

**Data Visualization Project - Atlas of Rural and Small-Town America**

David Boules - 900231172

Abdelrahman Nawara - 900231947

Department of Data Science, The American University in Cairo

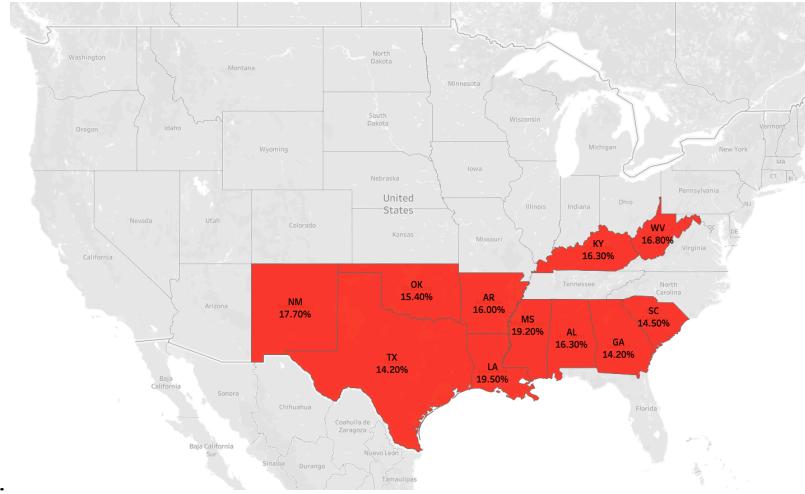
DSCI 2411: Data Visualization

December 8, 2024

## **1. Introduction**

The American Community Survey (ACS) is one of many sources of real-world government data that is easily accessible and open to the public. It is a yearly survey that is conducted to collect general information about all citizens in the United States, including information on employment, education and income (Census Bureau, 2024). The dataset explored in this project: ‘The Atlas of Rural and Small-Town America,’ is a dataset curated by the Economic Research Service (ERS) which utilizes information from the ACS and other federal sources, combined to derive insights about the economic, demographic, environmental and social forces affecting U.S. citizens, to shed light on the various challenges faced by communities in the U.S (2023).

The specific challenge chosen to be addressed in this project is poverty. In this dataset, an individual is defined as being in poverty if they live in families with an income below their poverty threshold, which is determined by certain factors such as family size and number of children (ERS, 2023). The main measure of poverty used in this project is a 5-year average poverty rate from 2017 to 2021. We observed that the upper quartile of states in terms of this metric were most of the southern states: **Louisiana, Mississippi, New Mexico, West Virginia, Kentucky, Alabama, Arkansas, Oklahoma, South Carolina, Georgia, and Texas:**



**Figure 1:** The upper-quartile of states in terms of 5-year average poverty rate (2017-2021)

To validate that this trend with southern states is not a coincidence, we referred to data published by the Center for American Progress (CAP - a public policy research organization) regarding yearly poverty rates per state, where the top states in terms of official poverty rate in 2023 were also the same ones observed in the Rural Atlas Dataset, with the exception of Texas and Georgia being replaced by Tennessee (another southern state) and New York (CAP, n.d.).

The aim of this study is to identify some of the factors that lead to high poverty rates in the United States to determine why the southern states in particular suffer from high poverty rates.

## 2. Dataset Description

### 2.1 General Description of the Dataset

The full dataset available on the ERS website contains six .csv files, along with a single .xlsx file, titled: ‘RuralAtlasData24.xlsx’ which is a compilation of the six

individual .csv files into one .xlsx file. The six individual files contain data under the categories: People, Jobs, County Classifications, Income and Veterans respectively, with the sixth file being a ‘Variable Name Lookup’ file.

Each category contains 3280 rows of data, with one row corresponding to the data for all attributes for one of the 3227 counties in the United States, along with numerical averages for each attribute for each of the 50 states (data for Puerto Rico, and Washington D.C. are also included, along with national averages for the U.S. for each attribute, bringing the total to 3280 observations). This is with the exception of ‘County Classifications,’ which only concerns categorical classifications of each county, and not for the states.

## 2.2 Category Descriptions

The Data in the collated ‘RuralAtlasData24.xlsx’ file falls under the following main categories:

- **People:** Demographic data from the ACS (ERS, 2023) including race/ethnicity, education, population size/change, and migration/immigration.
- **Jobs:** Economic data from the Bureau of Labor Statistics (ERS, 2023), a government agency that collects data concerning economic activity, including information on general employment & unemployment, as well as employment by industry.
- **County Classifications:** Categorical classifications for each county on information such as economic dependence, metro status, persistent poverty and persistent child poverty.

- **Income:** Data on median household income, poverty (including deep poverty and child poverty) and per capita income.
- **Veterans:** Data on veterans, including race/ethnicity, income, employment status, and service period (this data was completely unused in our analyses).

### **2.3 Data Cleaning/Pre-Processing Procedures**

We opted to create and work with two new separate .csv files: One containing the data for each county, and another containing the averaged data for each state (the latter does not contain any of the data under ‘County Classifications’ as previously mentioned). This process was done by importing the ‘RuralAtlasData24.xlsx’ file into Python and extracting the necessary rows and columns for each state/county into two separate data frames, which were exported as two separate .csv files: ‘States.csv’ and ‘Counties.csv’ (included in the final .zip file).

Regarding outliers, there were often some outliers for each key feature analyzed, however these were only excluded from visualizations manually in Tableau or Python if they were extreme. There were also many missing values, however we could not treat them using interpolation/extrapolation techniques due to each county’s data varying vastly from one another.

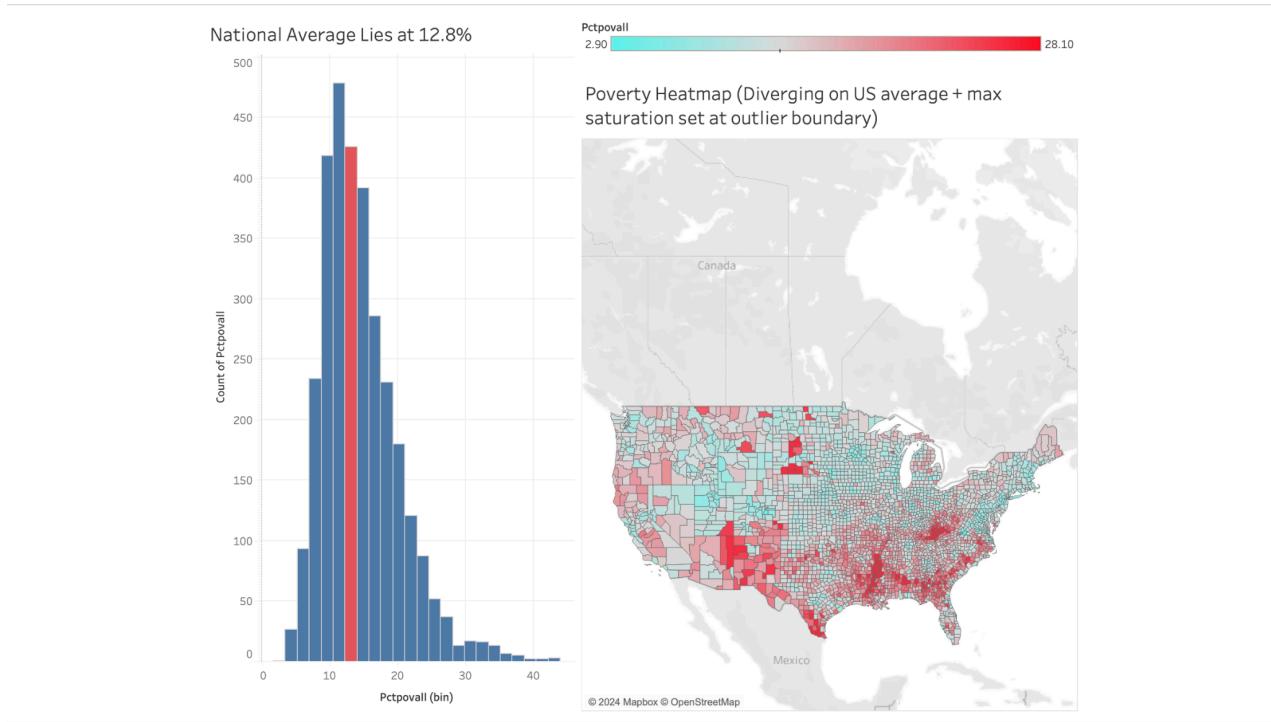
U.S. states are generally divided by region as: Northeastern, Midwestern, Southern, or Western states by the U.S. Census Bureau (2021). In our analyses, each state and county was given such a geographical classification since no geographical features appeared in the dataset curated by the ERS.

