Draft

STA 210 - Project

Bayes' Harem - Christina Wang, Kat Cottrell, David Goh, Ethan Song

```r
library(tidyverse)
library(tidymodels)
library(knitr)
library(ggfortify)
library(GGally)
```

```r
abortion_data_full<-read_csv(here::here("data/abortion-attitudes",
                                        "wvs-usa-abortion-attitudes-data.csv"))
```

Introduction and Research Question

Motivation

Understanding public attitudes on divisive political issues is an important way for political leaders to mobilize voters and for lawmakers to draft laws that represent their constituents. While it is simple to poll constituents' positions on an issue, it can be challenging to assess the complex factors that influence and predict those stances.

Abortion is one such divisive issue in the United States, with strong organization of pro-choice and pro-life groups in states across the country in support of legislation for their respective sides (Ziegler, 2020). However, following the Supreme Court's 1973 decision in Roe v. Wade, a ruling that has historically protected women's rights to abortion without excessive government restriction in America, the issue has risen in political salience. Both the pro-choice and pro-life movements have gained national prominence, and the two major political parties have polarized around the issue, with the Democratic Party in favor of and the Republican Party against policies legalization and increased access to abortion (Weinberger 2022). Abortion has also increasingly become a key issue that voters consider, with an increasing share of Americans identifying as "single-issue voters" regarding abortion (Brenan 2020).

In May 2022, a leaked draft opinion revealed that the US Supreme Court is prepared to overturn Roe v. Wade. Overturning Roe would dramatically change the trajectory of abortion politics in the US. Unless the US Congress passed a national policy, states would be able to decide whether or not to legalize abortion and gain much greater leverage in regulating access to the procedure (Weinberger 2022).

Given the potential overturning of Roe and the polarizing nature of the issue, it is important to understand how the American public feels about whether abortion should be legal or not, how accessible the procedure should be, and what factors influence these opinions. Understanding public opinion on the issue will ensure that political leaders are able to mobilize the correct constituencies, and that policy experts are able to pass policies on this issue that accurately reflect the preferences of the American people.

Data description

We will conduct our analysis on data collected about Americans' abortion attitudes, demographic information, and other ideologies, using an EDA-informed predictive model. The data are a nationally representative sample of the American people; hence we can infer that all the observations are independent and the variables may show a linear relationship. Given that this issue has become highly polarized by political party, we predict that liberal political attitudes and youth will correlate with belief that abortion is more justified.

The dataset observes "attitudes on the justifiability of abortion in the United States across six waves of World Values Survey data" (README.md) and some basic qualities of the respondents.

Observations include:

- WVS country code
- Generational wave (1982, 1990, 1995, 1999, 2006, or 2011)
- Justifiability of abortion (1-10)
- Age (17 to 96)
- College graduate (1 for yes)
- Female (1 for women) - Unemployed (1 = currently unemployed)
- Ideology (1-10 for left-right)
- Financial satisfaction (1-10 for least-most)
- WVS post-materialist index (-1 = materialist. 2 = mixed. 3 = post-materialist)
- Child autonomy index (-2 to 2 for obedience and religious faith-determination and independence)

- Trust (1 = believes most people can be trusted)
- Importance of God (1-10)
- Opinion of respect for authority (-1-1 for bad to good)
- National pride (1 = very proud to be an American)

The data were collected as part of the World Values Survey, which is administered every few years and collects information about people's values and beliefs worldwide. The survey aims to get a nationally representative sample of a minimum of 1200 for most countries, and the data are collected via face-to-face interviews at the respondents' homes. The data included in this set specifically include responses from 6 waves of the survey (administered over the period 1982-2011). The responses included in this set are from people in the United States, and it examines their attitudes towards abortion.

Research Question and Response Variable

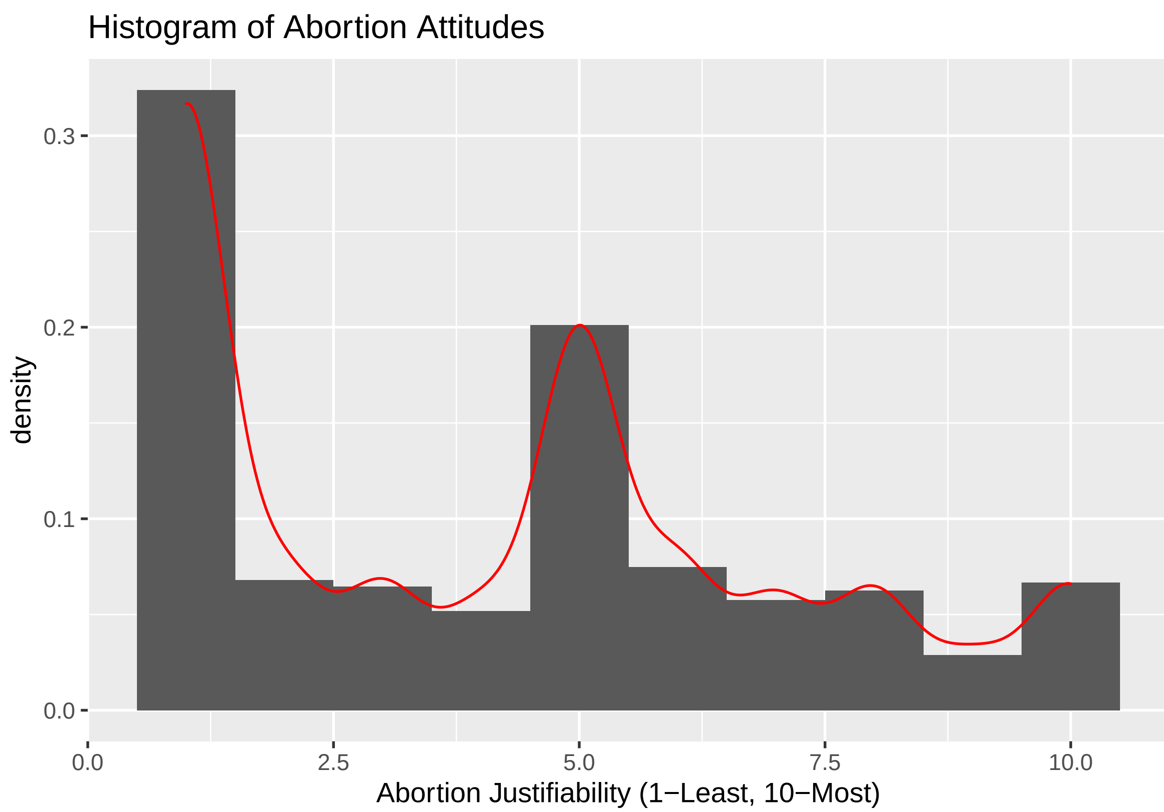
In this project, we investigate the correlation between attitudes on the justifiability of abortion with demographic factors (such as age, gender, and education level) and personal attitudes towards other issues (such as political ideology, importance of religion, and respect for authority) in a representative sample of American citizens from 1982-2011.

The response variable, `Justifiability of abortion`, is a numerical measure on a scale of 1 to 10 on the individual person's attitude toward whether abortion is justifiable or not. Individuals responded 1 for "abortion is never justified" and 10 for "abortion is always justified.

