Predicting Voter Turnout

Jake Wislek

VALPARAISO UNIVERSITY

Introduction

As a student studying Political Science alongside Statistics at Valparaiso University, I became incredibly interested in political behavior, especially the reasons people vote. Upon further research, I discovered what political scientist Richard Timpone describes as "hurdles" to the voting registration process in his article *Structure, Behavior, and Voter Turnout in the United States.* Timpone attributes low voter turnout rates to stricter rules around voter registration. This project explores other possible "hurdles" people must overcome to vote and determines if they have a significant impact on voter turnout.

To determine voter turnout, the Voter Eligible Population (VEP) will be used. VEP measures only people who can vote based on their state's rules around voting. For instance, some convicted felons cannot vote in certain states, so they would not be considered in the VEP. This allows me to control for deterministic situations that hinder voting and consider just people's willingness to vote. The VEP was provided by the United States Election Project, an organization that has been tracking VEP in each state and D.C. since 1980.

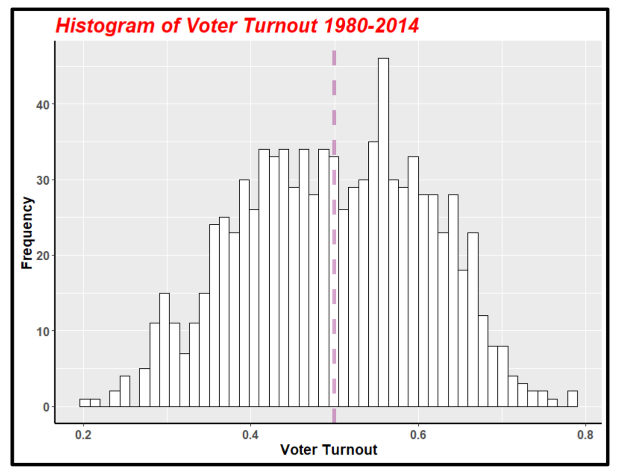
For this project, I considered the following variables for predicting voter turnout: Population, Property Crime Rates, Violent Crime Rates, Number of Federal Representatives a State has, Proportion of the House of Representatives a State Represents, Voter I.D. Laws, Whether or not there's a Presidential Election that year, the Party of the Governor, the Party of the President, the Majority Party in the House of Representatives, Whether or not the Governor's Party is the Same as the President's Party, and Whether or not the Governor's Party is the Same as the Majority Party in the House of Representatives.

Data Collection

Data was collected from a variety of sources. The Population came from the United States Election Project, the Property and Violent Crime Rates came from the U.S. Department of Justice, and the Number of Federal Representatives a State has and the Proportion of the House of Representatives a State Represents came from 270toWin.com. The Voter I.D. Laws were provided by the National Conference of State Legislators and Ballotpedia, and the rest were provided by lists on Wikipedia.

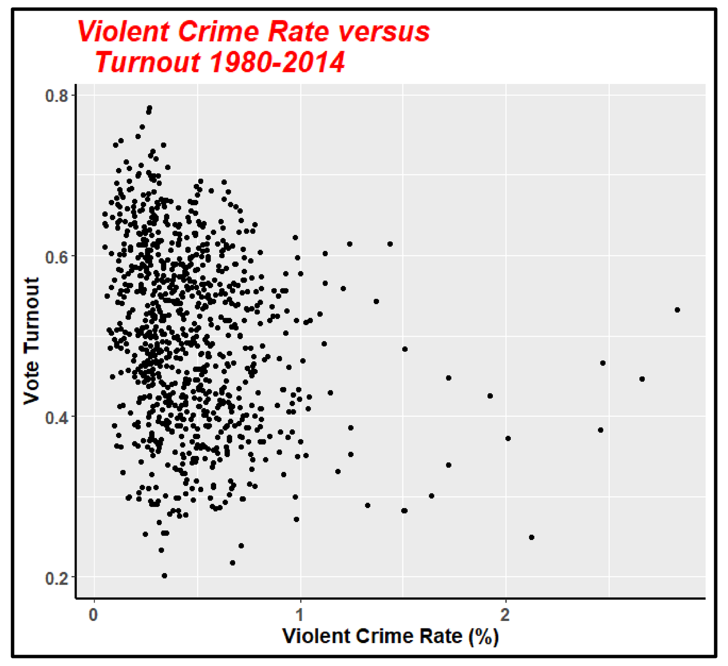
I merged all of the data from the various sources into a single Excel spreadsheet. I performed data cleaning to recode the Voter I.D. Laws into a ranked variable, with 1 indicating most strict and 5 indicating most lenient, and I recoded the other categorial variables into appropriate dummy variables. The statistical software R was used for the data analysis. The dates ranged from 1980 to 2014, with data recorded only for federal elections in even-numbered years. Since we have information for 50 states plus D.C. over 18 elections, the sample size is:
n = 51*18 =918.