More details on feature engineering and data preprocessing are available in the .ipynb file associated with our project in the .zip file.

## 2.4 Data Summarization

### 2.4.1 Poverty Measures

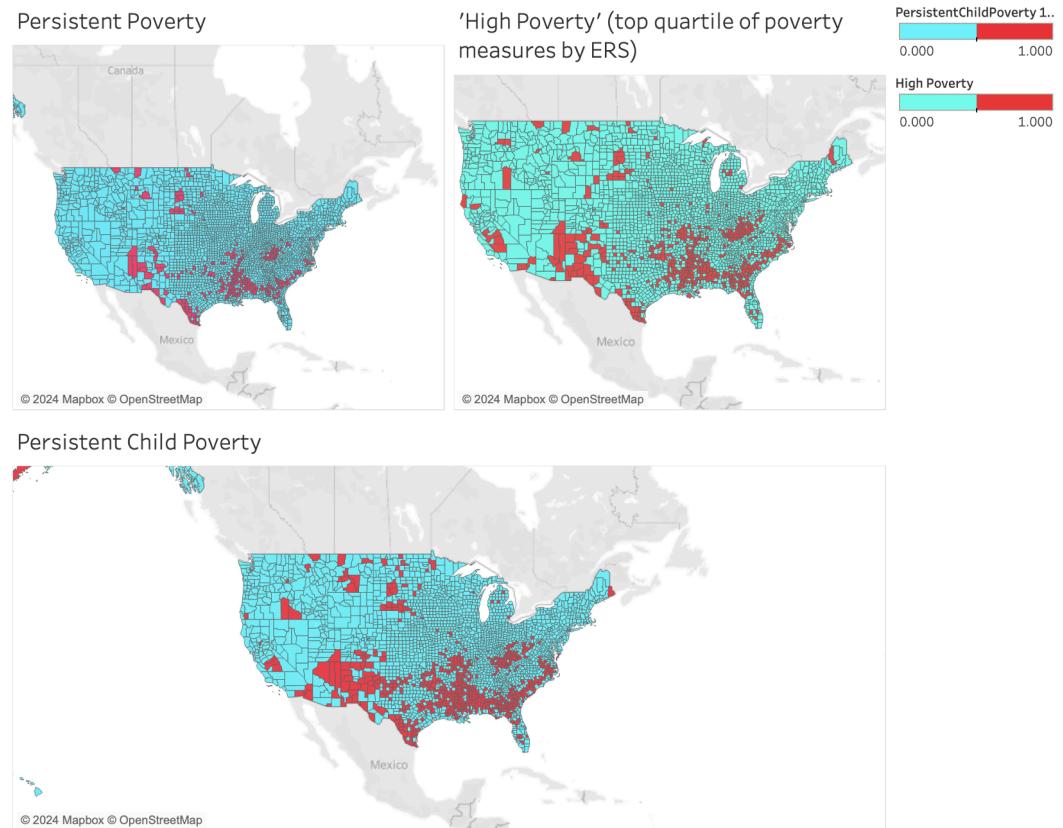
As previously mentioned, the main metric used for poverty in this study is the 5-year average poverty rate from 2017-2021.



**Figure 2 (left) & Figure 3 (right)** - Distribution of the average poverty rate across counties (left) and a heatmap of the average poverty rate for each state (right)

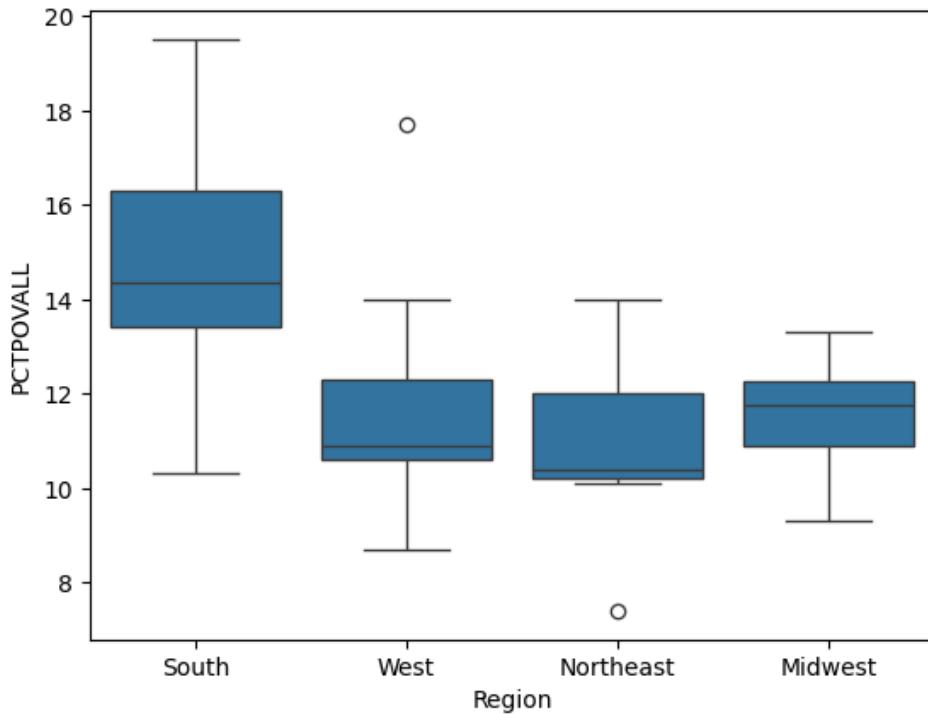
Figure 3 shows that there is some extreme poverty in the counties part of the southern states identified above, matching the trend of southern states having high poverty rates. However, this is not the only measure of poverty available in this dataset.

We also validated our measure of poverty by comparing the results of our initial heatmap to other heatmaps of different measurements of poverty:



**Figure 4:** Classification of counties by whether they are suffering from persistent poverty or persistent child poverty, and if they are a ‘high poverty’ county.

The same clusters of counties can be seen on all 3 maps in Figure 4, which match the clusters seen in the initial heatmap of average poverty rates. There is a clear presence of poverty in the southern areas. This observation is further supported by the following plot:

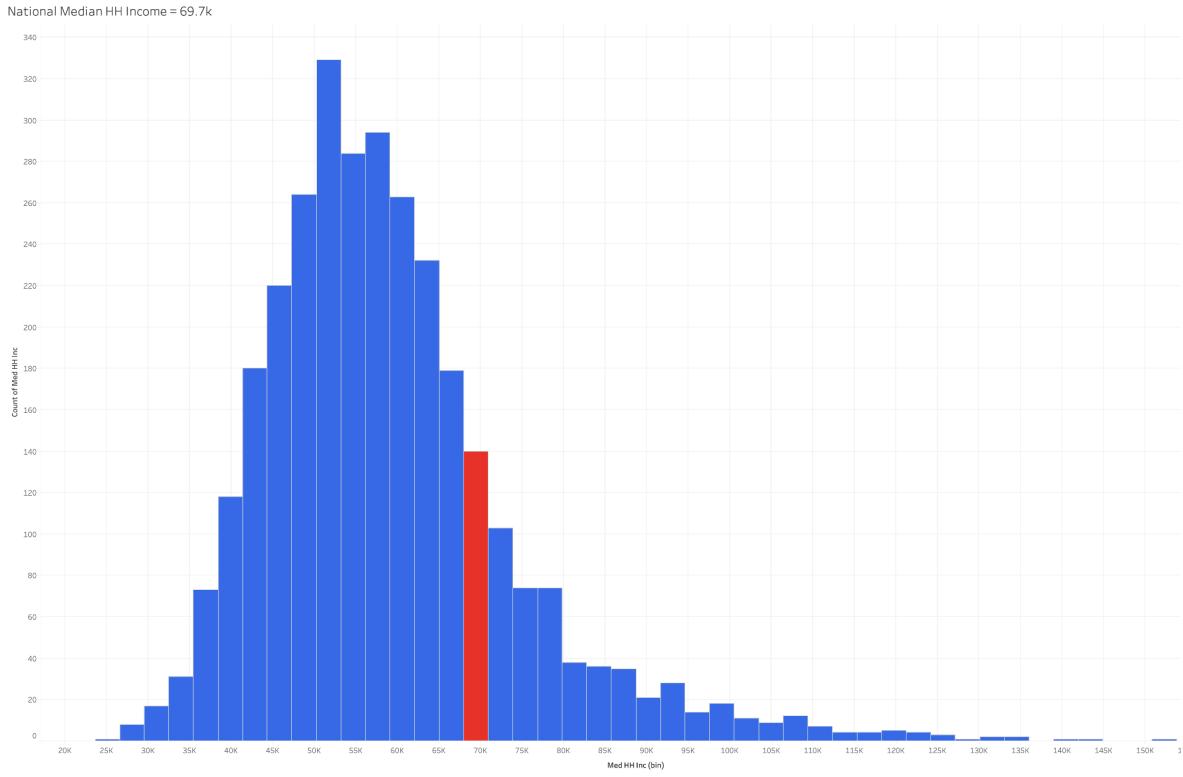


**Figure 5:** Distribution of average poverty rate per state, by geographical region

Figure 5 shows that numerically, there is a rather significant discrepancy between the average poverty rates in southern states when compared to states in other regions, ensuring that numerically, our observation is correct.

#### 2.4.2 Income Features

Income is an important feature when it comes to assessing poverty, and average incomes can vary across state and county-level. Therefore, it is important to monitor how income changes between each state/county:



**Figure 6:** Distribution of Median Household Incomes across U.S. counties (national average lies in red bar)

The data on median household incomes across U.S. counties appears to follow a semi-normal distribution that has a positive skew, where most counties appear to have a median household income below the national median of approximately \$70k.

### 3. What are the factors that might affect poverty?

#### 3.1 Education

Initial Assumption: Education level has an inverse relationship with poverty rate.



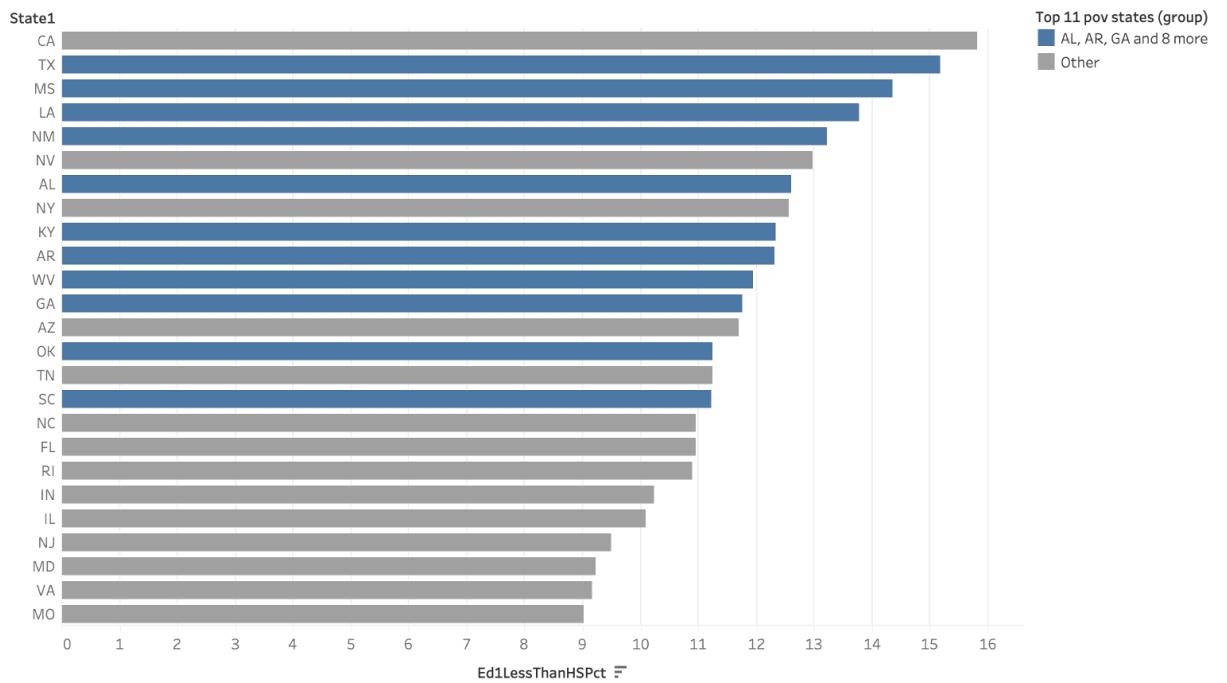
The plot of sum of Pctpovall for Ed1LessThanHSPct. The data is filtered on Ed1LessThanHSPct, which excludes 81.553398058.

**Figure 7:** Scatter plot describing the correlation between average poverty rate and percentage of county population with a low education level ( $r = 0.64$ )

**"Ed1LessThanHSPct"** is the measurement of low education level identifying Percentage of the state population 25 years old or older whose highest completed level of schooling was 12th grade or lower and who do not have a high school diploma or equivalent.

It appears that the higher the percentage of low education in a state the higher chance that this state suffers from high poverty.

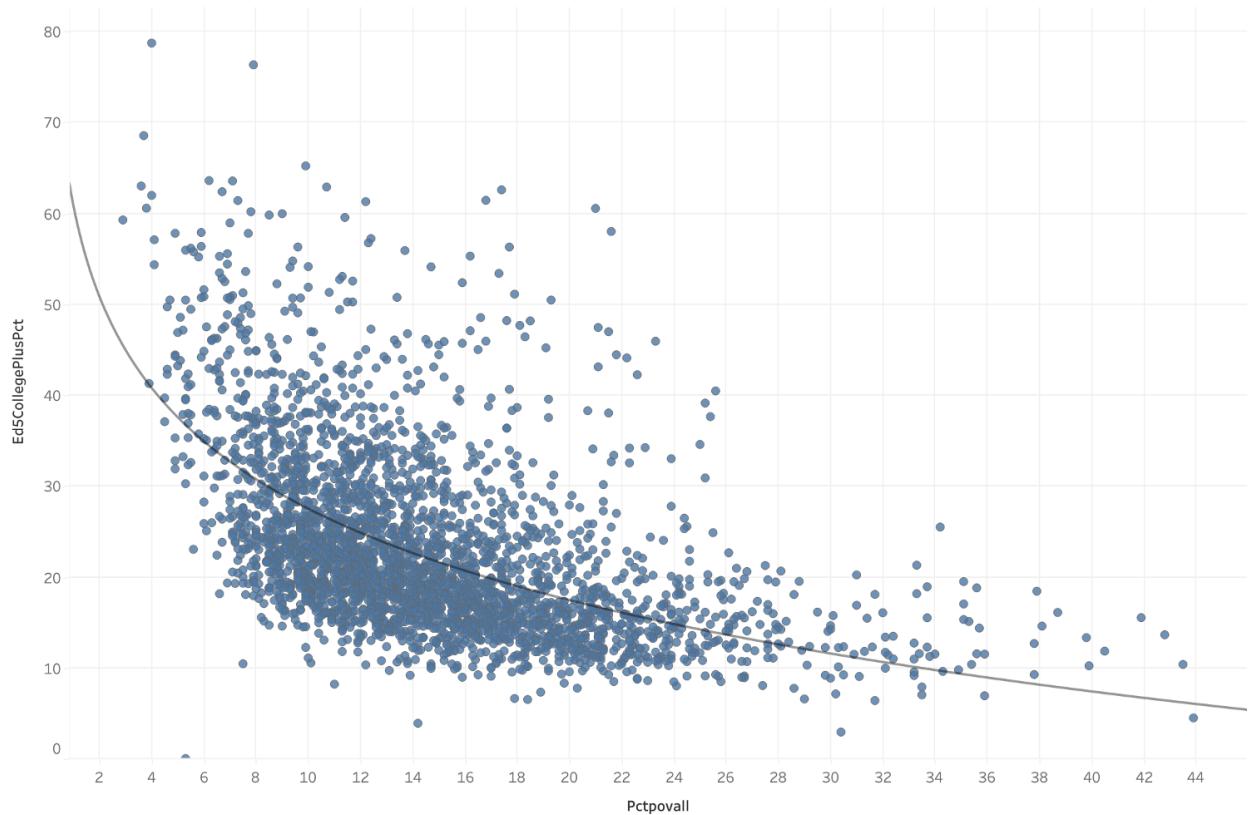
Ed1 - Less than High School Diploma (Top 25)



Sum of Ed1LessThanHSPct for each State1. Color shows details about Top 11 pov states (group). The view is filtered on State1, which keeps 25 of 50 members.

