Draft

STA 210 - Project

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

Introduction and Research Question

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 may be easier 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. Both pro-choice and pro-life groups have a history of mobilizing 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 which protected an individual's right to have an abortion before fetal viability, 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 legalizing and increasing access to abortion (Weinberger 2022). Abortion has also increasingly become a key issue that voters consider when making their choice at the ballot box, with an increasing share of Americans identifying as "single-issue voters" regarding abortion (Brenan 2020).

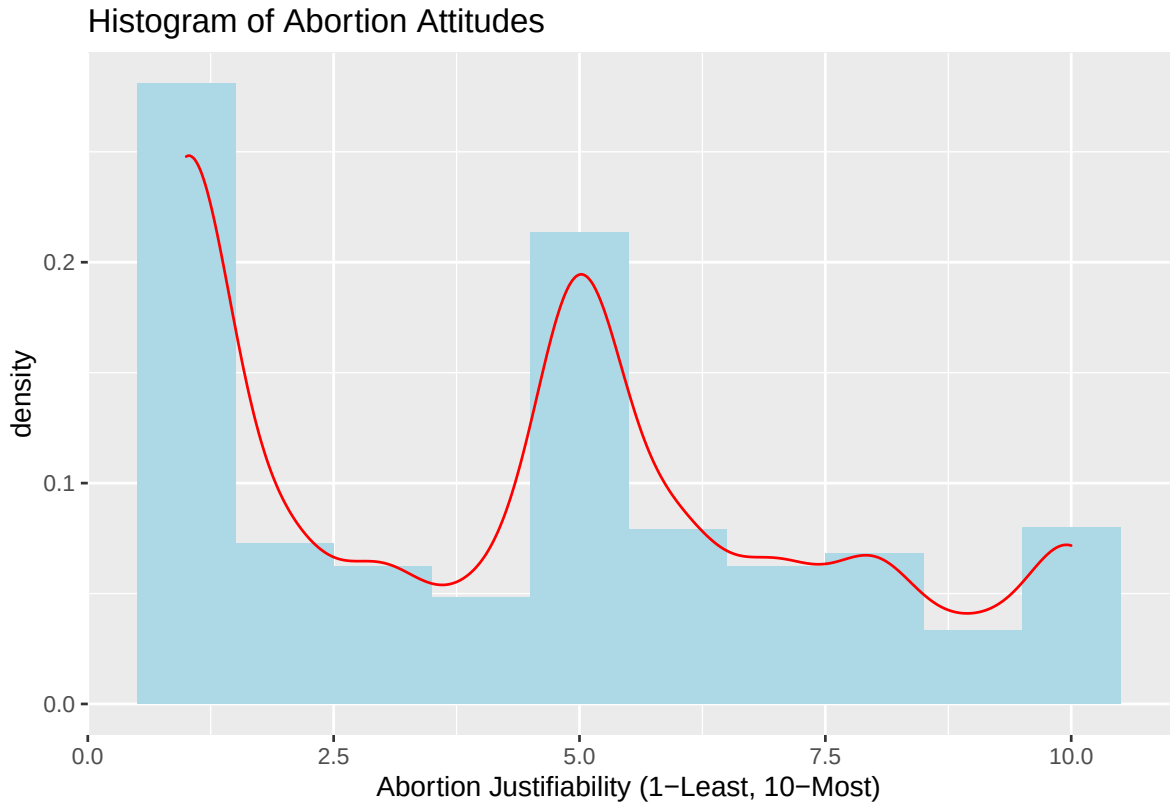
In May 2022, a leaked draft opinion revealed that the US Supreme Court is prepared to overturn Roe v. Wade. Overturning Roe would dramatically change the trajectory of abortion politics in the US. Unless Congress passes a national policy, states would be able to decide whether or not to legalize abortion 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 which 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.

The dataset used in this analysis observes attitudes on the justifiability of abortion among respondents in in the United States across six "waves" of the World Values Survey (1982-2011), which is administered every few years and collects information about people's values and beliefs worldwide, alongside basic demographic characteristics. The data are collected via face-to-face interviews at the respondents' homes. In this data set, 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.

