Midertm Project: Examining Relationships Between Poverty Rate, Race, and Voting Behavior

Anna Cook

12/10/2020

Abstract

The purpose of this analysis is to explore the relationships between race/ethnicity, poverty rate, and voting tendencies across states in the U.S. in recent years. The research question consists of two parts: 1) To assess the relationship between poverty rate and race/ethnicity across U.S. states, and 2) to assess the relationship between vote share in presidential elections and poverty rate across states. Two multilevel linear models were fit to the data with random intercepts for each state. The results show no significant association between the outcome variable (poverty rate), and the predictors (race/ethnicity and . However, there are important limitations to consider.

Introduction

Amid the Covid-19 pandemic as well as Donald Trump's presidency, the issues of wealth and racial disparities in the United States have become more exposed, and appear to be more prevalent now than ever before. The analysis presented here is aimed at better understanding how these issues are related to one another, by examining poverty rates for different racial/ethnic groups across states in the U.S., as well as recent presidential election results for those states. The research question is two-fold: First, I am interested in determining whether there is a relationship between poverty rate and racial/ethnic groups across states. For example, are there racial groups who are economically disadvantaged in some states but not in others? Second, I am interested in determining whether the poverty rate in various states is associated with the voting behavior in those states. The data was collected from two different sources. The first set of data was collected from diversitydatakids.org and contains information on the poverty rate for various racial/ethnic groups, organized by U.S. state and year (2009-2017). Although the original dataset included many racial/ethic groups, I chose to focus my analyses on only 6 of those: Native American, Black, Hispanic, White, Asian, and the total across the whole state's population. Figure 1 shows trends in poverty rates for two different states, Alabama and New York. Although these states greatly in both their politics and demographics, they show similar patterns amongst poverty rates, with the White, Asian, and Total groups having relatively low poverty rates, while the Native American, Black, and Hispanic groups had higher poverty rates. The full set of 50 states is shown in Figure ___ in the appendix.

The second set of data was collected from dataverse.harvard.edu and contains information on U.S. presidental election results, organized by state and election year. The original dataset included elections dating back to 1976, but I filtered the dataset to include only 2012 and 2016 election since those are the only presidential election years are fell within the year range of the poverty dataset (2009-2017). Although there are only two elections in this final dataset, I believe these are important elections to analyze, as they were vastly different from one another. In 2012 Barack Obama, who was the first Black President of the United States, was re-elected for his second term, defeating Republican Mitt Romney. In 2016 Donald Trump was elected, running against Democrat Hillary Clinton, who made history as the first female presidential nominee. Not only did the office of the president change along party lines between these two elections, but it changed along racial and ideological lines as well. Therefore, these are important elections to explore.

Figure 2 shows the percentage of vote share for the Democratic and Republican candidates for the 2012 and
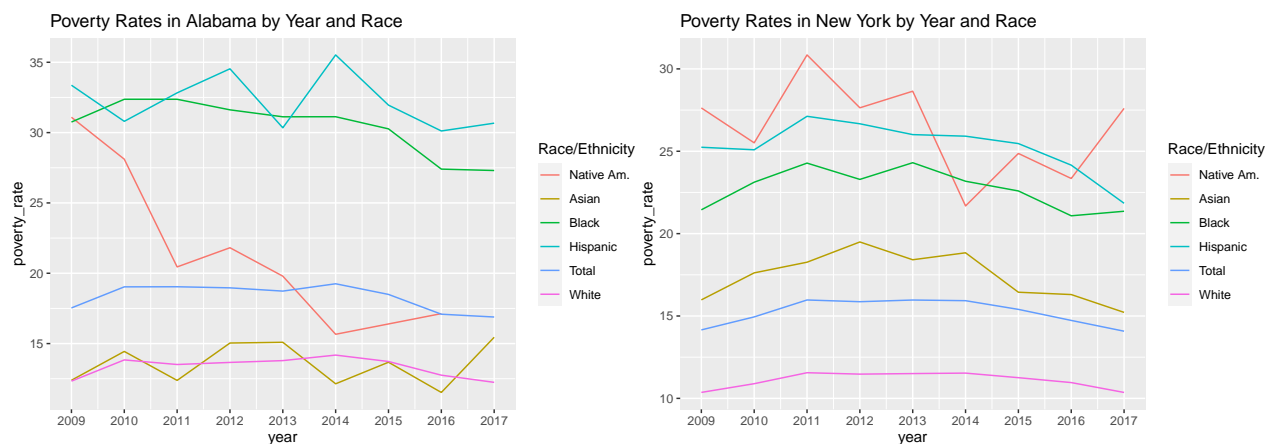
Figure 1: Trends in poverty rate amongst racial/ethnic groups in Alabama (left) and New York (right) from 2009 to 2017

