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 models were fit to the data to explore these questions. The results show that 1) poverty rate varies greatly between different racial/ethnic groups, and 2) there is a positive linear relationship between poverty rate and vote share, such that states with higher overall poverty rates tend to have higher vote share percentages for Republican Candidates. While these results are interesting, there are important limitations to consider

Introduction

Overview

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.

Poverty/Race Data

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 among 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 6 in the appendix.

Elections Data

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 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 7 in the appendix.

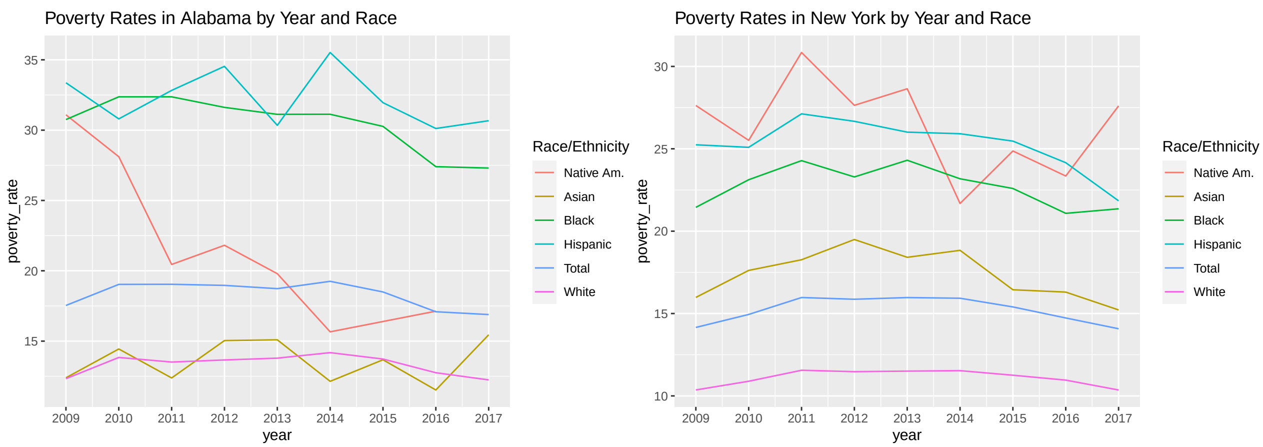

Figure 1: Trends in poverty rate amongst racial/ethnic groups in Alabama (left) and New York (right) from 2009 to 2017

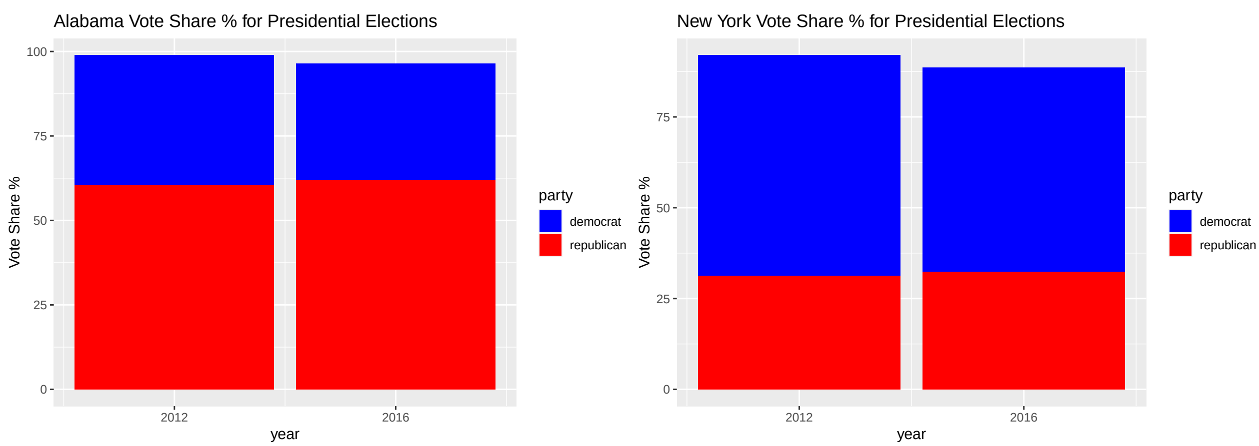

Figure 2: Percentage of Vote Share in Alabama (left) and New York (right) for Democrat and Republican Candidates in 2012 and 2016 Elections

Merged Data

As part of the data cleaning and EDA process, the two sets of data had to be merged in order to examine a relationship between poverty rate and voting behavior. There appears to be a slight positive linear trend, where the higher the poverty rate in a state, the higher the vote share for the republican candidate (see Figure 3), and therefore, the lower the vote share for the democratic candidate. This relationship will be explored further in the regression analysis below.

