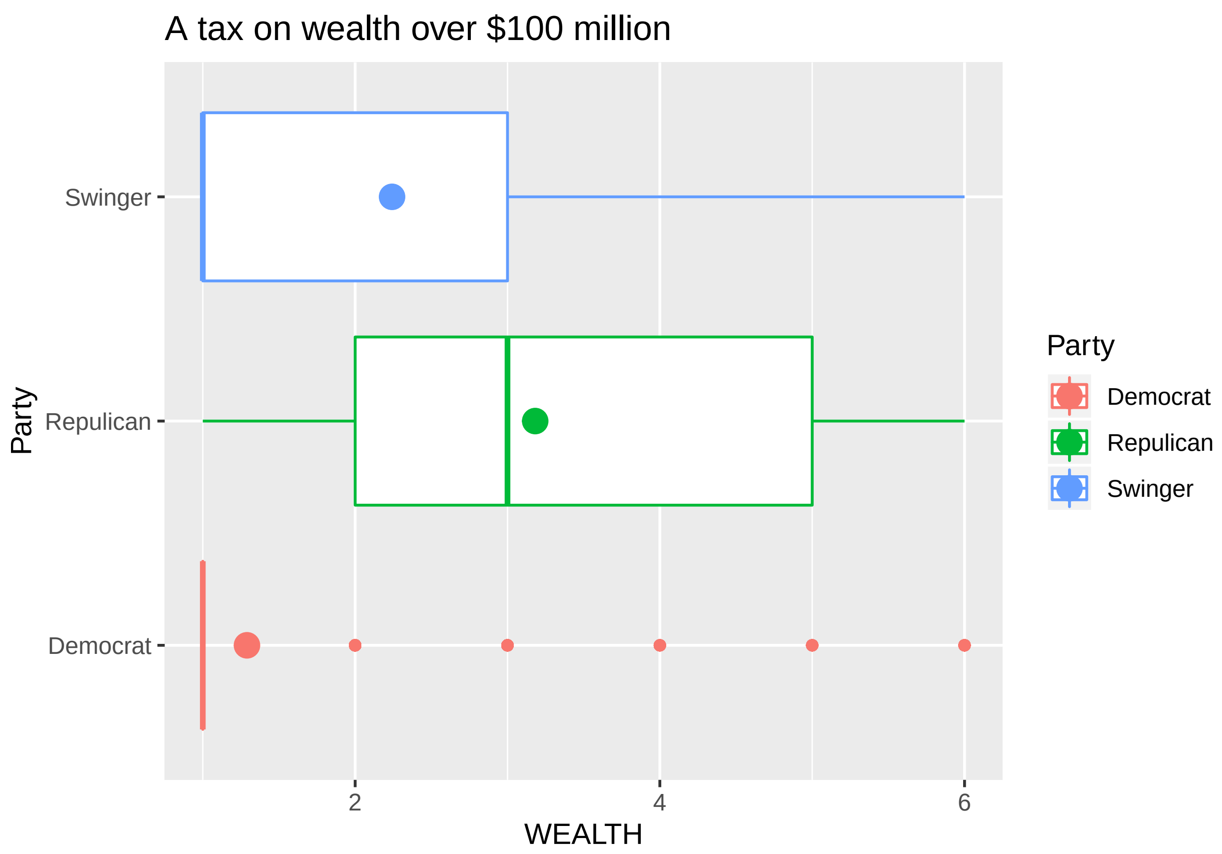
```
ggplot(df, aes(y = M4A, x = Party, color = Party)) + geom_boxplot() + coord_flip() + stat_summary(fun.y
```



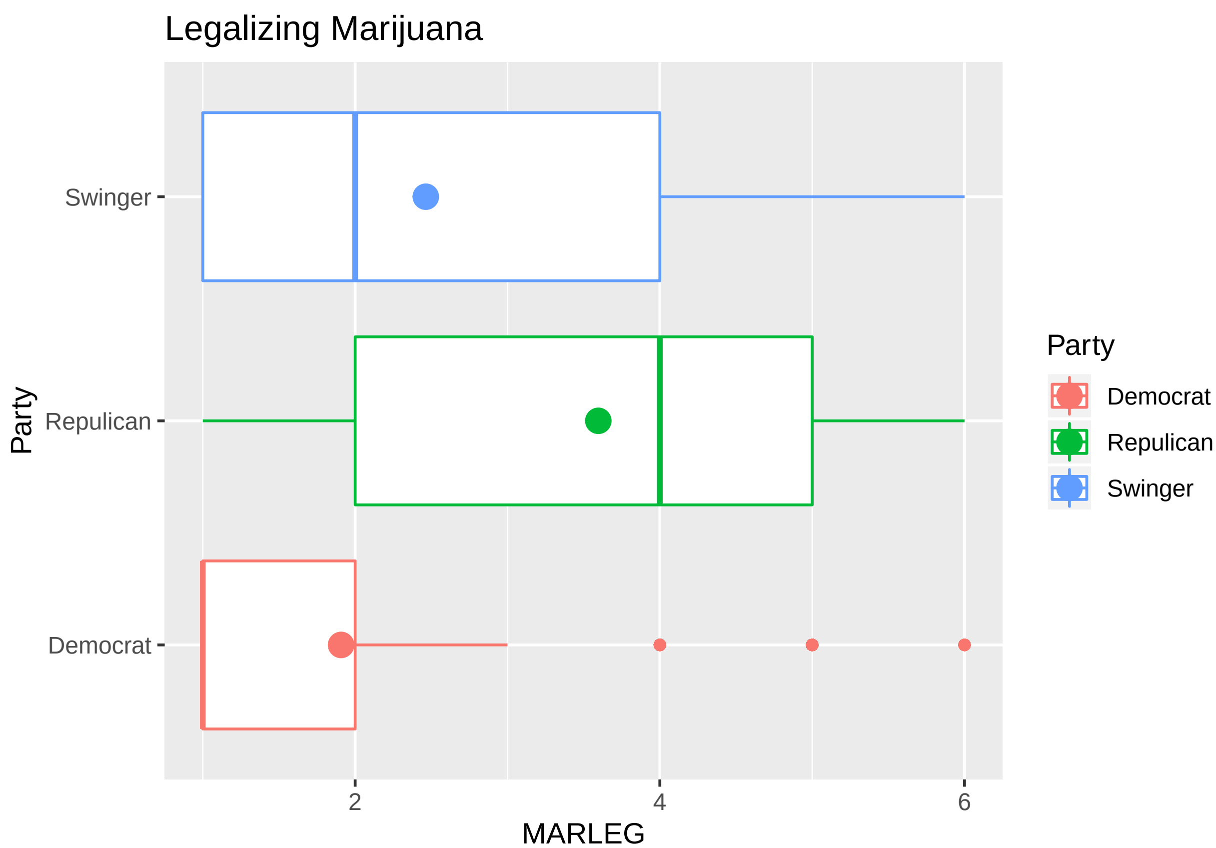

```
ggplot(df, aes(y = GREENJOB, x = Party, color = Party)) + geom_boxplot() + coord_flip() + stat_summary(
```