2016 elections in Alabama and New York. This figure shows a strong contrast between voting behavior in these two states. The full set of 50 states is shown in figure ___ in the appendix.
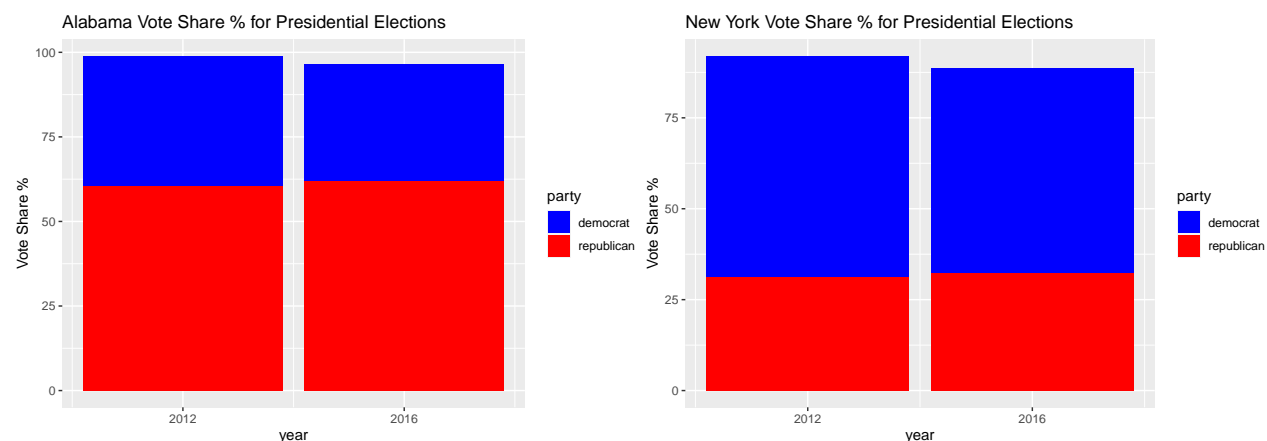

Figure 2: Percentage of Vote Share in Alabama (left) and New York (right) for Democrat and Republican Candidates in 2012 and 2016 Elections

Modeling

To examine an association between race/ethnicity and poverty rate for different states, I fit a multilevel linear model with intercepts varying for each state. The outcome variable is poverty rate, and the predictor variable was race/ethnicity. Next, to examine an association between voting behavior and poverty rate, I fit a second multilevel linear model, again with intercepts varying for each state. In this model, the output variable is the Republican vote share (calculated as votes for republican candidate as a percentage of total votes). The predictor variable was poverty rate. In order to validate these models, I used a posterior predictive check to compare the data's distribution to that of predictive simulations, in addition to examining the residual plot for each model. The results of these validation methods are described below. The output of these models and residual plots are shown in the appendix, and the results are described below.

Results

Discussion

One major limitation to this study is that the sample sizes are too small. Because there is only one poverty rate per racial/ethnic group per state per year, there isn't enough data to fit reliable models when using all of those variables as predictors. This probably may be partially alleviated by collapsing across states, but this wouldn't allow us to see how patterns vary by state. Another limitation is that while I am interested in voting behavior, my dataset only contained information on presidential elections which are held once every four years. This also leads to too small of sample sizes since the datasets only go back to 2008. The analyses may have been more reliable with more information; for example, if the dataset had contained voting behavior for more years, or on local elections as well as presidential elections. Both of these limitation led to poor performance of any of the models I tried fitting. Using different sets of predictors, collapsing across year, and using different types of models only led to very small changes, and in the end, no model fit the data very well. This was an important learning experience for me, as I now have a better sense of what to look for when I am searching for sufficient data to answer my research questions in the future.

