An Analysis of 2020 Poverty Rates by County

Shruti Biradar, Carolyn Fiore, Kevin Russell

Summary

According to the U.S. Census, the official poverty rate in 2020 was 37.2 million people or 11.4% of the U.S. population [1]. 8.6% of Americans did not have any health insurance for the entire year, despite the onset of the COVID-19 pandemic [1]. We want to explore the demographics of American poverty and find variables that are good predictors of the poverty rate. This would be beneficial to policymakers and legislators writ large. For this project, analysis will be conducted by county to determine whether race, sex, the gap between minimum wage and living wage, dominant employment type (manufacturing, service, or government), unemployment rates, inflation, and various GDP metrics are accurate predictors of poverty level. The goal will be to produce a model that can accurately predict the poverty rate of a county based on the analysis of the most highly correlated predictors.

A significant portion of the project will be cleaning and joining datasets to tidy data by county. Poverty data will be disaggregated by race/ethnicity and gender and compared to data from the Bureau of Labor Statistics, MIT Living Wage Calculator, US Census Bureau, and the Bureau of Economic Analysis. We will analyze the majority of these counties to build a prediction model, and then use the remainder of the counties as test data to see if the model can accurately predict poverty rate using the highest correlated predictors.

Several datasets were collected. The first several are from the US Census, called Poverty Status In The Past 12 Months By Sex By Age [2]. This data is collected by county and there is a different dataset for each racial/ethnic group demographic. The data in these datasets was collected by the 2020 U.S. Census. These require tidying and joining. The second dataset contains the 2020 living wage data by county for multiple family compositions and is from the MIT Living Wage Calculator created by Dr. Amy Glasmeier [3]. This data has been augmented with the applicable state minimum wage data, collected from the Department of Labor [4], and the two variables have been used to calculate the gap between the state minimum wage and the county's living wage, or cost of living. We have used the data for a single worker with 0 children to calculate this, as it is the smallest and simplest family unit to analyze. The third dataset is the 2020 Local Area Unemployment (LAU) Dataset from the Bureau of Labor and Statistics [5]. This dataset captures unemployment rates by county for 2020 and is a very tidy dataset. The fourth and fifth datasets are from the Bureau of Economic Analysis and Bureau of Labor Statistics, respectively, and are the measures of Gross Domestic Product (GDP) [6,7] and Consumer Price Index (CPI) for 2020 [8]. These datasets are sparser and have approximately 40 measures of missing data to account for. We have begun the process of tidying these datasets and will also conduct various joins to create a tidy dataset for our analysis.

Proposed plan

Each dataset will need to be cleaned and manipulated using tidyverse and dplyr. We will need to use various joins to create a dataset that we can use to create our model. 3 of our datasets have multiple instances of missing data, so we will account for that by imputation. We will use ggplot2 to create visualizations for analysis and for presentation. We may also need to use a by-county chloropeth if we have results that are well-depicted in map format. We plan to build our model using linear regression and use the residual plots, cross-validation, and stepwise model selection to classify our variables that are potential predictors, and to use classification to try and predict the poverty rate in several counties by using the highest correlated predictors.

We do anticipate challenges with the tidying and data analysis for this project. A primary challenge will be missing data, with some datasets not including data for some counties or simply not including the county. If significant data is missing for a given county, we will not include it in the overall analysis. If only minimal data is missing, we will consider using imputation from the data for surrounding counties. Another possible challenge is that, based on state or local governance policies, variables may be a good predictor in some states but not others; or the same variable may have a different impact by state. To account for this, we will consider a separate training set for each state or a separate training set for states with Republican governments and those with Democratic governments, as the different political parties enact public policy regarding poverty in very different manners.

Preliminary results

We conducted basic analysis on each data set to prove we can import and manipulate them for this project. Figure 1 uses the U.S. Census Poverty dataset [2] and shows 2020 poverty rates for Massachusetts by sex and race. Figure 2 uses the Living Wage Calculator data [3] and DOL minimum wage data [4] to show the gap, by county, between minimum wage and living wage. Figure 3 uses the GDP [6,7] and CPI dataset [8] to show relationships between GDP and CPI across the US. Figure 4 uses the LAUS dataset [5] and groups county data by state to show a boxplot of 2020 unemployment rates, categorized by political party governmental control as noted by [9].