Exploratory Data Analysis

We create a visualization and summary statistics for the response variable.

```
ggplot(abortion_data_full, aes(x = aj)) +
  geom_histogram(binwidth = 1, aes(y=..density..)) +
  geom_density(color = "red") +
  labs(title = "Histogram of Abortion Attitudes",
       x = "Abortion Justifiability (1-Least, 10-Most)")
```



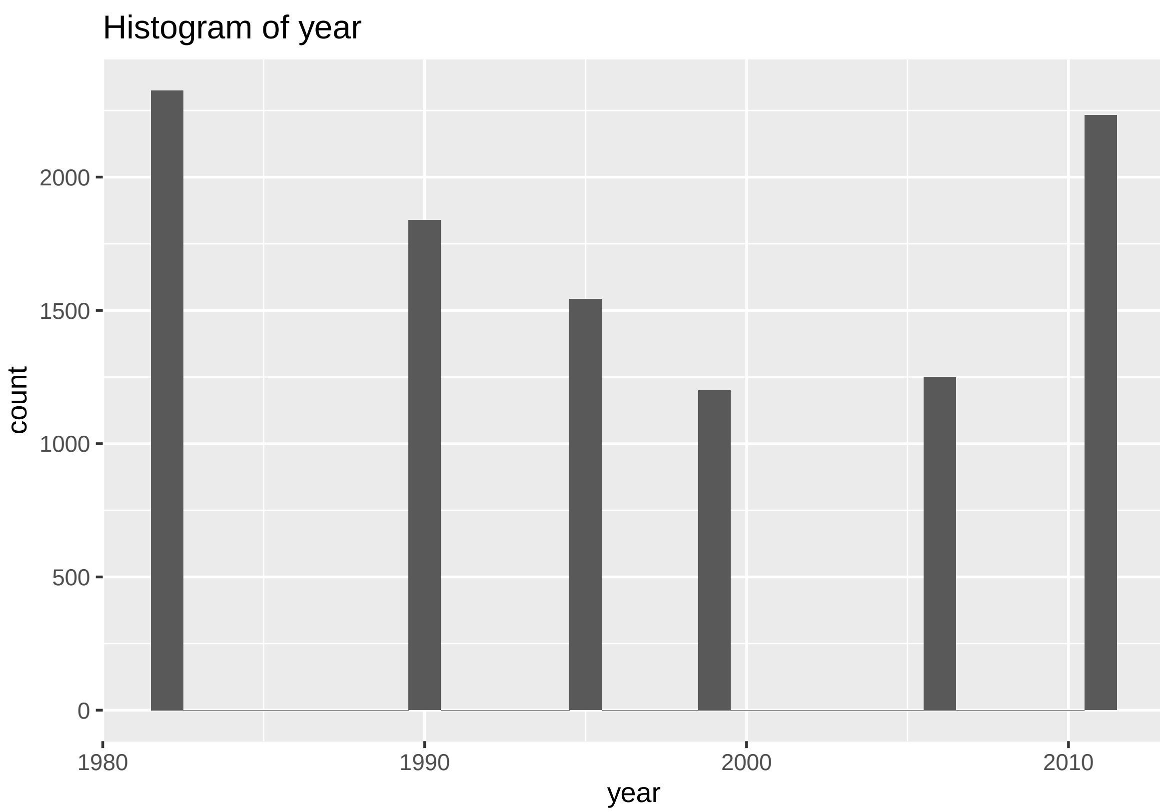

```r
summary(abortion_data_full$aj)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
  1.000   1.000   4.000   4.147   6.000  10.000     299 
```

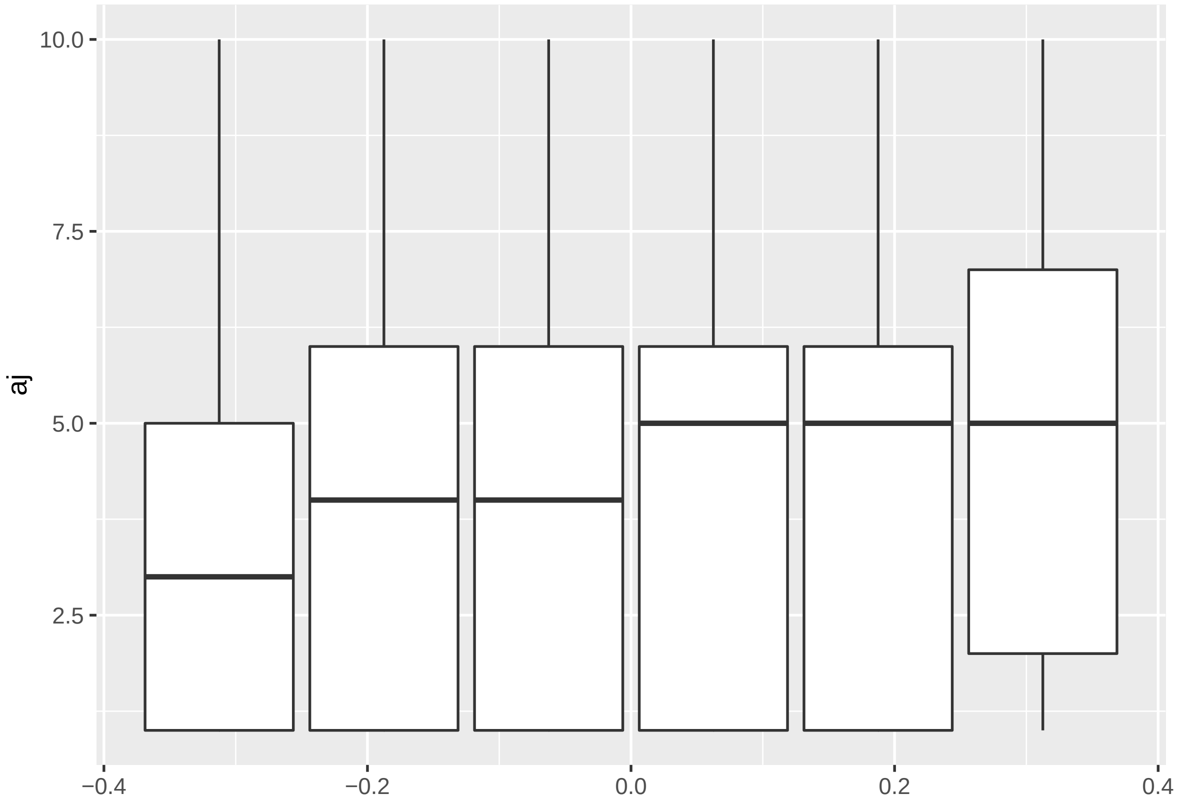
We can observe from this histogram that the distribution of the outcome variable is not a bell shape, and it is trimodal. This is likely because the question's phrasing is similar to a yes/no question, but respondents were asked to give their level of agreement on a scale of 1-10. This may result in our model not being a good fit for the data if we attempt a multiple linear regression model. We have two backup plans for this, if our MLR eventually has a poor performance. First, we can truncate this data into a categorical outcome variable such as (Agree, Disagree, Undecided), and conduct a binomial or multinomial logistic regression. Second, we can filter our population based on various characteristics (if we have good reason to do so) so that our outcome data follows a bell shape.

We now extend our exploratory data analysis (EDA) to some predictor variables of interest; namely, the year, ideology, Child Autonomy Index, Importance of God, Respect for Authority and National Pride predictors. The EDA for each variable comprises a histogram and a boxplot of the response variable grouped by predictor value. Additionally, we have a jitter plot to explore the potential for an interaction effect between year and importance of God.

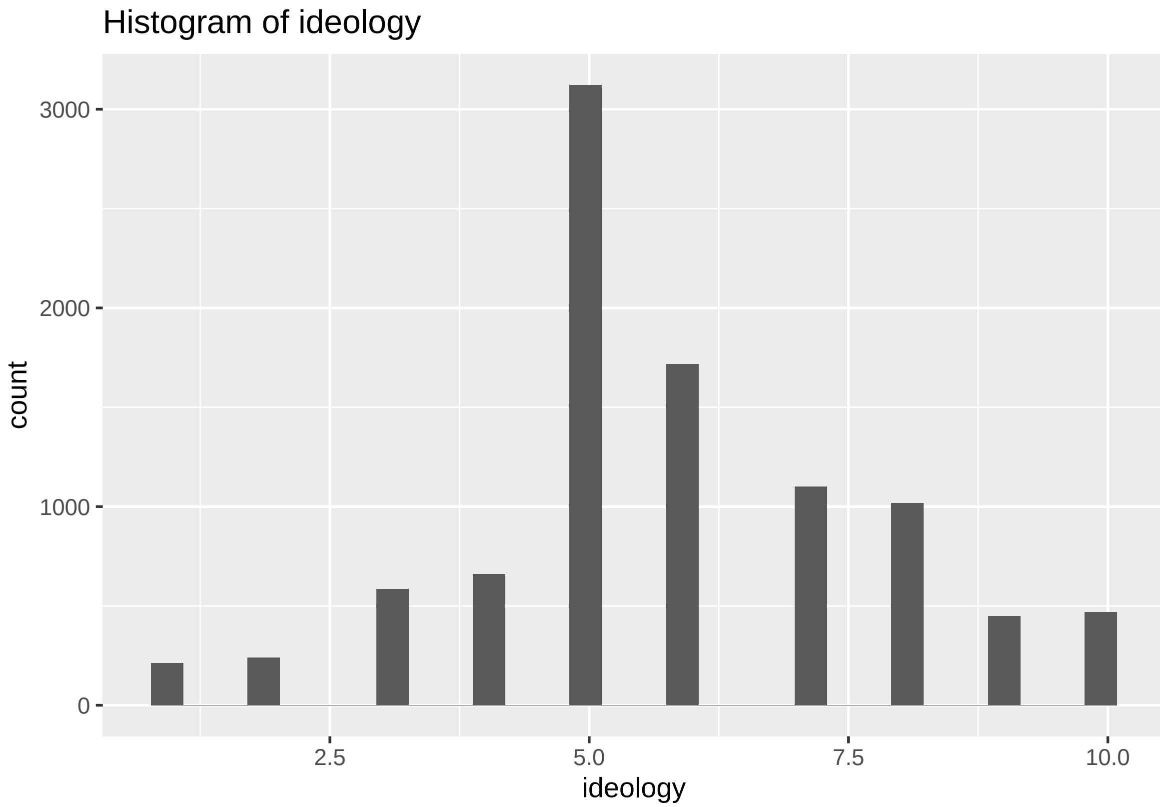
```
#Year of survey
ggplot(abortion_data_full, aes(x = year)) +
  geom_histogram() +
  labs(title = "Histogram of year")
```



