Politics Dataset

Sociology 312, University of Oregon
Prof. Gullickson

Overview

This data comes from the 2012 American National Election Study (ANES). The ANES is a survey of the American electorate that is conducted every two years. The study collects information on a variety of political attitudes and voting behaviors. For our purposes, we are going to primarily look at respondent's attitudes on three issues: (1) immigration, (2) gay marriage, and (3) global warming.

Unlike our other datasets, some of the variables originally include significant number of missing values. In most cases, this is because the respondent refused to answer the question or didn't understand the question enough to feel comfortable answering the question. To simplify the analysis here, I have used some advanced techniques to impute valid values for all of the missing valuese.

Table 1 provides summary statistics for all the quantitative variables in the politics data.

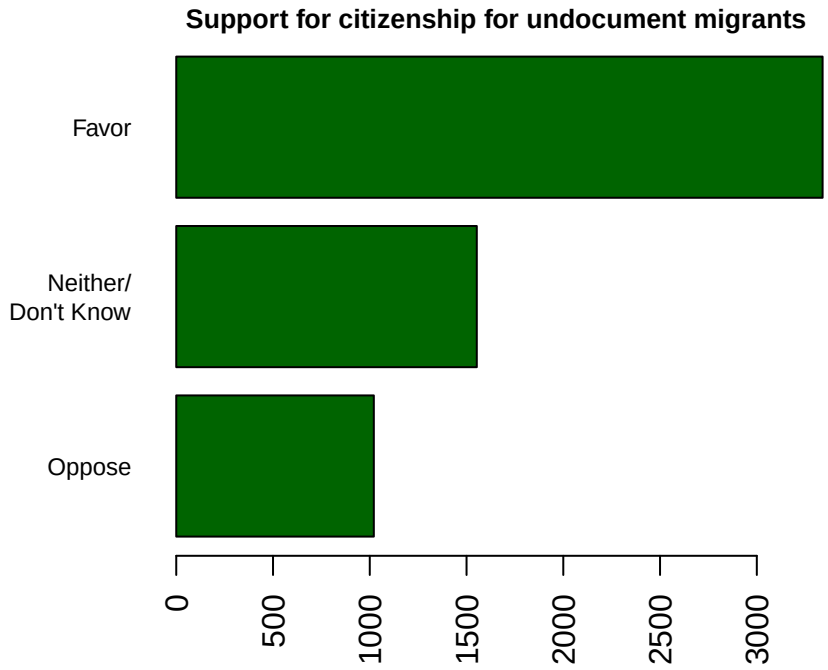
Table 1: Summary statistics for all quantitative variables in the politics dataset

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
age	5914	49.441	16.823	17	35	51	62	90
income	5914	58.704	54.148	2.500	21.250	42.500	85.000	250.000