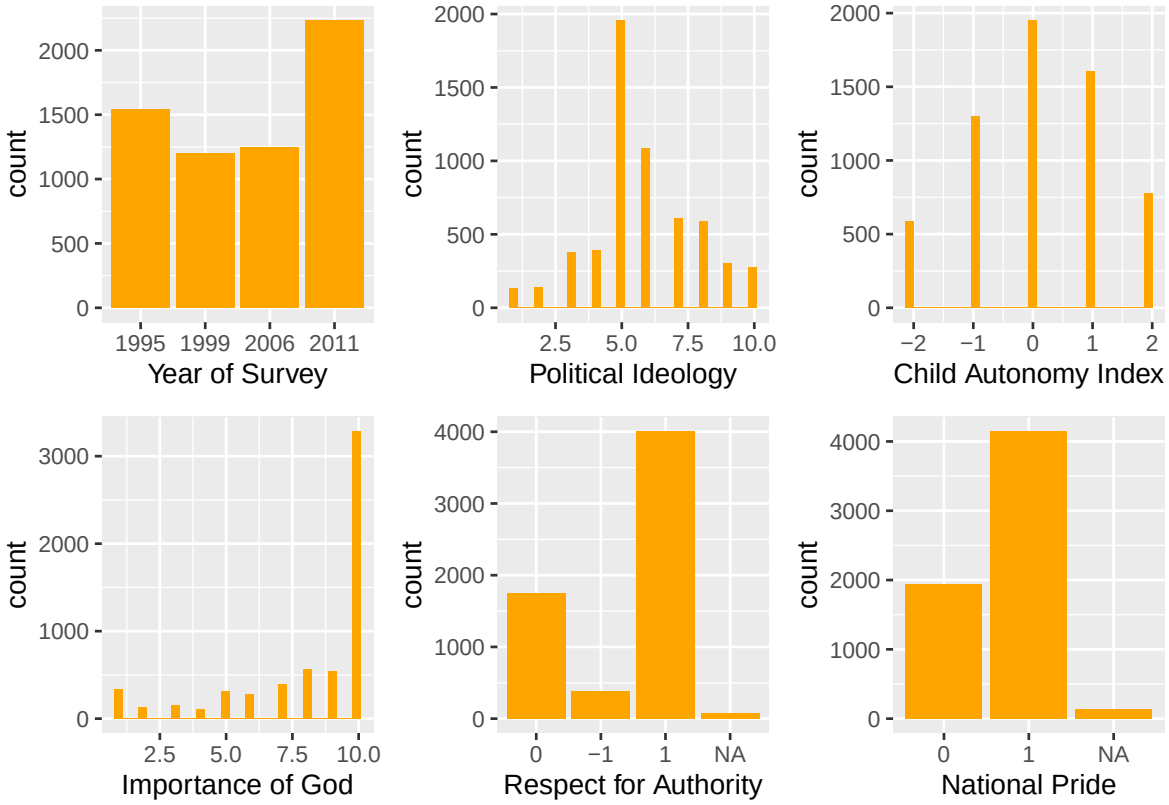
Our research question is as follows: "Do an individual's political ideology and their personal attitudes/preferences towards other issues, such as the importance of religion in their life and their respect for authority, among others, predict their attitude on the justifiability of abortion?" We will attempt to answer this question using an EDA-informed predictive model. Given that this issue has become highly polarized by political party, we predict that liberal ideology and liberal-leaning attitudes on other issues will correlate with belief that abortion is more justified.