**Figure 8:** Ranking top half of states having the greatest proportion of people with a low education level

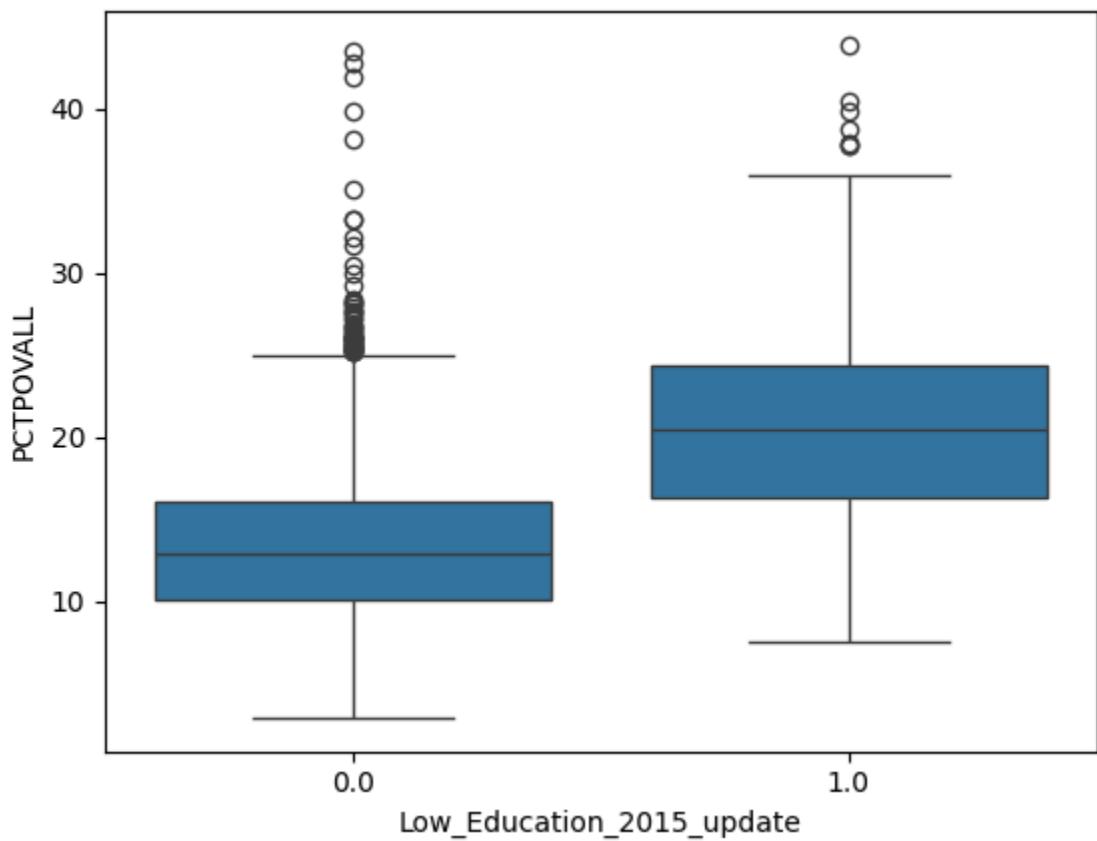
To check whether or not the southern states of interest are affected by this measure, a bar plot was implemented to illustrate whether there is a match between the top states having high poverty rates and low education levels. All 11 states in the upper quartile in terms of average poverty rate appear in the top 16 states with the highest 'Ed1LessThanHSPct' measure, with 7 of them found in the top 10 states for this measure.



Pctpovall vs. Ed5CollegePlusPct.

**Figure 9:** identifies the relationship between Poverty percentage in a state and higher education percentage in counties. ( $r = -0.5$ )

Unlike Ed1 identifying low education level percentage, '**Ed5CollegePlusPct**' defines the percentage of state population 25 years old or older with bachelor's degree or higher. Identifying the relationship between higher education and poverty percentage per each county to ensure the validity of the aforementioned claim. It was found that counties with higher percentage of higher education levels are more likely to have less percentage of poverty, following a rather logarithmic trend.

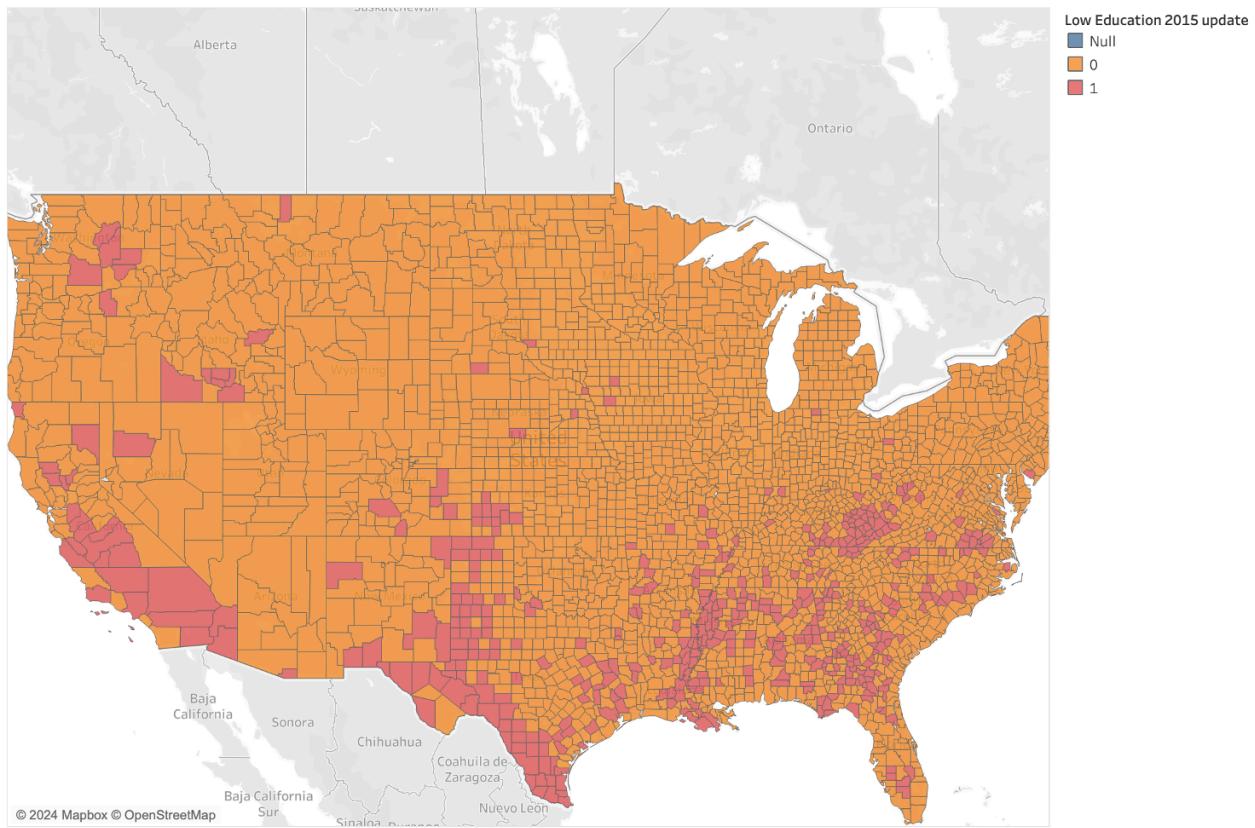


**Figure 10:** Boxplot for ‘Low Education’ classification groups with respect to average poverty rates ( $r = 0.48$ )

Ensuring the effect of education on poverty, Low\_Education\_2015 is another variable discussing low education level as it Classifies counties by level of education, where 1=low-education county; 0=all other counties; a county was classified as low-education if 25 percent or more of residents 25 years old or older had neither a high school diploma or GED.

