2024

Poverty Rate Across US Counties

STAT 530- Applied Regression Analysis

PROFESSOR: KESHAV POKHREL

GROUP MEMBERS:

1. ANISH KOLAPARTHI
2. DEVRAJ AMIN
3. POOJA GURAV

Introduction

Poverty is an enduring and intricate socioeconomic issue that surpasses geographical limits and impacts individuals, families, and communities globally. Although there have been notable improvements in economic development and social welfare initiatives, poverty continues to be a critical concern, affecting millions of individuals worldwide who live below the poverty threshold. Gaining insight into the fundamental factors that cause poverty is essential for policymakers, academics, and practitioners who want to adopt successful interventions and policies to reduce poverty and promote sustainable development.

Historically, studies on poverty have primarily concentrated on metrics based on money, such as the poverty line determined by household earnings or levels of consumption. However, poverty is a complex phenomenon that is shaped by a multitude of circumstances that extend beyond just wealth. Factors such as education, work, housing, health, and access to basic services have a significant impact on how susceptible individuals and communities are to poverty. Hence, it is crucial to adopt a comprehensive approach that taken into account the complex interaction of several socio-economic factors in order to gain a thorough knowledge of poverty and successfully address it. Researchers can acquire a comprehensive picture of the social, economic, and health dynamics influencing communities across the United States by evaluating such aspects in conjunction with demographic and economic data. This nuanced perspective is vital for educating evidence-based decision-making and encouraging equitable development policies at the local, state, and national levels.

Poverty is a continuing concern that requires a greater understanding of its determinants and practical measures for reduction despite major worldwide efforts and investments in its eradication. Intricate socioeconomic elements impacting poverty outcomes are frequently ignored in existing studies, and typical income-based measurements do not fully represent the complexity of poverty dynamics. Our report effort does a thorough regression analysis of poverty rates within United Nations (UN) counties in 2017 to remedy these disparities. We investigate the associations between poverty rates and socioeconomic variables such as unemployment rates, median household income, per capita income, metropolitan status, and median education level by using multiple linear regression and logistic regression approaches. Utilizing an extensive dataset that includes demographic trends, housing attributes, labor market metrics, and sociocultural elements, our research endeavors to identify noteworthy determinants of poverty and furnish empirically supported suggestions for policymakers and practitioners involved in poverty mitigation endeavors worldwide.

Among the varied landscapes of the United States, the prevalence of poverty stands out as a critical aspect of community narratives. From bustling urban centers to quiet rural corners, economic disparity shapes the experiences of residents. Factors such as homeownership rates, unemployment, and educational attainment intertwine to influence poverty levels across counties. While metropolitan areas offer abundant job prospects, rural regions face greater challenges in economic development. However, amidst these differences, there's room for optimism. Initiatives aimed at reducing poverty, such as comprehensive smoking bans and investments in education and employment, demonstrate a commitment to fostering resilient communities across the nation.

Data Description

The dataset comprises information from 3142 counties across the United States. It offers a comprehensive overview of various socio-economic indicators and demographic factors for each county.

The United States Counties dataset is a valuable resource for studying the diversity and dynamics of American communities. Researchers can analyze demographic transitions and identify locations that are growing or declining by analyzing population changes from 2000 to 2017. Poverty and homeownership rates are included in the dataset, providing insight into county residents' economic well-being and housing stability. Furthermore, unemployment rate data provides a snapshot of local labor market circumstances, which helps to assess employment prospects and economic vibrancy across regions. This extensive demographic

and economic data allow analysts to identify areas of need, focus interventions, and assess the efficacy of policies targeted at improving socioeconomic outcomes.

Furthermore, the information provides insight into other elements that influence community well-being and quality of life. The identification of counties that include metropolitan regions offers context for urbanization patterns and economic activity concentrations. Median education levels and per capita income serve as indices of human capital and economic prosperity in each county. Furthermore, the dataset's inclusion of smoking ban status reflects public health activities and regulations targeted at lowering tobacco use and increasing population health outcomes. Researchers can acquire a comprehensive picture of the social, economic, and health dynamics influencing communities across the United States by evaluating such aspects in conjunction with demographic and economic data. This nuanced perspective is vital for educating evidence-based decision-making and encouraging equitable development policies at the local, state, and national levels.