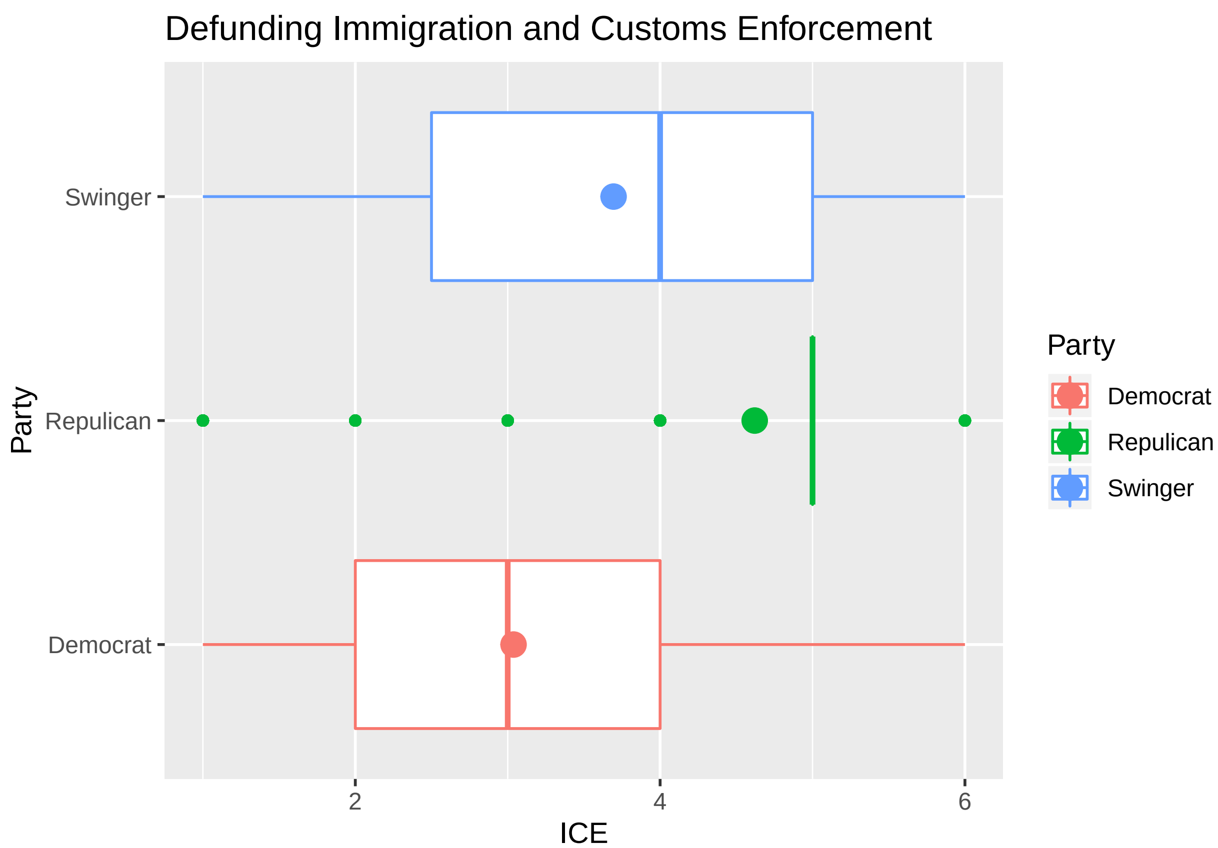

```
ggplot(df, aes(y = WEALTH, x = Party, color = Party)) + geom_boxplot() + coord_flip() + stat_summary(fu
```



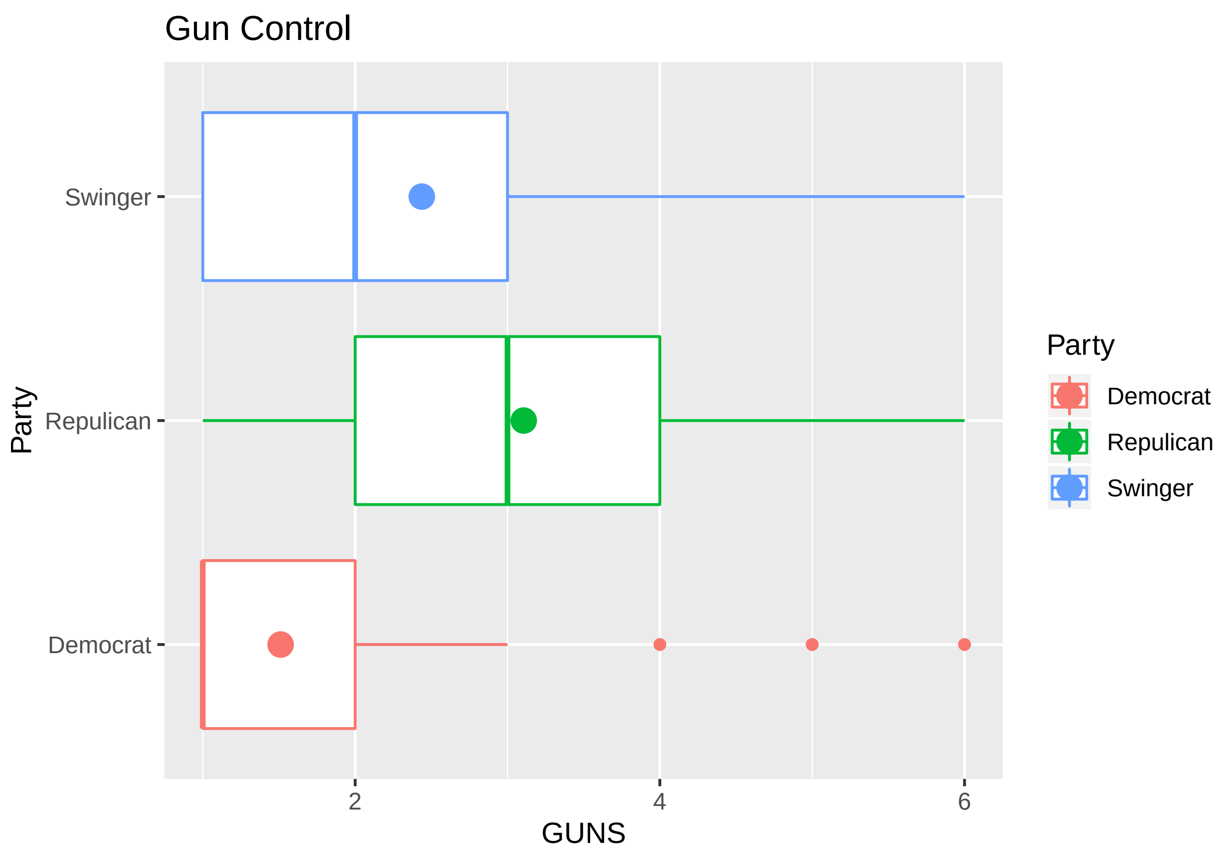

```
ggplot(df, aes(y = MARLEG, x = Party, color = Party)) + geom_boxplot() + coord_flip() + stat_summary(fu
```




```
ggplot(df, aes(y = ICE, x = Party, color = Party)) + geom_boxplot() + coord_flip() + stat_summary(fun.y
```




```
ggplot(df, aes(y = GUNS, x = Party, color = Party)) + geom_boxplot() + coord_flip() + stat_summary(fun.y
```



As was expected, it seems that swing voters represent a moderate group of voters when compared to both Democrats and Republicans.

- For each issue, the mean of the swing votes lies in between that of the Democrats and the Republicans. The difference between the mean swing vote and each party's mean vote is approximately equal for each issue. This indicates that swing voters tend to the middle of the distribution, largely separating the Democrats and Republicans.
- However, the overall distribution of the swing votes lies much closer to that of the Democrat votes for a tax on wealth over $100 million and legalizing marijuana. Therefore, it seems that swing voters think more like Democrats on these two issues.

Section 3: What predicts being a swing voter?

There are two sets of variables that can determine swing voters from non-swing voters. First, the six issue variables will be examined, then the populism variables. After that, the best model among each variable group will be determined.

Issue Variables

First, to get a glimspe at the data, a pairs plot was constructed to see how individual variables were related. Wealth and Medicare for All had the highest correlation. As a result, it was a assumed there was an interaction between them. The rest of the variables had similar correlations, but not quite as high as the interaction terms. Due to this, it was anticipated that all of the variables were important for predicting swing voters.