Variable Descriptions

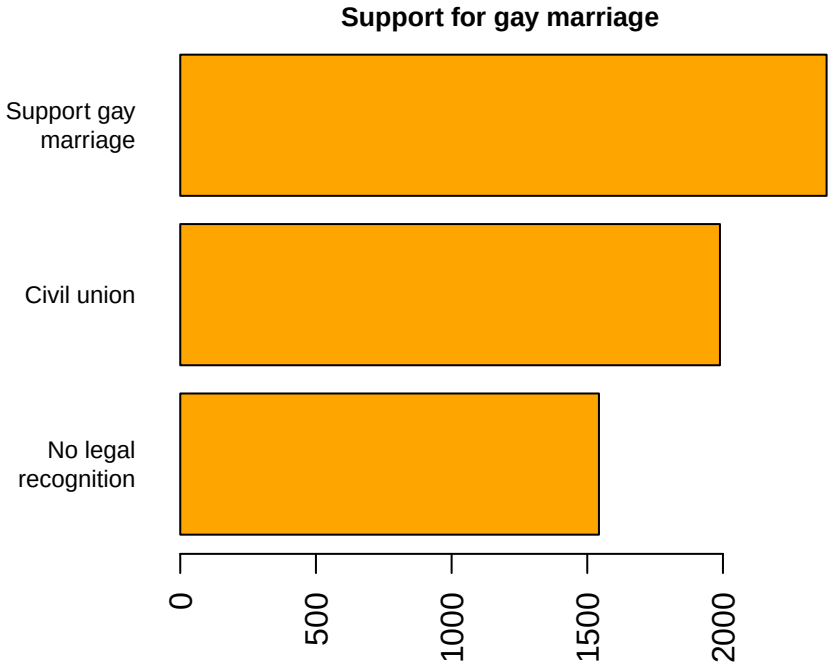
immigcitizen

Respondents were asked whether they would support a proposal to grant citizenship to some (not necessarily all) undocumented immigrants. Respondents could either favor, oppose, or do neither. There are 10 missing values for this response.



gaymarriage

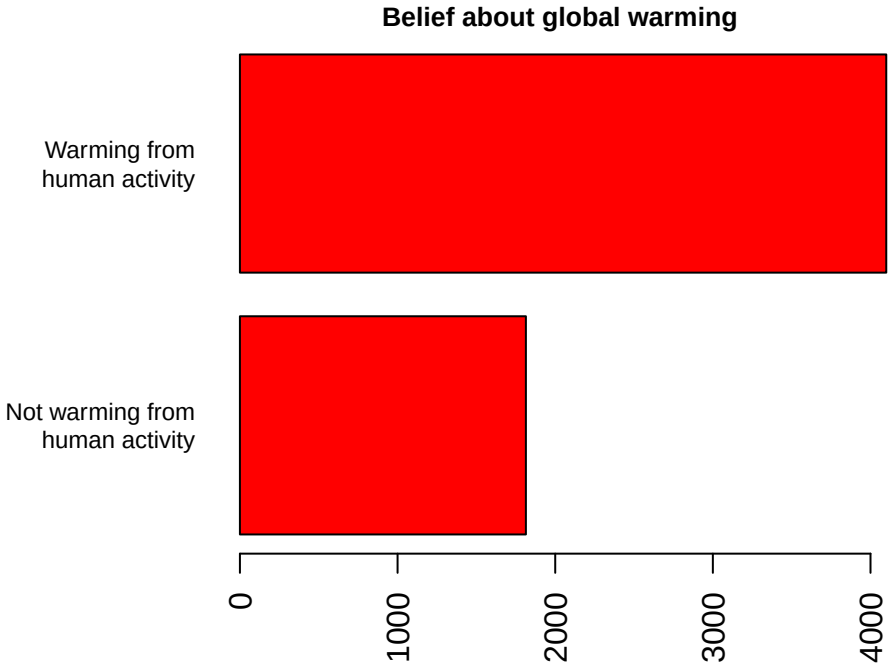
Respondents were asked for their position on gay marriage and were given the choices of "no legal recognition", "civil union (but no marriage)", "support gay marriage." There were 80 missing values.



globalwarm

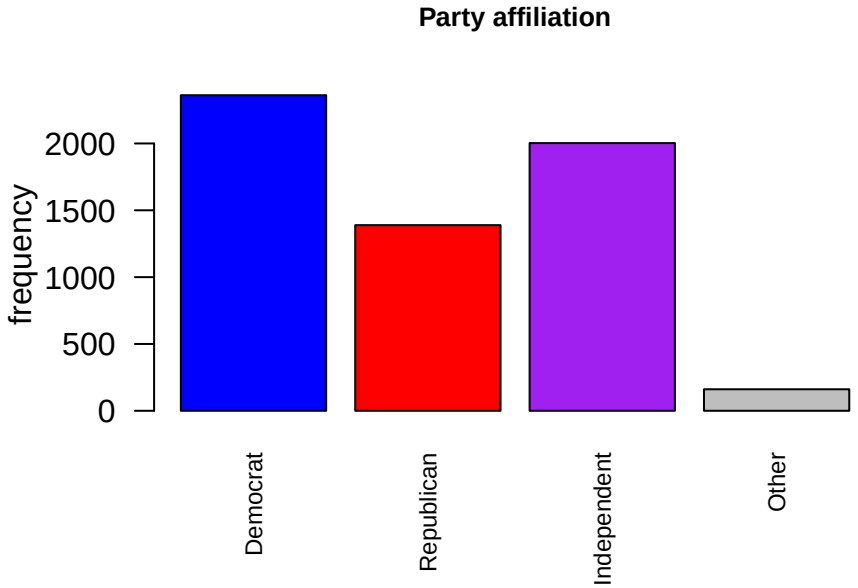
I constructed this variable from two separate questions. The first question asks whether respondents think that global warming has been happening with the options being that it "probably has" or "probably has

