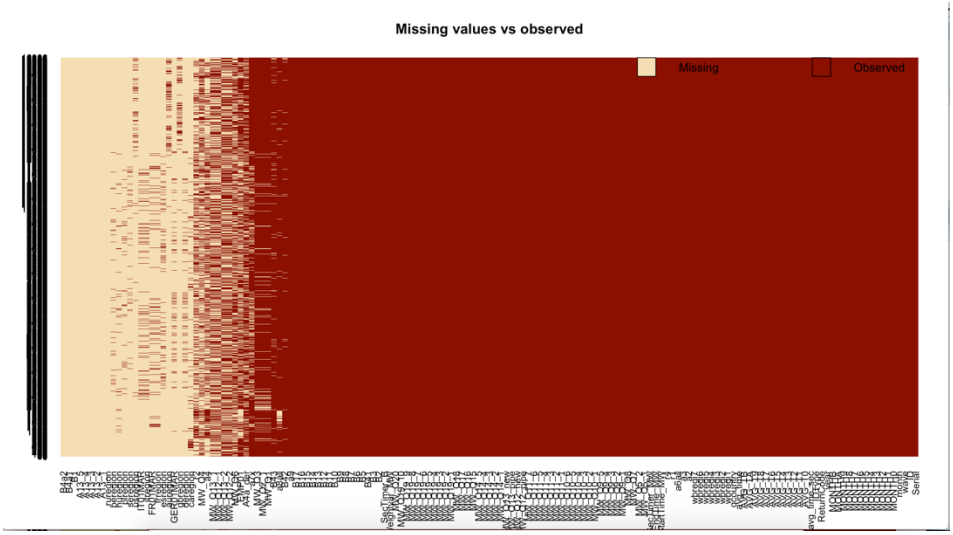
Report on the analysis of the Nativism data

Step 1: Cleaning and making sense of data

1. Generally, the data is pretty clean and tidy. As shown in the following picture, the yellow pixels mean the losing value and the red mean complete value. The variables of our interests are all complete with no Null value at all.



2. The variables that I used are:
    1. Nativism data: Q09_1 – Q09_5
    2. Sys_broke: the "system is broken" variable; Q10
    3. Confidence: the confidence in the society and its public institutions; Q11
    4. Open: the opinion towards an open economy and society; Q12 & Q 13
    5. Gov_effect: the level of governmental intervention; Q 14
    6. Demographic variables: gender and education. There is no variable evaluating the ethnicity.

There are some other variables that can be played with such as QB3 about the level of satisfaction about the economic situations. I did not include it in the model. But it can be further explored.