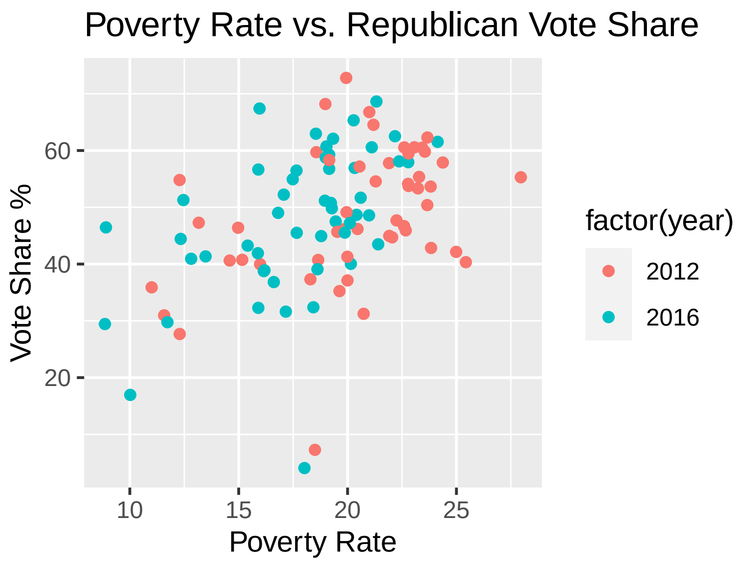

Figure 3: Poverty Rate vs. Percentage of Vote Share for Republican Candidates in 2012 and 2016 Elections

Modeling

To examine an association between race/ethnicity and poverty rate for different states, I first fit a multilevel linear model with intercepts varying for each state. The outcome variable is poverty rate, and the predictor variable was race/ethnicity. By allowing the intercepts to vary, we will be able to explore differences in baseline poverty rates for individual states. Next, to examine an association between voting behavior and poverty rate, I fit a simple linear regression model. In this model, the output variable is the Republican vote share (calculated as votes for republican candidate as a percentage of total votes for each state). The predictor variable was poverty rate for each state. In order to validate these models, I used a posterior predictive check to compare the data distribution to that of predictive simulations, in addition to examining the residual plot for each model. The results of these models and validation methods are described below, and the outputs are shown in the appendix (see Figures 8 and 9).

Results

Multilevel Linear Model

The results of the multilinear model assessing the relationship between poverty rate and race/ethnicity are shown in Figure 8 in the appendix. The results for the fixed effects show that most racial/ethnic groups tend to be lower in poverty rate compared to the Native American Group. For example, the coefficient for Whites is -15.34, indicating that on average, the poverty rate for White people in the U.S. is 15.34 percentage points lower than that of Native Americans. The only racial group that showed a slightly higher poverty rate on average was Black people. For the random effects (state intercepts), each state varied in their intercepts with no clear pattern in the variation. Because there are 50 states, the random effects will not be discussed further, but these effects are shown in the appendix (see Figure 8). The residual plot is shown in Figure 4, and shows a cone-like shape, with some large residual values. This suggests that the model is not fitting the data very well. This issue is discused further below.

Simple Linear Model

The results of the simple linear model assessing the relationship between voting behavior and poverty rate are shown in figure 9 in the appendix. These results show that there is a positive, linear relationship between poverty rate and vote share for republican candidates (see Figure 5), which confirms the positive relationship suggested by the EDA. The coefficient for the poverty rate is 1.4, suggesting that on average the vote share