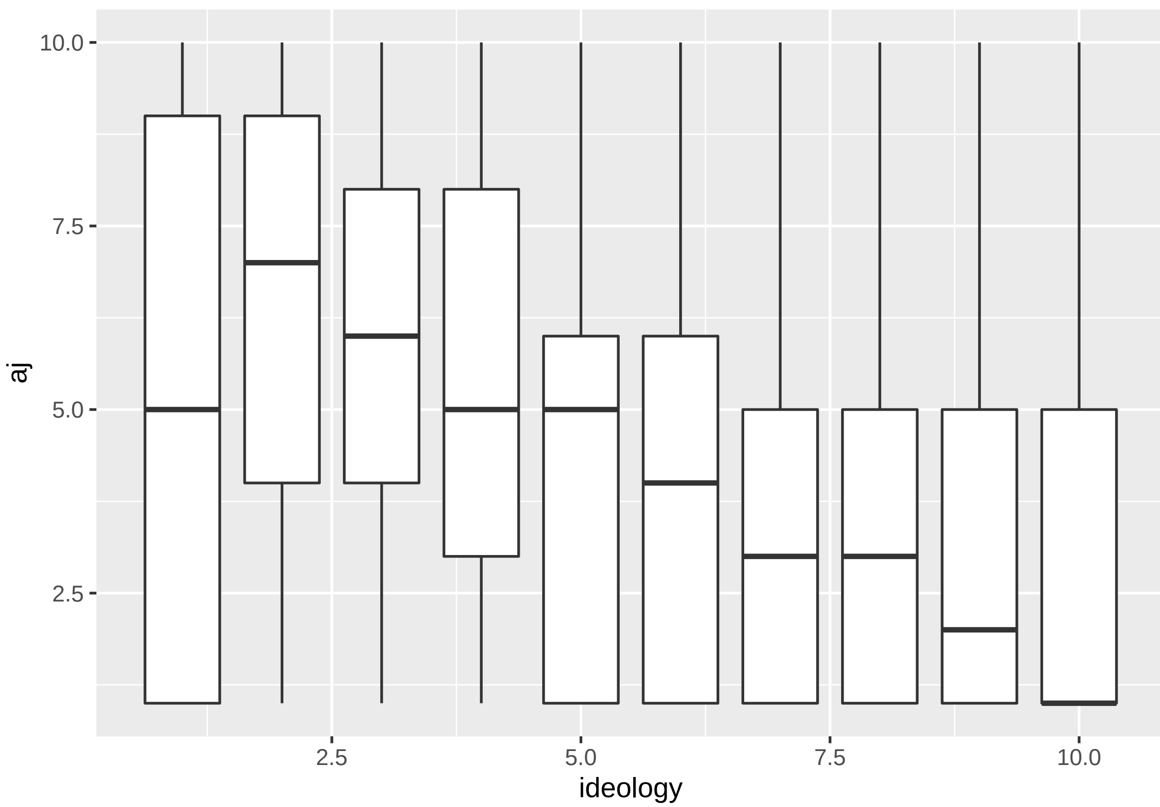

```
ggplot(abortion_data_full, aes(group = year, y = aj)) +
  geom_boxplot()
```



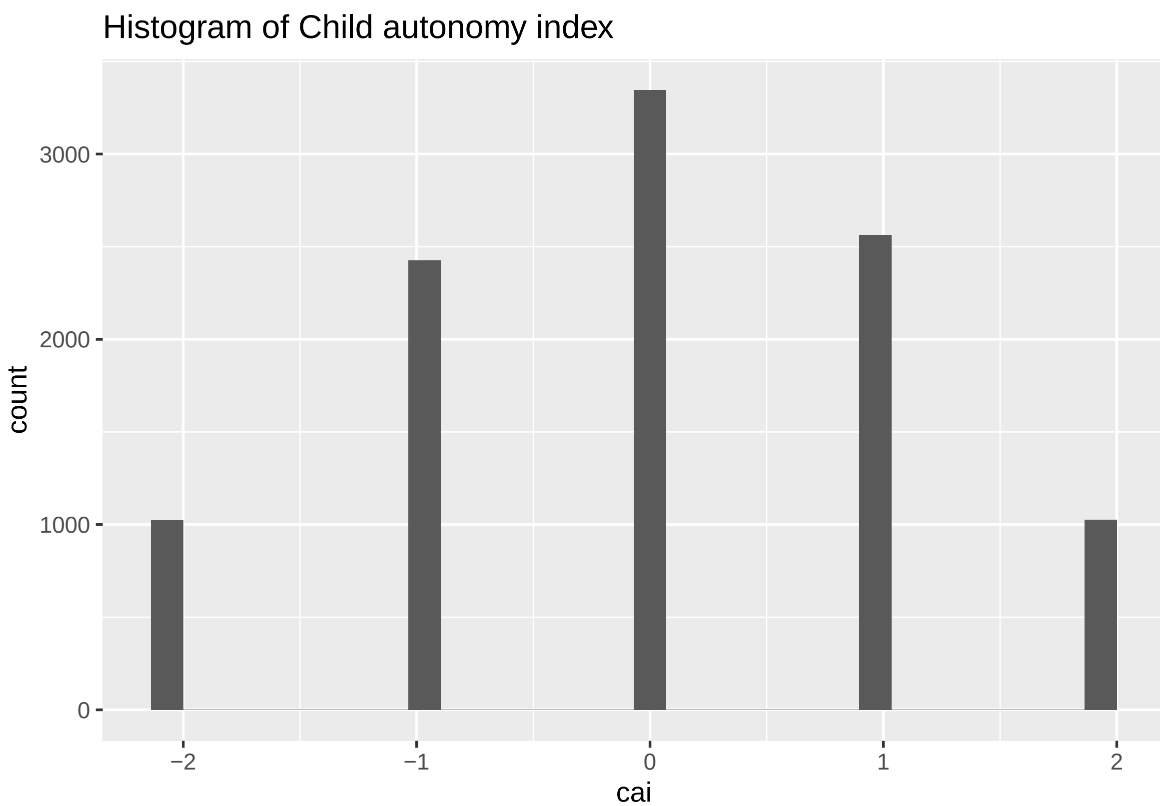

```
# ideology
ggplot(abortion_data_full, aes(x = ideology)) +
  geom_histogram() +
  labs(title = "Histogram of ideology")
```