not." The second question asks whether respondents thought that global warming was caused by human activity (either entirely or partially). I combine these into a single dichotomous variable where individuals either think the earth is warming from human activity or that it is not warming from human activity, where the latter category includes people who think it isn't warming at all and people who think it is warming but not because of human activity. There are 142 missing values for this variable.



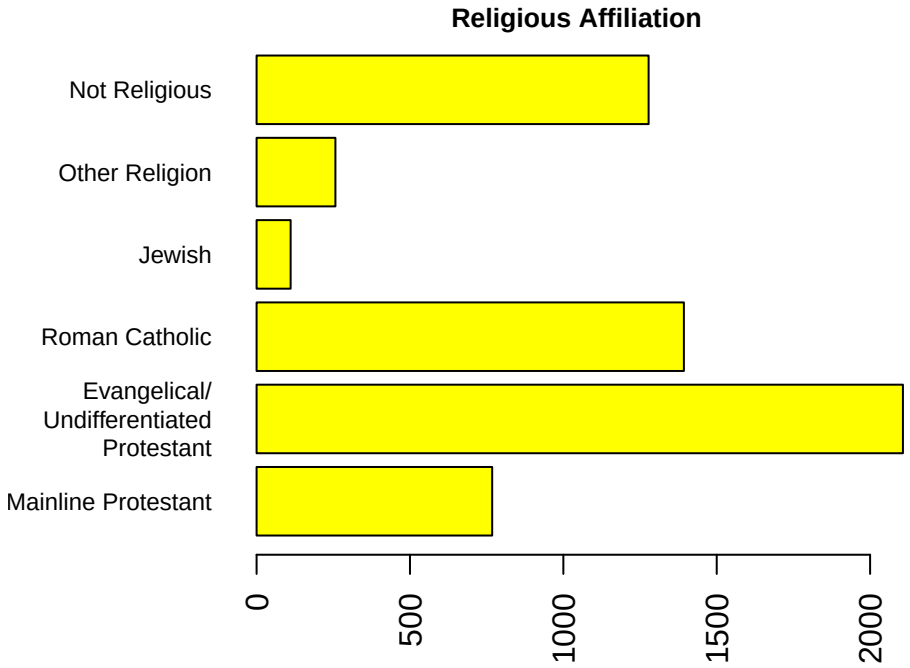
party

The political party with which the respondent identifies. The original data include a category for "no preference" that I have collapsed with the category of "independent."