Distribution of Voter Turnout



The histogram visualizing Voter Turnout between 1980 and 2014 shows that the data is approximately normal. There appears to be little to no skewness, and the median and mean are both close at 0.5010 and 0.4996, respectively. The minimum for voter turnout is 0.2020 and the maximum is 0.7840, indicating that there is quite a bit of variability in voter turnout across the states and across the years. I seek to predict what influences that variability in voter turnout.

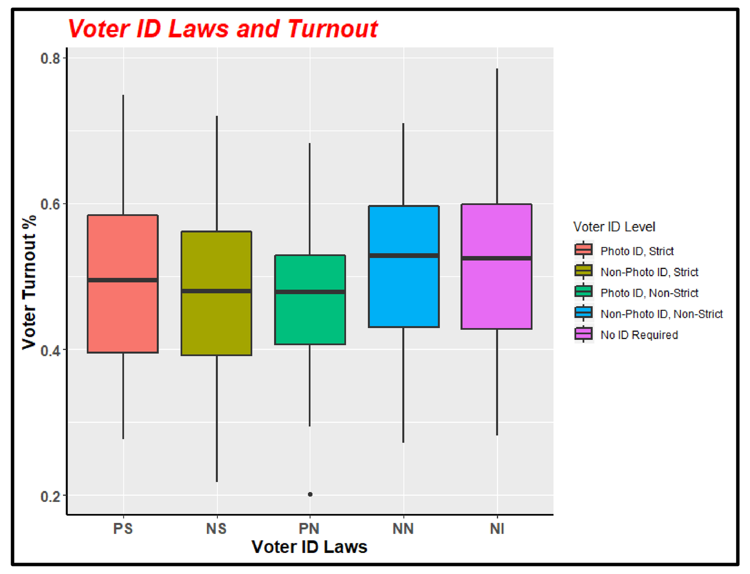
Quantitative Predictor Variables



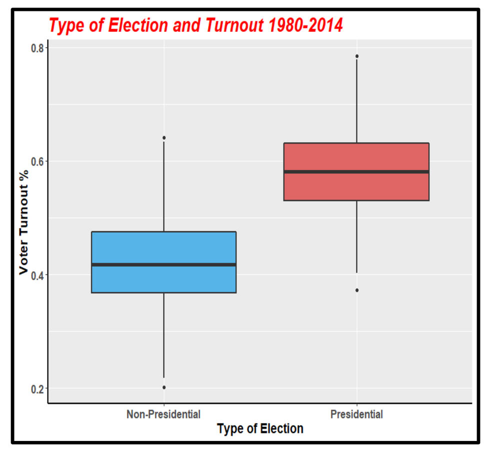
The scatterplot of Voter Turnout versus Violent Crime Rate shows a moderate, negative relationship between the two variables. The R^2 value is -0.2395, indicating 23.95% of the variation in Voter Turnout between 1980 and 2014 can be explained by Violent Crime Rates. Violent Crime Rates yielded the largest R^2 value amongst all the other quantitative variables for predicting Voter Turnout. The following chart gives the R^2 values for the other variables.

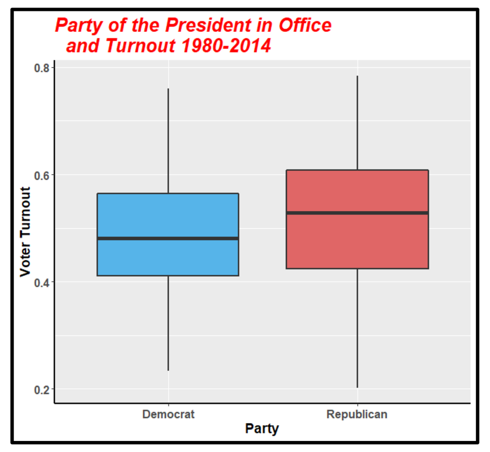
Variable	R^2
Population	-0.1172
Property Crime Rates	-0.1359
Number of Representatives	-0.1113
Proportion of Congress Represented	-0.1095