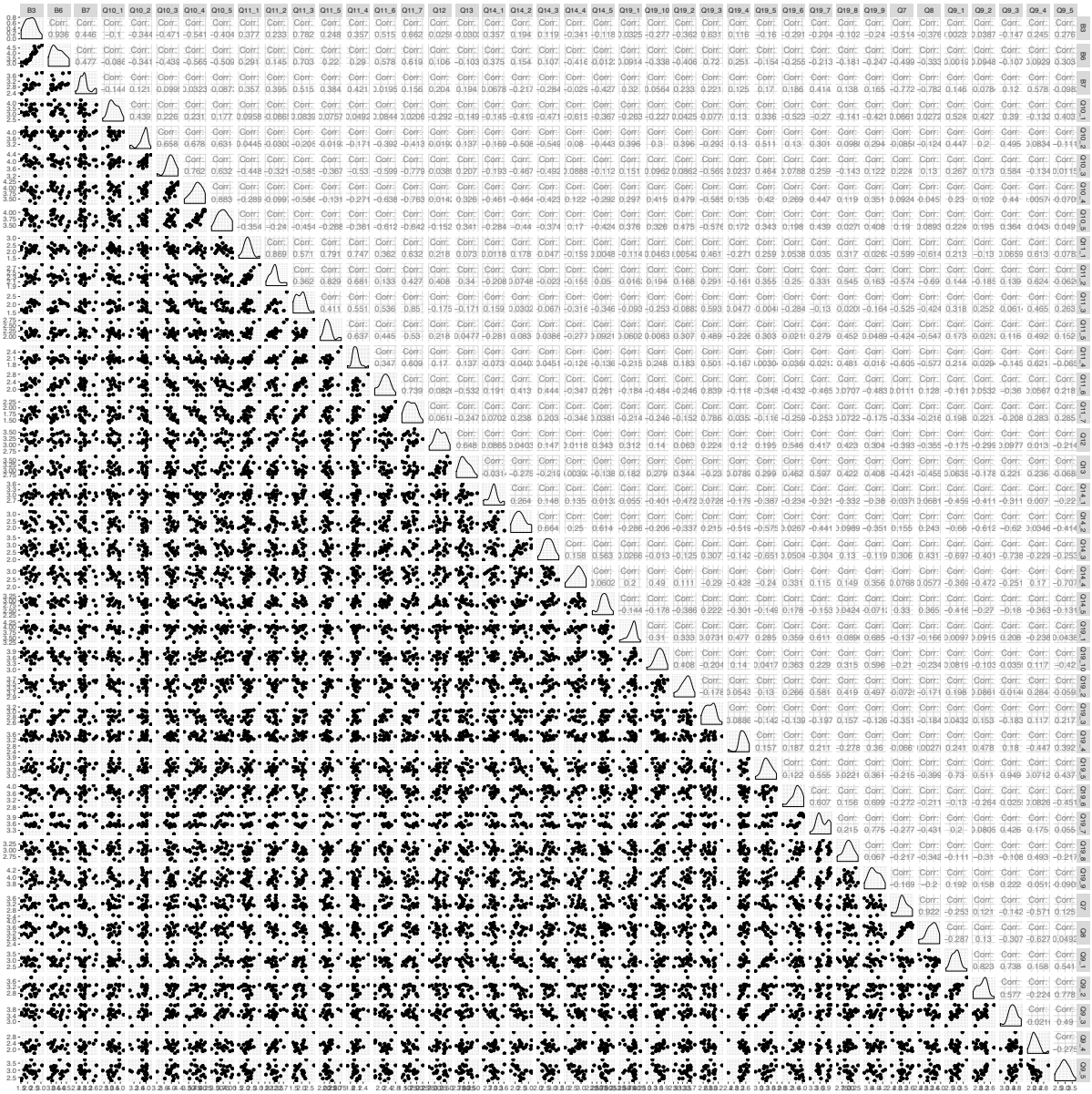
The data set is not the same with the data used in the Ipsos report.

For the independent variable, there is no questions about "Abortion," "Fear of others," "Worse off than parents," "Government intervention" and "American dream." And there are some minor changes of the questions in other latent variables.

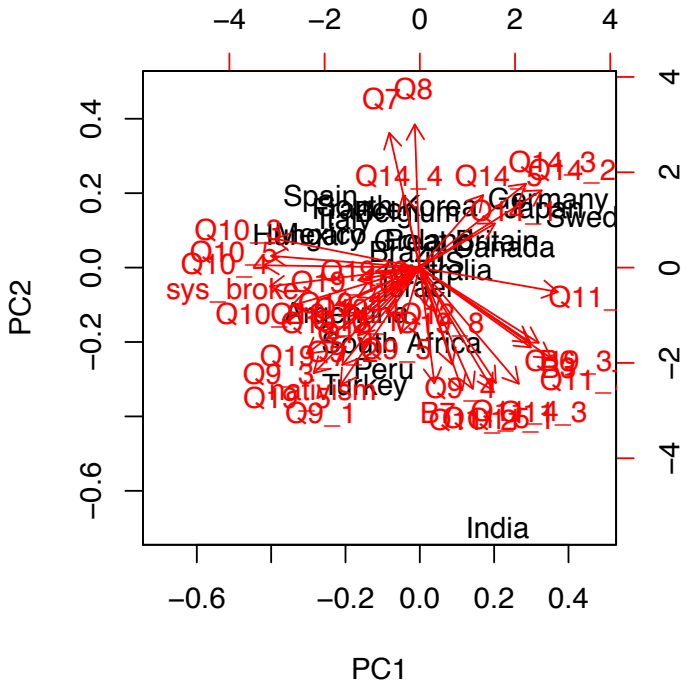
For the dependent variable, there is no questions about approval of Trump, Trump among Republicans or identification of Republicans. Therefore, in the model that I used, I only used "System is broken" as my response variable. Although there are other potential response variables such as "confidence in the society" or "economic projection", I think the "System is broken" might be the most interesting one.