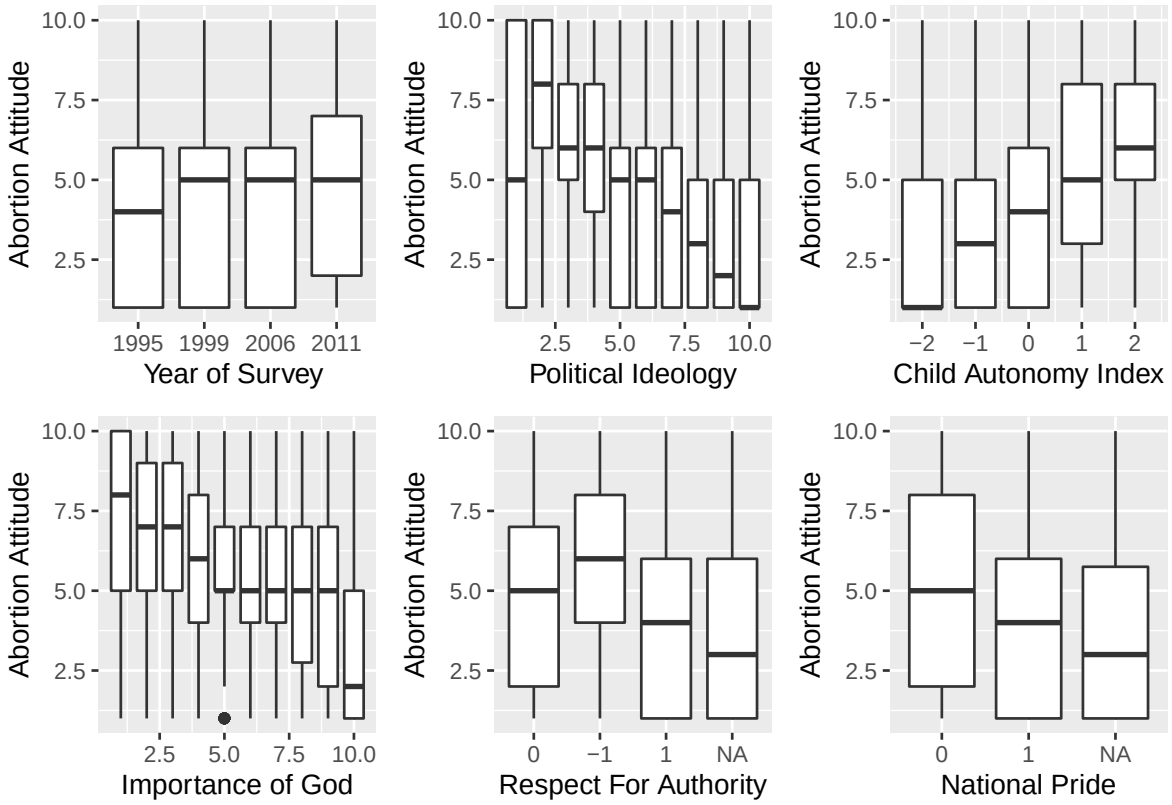
Exploratory Data Analysis

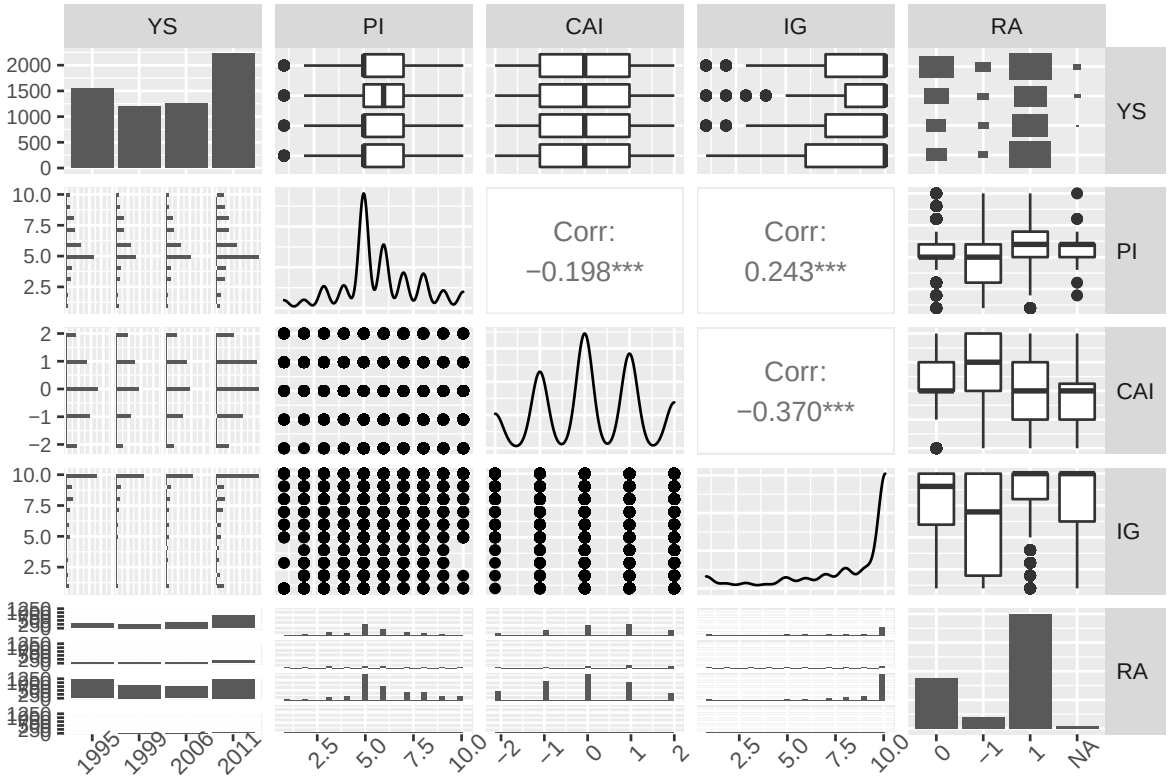


From the above visualization of the distribution of attitudes toward the justifiability of abortion, we observe that it is not a bell shape, but rather trimodal. This may be because the question's phrasing is similar to a yes/no question, but respondents were asked to give their level of agreement on a scale of 1-10. This may worsen the ability of a multiple linear regression (MLR) model to fit the data. Consequently, further research may choose to truncate the attitude on the justifiability of abortion into a categorical variable such as (Agree, Disagree, Undecided), conducting a binomial or multinomial logistic regression thereafter.

We now extend our exploratory data analysis (EDA) to the predictor variables of interest: year of survey, ideology, Child Autonomy Index, Importance of God, Respect for Authority, and National Pride. The EDA for each variable comprises a histogram and a boxplot of the response variable, grouped by value of the predictor. We also provide a correlation matrix to detect any multicollinearity between our predictor variables, which would increase the uncertainty of our model's parameters.




Abortion Attitude
10.0
7.5
5.0
2.5
1995 1999 2006 2011
Year of Survey
2.5 5.0 7.5 10.0
Political Ideology
−2 −1 0 1 2
Child Autonomy Index
2.5 5.0 7.5 10.0
Importance of God
0 −1 1 NA
Respect For Authority
0 1 NA
National Pride


YS
PI
CAI
IG
RA
Corr:
−0.198***
Corr:
0.243***
Corr:
−0.370***
1995
1999
2006
2011
NA

Methodology

Given that our response variable, attitude towards the justifiability of abortion, is measured on a numeric scale from 1 to 10, and that there are multiple predictor variables being tested against the null hypothesis, we will conduct multiple linear regression (MLR) to model the effect of these predictors on the variation in abortion attitudes.