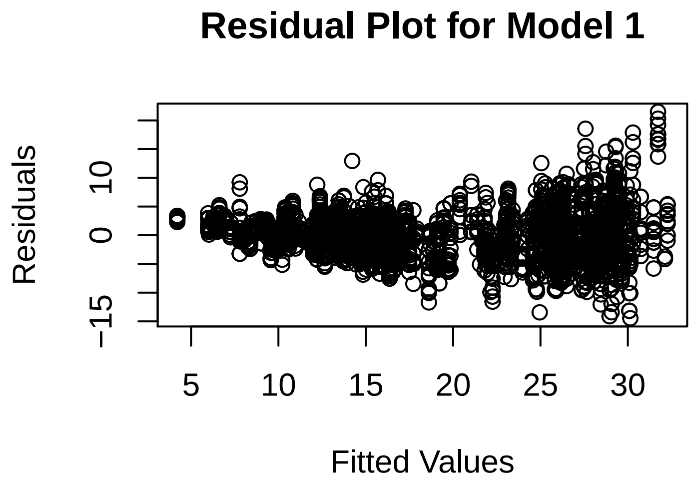

Figure 4: Residual plot for multilevel model shows a cone-like shape, suggesting non-constant variance

for Republican candidates increases by 1.4 percentage points for every 1 percentage point increase in poverty rate. The implications and limitations of these results are discussed further in the next section.

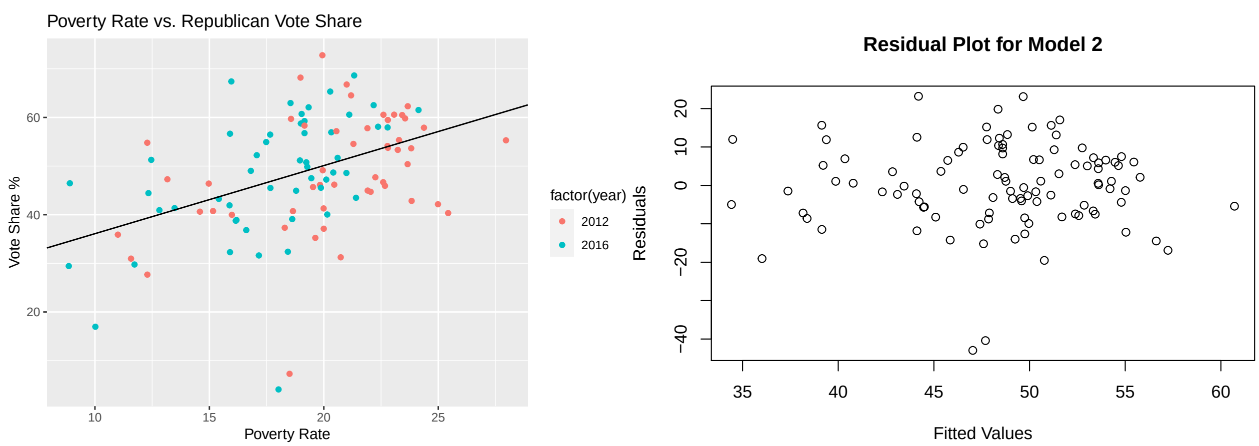

Figure 5: Black line represents regression line from simple linear model

Discussion

Multilevel Linear Model

The results of the multilevel linear model suggest that there are big differences in poverty rates between different racial/ethnic groups, but that these differences vary across U.S. states. On average, it appears that Black and Native people are the most economically disadvantaged in the U.S., while White people tend to have much lower poverty rates and Asian and Hispanic people tend to fall somewhere in the middle. While these results are not surprising, it is important to recognize the extent of the differences in poverty rate between different racial/ethnic groups. For example, according to this model, White people have a poverty rate that is on average 15.34 percentage points lower than Native Americans. This is a huge difference, and brings to light the importance of working toward equity for marginalized groups. However, this result must be taken with a grain of salt, as the residual plot (see Figure 4) and posterior predictive check (see Figure 8) show that the model is not an ideal fit for the data.

Simple Linear Model

The results of the simple linear model suggest that there is a positive linear relationship between poverty rate and the percentage of vote share for Republican candidates in U.S. presidential elections. This also suggests that we should expect a negative linear relationship between poverty rate and the percentage of vote share for Democratic candidates. While we cannot make any causal claims about this result, it sparks curiosity as to what leads to this relationship. For example, are there systematic difference in party ideology such that Republican candidates appeal more to individuals in states with higher poverty rates? These are factors that political candidates could take into account when running their campaigns to try to secure more votes. The residual plot (see Figure 5) shows an equal spread around 0, with a few noticeable outliers. For both of these models, there are important limitations to consider.