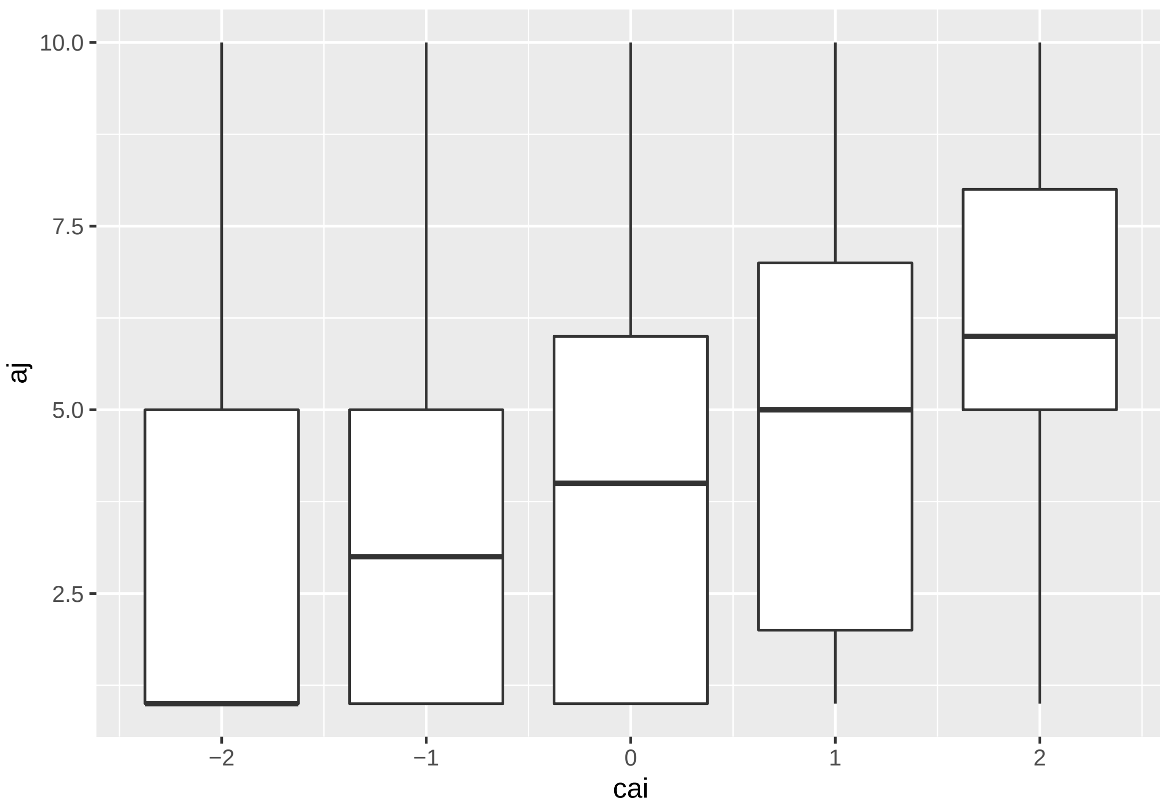

```
ggplot(abortion_data_full, aes(x = ideology,group = ideology, y = aj)) +
  geom_boxplot()
```



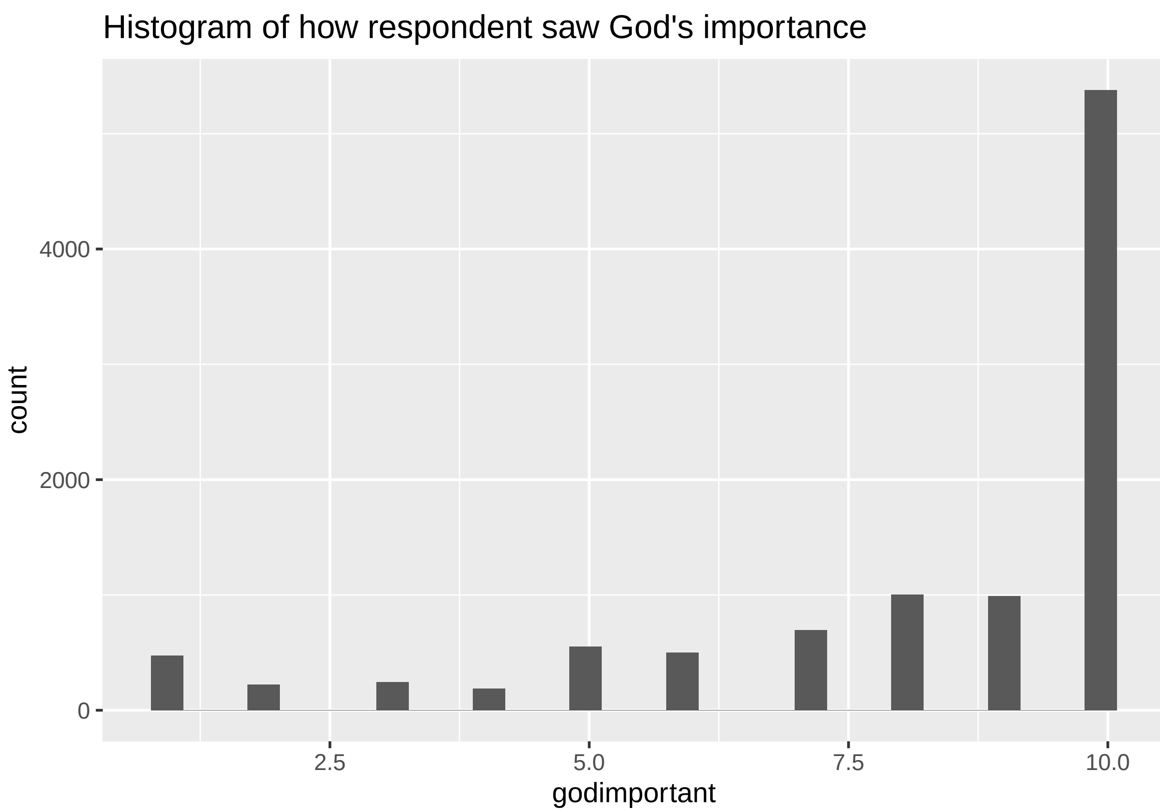

```r
# Child autonomy index
ggplot(abortion_data_full, aes(x = cai)) +
  geom_histogram() +
  labs(title = "Histogram of Child autonomy index")
```



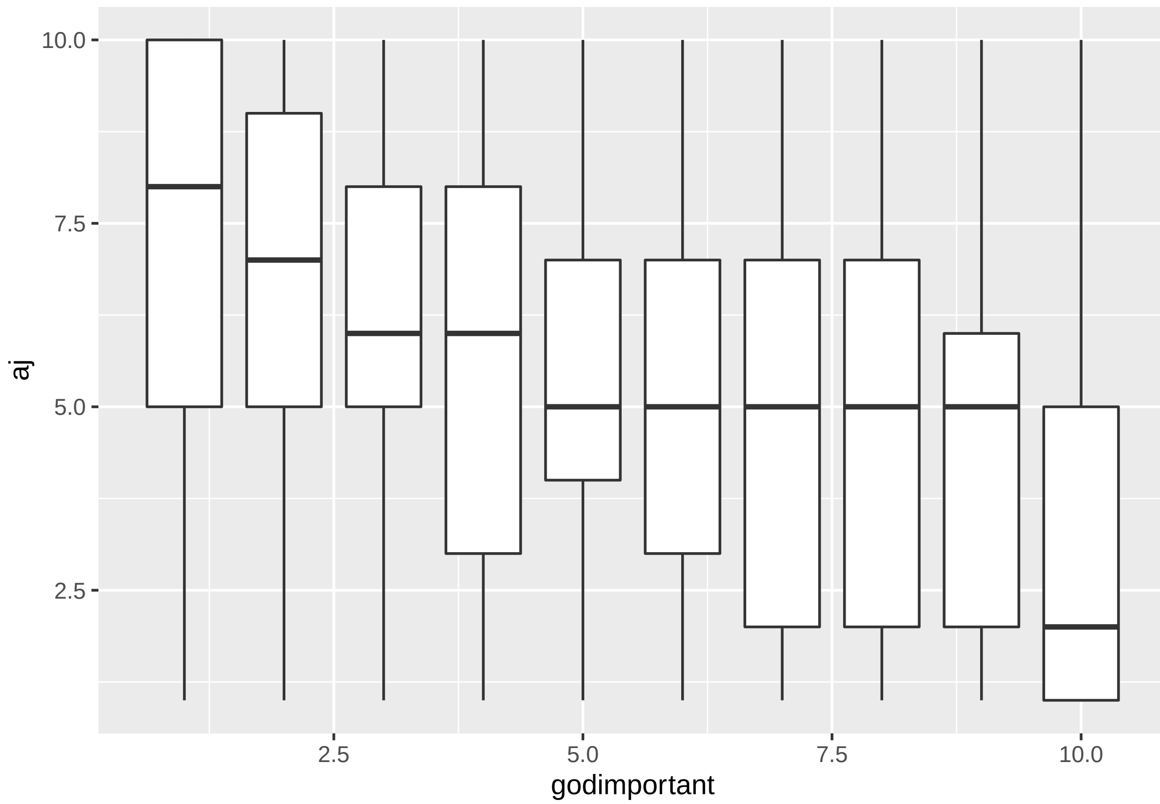

```
ggplot(abortion_data_full, aes(x = cai, group = cai, y = aj)) +
  geom_boxplot()
```