It is shown that the 1.0 group had larger values for average poverty rate, than the 0.0 group, with a greater median, having its lower quartile higher than the 0 values upper quartile. This shows a significant difference between counties with 25% or more of low education and counties that have less than 25% of population with low education.

Low\_Education



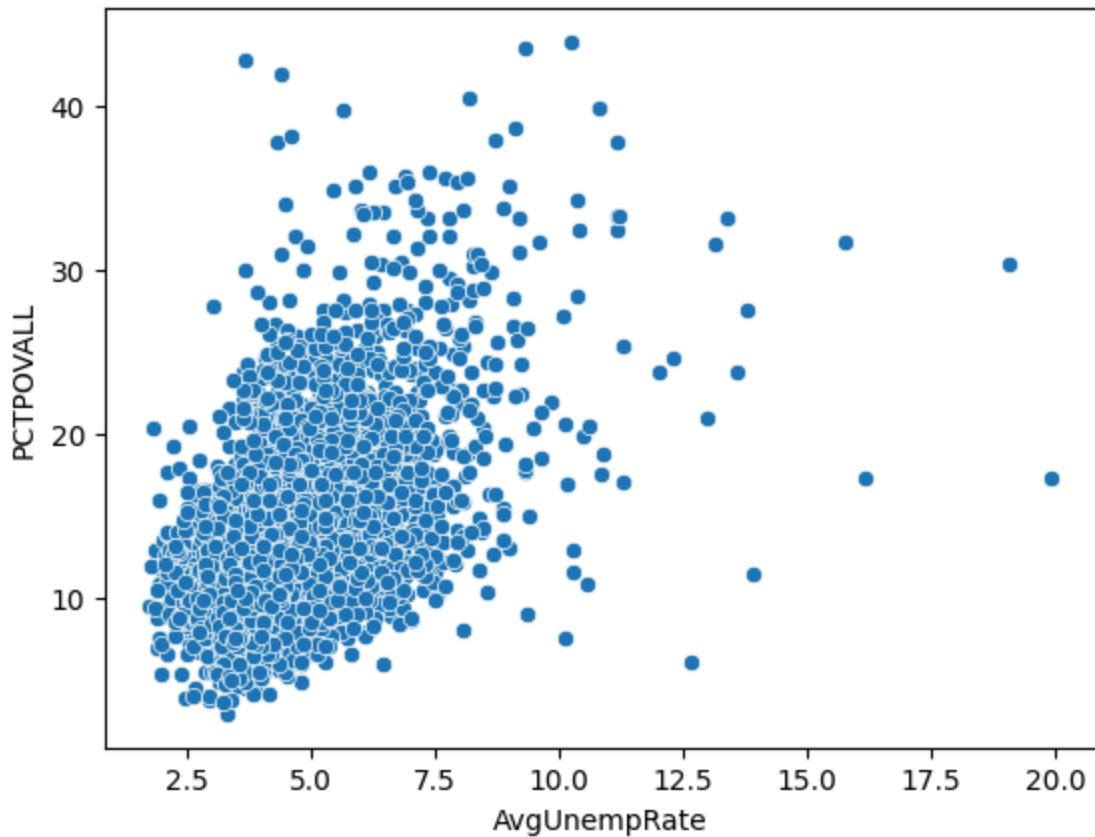
Map based on Longitude (generated) and Latitude (generated). Color shows details about sum of Low Education 2015 update. Details are shown for State1 and County1.

**Figure 11:** heat map of USA in counties values of Low Education 0 and 1 identifying its state of having 25% of low education of its population or not ( $r = 0.48$ )

Figure 11 displays a similar trend over the southern counties of the United States being uneducated, aligning with all the previous figures for educational factors. Overall, there is a trend that the more educated an individual is, the less likely they are to suffer from poverty, and vice versa. This could be due to the fact that a more educated individual is more likely to be employed, which reduces the risk of poverty.

### 3.2 Unemployment

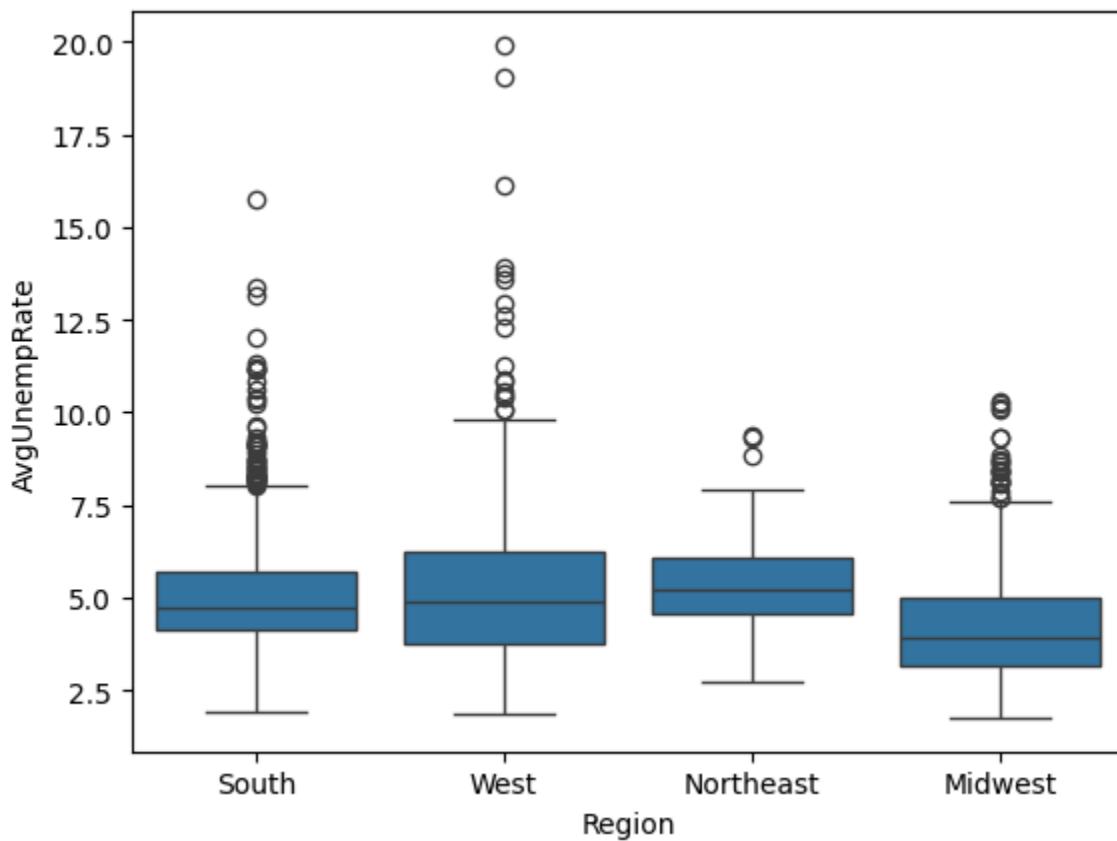
Initial Assumption: Higher unemployment rates correlate with higher poverty rates



**Figure 12:** scatter plot visualizing the relationship between Unemployment rates within counties VS poverty percentages ( $r = 0.5$ )

“**AvgUnempRate**” illustrates the number of unemployed people who were looking for work, 16 years and older, as a percent of the total county labor force using the average of gathered data starting from 2017 to 2021. This feature was not included in the initial dataset; it was created by taking the average of the unemployment rates from 2017 to 2021 for each state and county, to align with the five-year average rate for poverty primarily used in this study.

Figure 12 shows a clear relationship between unemployment and poverty within counties which supports the main claim. This makes sense, because being employed makes an individual less susceptible to falling into poverty, therefore as the unemployment rate rises, one can expect the poverty rate to rise simultaneously.