For each model, the absolute difference between mean probabilities for swing voters and non swing voters were calculated. The model with the highest difference indicates a clearer separation between the two types of voters. THe model with the highest difference between the means is considered to be the best performing model as a result. Below shows the probability scores after adding each variable, as well as determining the most accurate model.

```
## Probabilities for one issue variable (Wealth):
##  0.2084428 0.216675 0.2251398 0.2338367 0.2427642

## [1] ""

##
## Probabilities for two issue variables (WEALTH + M4A):
##  0.1709065 0.1691748 0.167457 0.1657532 0.1640633

## [1] ""

##
## Probabilities for three issue variables (WEALTH + M4A + MARLEG):
##  0.2462047 0.248449 0.250707 0.2529785 0.2552637

## [1] ""

##
## Probabilities for four issue variables (WEALTH + M4A + MARLEG + GUNS):
##  0.1861425 0.1848046 0.1834741 0.1821511 0.1808354

## [1] ""

##
## Probabilities for five issue variables (WEALTH + M4A + MARLEG + GUNS+ GREENJOB):
##  0.1704236 0.1686309 0.1668533 0.1650907 0.163343

## [1] ""

##
## Probabilities for six issue variables (WEALTH + M4A + MARLEG + GUNS+GREENJOB + ICE):
##  0.2350682 0.2365042 0.2379462 0.2393942 0.2408483

##      Non-Swinger Swinger Difference

## # Variables 1 0.7915572 0.2084428 0.5831144
## # Variables 2 0.783325 0.216675 0.56665
## # Variables 3 0.7748602 0.2251398 0.5497203
## # Variables 4 0.7661633 0.2338367 0.5323267
## # Variables 5 0.7572358 0.2427642 0.5144716
## # Variables 6 0.8290935 0.1709065 0.6581869
```

```
##
## The best model that shows the greatest difference
##     between swing and non-swing votes the model with 6 variables
```

Final 6 variable equation:

```
votes = -1.99244285 + 0.29325492 + 0.18323905 + (-0.18073493) + 0.11441703  + 0.16567127 + (-0.07662718)
```

For the issue variables, the model with all six variables seemed to have the highest separation with a difference of 0.6581869. This could show that all of the issues are important when determining a swing voter from a non-swing voter.

Populism Variables

The same steps were conducted for the populism variables. The only difference was that there wasn't a significant interaction term among the variables. Below shows the probabilities and the selection for the most accurate model.

```
## Probabilities for one populism variable (POP_1):
##  0.3109913 0.2642393 0.2222485 0.1852505 0.1531985

## [1] ""

## Probabilities for two populism variables (POP_1+ POP_2):
##  0.3245092 0.3147179 0.3050887 0.2956269 0.2863377

## [1] ""

## Probabilities for three populism variables (POP_1 + POP_2 + POP_3):
##  0.3163266 0.3224769 0.3286894 0.3349623 0.3412941

## [1] ""

##      Non-Swinger Swinger Difference

## # Variables 1 0.6890087 0.3109913 0.3780175
## # Variables 2 0.7357607 0.2642393 0.4715214
## # Variables 3 0.7777515 0.2222485 0.5555029

##
## The best model that shows the greatest difference
##     between populism variables is the model with 3 variables
```

Final 3 populism variable equation:

```
votes = -1.06513692 + (-0.20446605) + (-0.04608700) + 0.06178131 * X
```

For the populism variables, the model with the three variables had the highest separation with a difference of 0.5555029. Again, this could show that encompassing all variables gives a holistic picture of the swing and non-swing voters' values.

Best Model

Among the best two models, the issue variables out perform it by roughly 18%. This could mean that issue variables are slightly better predictors than populism variables, as they point to concrete issues rather than abstract beliefs and ideals.

Appendix

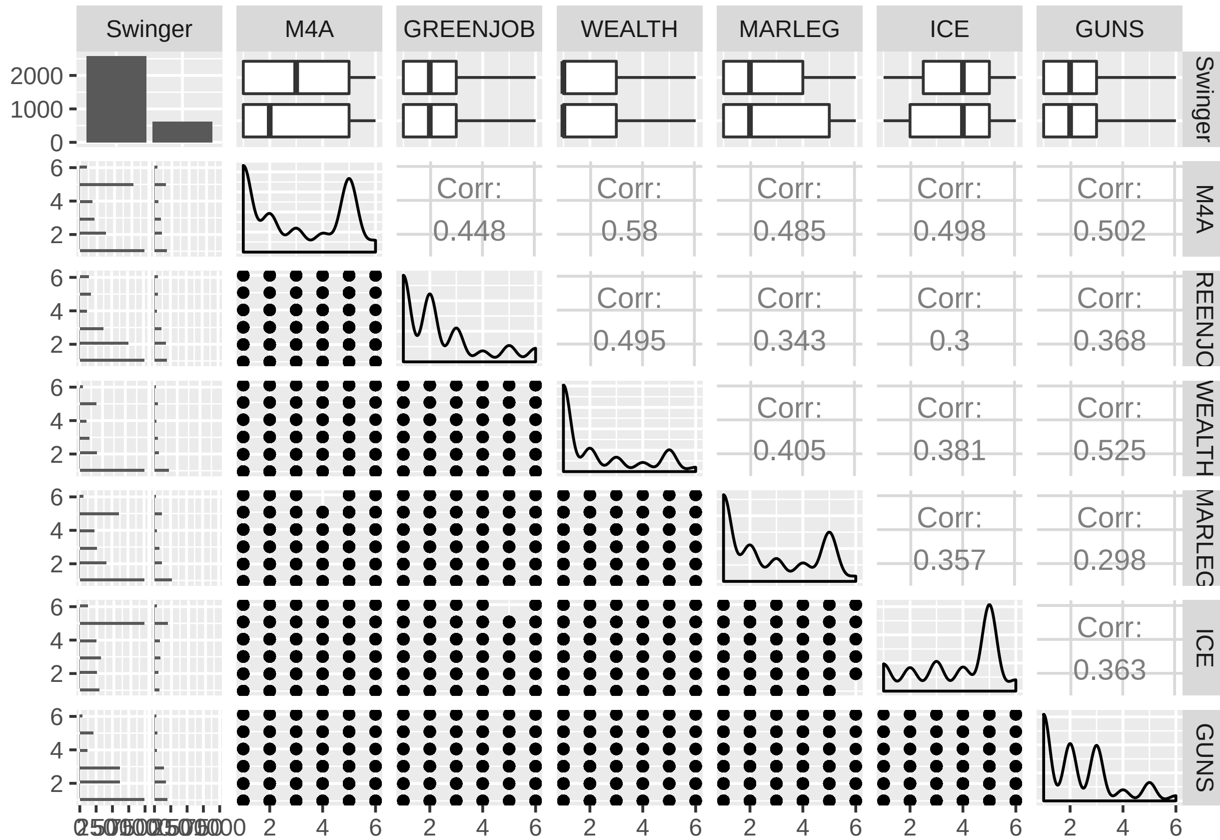
Section 3

GGPairs plot to detect correlation and possible interactions with issue variables:

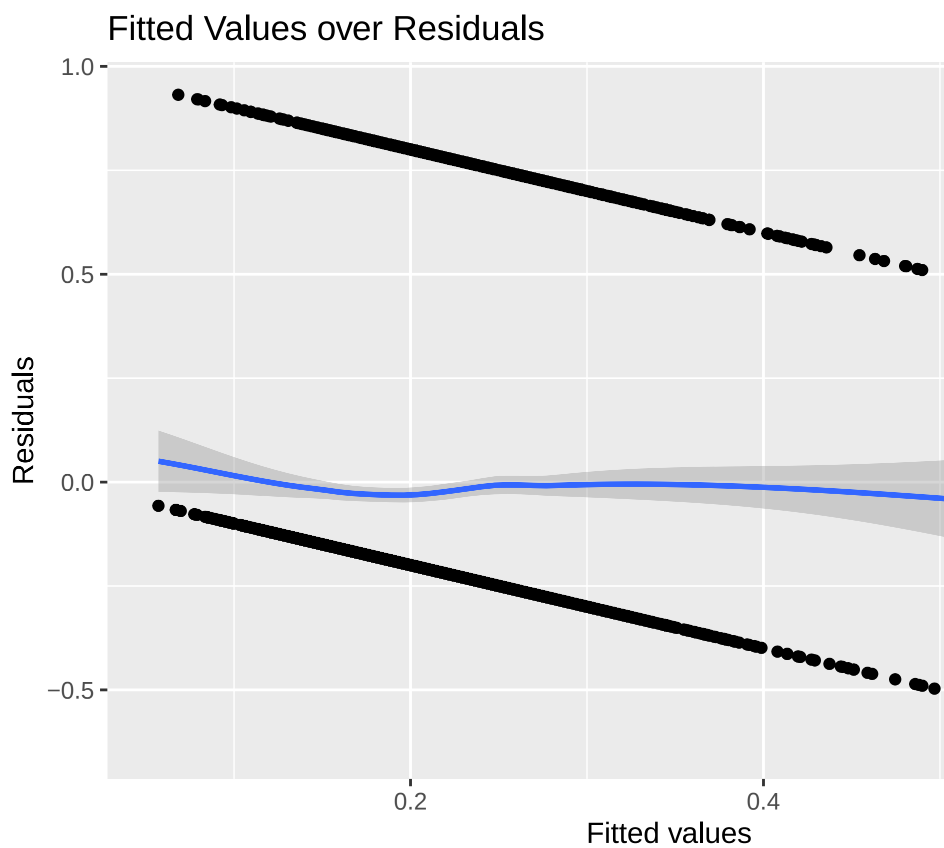
```
## Warning: Removed 19 rows containing non-finite values (stat_boxplot).

## Warning: Removed 28 rows containing non-finite values (stat_boxplot).

## Warning: Removed 29 rows containing non-finite values (stat_boxplot).

## Warning: Removed 14 rows containing non-finite values (stat_boxplot).

## Warning: Removed 17 rows containing non-finite values (stat_boxplot).

## Warning: Removed 19 rows containing non-finite values (stat_boxplot).

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 19 rows containing non-finite values (stat_bin).

## Warning: Removed 19 rows containing non-finite values (stat_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 41 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 40 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 25 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 31 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 31 rows containing missing values

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 28 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 41 rows containing missing values (geom_point).

## Warning: Removed 28 rows containing non-finite values (stat_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 48 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 36 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 40 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 39 rows containing missing values

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 29 rows containing non-finite values (stat_bin).

## Warning: Removed 40 rows containing missing values (geom_point).

## Warning: Removed 48 rows containing missing values (geom_point).

## Warning: Removed 29 rows containing non-finite values (stat_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 34 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 41 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 38 rows containing missing values

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 14 rows containing non-finite values (stat_bin).

## Warning: Removed 25 rows containing missing values (geom_point).

## Warning: Removed 36 rows containing missing values (geom_point).

## Warning: Removed 34 rows containing missing values (geom_point).

## Warning: Removed 14 rows containing non-finite values (stat_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 25 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 25 rows containing missing values

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 17 rows containing non-finite values (stat_bin).

## Warning: Removed 31 rows containing missing values (geom_point).

## Warning: Removed 40 rows containing missing values (geom_point).

## Warning: Removed 41 rows containing missing values (geom_point).

## Warning: Removed 25 rows containing missing values (geom_point).

## Warning: Removed 17 rows containing non-finite values (stat_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 30 rows containing missing values

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 19 rows containing non-finite values (stat_bin).

## Warning: Removed 31 rows containing missing values (geom_point).

## Warning: Removed 39 rows containing missing values (geom_point).

## Warning: Removed 38 rows containing missing values (geom_point).

## Warning: Removed 25 rows containing missing values (geom_point).

## Warning: Removed 30 rows containing missing values (geom_point).

## Warning: Removed 19 rows containing non-finite values (stat_density).
```

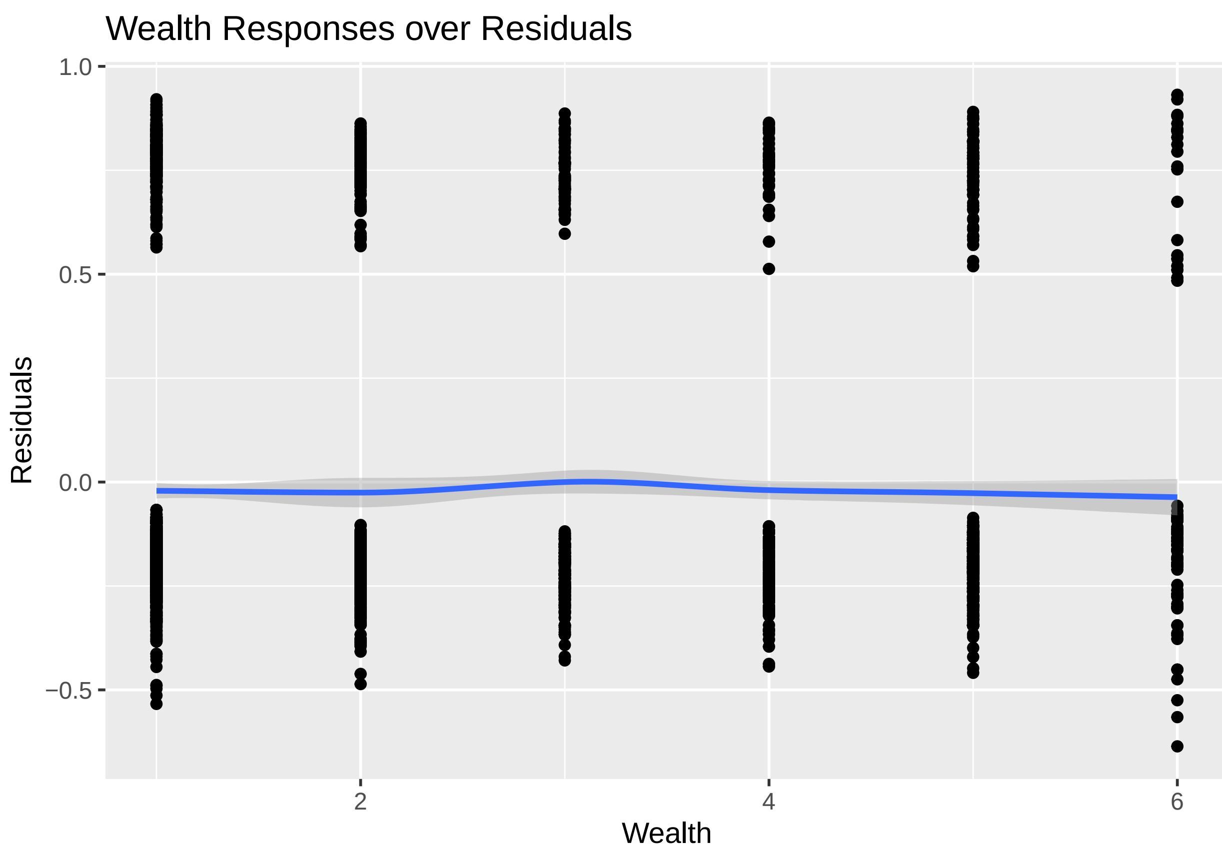


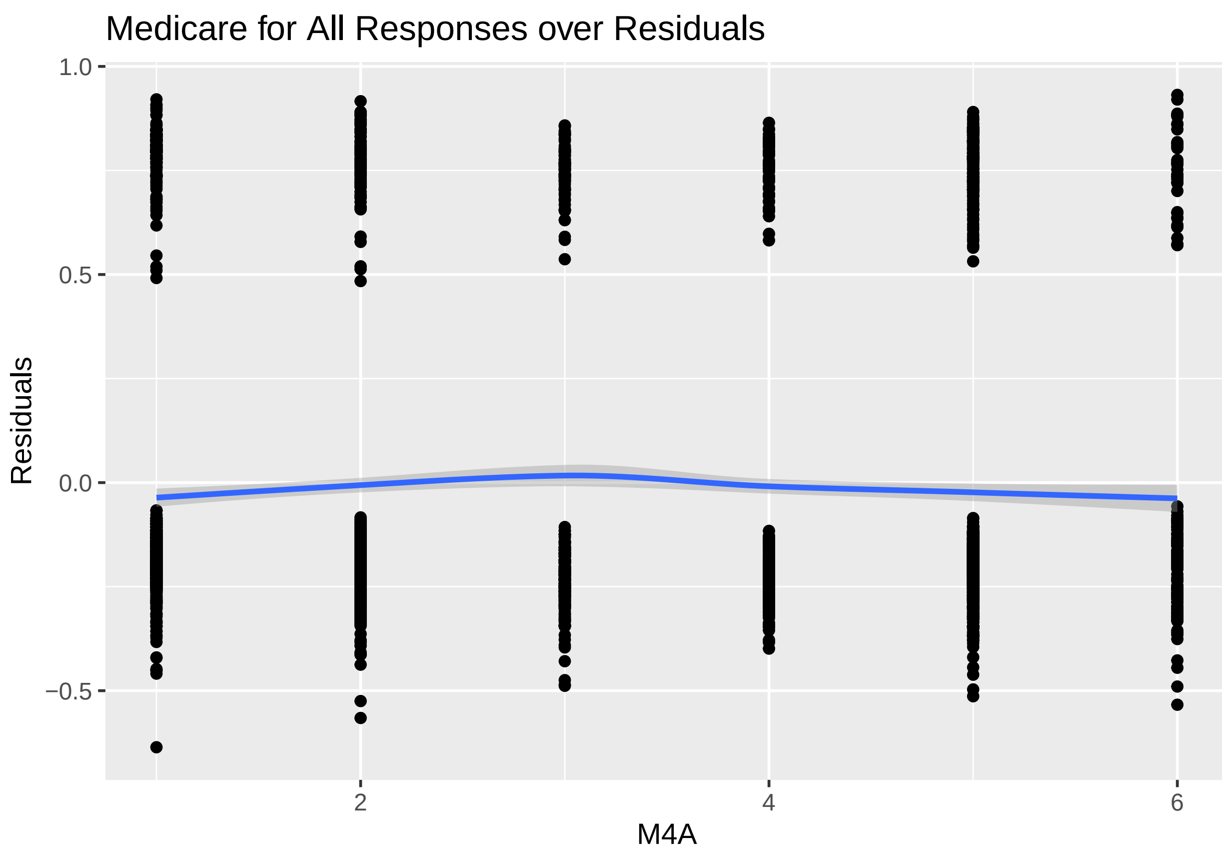
Swinger
M4A
GREENJOB
WEALTH
MARLEG
ICE
GUNS
Corr: 0.448
Corr: 0.58
Corr: 0.485
Corr: 0.498
Corr: 0.502
Corr: 0.495
Corr: 0.343
Corr: 0.3
Corr: 0.368
Corr: 0.405
Corr: 0.381
Corr: 0.525
Corr: 0.357
Corr: 0.298
Corr: 0.363