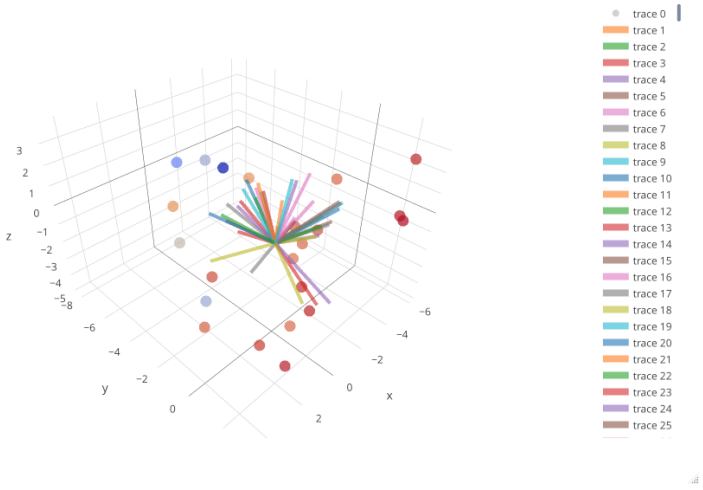
Step 2: Exploratory data analysis

The relationship between variables and the distribution of each variable (question) is shown in the following plot.