**Figure 13:** Boxplot for each of the four regions of the USA and their spread over unemployment rates.

As the southern states were proved to be the states which suffer from severe poverty, it is shown that there is almost no difference between the 4 regions having the similar spread with close medians. This means that while unemployment rate is a contributing factor to poverty as per Figure 12, it does not directly explain why southern states in particular suffer from higher poverty rates. High unemployment rates may have some indirect effect on the poverty rates in the southern states.

### 3.3 Race

Assumption: Dominance of a particular racial group (other than white) within a county may lead to a change in the county's poverty rate.



**Figure 14:** Scatter plots of a county's population of each prominent race in the US in relation to average poverty rates [14.a,  $r = 0.5$  | 14.b,  $r = -0.45$  | 14.d,  $r = 0.27$ , all after filtration above/below median accordingly]

**HispanicPct2020** shows the Percentage of county population that self-identifies as Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, or Other Asian and that does not self-identify as Spanish, Hispanic, or Latino; does not include people who report more than one race in 2020.

**BlackNonHispanicPct2020** shows Percentage of the county population that self-identifies as Black or African-American and that does not self-identify as Hispanic; does not include people who report more than one race in 2020.

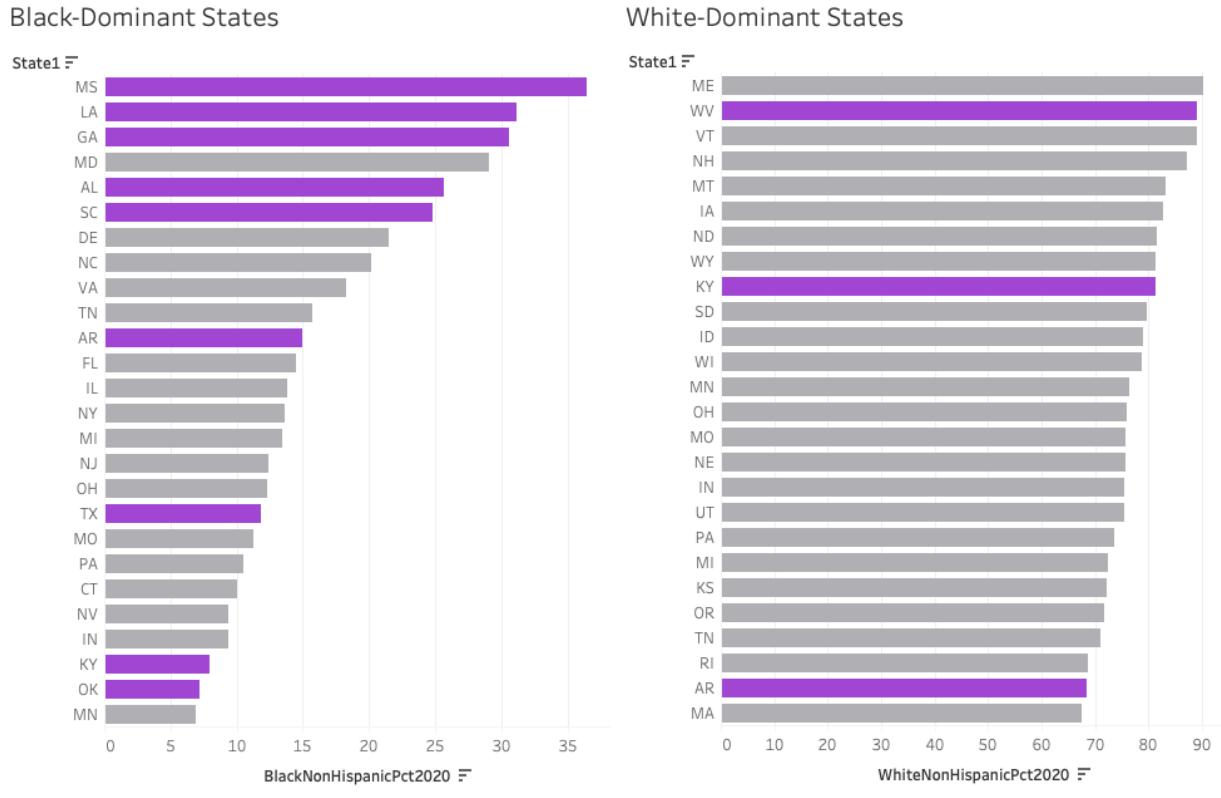
**WhiteNonHispanicPct2020** shows the percentage of county population that self-identifies race as White and that does not self-identify as Hispanic; does not include people who report more than one race in 2020.

**AsianNonHispanicPct2020** shows the percentage of county population that self-identifies as Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, or Other Asian and that does not self-identify Hispanic; does not include people who report more than one race in 2020.

To illustrate the strong correlation between White/Afro-American/Hispanics and poverty, values below the median of the black and hispanic population percentage were removed to erase the cluster of the observations with small black and hispanic populations to show how a large black and hispanic population affects poverty rates. This was also done for the white percentage too, values above the median were removed in this case to observe how counties with smaller white populations fare against poverty. No correlation was found for varying percentages of Asian populations, so analysis was stopped here:

As illustrated in figure 14, race showed some correlation with poverty in US counties:

- For Blacks and Afro-Americans, it showed a direct relation with poverty rates.
- For Whites, it showed an inverse relation with poverty rates.
- For Hispanics, a rather weak but direct relation with poverty rates was observed (weak, so analysis for the Hispanic population stopped here)



**Figure 15:** Bar plots identifying bottom countries in white population majority and top countries in black population majority.

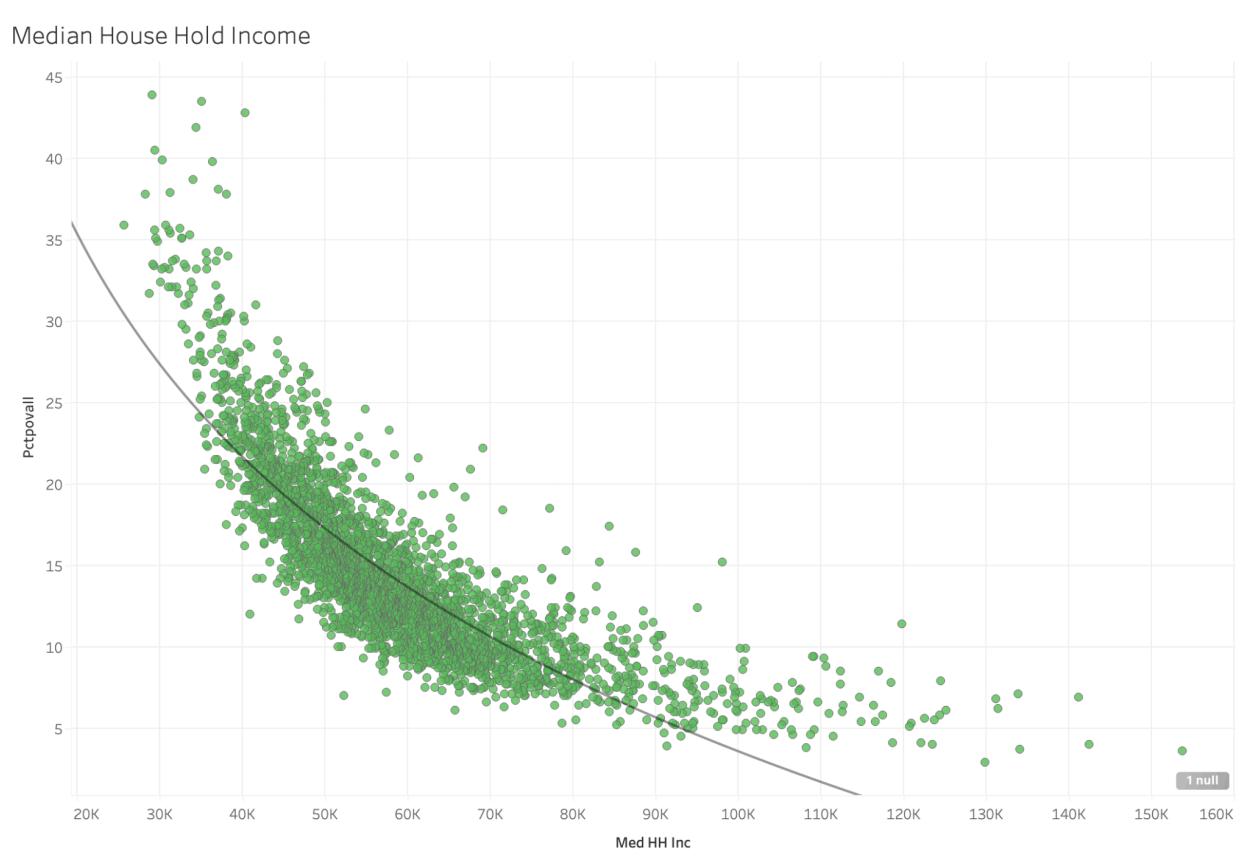
The bar plots in Figure 15 identify the states with the largest black/white population as a percentage of the total population, filtering on the top half of states. Five of the six southern states of interest dominate the top 6 largest black populations, whereas only three of the states make an appearance in the top 25 states in terms of white population.