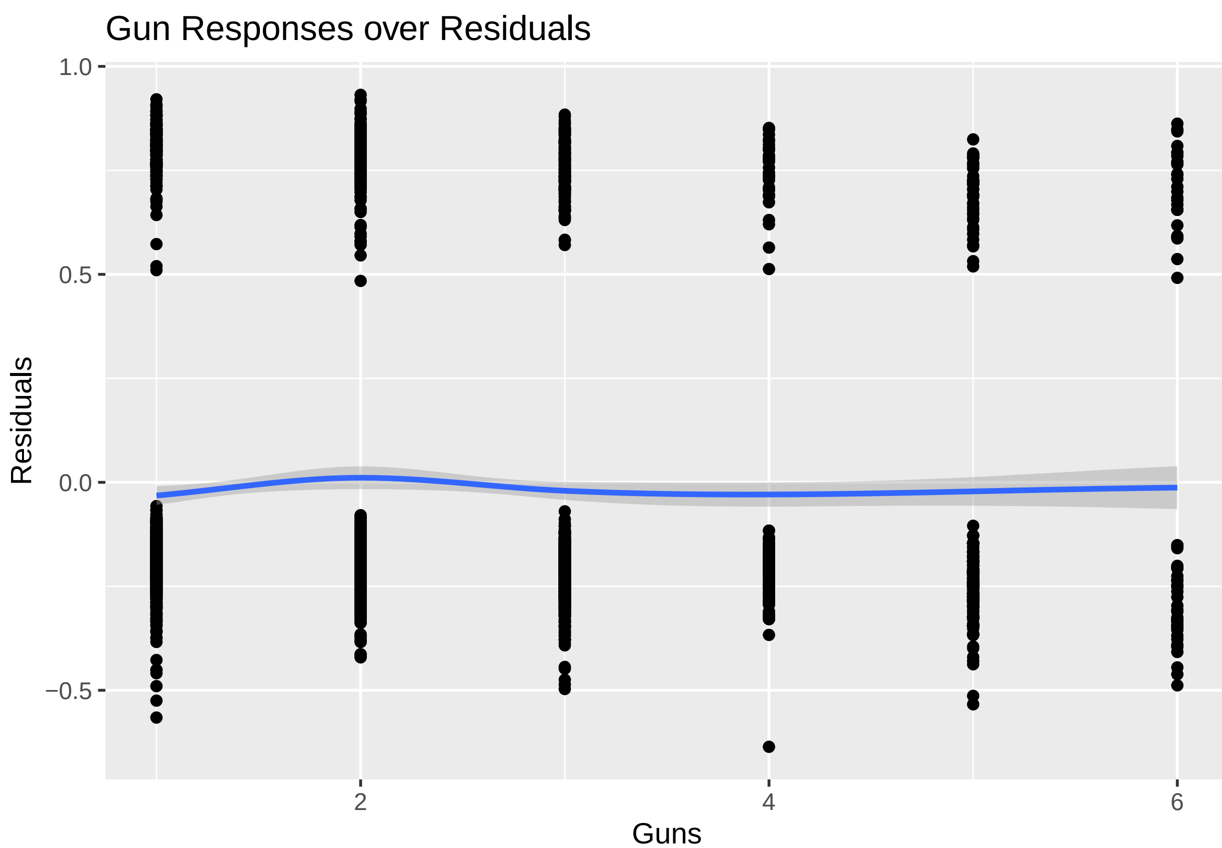
Results of fitting the issue variable models:

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 2

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number -0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used
## at 2