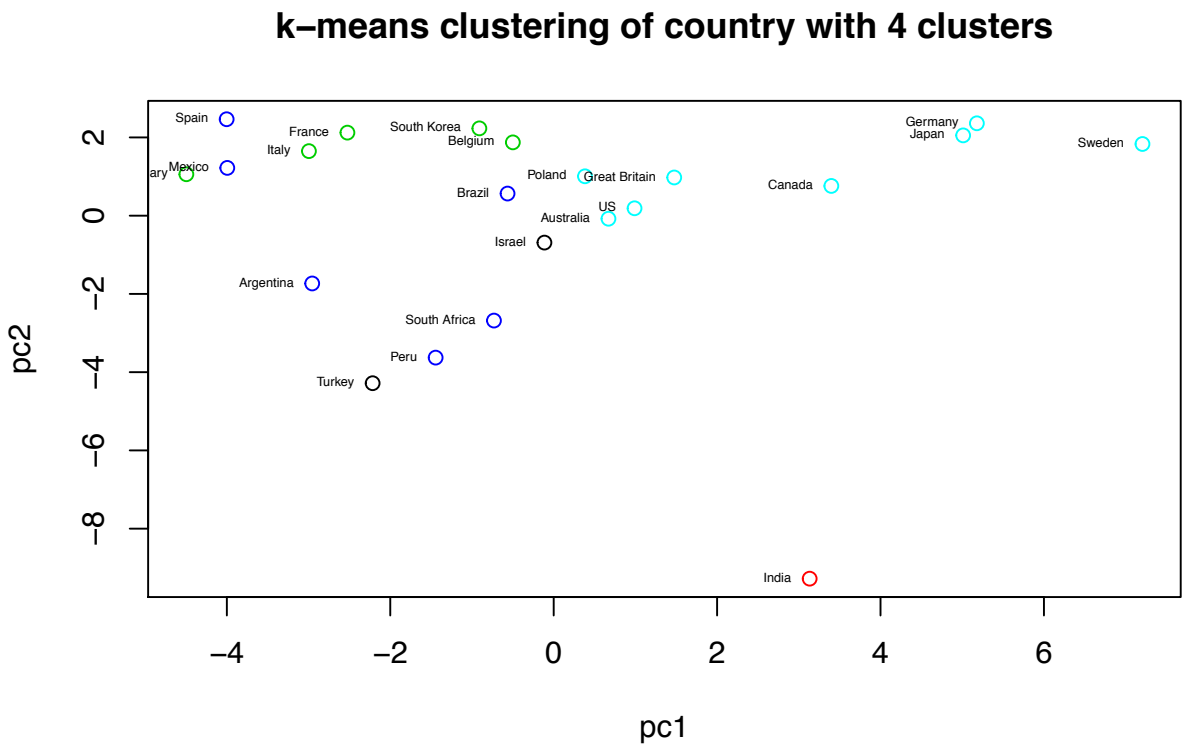


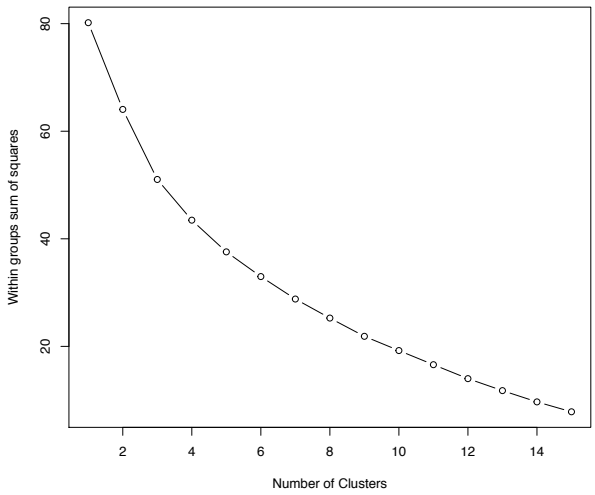
To further find out the relationship between variables. The principle component analysis result is shown below:



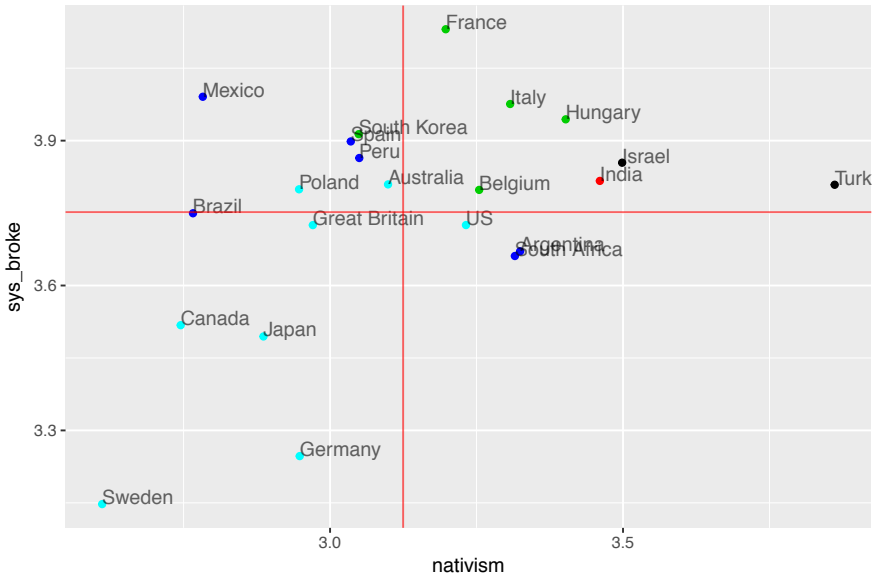


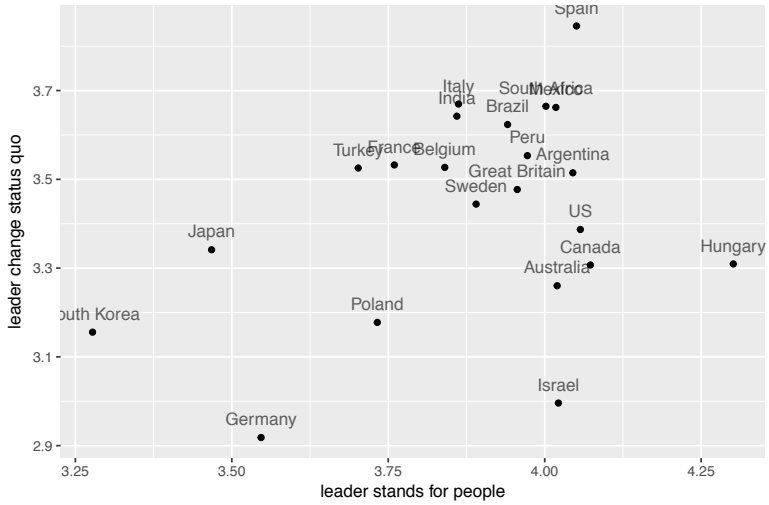
A cluster analysis is done to further distinguish the groups of different countries. After the analysis and a few trials, I clustered them into 5 clusters (mainly because the specialty of India).