Observations in Figures 14 and 15 would lead to the conclusion that the dominance of a particular race can have an effect on poverty rates, although not as strong as the previous two features in education and unemployment. Generally, the larger the black population, the higher the unemployment rate. This could be due to the fact that historically, black civilians face

exclusion from opportunities such as education and employment, leaving them more susceptible to falling into poverty.

### 3.4 Income

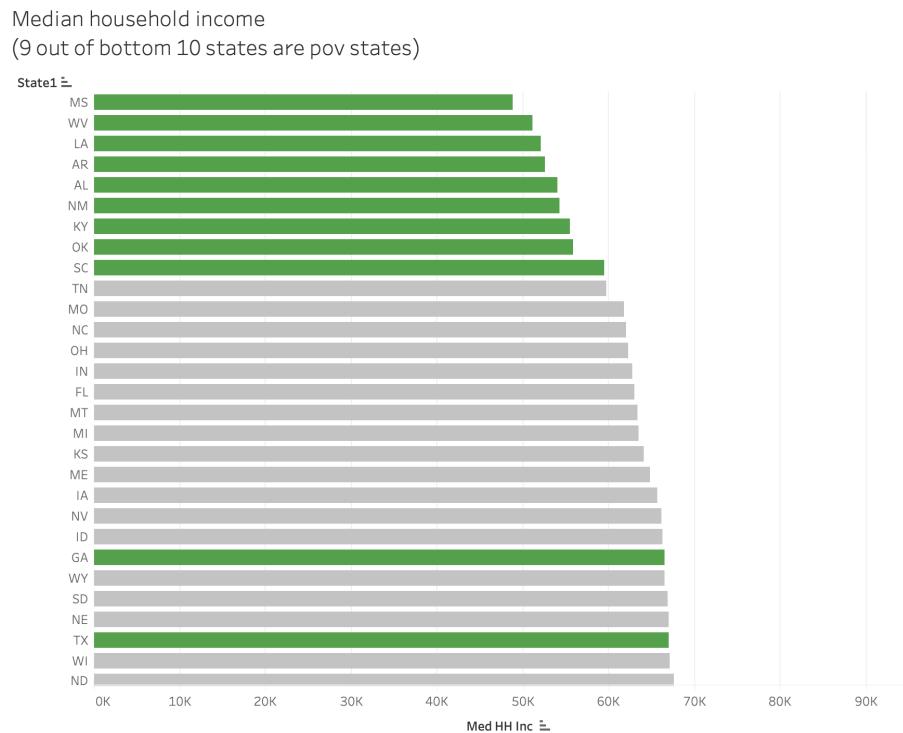
Assumption: Counties with a lower median household income are more likely to suffer from high poverty rates



**Figure 16:** Scatter plot for median household income on average poverty rates ( $r = -0.77$ )

**Med HH Inc** shows median income by household: Income level that divides county households in half, one half with income above the median and the other half with income below the median; includes income of all household members 15 years old or older.

Figure 16 indicates that there is an inverse relationship between a household's income, following a rather logarithmic trend.



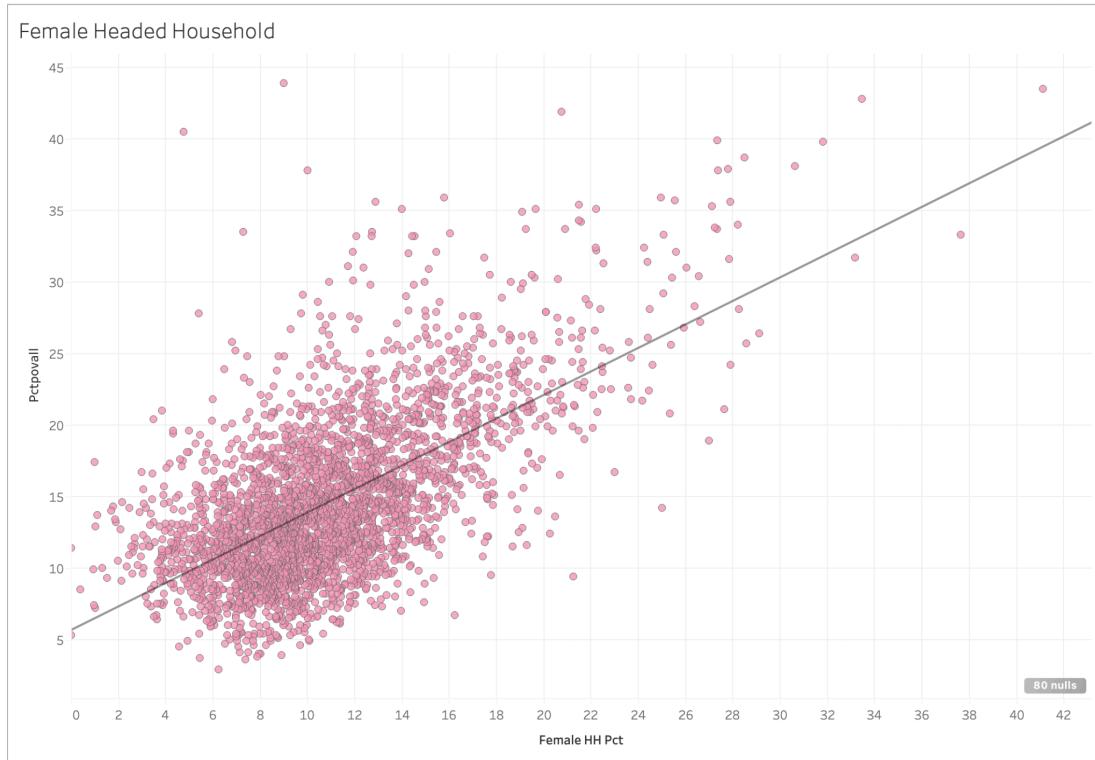
**Figure 17:** Identification of the states with the lowest median household income.

Figure 17 shows that nine out of the bottom 10 states with the lowest median incomes are part of the upper quartile of states with the highest poverty rates, further supporting the notion that income correlates inversely with poverty.

This is logical because being in a state of poverty is defined as having an income below a certain income threshold, along with the fact that having a low income makes it difficult to keep up with costs of living.