References

1. U.S. Census Bureau. "Income, Poverty and Health Insurance Coverage in the United States: 2020". Publication: 14 September, 2021. [Online]. Available: https://www.census.gov/newsroom/press-releases/2021/income-poverty-health-insurance-coverage.html#:~:text=The%20official%20poverty%20rate%20in,million%20more%20than%20in%202019 [accessed 17 March, 2022].
2. American Community Survey, U.S. Census Bureau. "Poverty Status In The Past 12 Months By Sex By Age". Publication: 2020. [Online]. Multiple datasets disaggregated by race/ethnicity available: https://data.census.gov/cedsci/table?q=United%20States&t=Income%20and%20Poverty%3ARace%20and%20Ethnicity&g=0100000US,%240500000&y=2020&tid=ACSDT5Y2020.B17001B [accessed 22 March, 2022].
3. Glasmeier, A. K., Department of Urban Studies and Planning, Massachusetts Institute of Technology. "MIT Living Wage Calculator." [Online] Available: https://livingwage.mit.edu/ [accessed 26 February, 2022].
4. Department of Labor. "Changes in Basic Minimum Wages in Non-Farm Employment Under State Law: Selected Years 1968 to 2021". [Online]. Available: https://www.dol.gov/agencies/whd/state/minimum-wage/history [accessed 27 February, 2022].
5. Bureau of Labor Statistics. "Local Area Unemployment Statistics". Publication: 2 March, 2022. [Online]. Available: https://www.bls.gov/lau/laucnty20.xlsx [accessed 25 March, 2022].
6. Bureau of Economic Analysis. "2020 GDP Dataset". [Online, three files]. Available: https://download.bls.gov/pub/time.series/cu/cu.series, https://download.bls.gov/pub/time.series/cu/cu.data.0.Current, https://download.bls.gov/pub/time.series/cu/cu.item [accessed 23 March, 2022].
7. Bureau of Economic Analysis. "Real Gross Domestic Product, by County, 2017-2020." [Online]. Available: https://www.bea.gov/sites/default/files/2021-12/lagdp1221.xlsx [accessed 21 March, 2022].
8. Bureau of Labor Statistics. "Average Price Data – Consumer Price Index (CPI)." [Online]. Available: https://www.bls.gov/cpi/tables/supplemental-files/home.htm [accessed 13 March, 2022].
9. Ballotpedia. "Current State Government Trifectas". [Online]. Available: https://ballotpedia.org/Gubernatorial_and_legislative_party_control_of_state_government [accessed 23 March, 2022].
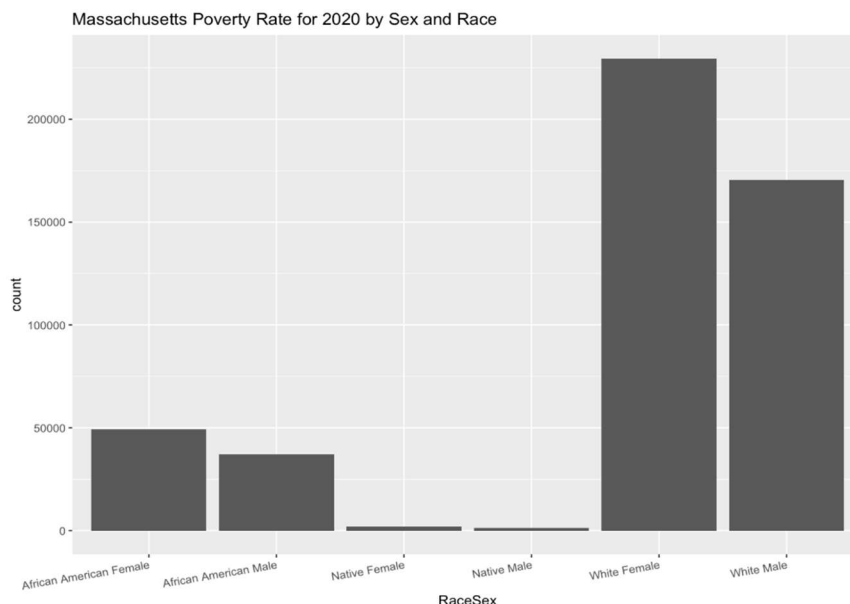
Supplemental Data:

Figure 1. Massachusetts Poverty Rates in 2020 by Sex and Race
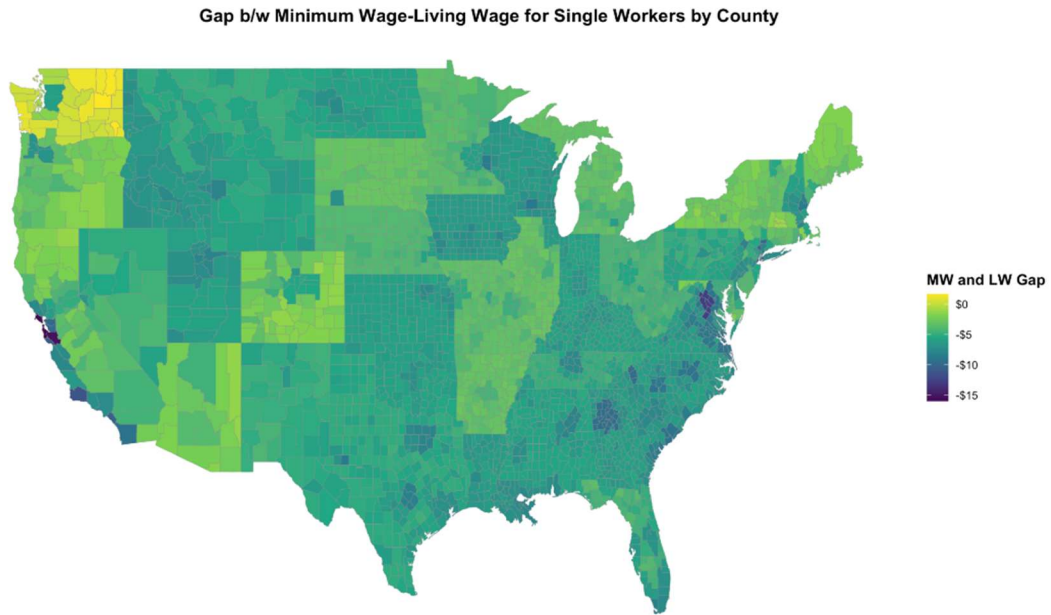
Figure 2. Gap between Minimum Hourly Wage and Hourly Living Wage by County
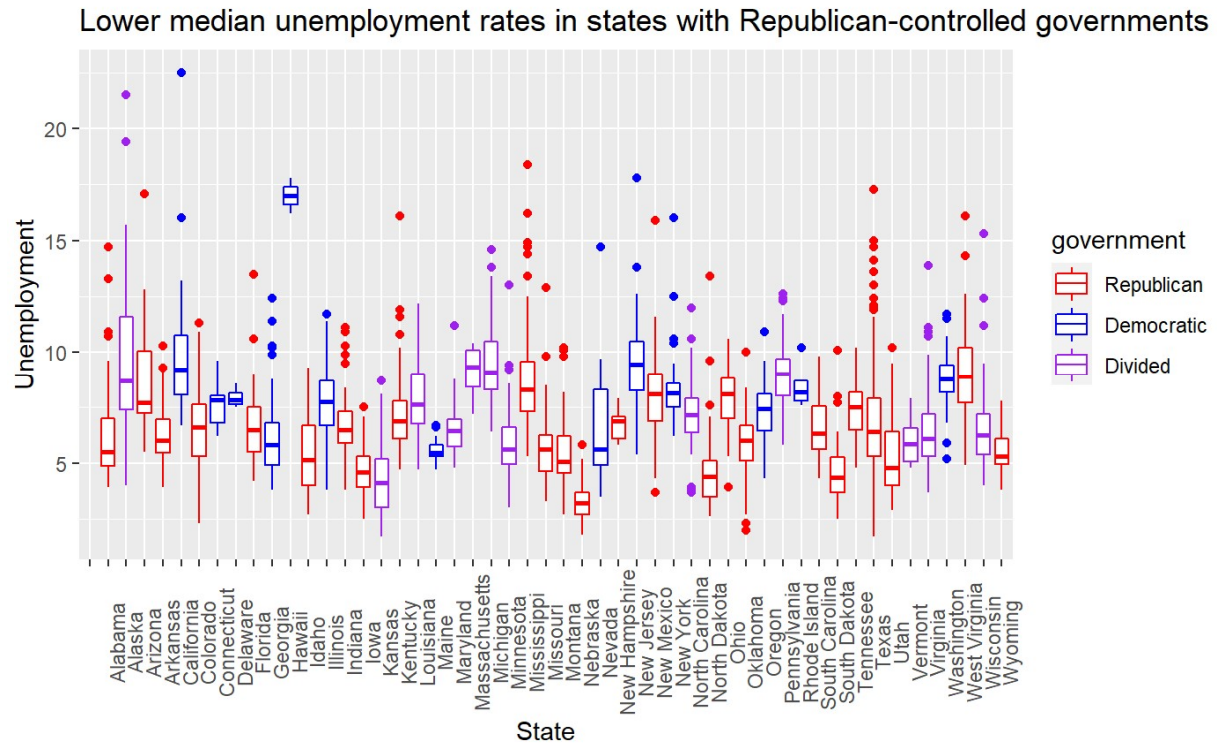

Figure 3. Regional Relationship between GDP and CPI

Figure 4. Median Unemployment Rates by County