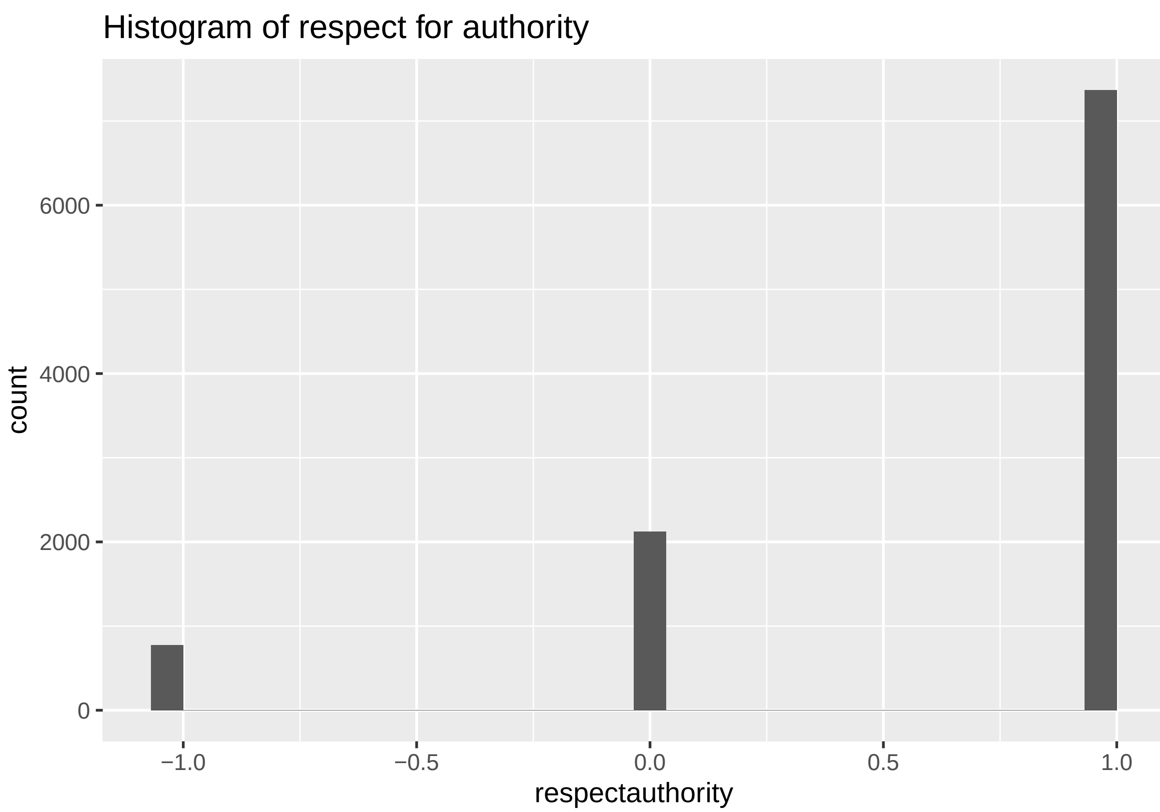

```
# Importance of God
ggplot(abortion_data_full, aes(x = godimportant)) +
  geom_histogram() +
  labs(title = "Histogram of how respondent saw God's importance")
```



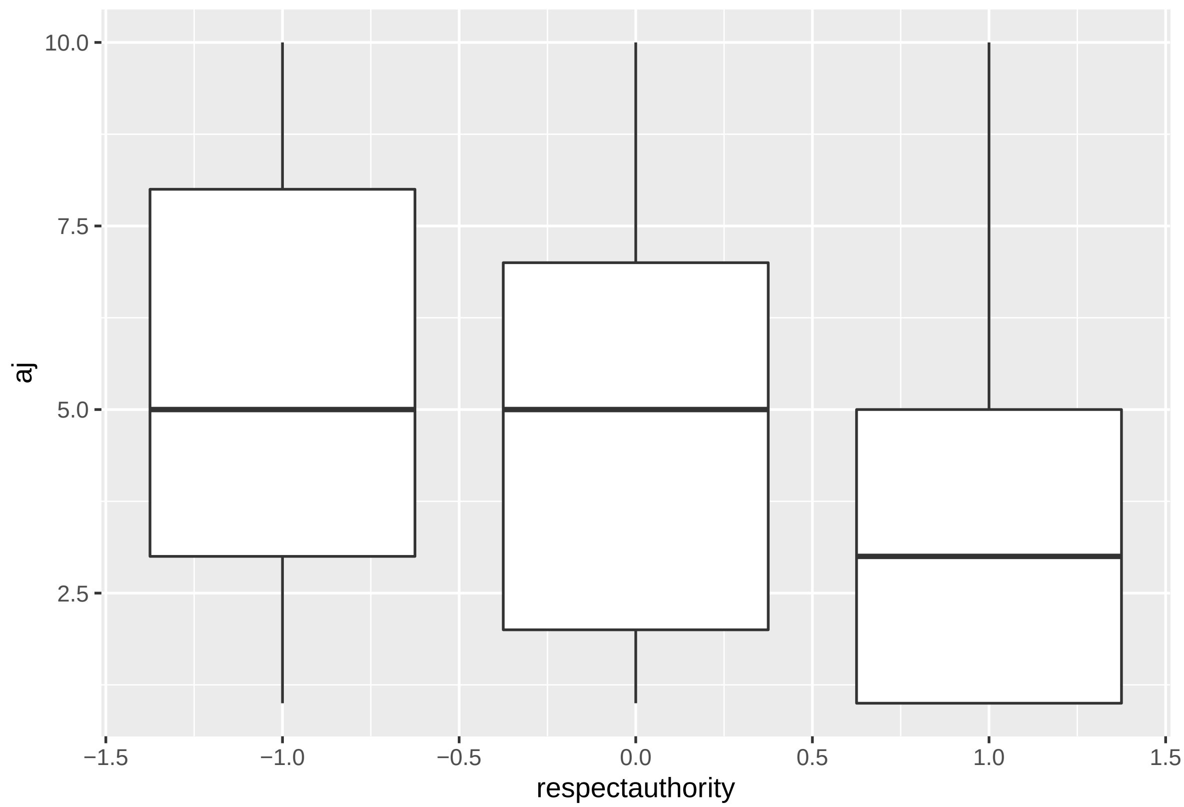

```
ggplot(abortion_data_full, aes(x = godimportant, group = godimportant,
                               y = aj)) +
  geom_boxplot()
```



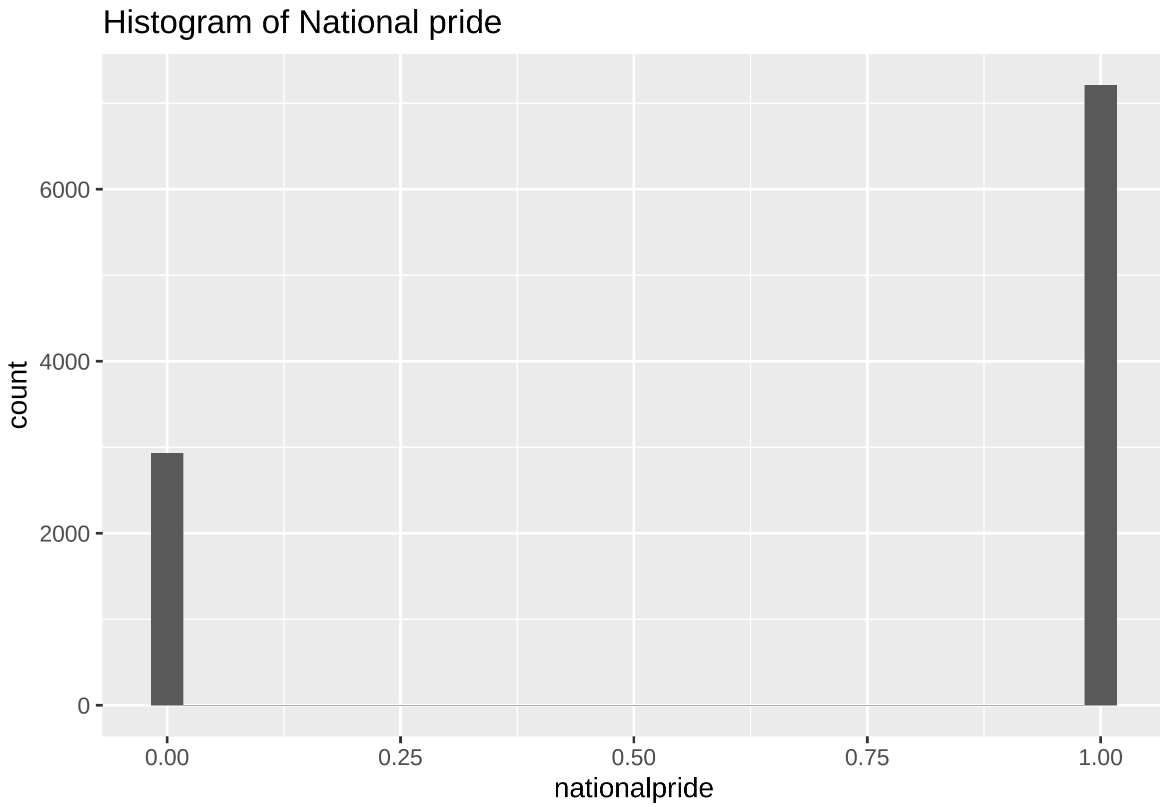

```r
# Respect for authority
ggplot(abortion_data_full, aes(x = respectauthority)) +
  geom_histogram() +
  labs(title = "Histogram of respect for authority")
```