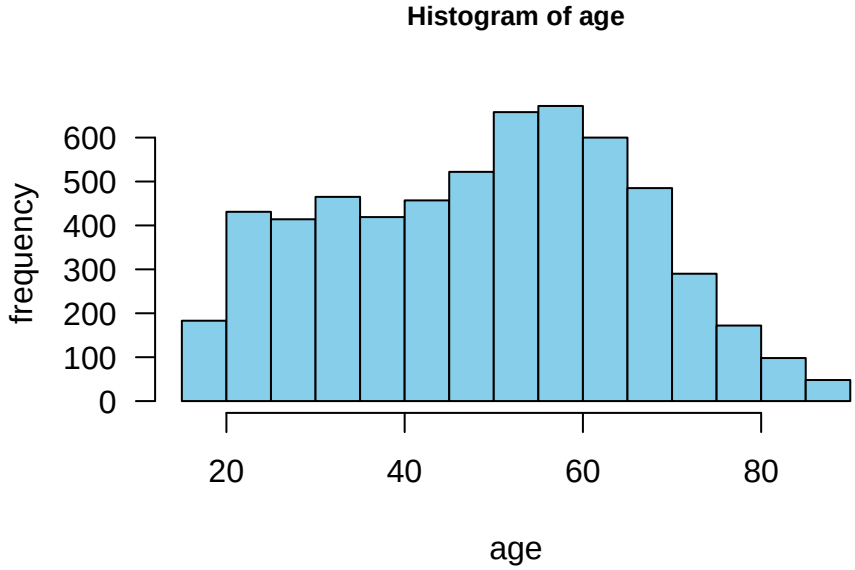
relig

The respondent's religion. To keep things simpler here, I have combined the category of evangelicial and "undifferentiated" protestant. The latter category includes protestants who are so-called "non-denominational." I have also collapsed black protestant denomination with mainline protestants. Thereare 86 missing values for this variable.



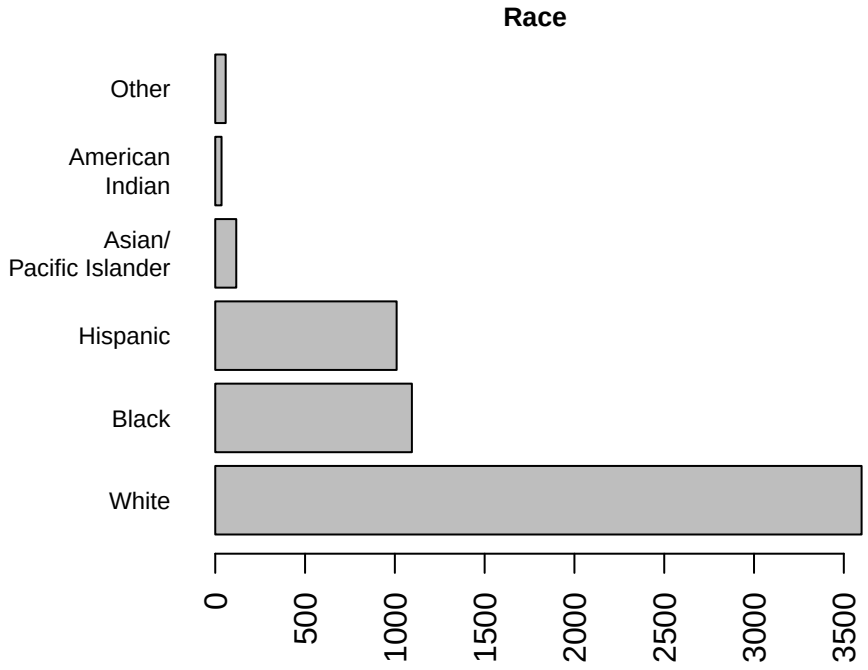
age

The age of the respondent. The ANES is focused on the voting age population. The youngest interviewee was 17 years old. Presumably this person was interviewed because they would turn 18 by the time of the actual election. There are 60 missing values for this variable.