Appendix

References

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Goodrich B, Gabry J, Ali I & Brilleman S. (2020). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.21.1 https://mc-stan.org/rstanarm.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Hadley Wickham, Jim Hester and Romain Francois (2018). readr: Read Rectangular Text Data. R package version 1.3.1. https://CRAN.R-project.org/package=readr

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr

Hadley Wickham (2020). tidyr: Tidy Messy Data. R package version 1.1.2. https://CRAN.R-project.org/package=tidyr

Stefan Milton Bache and Hadley Wickham (2014). magrittr: A Forward-Pipe Operator for R. R package version 1.5. https://CRAN.R-project.org/package=magrittr

Data Collected From: http://data.diversitydatakids.org/dataset/17001_1_p-poverty-rate--percent--by-race-ethnicity/resource/b7cd5119-acd7-4514-8b65-a6d22608994d

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/42MVDX

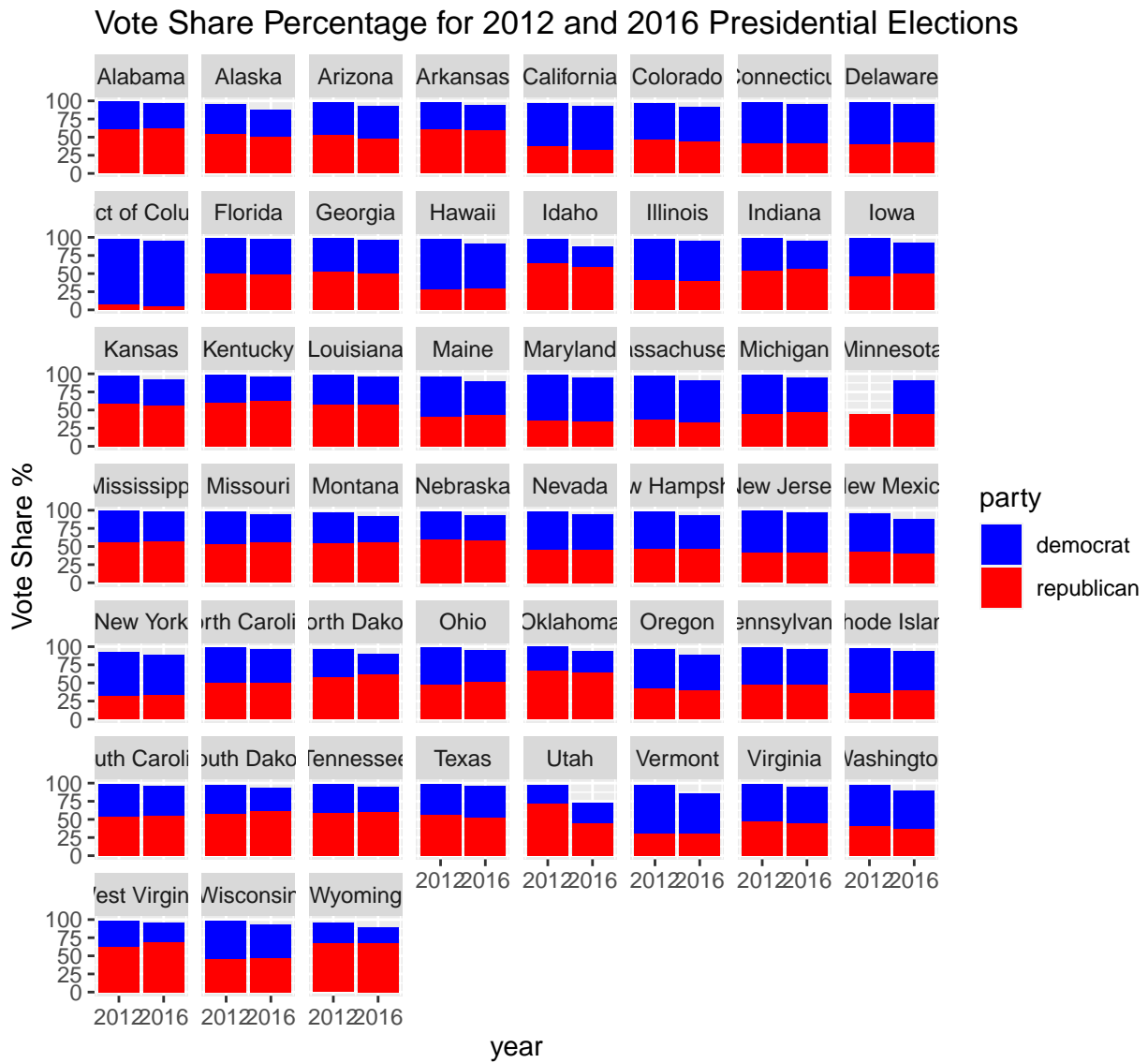Figure 3: Trends in poverty rate amongst racial/ethnic groups from 2009 to 2017

Figure 4: Vote share percentage in 2012 and 2016 U.S. presidential elections. Note: these figures do not include third party vote shares.