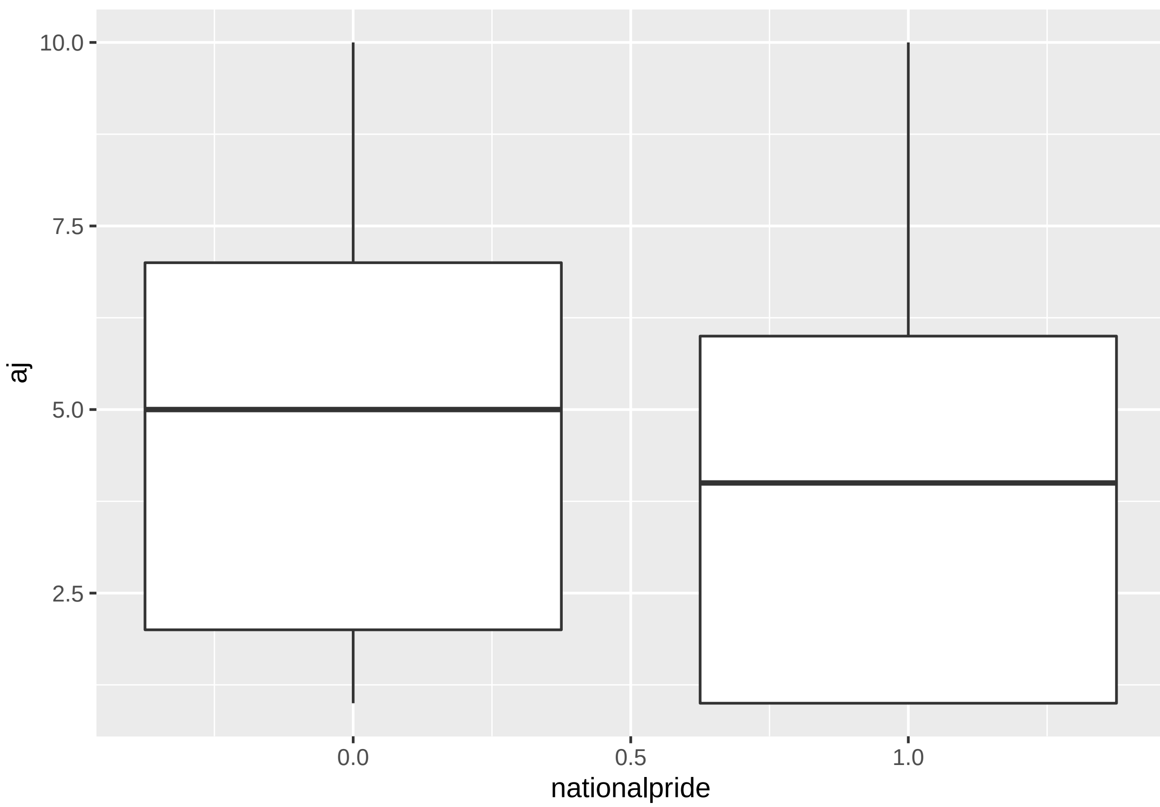

```
ggplot(abortion_data_full, aes(x = respectauthority, group = respectauthority,
                               y = aj)) +
  geom_boxplot()
```




```r
# National Pride
ggplot(abortion_data_full, aes(x = nationalpride)) +
  geom_histogram() +
  labs(title = "Histogram of National pride")
```

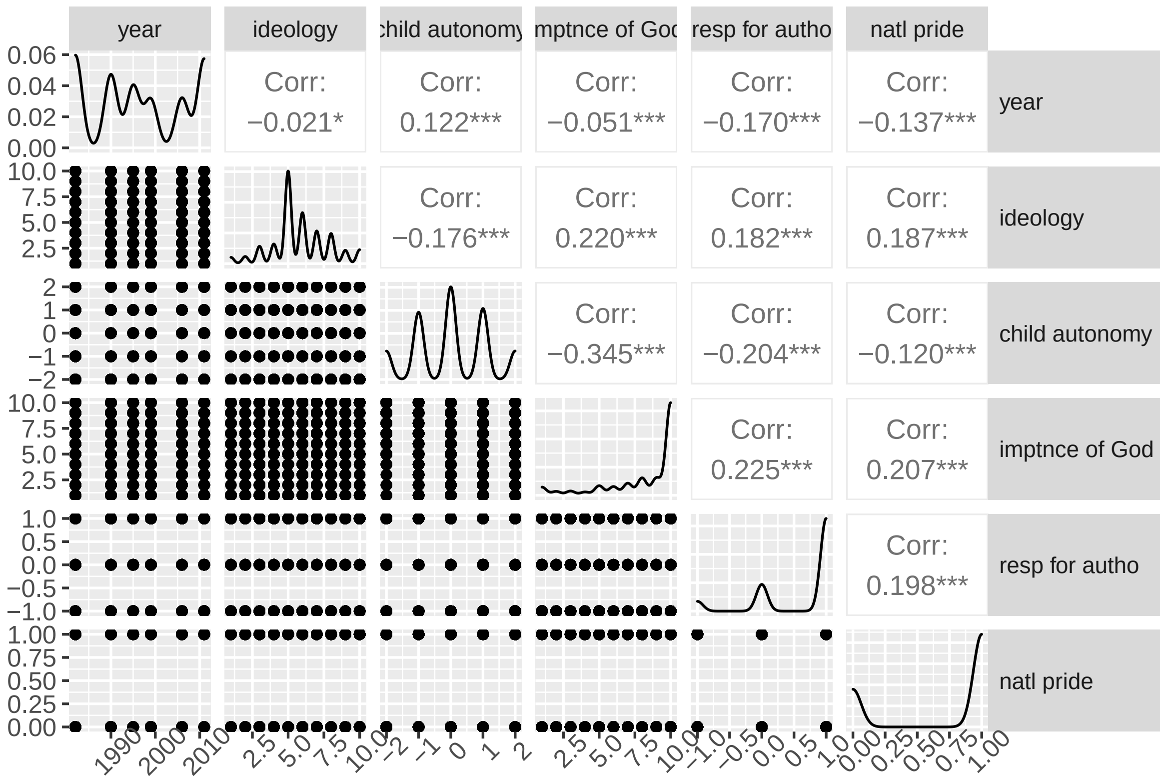



```
ggplot(abortion_data_full, aes(x = nationalpride, group = nationalpride, y = aj)) +
  geom_boxplot()
```



We now test if any of the predictors are strongly correlated with each other.

```
ggpairs(abortion_data_full,
  columns = c("year", "ideology", "cai", "godimportant", "respectauthority",
              "nationalpride"),
  columnLabels = c("year", "ideology", "child autonomy", "imptnce of God",
                   "resp for autho", "natl pride")) +
   theme(
    axis.text.y = element_text(size = 10),
    axis.text.x = element_text(angle = 45, size = 10),
    strip.text.y = element_text(angle = 0, hjust = 0)
    )
```



From these correlation matrices, we can conclude that the highest correlations (above 0.2) are those between `godimportant` and `cai` 0.370 `godimportant` and `respectauthority` 0.244 `godimportant` and `ideology` 0.243 `godimportant` and `nationalpride` 0.216 `nationalpride` and `respectauthority` 0.206 `respectauthority` and `cai` 0.205

METHOD

We will conduct a Multiple Linear Regression (MLR) on the attitudes on the justifiability of abortion against several other predictor variables found in this data set. This is because our outcome variable, the attitude towards abortion, is measured on a numeric scale from 1 to 10, and there are multiple predictor variables that we feel are potentially correlated with it.