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 1

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal
## condition number -0
```

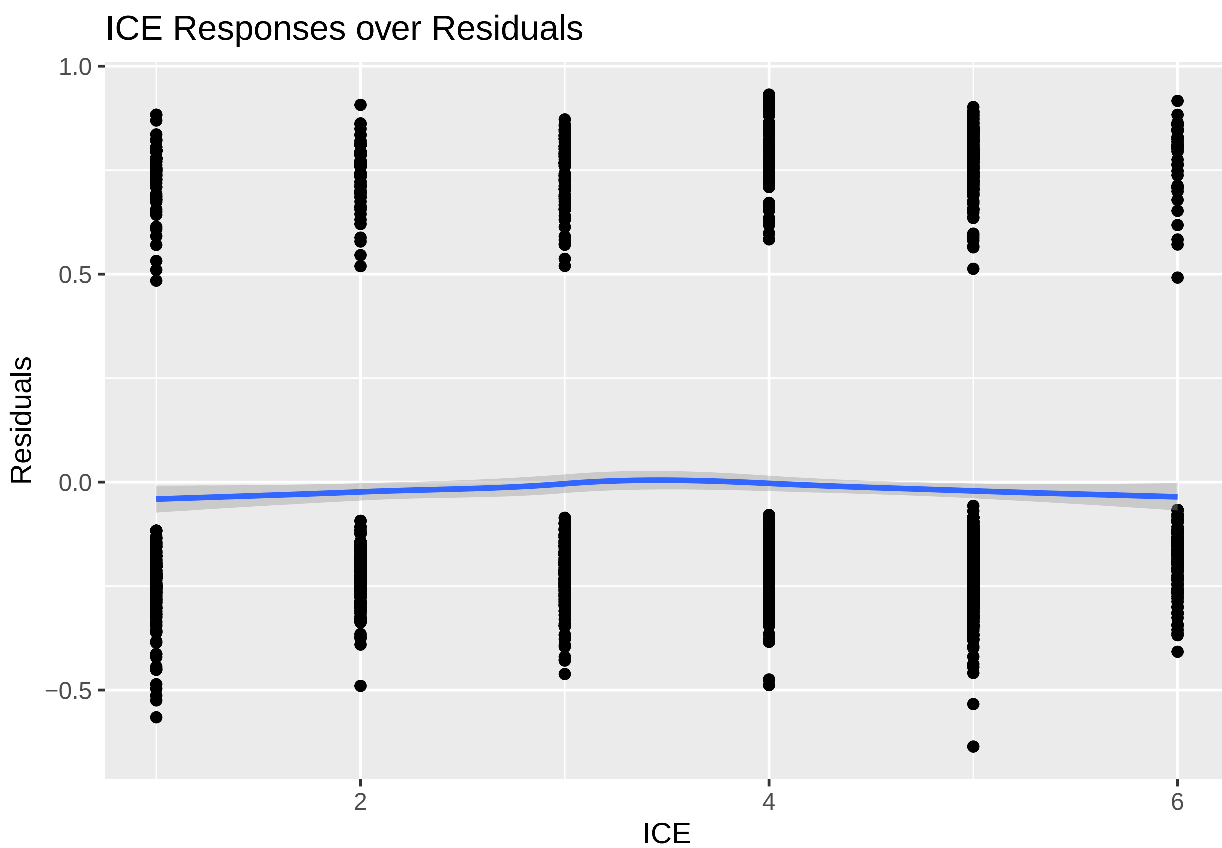

Wealth Responses over Residuals
1.0
0.5
0.0
−0.5
Residuals
2
4
6
Wealth



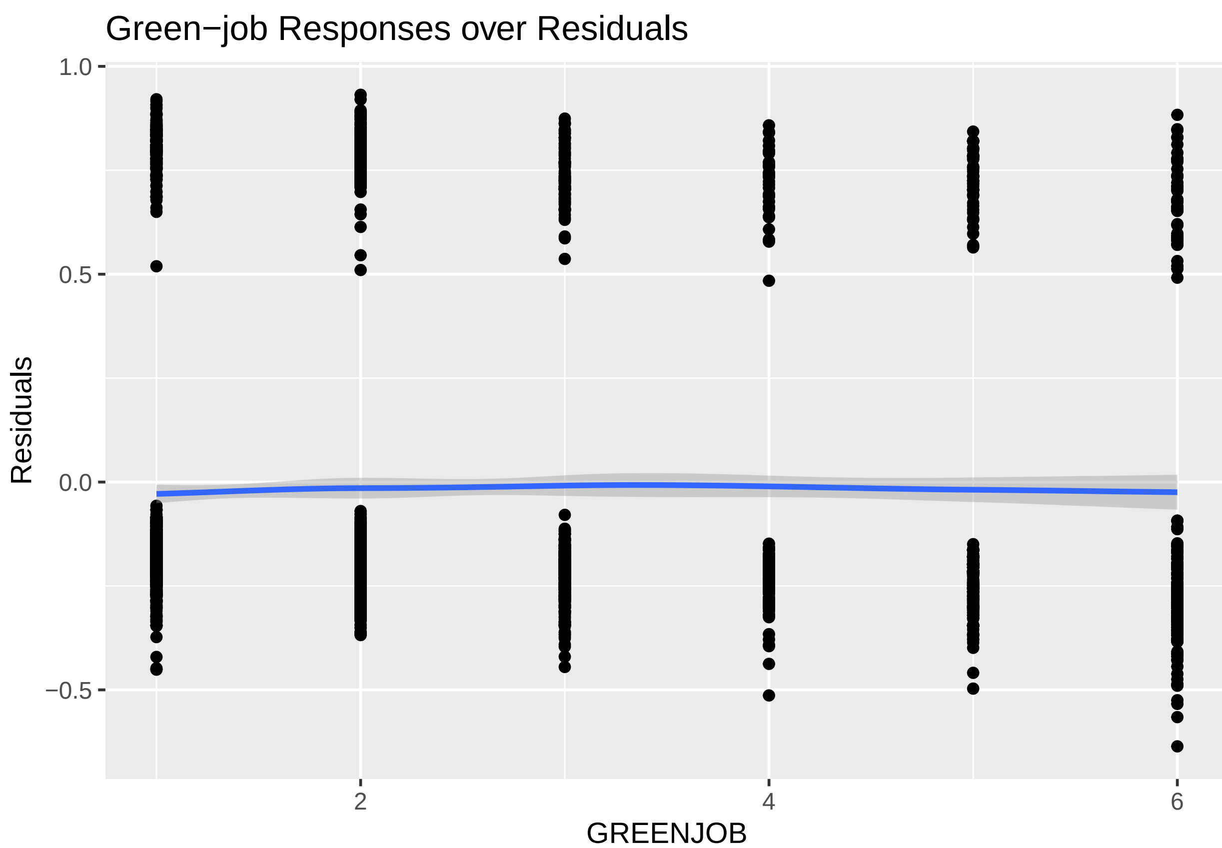

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 2

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number -0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used
## at 2

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 1

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal
## condition number -0
```