Limitations and Future Directions

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 simply isn't enough data to fit reliable models. 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 2009. 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 (see residual plots in Figures 4, 5, and posterior predictive checks in Figures 8 and 9). Using different sets of predictors, collapsing across year, and using different types of models only led to very small changes in model fit, 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

Multilevel Linear Model Output and Validation

```
## Warning: Bulk Effective Samples Size (ESS) is too low, indicating posterior means and medians may be
## Running the chains for more iterations may help. See
## http://mc-stan.org/misc/warnings.html#bulk-ess
```

```
##             (Intercept) race_ethnicityasian_est race_ethnicityblack_est
##              27.0744903             -13.5808437               0.1763293
##  race_ethnicityhisp_est race_ethnicitytotal_est race_ethnicitywhite_est
##              -2.7595164             -12.6380554             -15.3470711
```

```
## $state
##                       (Intercept)
## Alabama                1.96619921
## Alaska                -5.72933261
## Arizona                2.37307177
## Arkansas               3.03628083
## California            -1.35635055
## Colorado              -2.52339577
## Connecticut           -3.97025641
## Delaware              -3.29297505
## District of Columbia -2.07140978
## Florida               -1.23407725
## Georgia                0.90563981
## Hawaii                -5.13960268
## Idaho                  1.36040531
## Illinois              -2.16655348
```