We will now discuss the data cleaning and predictor selection process. For observations, we removed the observations that were part of "generational waves" before 1995, because many survey questions were only added from 1995 onward, and observations before 1995 have a high number of null values.

We removed the variable "wvsccode" as it is the same for all the observations. The WVSC Code depends on the country the surveys were conducted in, and all surveys in this data set were conducted in the United States of America.

```
abortion_data <- abortion_data_full %>%
  filter(year >= "1995") %>%
  select(-starts_with("wave"), -starts_with("wvsccode"))
```

We chose to consider predictor variables that, from our exploratory data analysis (EDA), looked like they were potentially correlated with the outcome. Because the predictor variables were discrete, our EDA was presented as a series of box plots. We observed the differences in median and quartile values of abortion attitude across different predictor values and ruled out variables with no visible hint of relationship.

This led to our first set of predictor variables, which we designate as Recipe1:

- Year
- Ideology
- Child Autonomy Index
- Importance of God
- Respect for Authority
- National Pride

We also wanted to choose predictor variables that are not strongly correlated with each other, to avoid multi-collinearity. From our EDA, we found that several predictors had correlations of 0.2 or higher with each other. Given these correlations, we decided to make a second recipe (`Recipe2`) with the following predictor variables, after excluding those with a high correlation with the `godimportant` variable:

- Year
- Importance of God

Finally, we considered that the distribution for the `godimportant` variable is heavily left-skewed in our data set, which we observed in the EDA. This may reduce the ability of a model that relies on the "importance of God" predictor to explain the variations we observe in the outcome. The same can be said of the `respectauthority` variable, with more than 8 times the observations indicating "1" compared to the observations indicating "-1".

Consequently, our third recipe (Recipe3) excludes `godimportant` and uses the other variables that are not correlated with each other more than 0.200:

- Year
- Ideology
- Child Autonomy Index
- National Pride

In the next section, we create and apply our recipes to fit MLR models.

Data Split

We'll now split each of the three "recipes" into a training set (75%) and a testing set (25%)

```
set.seed(206)
abortion_split <- initial_split(abortion_data)
abortion_training <- training(abortion_split)
abortion_testing <- testing(abortion_split)
```

Fit Models

Now that the data is split, we will specify the models and create recipes for each.

```
abortion_spec <- linear_reg() %>%
  set_engine("lm")
```

Now that our model is specified, we will create 3 recipes, each with the corresponding predictors listed above.

The first recipe will be predicting abortion attitudes from `year`, `ideology`, `cai`, `godimportant`, `respectauthority`, and `nationalpride`. We will center the variable `year` so that we can have a meaningful intercept, create dummy variables when needed, and eliminate any zero variance predictors.

```
abortion_rec1 <- recipe(aj ~ year + ideology + cai + godimportant +
                        respectauthority, nationalpride,
                        data = abortion_data) %>%
  step_center(year) %>%
  step_dummy(all_nominal_predictors()) %>%
  step_zv(all_predictors())

abortion_rec1
```

```
Recipe

Inputs:

      role #variables
   outcome          1
 predictor          5

Operations:

Centering for year
Dummy variables from all_nominal_predictors()
Zero variance filter on all_predictors()
```

Our second recipe will be predicting abortion attitudes from just `year` and `godimportant`. We will once again center the variable `year` so that we can have a meaningful intercept, create dummy variables when needed, and eliminate any zero variance predictors.

```
abortion_rec2 <- recipe(aj ~ year + godimportant,
                        data = abortion_data) %>%
  step_center(year) %>%
  step_dummy(all_nominal_predictors()) %>%
  step_zv(all_predictors())

abortion_rec2
```

```
Recipe

Inputs:

      role #variables
   outcome          1
```

```
 predictor          2

Operations:

Centering for year
Dummy variables from all_nominal_predictors()
Zero variance filter on all_predictors()
```

Our third and final recipe will predict abortion attitudes from `year`, `ideology`, `cai`, and `nationalpride`. We will take the same steps from the last 2 recipes.

```r
abortion_rec3 <- recipe(aj ~ year + ideology + cai + nationalpride,
                        data = abortion_data) %>%
  step_center(year) %>%
  step_dummy(all_nominal_predictors()) %>%
  step_zv(all_predictors())

abortion_rec3
```

```
Recipe

Inputs:

      role #variables
   outcome          1
 predictor          4

Operations:

Centering for year
Dummy variables from all_nominal_predictors()
Zero variance filter on all_predictors()
```

With the recipes made, we are ready to create our workflows.

```r
abortion_wflow1 <- workflow() %>%
  add_model(abortion_spec) %>%
  add_recipe(abortion_rec1)

abortion_wflow2 <- workflow() %>%
  add_model(abortion_spec) %>%
```

```
  add_recipe(abortion_rec2)

abortion_wflow3 <- workflow() %>%
  add_model(abortion_spec) %>%
  add_recipe(abortion_rec3)
```

And finally, we will fit each of the models above

```
abortion_fit1 <- abortion_wflow1 %>%
  fit(data = abortion_training)
tidy(abortion_fit1) %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	8.500	0.162	52.393	0.000
year	0.029	0.006	4.754	0.000
ideology	-0.284	0.021	-13.741	0.000
cai	0.469	0.037	12.766	0.000
godimportant	-0.288	0.016	-18.224	0.000
respectauthority	-0.190	0.068	-2.795	0.005

```
abortion_fit2 <- abortion_wflow2 %>%
  fit(data = abortion_training)
tidy(abortion_fit2) %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	7.914	0.125	63.183	0
year	0.035	0.006	5.782	0
godimportant	-0.429	0.015	-29.349	0

```
abortion_fit3 <- abortion_wflow3 %>%
  fit(data = abortion_training)
tidy(abortion_fit3) %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	6.653	0.133	49.865	0
year	0.034	0.006	5.466	0
ideology	-0.348	0.021	-16.386	0
cai	0.704	0.036	19.641	0
nationalpride	-0.402	0.090	-4.488	0

Model Comparison and Selection

Now that we have all of our models fit and displayed, we need to interpret the results and determine the best model.

Let's start by seeing how well our models predict the abortion attitudes

```
abortion_training_pred1 <- predict(abortion_fit1, abortion_training) %>%
  bind_cols(abortion_training %>% select(aj, year, ideology, cai,
                                         godimportant, respectauthority,
                                         nationalpride))
abortion_training_pred1
```

```
# A tibble: 4,667 x 8
   .pred    aj  year ideology   cai godimportant respectauthority nationalpride
   <dbl> <dbl> <dbl>    <dbl> <dbl>        <dbl>            <dbl>         <dbl>
 1  5.46     9  2011        9     2            5                1             1
 2  5.01     5  1999        4     1           10               -1             1
 3 NA        3  2006       NA     0            9                1             0
 4  6.03     5  2011        5     1            6                0             0
 5  3.87     3  1995        6    -1            7                1             1
 6  4.41     1  2011        5     0           10                0             1
 7  3.28     1  2011        5    -2           10                1             1
 8  5.56     5  2011        4     0            7                0             1
 9  6.36     1  1999        5     1            3                1             0
10  6.15    10  1995        5     1            4                0             1
# ... with 4,657 more rows
```

```
abortion_training_pred2 <- predict(abortion_fit2, abortion_training) %>%
  bind_cols(abortion_training %>% select(aj, year, godimportant))
abortion_training_pred2
```

```
# A tibble: 4,667 x 4
   .pred    aj  year godimportant
   <dbl> <dbl> <dbl>        <dbl>
 1  6.03     9  2011            5
 2  3.46     5  1999           10
 3  4.14     3  2006            9
 4  5.60     5  2011            6
 5  4.61     3  1995            7
 6  3.88     1  2011           10
 7  3.88     1  2011           10
 8  5.17     5  2011            7
 9  6.46     1  1999            3
10  5.89    10  1995            4
# ... with 4,657 more rows
```

```
abortion_training_pred3 <- predict(abortion_fit3, abortion_training) %>%
  bind_cols(abortion_training %>% select(aj, year, ideology, cai,
                                         nationalpride))
abortion_training_pred3
```

```
# A tibble: 4,667 x 6
   .pred    aj  year ideology   cai nationalpride
   <dbl> <dbl> <dbl>    <dbl> <dbl>         <dbl>
 1  4.78     9  2011        9     2             1
 2  5.40     5  1999        4     1             1
 3 NA        3  2006       NA     0             0
 4  5.87     5  2011        5     1             0
 5  3.16     3  1995        6    -1             1
 6  4.76     1  2011        5     0             1
 7  3.35     1  2011        5    -2             1
 8  5.11     5  2011        4     0             1
 9  5.46     1  1999        5     1             0
10  4.92    10  1995        5     1             1
# ... with 4,657 more rows
```

A quick glance at our data might seem that the models are not that strong at predicting the abortion attitude (aj) of a given observation. We will look into this further now with cross-validation, then a test in AIC and BIC statistics.