### 3.5 Family composition

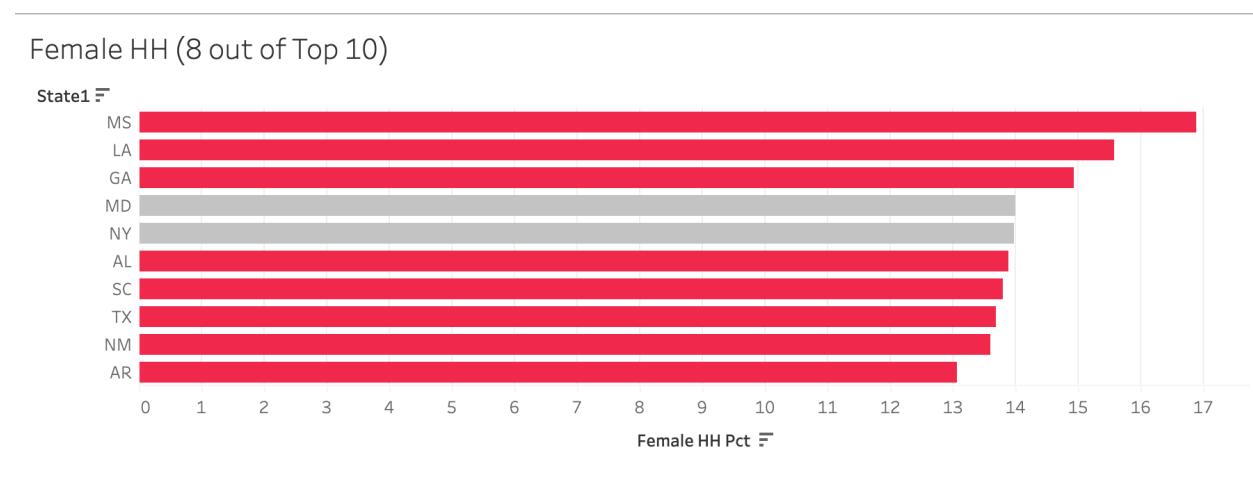
Assumption: Having incomplete families with only females heading them may lead to higher poverty rates.



**Figure 18:** Scatter plot of the percentage of female-headed households on average poverty rates ( $r = 0.63$ )

**Female HH pct** identifies the percentage of county households headed by a female with no husband present from 2017 to 2021.

Having a higher proportion of female headed households in a county shows a strong relationship with higher poverty rates, as per Figure 16.

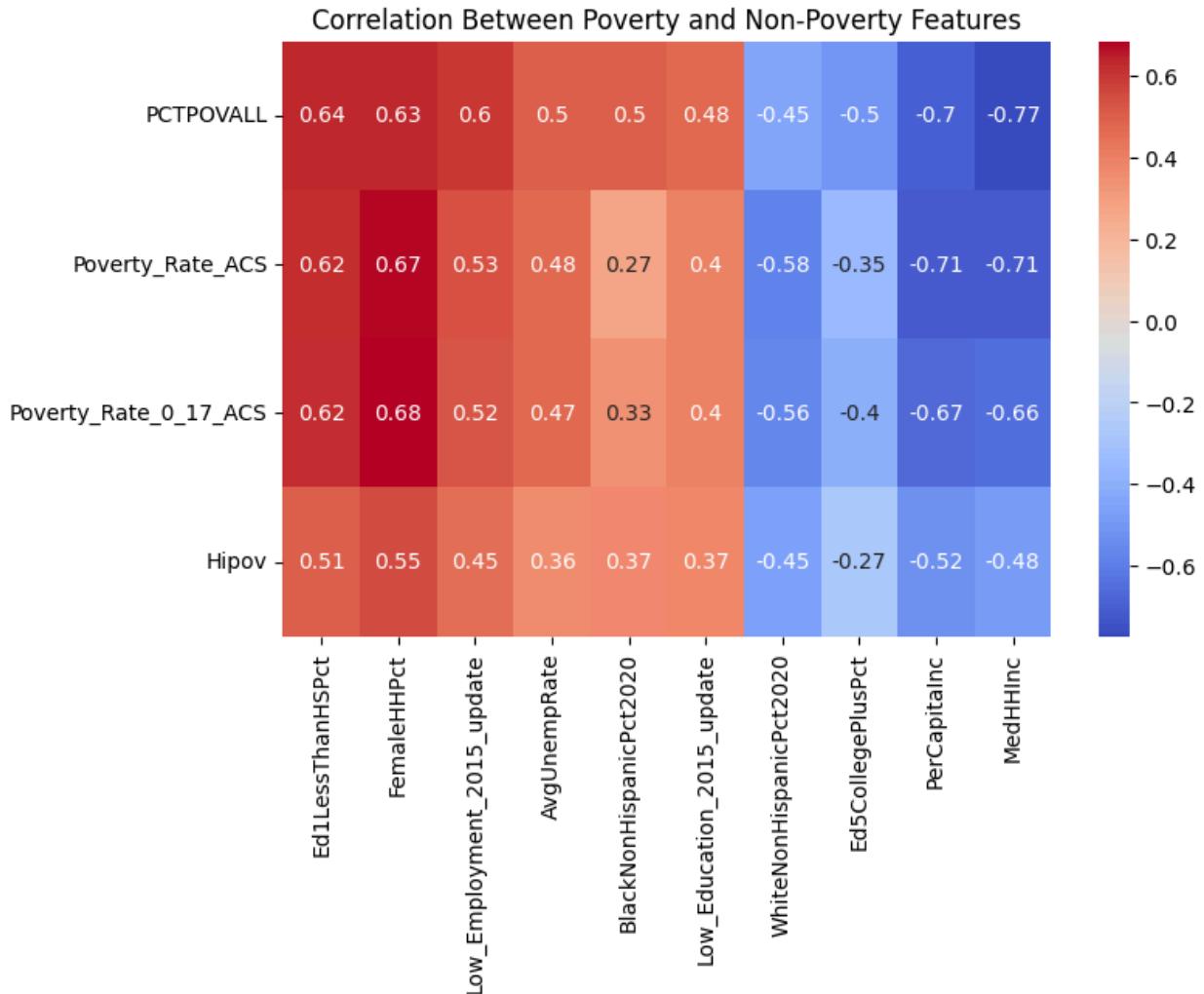


**Figure 19:** Identifies the top 10 states having the highest percentage of female-headed households

As illustrated in the Figure, 17 out of the top 10 states with the highest percentage of female-headed households, 8 states are from the upper quartile of states in terms of average poverty rate, which further ensures the aforementioned claim.

This is the case for several reasons related to income, including the gender pay gap, which sees women get paid less, on average, than men. The observed trend may also be due to the fact that the household's income is likely only coming from one source in the wife's income, rather than a standard household that may see income coming from both the wife's and husband's jobs, bringing the household income down, which as previously discussed, leads to higher poverty rates.

### 3.6 Correlations between poverty and non-poverty factors (just as a reference)



**Figure 20:** Correlation matrix between poverty measures and independent variables of interest

Figure 20 ensures that the correlations we identified are statistically significant to some extent (Poverty\_Rate\_ACS and Poverty\_Rate\_0\_17\_ACS are two additional measures of poverty which measure the general poverty over the same period of time by the ACS, and among all people under 18, respectively).

#### 4. Conclusion

Poverty is one of the most horrific challenges that faces humanity globally. 5 aspects were hypothesized to correlate with poverty rates in the US which have been explored in depth. Using multiple variables and visualizing methods, we found that the reason southern states suffer from higher poverty rates than other regions is due to the following:

- **Education:** The more uneducated individuals in a county, the higher the expected poverty rate. The southern states appeared to dominate the top states with the highest percentage of populations without a high school diploma or equivalent.
- **Dominance of a Race:** States with a relatively large Black population had higher poverty rates, and White majority counties/states had lower rates of poverty. Dominance of other races showed no significant results. Most of the southern states had an above average Black population.
- **Income:** Low income households are more susceptible to poverty. Southern states had the lowest median household incomes, on average.
- **Family Composition:** States with the highest proportion of female-headed households experienced the highest poverty rates. Most of the southern states had an above average proportion of female-headed households.
- **(Unemployment:** While it was observed that higher unemployment rates do lead to higher poverty rates, there was no direct evidence that higher unemployment rates lead to higher poverty rates in southern states in particular, as this was a general trend across all regions)

## 5. References

- Center for American Progress. (n.d.). *Poverty data map tool*. Retrieved December 8, 2024, from  
<https://www.americanprogress.org/data-view/poverty-data/poverty-data-map-tool/>
- U.S. Census Bureau. (2024). *About the American Community Survey (ACS)*. Retrieved December 8, 2024, from  
<https://www.census.gov/programs-surveys/acs/about.html>
- U.S. Census Bureau. (2021). *Guidance for Economic Census Geographies Users - Geographic Levels*. Retrieved December 8, 2024, from  
[https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html#par\\_textimage\\_30](https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html#par_textimage_30)
- United States Department of Agriculture Economic Research Service. (2023). *Atlas of Rural and Small-Town America*. Retrieved December 8, 2024, from  
<https://www.ers.usda.gov/data-products/atlas-of-rural-and-small-town-america/>