Figure 6: Trends in poverty rate amongst racial/ethnic groups from 2009 to 2017

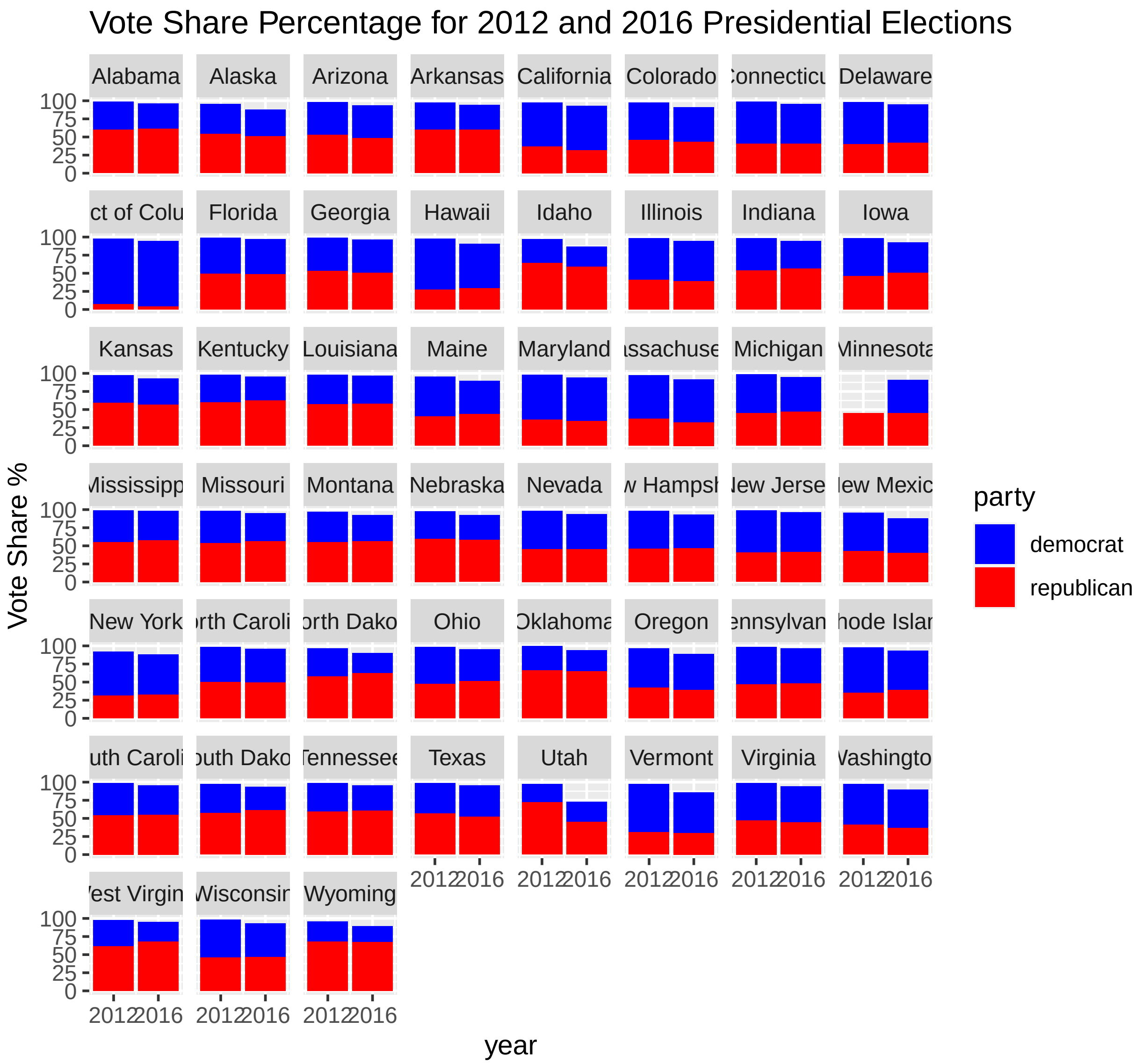

Figure 7: Vote share percentage in 2012 and 2016 U.S. presidential elections. Note: these figures do not include third party vote shares.

```
## Indiana              2.66264541
## Iowa                 1.46477918
## Kansas              -1.27441746
## Kentucky             4.20046523
## Louisiana            1.85469443
## Maine                3.02295056
## Maryland            -7.54256427
## Massachusetts       -1.19344060
## Michigan             1.33361467
## Minnesota            0.91489486
## Mississippi          5.02904866
## Missouri             0.54815396
## Montana              2.14280479
## Nebraska             2.17546561
## Nevada              -1.37943725
## New Hampshire       -4.47325215
## New Jersey          -5.27808925
## New Mexico           2.83994872
## New York            -0.05625286
## North Carolina       1.78384041
## North Dakota         0.45942792
## Ohio                 1.91196077
## Oklahoma             0.99902984
## Oregon               1.90897131
## Pennsylvania         1.60079138
## Rhode Island         0.70146752
## South Carolina       1.33364380
## South Dakota         4.61349796
## Tennessee            2.29676636
## Texas               -0.92992726
## Utah                 0.21837388
## Vermont             -1.46231177
## Virginia            -5.46190844
## Washington          -1.30026831
## West Virginia        3.46707131
## Wisconsin            1.97825997
## Wyoming             -2.36360978
##
## with conditional variances for "state"
```

Simple Linear Model Output and Validation

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      vote_pct ~ poverty_rate
##  observations: 102
##  predictors:   2
## ------
##              Median MAD_SD
## (Intercept)  22.2    5.7
## poverty_rate  1.4    0.3
##
## Auxiliary parameter(s):
##       Median MAD_SD
```

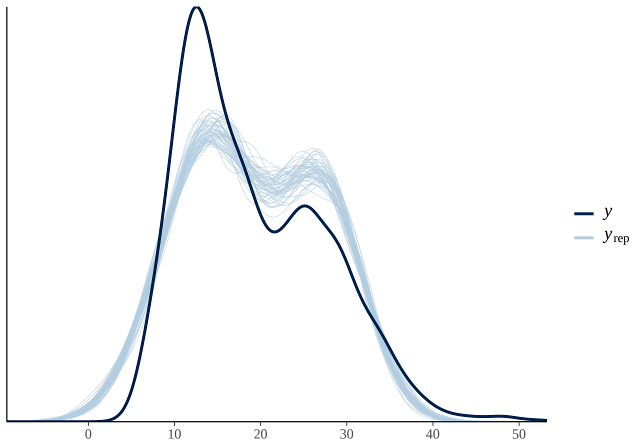

Figure 8: Output and posterior predictive check for multilevel linear regression model. Posterior density plot shows that the model fit is not ideal.

```
## sigma 11.1    0.8
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

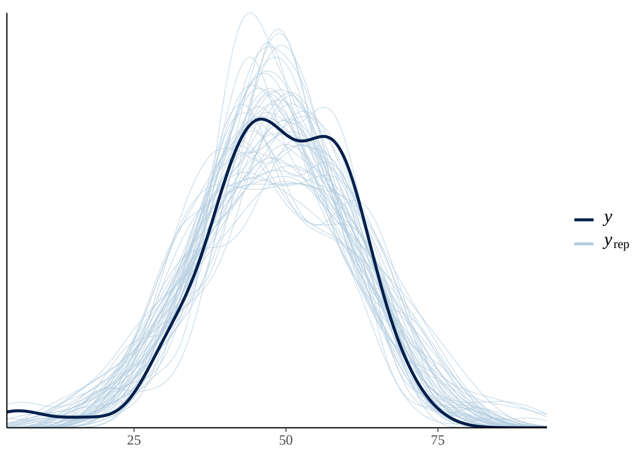

Figure 9: Output and posterior predictive check for simple linear regression model. Posterior density plot shows that the model fit is not ideal.

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