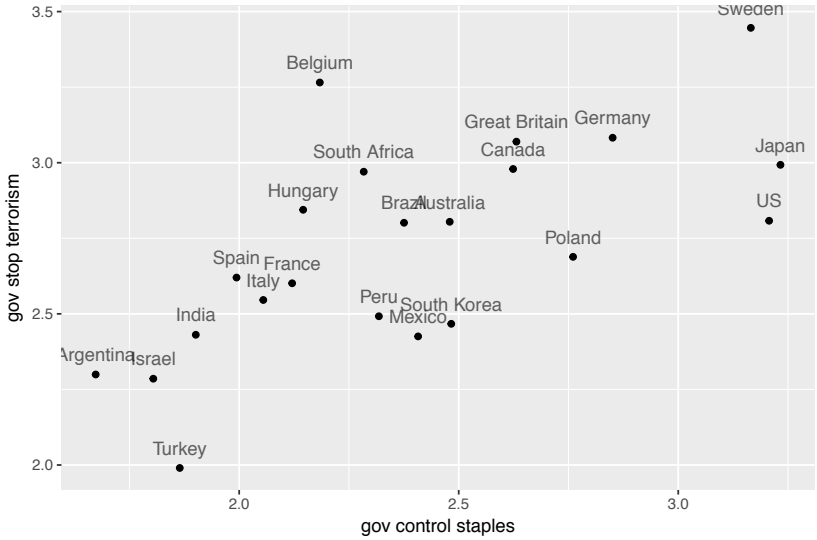




Some visuals of the scatterplots of the countries and their corresponding values:







Step 3: Modeling the data

I used ordered logistic regression to model the data. The “system is broken” is the response variable. To understand the differences between countries, I modeled 5 countries in total with each of them from one particular cluster.

I compile the results as follows:

```
> summary(US_model)
Call:
polr(formula = Sys_broke ~ (nativism + age + education + gender +
    confidence + open + gov_effect), data = US, Hess = TRUE)

Coefficients:
                    Value Std. Error t value
nativism         0.484983   0.072231  6.7143
age             -0.003455   0.005258 -0.6570
educationMedium  0.350039   0.175358  1.9961
educationHigh    0.296020   0.165990  1.7834
genderFemale     0.030330   0.138671  0.2187
confidence      -0.946733   0.127441 -7.4288
open             0.055167   0.074227  0.7432
gov_effect      -0.666080   0.104214 -6.3915

Intercepts:
                                    Value   Std. Error t value
disagree|neither agree or disagree -6.0082  0.7102    -8.4602
neither agree or disagree|agree    -2.3304  0.6749    -3.4530

Residual Deviance: 1417.531
AIC: 1437.531
>
```

The coefficient table is as below:

	Value	Std. Error	t value	p value	Odds_ration
nativism	0.4850	0.0722	6.7143	0.0000	1.6241
age	-0.0035	0.0053	-0.6570	0.5112	0.9966
educationMedium	0.3500	0.1754	1.9961	0.0459	1.4191
educationHigh	0.2960	0.1660	1.7834	0.0745	1.3445
genderFemale	0.0303	0.1387	0.2187	0.8269	1.0308
confidence	-0.9467	0.1274	-7.4288	0.0000	0.3880
open	0.0552	0.0742	0.7432	0.4573	1.0567
gov_effect	-0.6661	0.1042	-6.3915	0.0000	0.5137
disagree|neither agree or disagree	-6.0082	0.7102	-8.4602	0.0000	1.6241
neither agree or disagree|agree	-2.3304	0.6749	-3.4530	0.0006	0.9966

```
>
```

Further stepAIC is run to improve the model (only on the training data):
The final step is shown as below:

```
                          Df    AIC
- education                2 1375.0
<none>                       1375.5
+ nativism:gender          1 1375.8
+ nativism:education       2 1376.2
+ education:gender         2 1376.6
+ nativism:age             1 1376.7
- age:confidence           1 1377.1
+ nativism:confidence      1 1377.2
+ age:gender               1 1377.3
+ open                     1 1377.3
+ gender:gov_effect        1 1377.3
+ age:gov_effect           1 1377.5
+ age:education            2 1378.2
+ education:gov_effect     2 1379.0
+ education:confidence     2 1379.2
- gender:confidence        1 1379.8
- nativism:gov_effect      1 1382.1
- confidence:gov_effect    1 1404.9

Step:  AIC=1374.99
Sys_broke ~ nativism + age + gender + confidence + gov_effect +
    confidence:gov_effect + nativism:gov_effect + gender:confidence +
    age:confidence
```