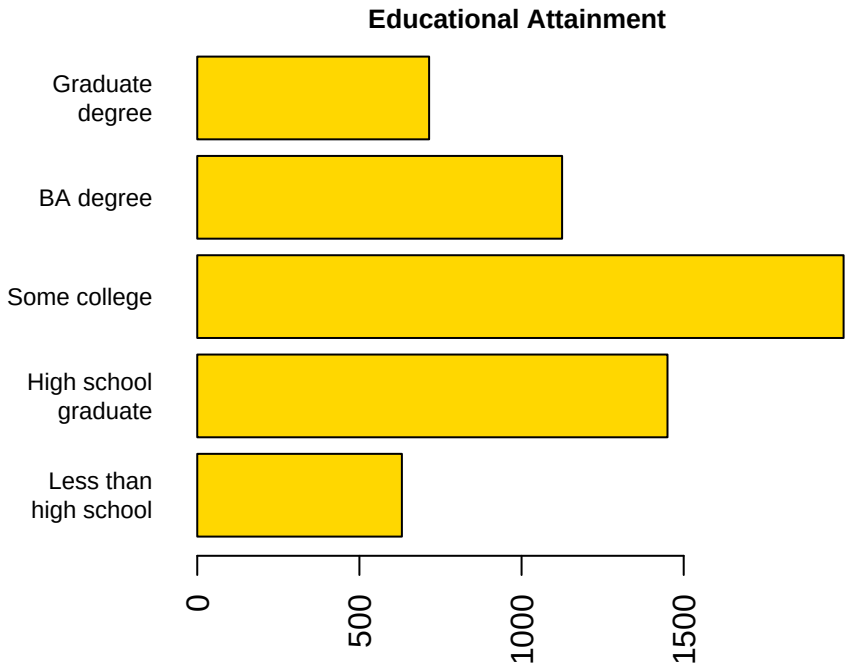
race

The race of the respondent. This is always a tricky category to code. The ANES uses a relatively simple and crude set of six categories and does not allow a respondent to check multiple races. There are 24 missing values on this variable.



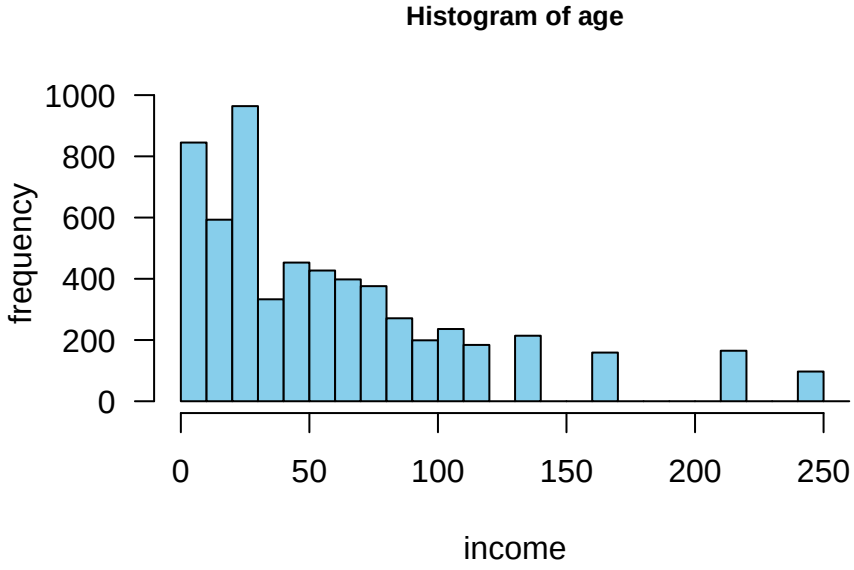
educ

The respondent's level of education. This is recorded as an ordinal variable. The "Some college" response indicates individuals who have attended college (including 2-year programs) but have not earned a BA. There are 50 missing values.



income

The family income of the respondentin 1000s of dollars. Respondents did not give actual dollar amounts here but rather indicated which bracket of income (e.g. $20,000-30,000) they fell within. I am treating this like a continuous variable by assigning each respondent the midpoint of the bracket as their actual income (e.g. $20,000-30,000 becomes $25,000). The final category of $250,000 or more is treated as $250,000. There are 520 missing values for income.



workstatus

The work status of the respondents. Respondents could either be working, unemployed, or out of the labor force. The last category refers to people who are not employed and not currently looking for work, whereas

unemployed indicates a person who is not employed an is currently looking for work. There are 13 missing values.