```
set.seed(206)
folds <- vfold_cv(abortion_training, v = 10)
```

```
abortion_fit_rs1 <- abortion_wflow1 %>%
  fit_resamples(folds)

abortion_fit_rs2 <- abortion_wflow2 %>%
  fit_resamples(folds)

abortion_fit_rs3 <- abortion_wflow3 %>%
  fit_resamples(folds)
```

```
cv_metrics1 <- collect_metrics(abortion_fit_rs1, summarize = FALSE)
cv_metrics2 <- collect_metrics(abortion_fit_rs2, summarize = FALSE)
cv_metrics3 <- collect_metrics(abortion_fit_rs3, summarize = FALSE)

cv_metrics1 %>%
  mutate(.estimate = round(.estimate, 3)) %>%
  pivot_wider(id_cols = id, names_from = .metric, values_from = .estimate) %>%
  kable(col.names = c("Fold", "RMSE", "R-squared"))
```

Fold	RMSE	R-squared
Fold01	2.606	0.213
Fold02	2.490	0.236
Fold03	2.575	0.263
Fold04	2.694	0.182
Fold05	2.428	0.274
Fold06	2.526	0.269
Fold07	2.500	0.288
Fold08	2.513	0.280
Fold09	2.391	0.261
Fold10	2.579	0.196

```
cv_metrics2 %>%
  mutate(.estimate = round(.estimate, 3)) %>%
  pivot_wider(id_cols = id, names_from = .metric, values_from = .estimate) %>%
  kable(col.names = c("Fold", "RMSE", "R-squared"))
```

Fold	RMSE	R-squared
Fold01	2.674	0.181
Fold02	2.634	0.146
Fold03	2.667	0.202

Fold	RMSE	R-squared
Fold04	2.757	0.139
Fold05	2.591	0.192
Fold06	2.687	0.179
Fold07	2.640	0.191
Fold08	2.611	0.227
Fold09	2.573	0.155
Fold10	2.705	0.115

```
cv_metrics3 %>%
  mutate(.estimate = round(.estimate, 3)) %>%
  pivot_wider(id_cols = id, names_from = .metric, values_from = .estimate) %>%
  kable(col.names = c("Fold", "RMSE", "R-squared"))
```

Fold	RMSE	R-squared
Fold01	2.739	0.132
Fold02	2.579	0.182
Fold03	2.681	0.197
Fold04	2.748	0.142
Fold05	2.510	0.230
Fold06	2.629	0.207
Fold07	2.572	0.250
Fold08	2.708	0.159
Fold09	2.483	0.202
Fold10	2.630	0.164

As with the predictions from before, the RMSE and R-squared values for each of these models do not show promising results. Of the three models, the first is seemingly the strongest model as it has the closest R-squared values to 1. We will continue our analysis still with AIC and BIC analysis.

```
glance(abortion_fit1) %>%
  select(AIC, BIC)
```

```
# A tibble: 1 x 2
      AIC    BIC
    <dbl>  <dbl>
1 19980. 20025.
```

```r
glance(abortion_fit2) %>%
  select(AIC, BIC)
```

```
# A tibble: 1 x 2
      AIC    BIC
    <dbl>  <dbl>
1 21492. 21517.
```

```r
glance(abortion_fit3) %>%
  select(AIC, BIC)
```

```
# A tibble: 1 x 2
      AIC    BIC
    <dbl>  <dbl>
1 20330. 20369.
```

From the above AIC and BIC values, we have further evidence that model 1 is the strongest model of the three. Model 1 has the lowest AIC and BIC values, which are usually penalized for having more predictors, but in this case the model is strong enough to overcome that penalty.

Inference: Confidence Interval and Hypothesis Test

Before we discuss results, we will take a quick look at the resulting confidence intervals from our fitted models and the resulting coefficients.

```r
tidy(abortion_fit1, conf.int = TRUE) %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	8.500	0.162	52.393	0.000	8.182	8.818
year	0.029	0.006	4.754	0.000	0.017	0.041
ideology	-0.284	0.021	-13.741	0.000	-0.325	-0.244
cai	0.469	0.037	12.766	0.000	0.397	0.541
godimportant	-0.288	0.016	-18.224	0.000	-0.319	-0.257
respectauthority	-0.190	0.068	-2.795	0.005	-0.323	-0.057

```r
tidy(abortion_fit2, conf.int = TRUE) %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	7.914	0.125	63.183	0	7.669	8.160
year	0.035	0.006	5.782	0	0.023	0.047
godimportant	-0.429	0.015	-29.349	0	-0.457	-0.400

```r
tidy(abortion_fit3, conf.int = TRUE) %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	6.653	0.133	49.865	0	6.391	6.914
year	0.034	0.006	5.466	0	0.022	0.047
ideology	-0.348	0.021	-16.386	0	-0.390	-0.306
cai	0.704	0.036	19.641	0	0.634	0.775
nationalpride	-0.402	0.090	-4.488	0	-0.578	-0.227

Results

From the analysis of the models through multiple different calculations, we know that of our three models, model 1 is the strongest, predicting abortion attitude from `year`, `ideology`, `cai`, `godimportant`, `respectauthority`, and `nationalpride`. Given that model 1 is our strongest model, it must also be noted that all 3 of our models had difficulty predicting an accurate attitude towards abortion of a given observation. This is most likely due to the data having a large number of very different observations given the same `aj` score. In model 1, however, our p-values for the coefficients given are all at an acceptable range except for that of `nationalpride`. This means that every other predictor was able to be helpful in predicting the response variable.

Discussion + Conclusion

Ultimately, our investigation in the correlation between attitudes on the justifiability of abortion and demographic factors + personal attitudes towards other issues presented some problems. Through our model creation and comparison, we are able to determine that the best combination of predictors for abortion attitudes are `year`, `ideology`, `cai`, `godimportant`, `respectauthority`, and `nationalpride`. While every one of these predictors are shown to be helpful in predicting the response variable, the model itself is generally weak although it is the best of our 3.

Unfortunately, given our data set, we are limited by the truthfulness and accuracy of each observation. While we can assume each observation is answered honestly and thoughtfully, there was the option to leave some answers empty, and some predictors are quite hard to place nicely into a 1-10 scale (take ideology for example). Steps that could be taken to counter this problem are to either categorize some of these 1-10 scales into even smaller categories (like 3/4) to possibly show better trends, or even more categories (like 1-50 scale) to increase accuracy of responses and also possibly reveal more trends.

With abortion being such a divisive issue in our society, it is incredibly difficult to pinpoint just a few predictors to try and accurately tell how any one person might feel about abortion. Knowing that, however, we do not expect that abortion attitudes are impossible to more accurately predict. We have shown through our modeling and analysis that some of the predictors presented are helpful in predicting our response variable. Given this, it is not unreasonable to assume that there might be even more helpful predictors to use in place, or in addition to the ones we have selected. For future work, we might be able to produce stronger models given additional demographic/personality predictors to work with.

Data dictionary

The data dictionary can be found here.

References

Brenan, M. (2020, July 7). One in Four Americans Consider Abortion a Key Voting Issue. Gallup. https://news.gallup.com/poll/313316/one-four-americans-consider-abortion-key-voting-issue.aspx

Weinberger, J. (2022, May 6). How we got here: Roe v. Wade from 1973 to today. Vox. https://www.vox.com/23055389/roe-v-wade-timeline-abortion-overturn-political-polarization

Ziegler, M. (2020, October 22). Abortion politics polarized before Roe. When it's gone, the fighting won't stop. The Washington Post. https://www.washingtonpost.com/outlook/2020/10/22/roe-polarize-abortion-politics/