We chose to consider the predictor variables our exploratory data analysis (EDA) suggested were potentially correlated with the response variable. 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, ruling out variables with no significant influence on the outcome. This led to the following predictor variables: Year of survey, Political Ideology, Child Autonomy Index, Importance of God, Respect for Authority, and National Pride.

*A note on data cleaning: Many survey questions were only added from 1995 onward, so observations before 1995 have a high number of missing values. For these reasons, observations prior to 1995 were removed.

Year of survey, Respect for Authority, and National Pride were changed into factors since they are categorical variables. Year has 4 levels (one corresponding to each year), Respect for Authority has 3 levels (0 being the baseline - neutral, -1 being respect for authority bad, and 1 being respect for authority good), and National Pride is a binary (0 being the baseline - having no feelings of national pride, and 1 being having feelings of national pride) Political Ideology, Child Autonomy Index, and Importance of God had numerical sample spaces and were therefore treated as numerical variables.*

Fitting the Model

We conducted a 75% - 25% data split into training and testing sets, using the random seed 206.

```
Warning: There are new levels in a factor: NA
There are new levels in a factor: NA
There are new levels in a factor: NA
```

Fitting the MLR, we obtain the following table of parameter point estimates.

term	estimate	std.error	statistic	p.value
(Intercept)		6.596	0.170	38.688	0.000
ideology		-0.280	0.021	-13.354	0.000
cai		0.468	0.037	12.622	0.000