Next, I only compiled the coefficient table as illustrations:

```
> round(Sweden_coeff_table, 4)
                                    Value Std. Error t value p value Odds_ratio
nativism                           0.3630     0.0968  3.7510  0.0002     1.4376
age                                0.0034     0.0076  0.4495  0.6531     1.0034
educationMedium                   -0.3290     0.2455 -1.3400  0.1803     0.7196
educationHigh                     -0.6289     0.2485 -2.5308  0.0114     0.5332
genderFemale                       0.2537     0.1966  1.2904  0.1969     1.2888
confidence                        -1.6701     0.2246 -7.4371  0.0000     0.1882
open                              -0.1871     0.1106 -1.6920  0.0907     0.8294
gov_effect                        -1.0011     0.1516 -6.6025  0.0000     0.3675
disagree|neither agree or disagree -8.2908    0.9893 -8.3806  0.0000     1.4376
neither agree or disagree|agree    -5.2489    0.9396 -5.5862  0.0000     1.0034
```

```
> round(US_coeff_table, 4)
                                    Value Std. Error t value p value Odds_ratio
nativism                           0.4850     0.0722  6.7143  0.0000     1.6241
age                               -0.0035     0.0053 -0.6570  0.5112     0.9966
educationMedium                    0.3500     0.1754  1.9961  0.0459     1.4191
educationHigh                      0.2960     0.1660  1.7834  0.0745     1.3445
genderFemale                       0.0303     0.1387  0.2187  0.8269     1.0308
confidence                        -0.9467     0.1274 -7.4288  0.0000     0.3880
open                               0.0552     0.0742  0.7432  0.4573     1.0567
gov_effect                        -0.6661     0.1042 -6.3915  0.0000     0.5137
disagree|neither agree or disagree -6.0082     0.7102 -8.4602  0.0000     1.6241
neither agree or disagree|agree    -2.3304     0.6749 -3.4530  0.0006     0.9966
>
```

```
> round(France_coeff_table, 4)
                                    Value Std. Error  t value p value Odds_ratio
nativism                           0.0730     0.0837   0.8715  0.3835     1.0757
age                                0.0069     0.0065   1.0554  0.2912     1.0069
educationMedium                    0.4882     0.2274   2.1465  0.0318     1.6294
educationHigh                      0.6672     0.2437   2.7377  0.0062     1.9487
genderFemale                       0.3004     0.1638   1.8346  0.0666     1.3504
confidence                        -1.2305     0.1532  -8.0343  0.0000     0.2921
open                               0.0192     0.0859   0.2238  0.8229     1.0194
gov_effect                        -1.2466     0.1478  -8.4315  0.0000     0.2875
disagree|neither agree or disagree -9.5654     0.8879 -10.7726  0.0000     1.0757
neither agree or disagree|agree    -5.6969     0.8096  -7.0368  0.0000     1.0069
>
```

```
> round(India_coeff_table, 4)
                                    Value Std. Error  t value p value Odds_ration
nativism                           0.5913     0.1257   4.7020  0.0000      1.8063
age                                0.0080     0.0093   0.8540  0.3931      1.0080
educationMedium                  -10.6980     0.3208 -33.3507  0.0000      0.0000
educationHigh                    -10.2969     0.3261 -31.5766  0.0000      0.0000
genderFemale                      -0.1920     0.2029  -0.9465  0.3439      0.8253
confidence                        -0.2299     0.1873  -1.2275  0.2196      0.7946
open                              -0.1842     0.0925  -1.9927  0.0463      0.8317
gov_effect                        -1.1592     0.1358  -8.5373  0.0000      0.3138
disagree|neither agree or disagree -16.0738     0.5847 -27.4912  0.0000      1.8063
neither agree or disagree|agree    -12.5147     0.5388 -23.2253  0.0000      1.0080
>
```

```
> round(Turkey_coeff_table, 4)
                                    Value Std. Error t value p value Odds_ration
nativism                           0.7057     0.1152  6.1253  0.0000      2.0252
age                                0.0157     0.0094  1.6743  0.3931      1.0158
educationMedium                   -0.7519     0.3364 -2.2353  0.0000      0.4715
educationHigh                     -0.4613     0.2465 -1.8715  0.0000      0.6304
genderFemale                       0.2836     0.2012  1.4094  0.3439      1.3279
confidence                        -0.7244     0.1498 -4.8365  0.2196      0.4846
open                              -0.0807     0.0896 -0.9009  0.0463      0.9225
gov_effect                        -0.6118     0.1296 -4.7227  0.0000      0.5424
disagree|neither agree or disagree -3.6109     0.7957 -4.5378  0.0000      2.0252
neither agree or disagree|agree    -0.7263     0.7769 -0.9349  0.0000      1.0158
```