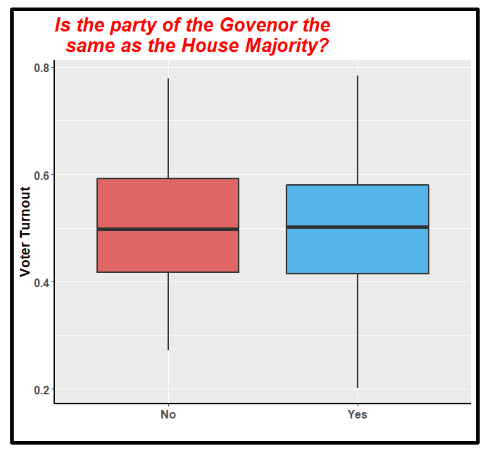
Categorical Predictor Variables



The side-by-side box plots on the left display the relationship between each level of Voter I.D. laws and Voter Turnout. Here, it is shown that it is not until the levels reach the two highest levels of lenience on Voter I.D. laws (NN and NI) that the median of Voter Turnout reaches over 0.5, indicating that people are more likely to vote when the Voter I.D. laws are more lenient. These box plots also show that within each level, the distribution of Voter Turnout is approximately normal.







The side-by-side box plots above give insight into the relationship between Voter Turnout and the Type of Election, the Party of the President, and whether the Party of the Governor is the same as the House Majority.

The side-by-side box plot for Type of Election shows that the median of voter turnout for Non-Presidential Elections is much smaller than the median for Presidential Elections. Meaning, people are more likely to vote if there is a Presidential Election in that Year. For the Party of the President, the median of voter turnout for Democrat Presidents is slightly smaller than the median of voter turnout for Republican Presidents. Therefore, people are slightly more likely to vote in an election if the current President in power is Republican. The distribution of voter turnout does not differ between whether the party of the current Governor in power is the same as the current House Majority Party. Meaning, this parameter does not have a strong influence on voter turnout by itself.

Regression Results

I created a multiple linear regression model using all 12 possible variables for predicting voter turnout. However, not all of the predictor variables were statistically significant, which hinders the ability to interpret the results. Hence, I performed backwards elimination, which is a process where the variable with the highest p-value is iteratively removed from the model until all remaining variables have statistical significance (p-value < 0.05). The resulting estimated regression equation from backwards elimination is:

Voter Turnout = 0.4333 – 0.1097 (Violent Crime Rate)
– 31.5439 (Proportion of Congress Represented)
+ 0.0719 (Number of Federal Representatives)
+ 0.01194 (Voter I.D. Strictness Level)
+ 0.1562 (1 if Presidential Election)
+ 0.02233 (1 if President is Republican)
– 0.01299 (1 if Party of Governor and House are the same)

$R^2 = 0.6158$, p-value $< 2.2e^{-16}$

Interpretation of Results:
An increase of 1% in Violent Crime Rate, with all other variables held the same, is predicted to decrease Voter Turnout by 0.1097.

An increase of 0.01 in the Proportion of Congress Represented, with all other variables held the same, is predicted to decrease Voter Turnout by 0.315439.

An increase of 1 Federal Representative, with all other variables held the same, is predicted to increase Voter Turnout by 0.0719.

An increase of 1 level in the Voter I.D. Strictness (which corresponds to being *less* strict) is predicted to increase Voter Turnout by 0.01194.

If there is a Presidential election that year, Voter Turnout is predicted to be 0.1562 higher than if there was not a Presidential election that year.

If the current President is Republican, Voter Turnout is predicted to be 0.02233 higher than if the current President was a Democrat.

If the Party of the Governor and House are the same that year, Voter Turnout is predicted to be 0.01299 lower than if the parties were different.

Results of other Models:

Model	R^2
Full	0.6155
Best Subsets with 8 Variables	0.6163

Conclusions and Future Work

I was able to identify seven variables that are significant for predicting voter turnout. In particular, I found that voter I.D. laws are a significant predictor, with stricter ID laws associated with lower voter turnout.

In the future, I would include more elections in my data set, particularly smaller, state-level elections. It is typically hard to generalize across an entire state, so looking at voter turnout at the county level instead of state level may be a more accurate representation of what I was studying. I would also include additional predictor variables, such as the average time it takes to vote, average distance from a precinct, and the time limit someone had to vote. Finally, a limitation of this study was violating the independence assumption of the regression model due to having repeated data for each state over different years in the data. To address this limitation, I could perform a mixed-effects model in the future.