

# Datathon Writeup

Linda Cao, Georgie Stammer, Enzo Moraes Mescall, Erik Mendes Novak

2021-11-07

## Introduction

Immediately after going over the data set our group was interested in section P of the questionnaire where individuals were asked a series of intense political questions and responded with their opinions. This provides a unique subjective insight of a population's perspective on their civic space. The data set provides information about the material conditions, and their perceived material conditions, of the surveyed and thus allows us to attempt to estimate the impact of an individual's material conditions on their political leanings. Notably, all the questions in section P are on a scale of 1 - 4 where answering 1 indicates support of a larger authoritarian government while answering 4.

**Research Question** To what extent do an individual's material conditions affect their leanings toward authoritarian form of government?

## Methodology

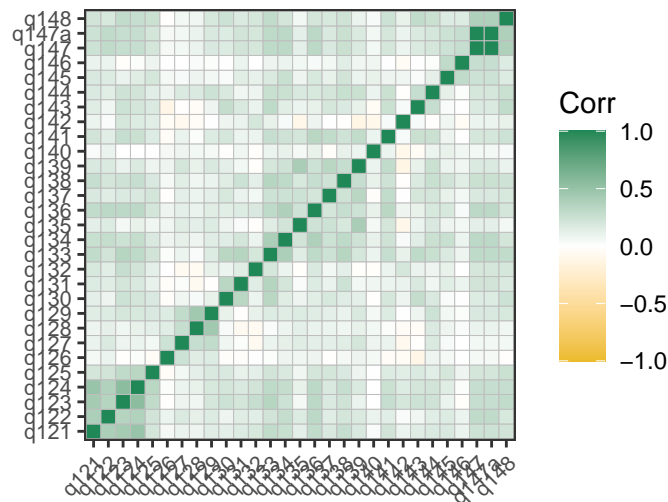
To measure an individual's political leanings based off of their answers from the questionnaire we took their responses from questions Q121 to Q148 which scale from 1 to 4 and took the arithmetic mean across selected questions.

//space for erik

```
corr_ideology <- cor(cleaned_ideology, method = c("spearman"), use = "complete.obs")

ggcorrplot(corr_ideology, ggtheme = theme_bw(), tl.cex = 8,
            colors = c("#ebbb2a", "white", "#208756")) +
  labs(title = "Correlation between questions")
```

Correlation between questions

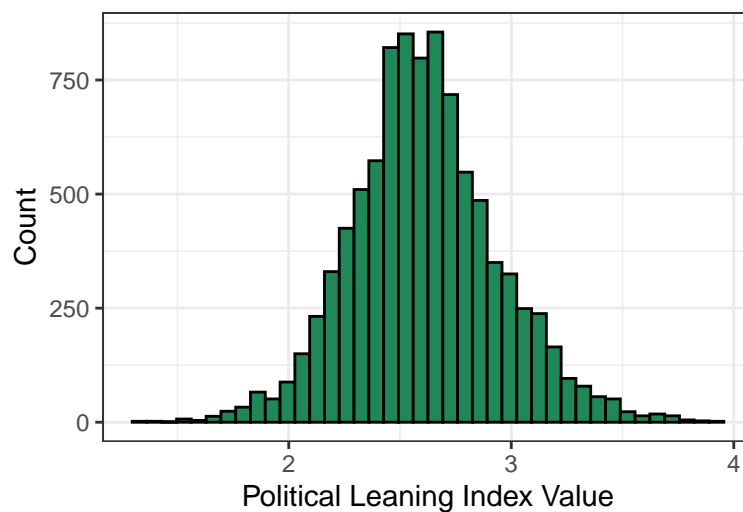


Next we look at the aforementioned questions in section P.

### Exploratory Data Analysis

```
ggplot(data = final_data, aes(x = avg)) +
  geom_histogram(fill = "#208756", color = "black", bins = 40) +
  labs(title = "Distribution of Political Leanings Index",
       x = "Political Leaning Index Value",
       y = "Count") +
  theme_bw()
```

Distribution of Political Leanings Index



```
kable(learning_index, digits = 3,
      caption = "Summary Statistics for Learning Index",
      col.names = c("Mean", "Max", "Median", "Min", "IQR", "SD"))
```

Table 1: Summary Statistics for Learning Index

Mean	Max	Median	Min	IQR	SD
2.613	3.926	2.593	1.333	0.415	0.33

The distribution is unimodal, looks vaguely normal and doesn't have any apparent outliers.

### **Data Analysis**

After cleaning up the data we created a linear model using the responses to questions SE001 through SE017

### **Conclusion**