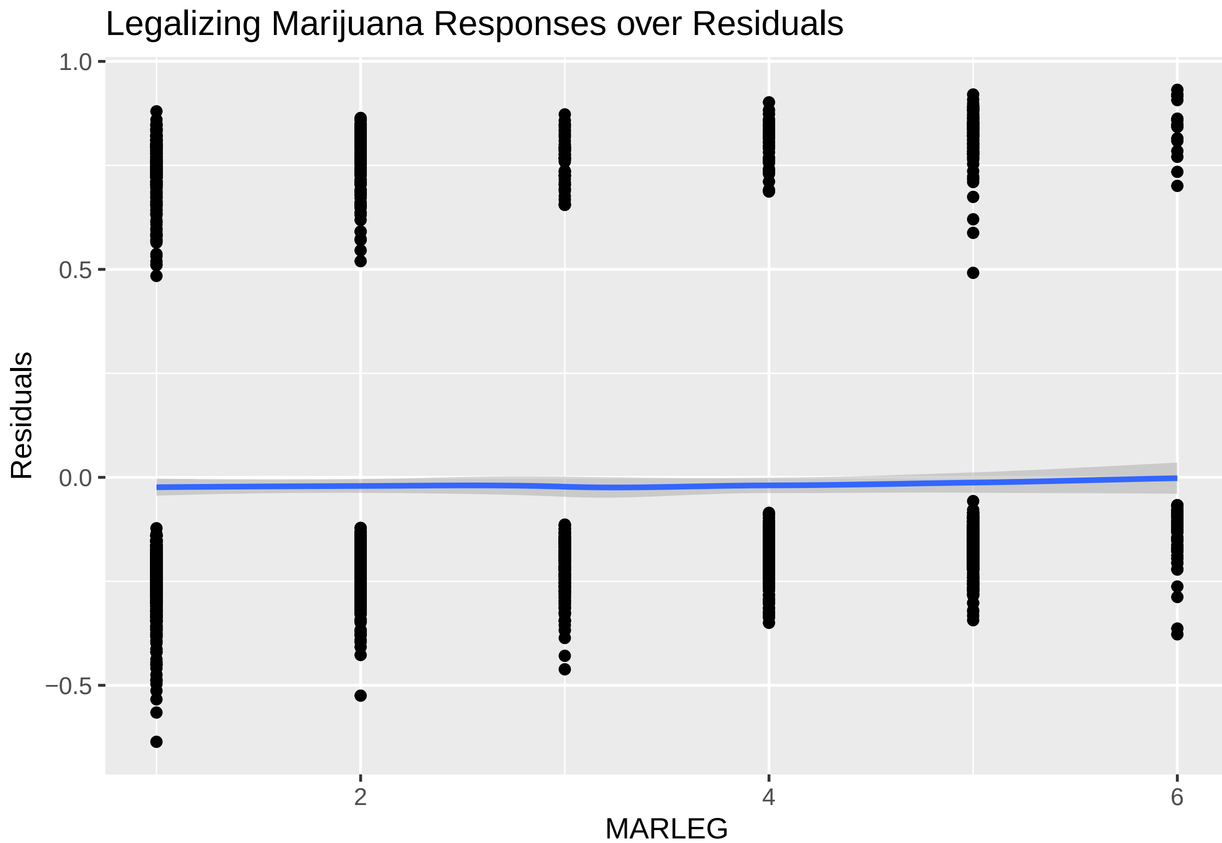


Gun Responses over Residuals
1.0
0.5
0.0
−0.5
Residuals
2
4
6
Guns



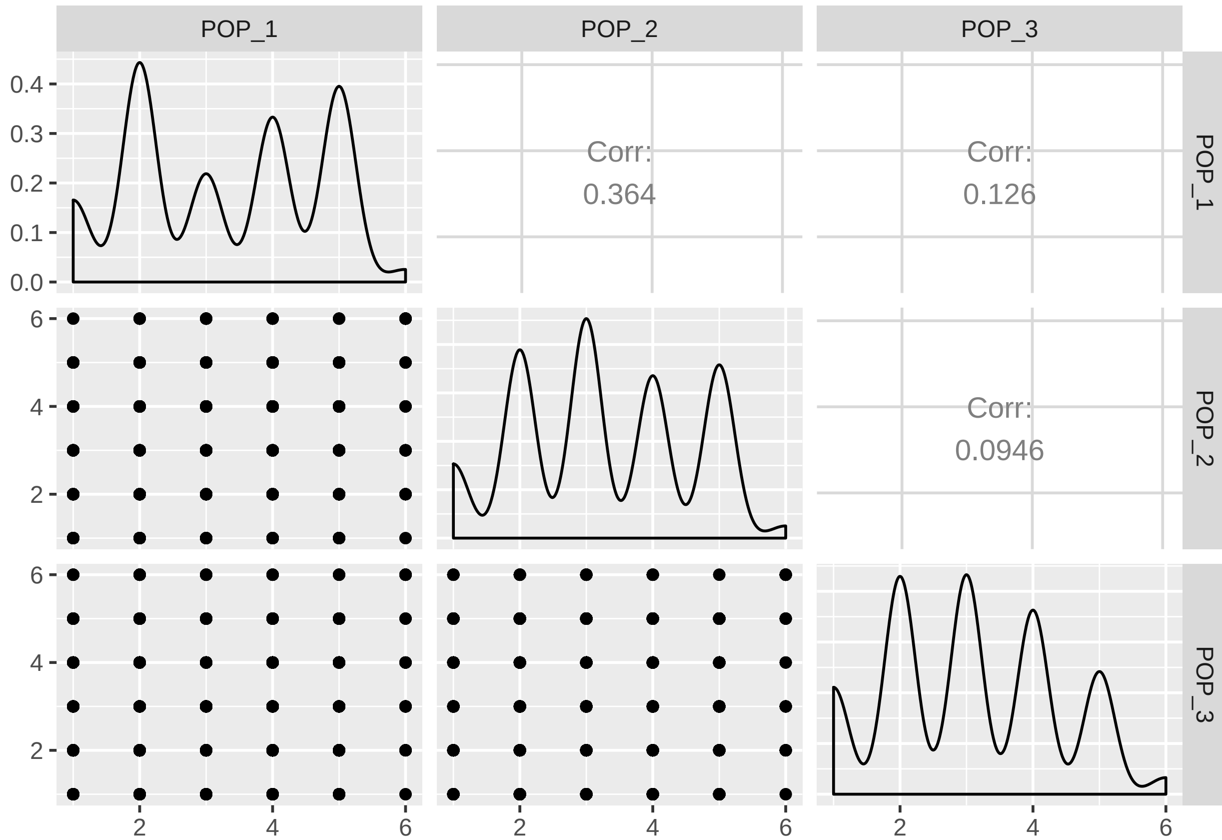

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 2

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used
## at 2

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 1

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal
## condition number 0
```

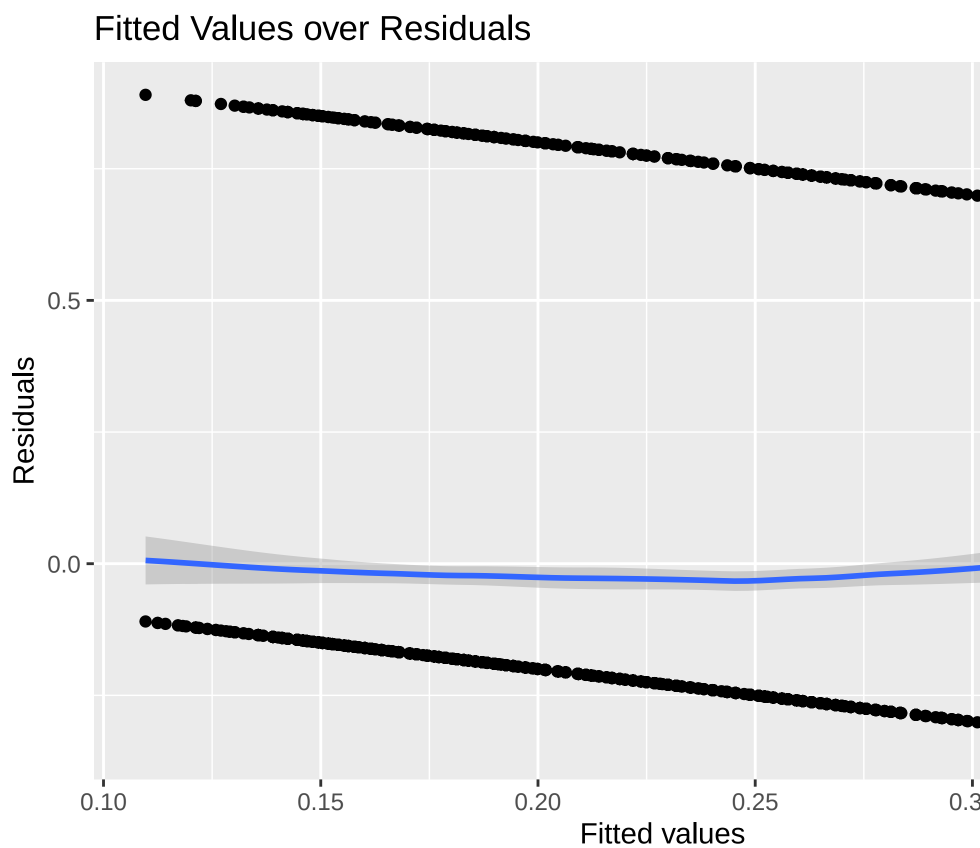

Green−job Responses over Residuals
1.0
0.5
0.0
−0.5
Residuals
2
4
6
GREENJOB



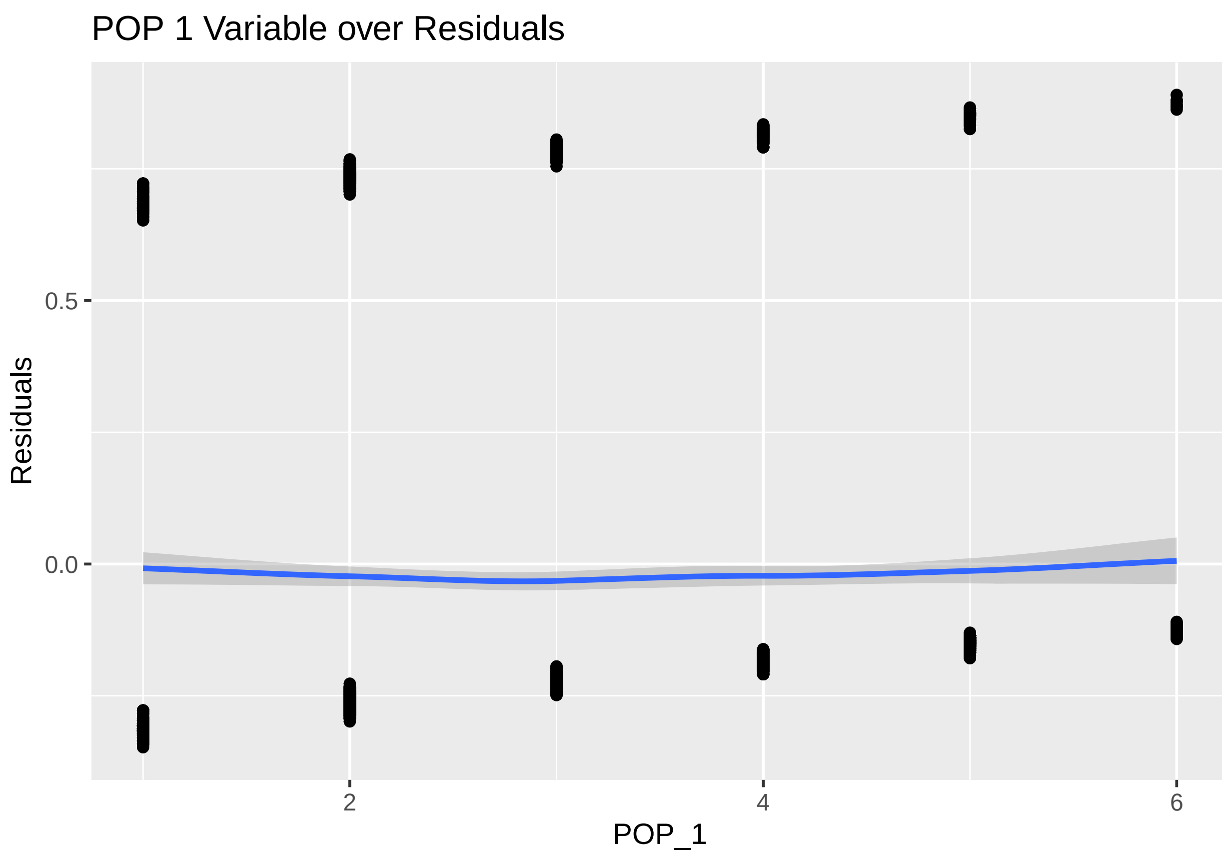
GGPairs for populism variables:

```
## Warning: Removed 21 rows containing non-finite values (stat_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 26 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 37 rows containing missing values

## Warning: Removed 26 rows containing missing values (geom_point).

## Warning: Removed 13 rows containing non-finite values (stat_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 28 rows containing missing values

## Warning: Removed 37 rows containing missing values (geom_point).

## Warning: Removed 28 rows containing missing values (geom_point).

## Warning: Removed 22 rows containing non-finite values (stat_density).
```

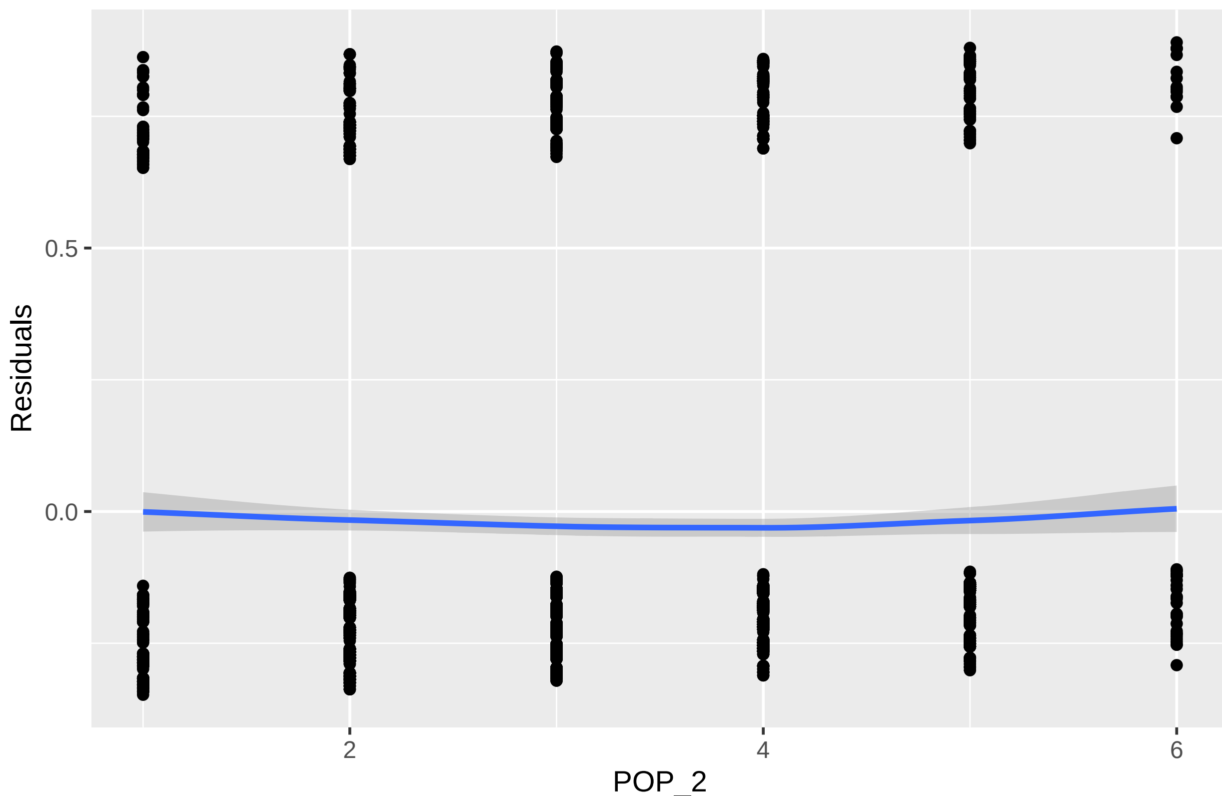

POP_1
POP_2
POP_3
Corr:
0.364
Corr:
0.126
Corr:
0.0946
0.0
0.1
0.2
0.3
0.4
2
4
6

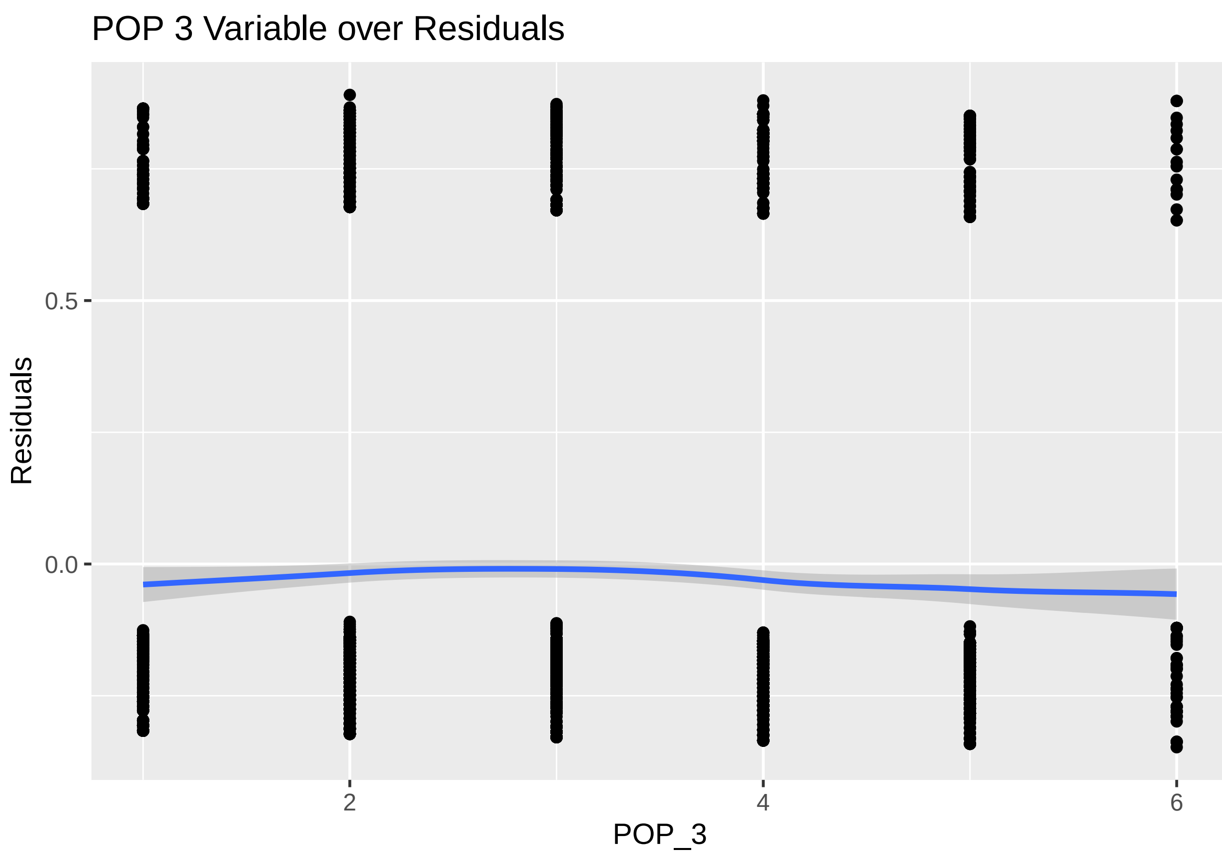


Results from fitting populism variables:


POP 1 Variable over Residuals
Residuals
0.5
0.0
2
4
6
POP_1

POP 2 Variable over Residuals




POP 3 Variable over Residuals
Residuals
0.5
0.0
2
4
6
POP_3