The data set contains 10 Numerical variables:

- pop2000: Population in 2000.
- pop2010: Population in 2010.
- pop2017: Population in 2017.
- pop_change: Population change from 2010 to 2017.
- poverty: Percentage of the population in poverty in 2017.
- homeownership: Home ownership rate for the period 2006-2010.
- multi_unit: Percentage of housing units in multi-unit structures for the period 2006-2010.
- unemployment_rate: Unemployment rate in 2017.
- per_capita_income: Per capita (per person) income for the period 2013-2017.
- median_hh_income: Median household income.

The data set also contains 5 Categorical Variables:

- name: County names.
- state: State names.
- smoking_ban: Describes the type of county-level smoking ban in place in 2010, with values "none", "partial", or "comprehensive".
- metro: Indicates whether the county contains a metropolitan area.
- median_edu: Median education level for the period 2013-2017

Summary Statistics

The table provides statistical data on various demographic and socio-economic indicators for different years, ranging from 2000 to 2017. It includes population figures for 2000, 2010, and 2017, showcasing population changes over time. Additionally, it features metrics such as poverty rates, homeownership rates, unemployment rates, per capita income, median household income, and a poverty high indicator. The data highlights considerable variations across these indicators, with significant changes observed in population, poverty rates, and home ownership percentages. The table offers insights into the dynamics of population growth, economic conditions, and social well-being over the specified period.

	pop2000	pop2010	pop2017	pop_change	poverty	Homeownership	multi_unt	unemployment_rate	per_capita_income	median_hh_income
Min.	444	460	457	-33.63	2.4	22.8	0	1.62	10467	19264
1stQu.	10638	10803	10653	-1.94	11.6	69.2	5.9	3.507	21494	40598
Median	23590	24767	24968	-0.01	15.6	74.4	9.25	4.33	24926	47132
Mean	84569	93788	99575	5.836	16.28	73.25	11.9	4.6	25758	49047
3rdQu.	56909	63126	64113	2.4925	19.8	78.44	15.4	5.312	29036	55109
Max.	9519338	9818605	10163507	37.19	48.2	91.3	98.5	19.07	69533	55109
SD	285387.5	310609.5	333062.1	4.1922	6.5430	7.5835	8.89738	1.67541	6207.778	12942.53
Variance	81446050	96478251	110930367	17.5746	42.8115	57.51058	79.16347	2.807	38536505	167509083

Table 1.1 – Descriptive Statistics

Literature Review

The United States Counties dataset provides rich insights into socio-economic dynamics, public health outcomes, and regional disparities across American counties. Researchers leverage population variables like pop2000, pop2010, and pop2017 to explore demographic shifts and migration patterns. Economic indicators such as poverty rates (poverty) and unemployment rates (unemployment_rate) illuminate income inequality and disparities in job access. Homeownership rates (homeownership) and multi-unit housing prevalence (multi_unit) reveal housing affordability and urbanization trends, crucial for understanding community dynamics. The metropolitan designation (metro) distinguishes areas of economic development, while education levels (median_edu) and income metrics (per_capita_income, median_hh_income) provide insights into socio-economic status. The smoking ban variable (smoking_ban) adds a public health dimension, reflecting policy effectiveness in mitigating smoking-related health risks. Integrating global perspectives from the United Nations (UN) and other sources enriches analyses, fostering a comprehensive understanding of socio-economic landscapes and informing policymaking for American counties and communities.

Research Question

What is the relationship between various socio-economic indicators such as poverty rate, unemployment rate, per capita income, homeownership rate, and population change across different states and metropolitan areas?

Exploratory Data Analysis (EDA)

The following graphs illustrates the distribution of numerical variables:

Poverty: Higher bars indicate a concentration of counties with lower log-transformed poverty rates, suggesting a skewed distribution towards lower poverty levels. The histogram exhibits left skewness, indicating a greater concentration of data points on the left side.

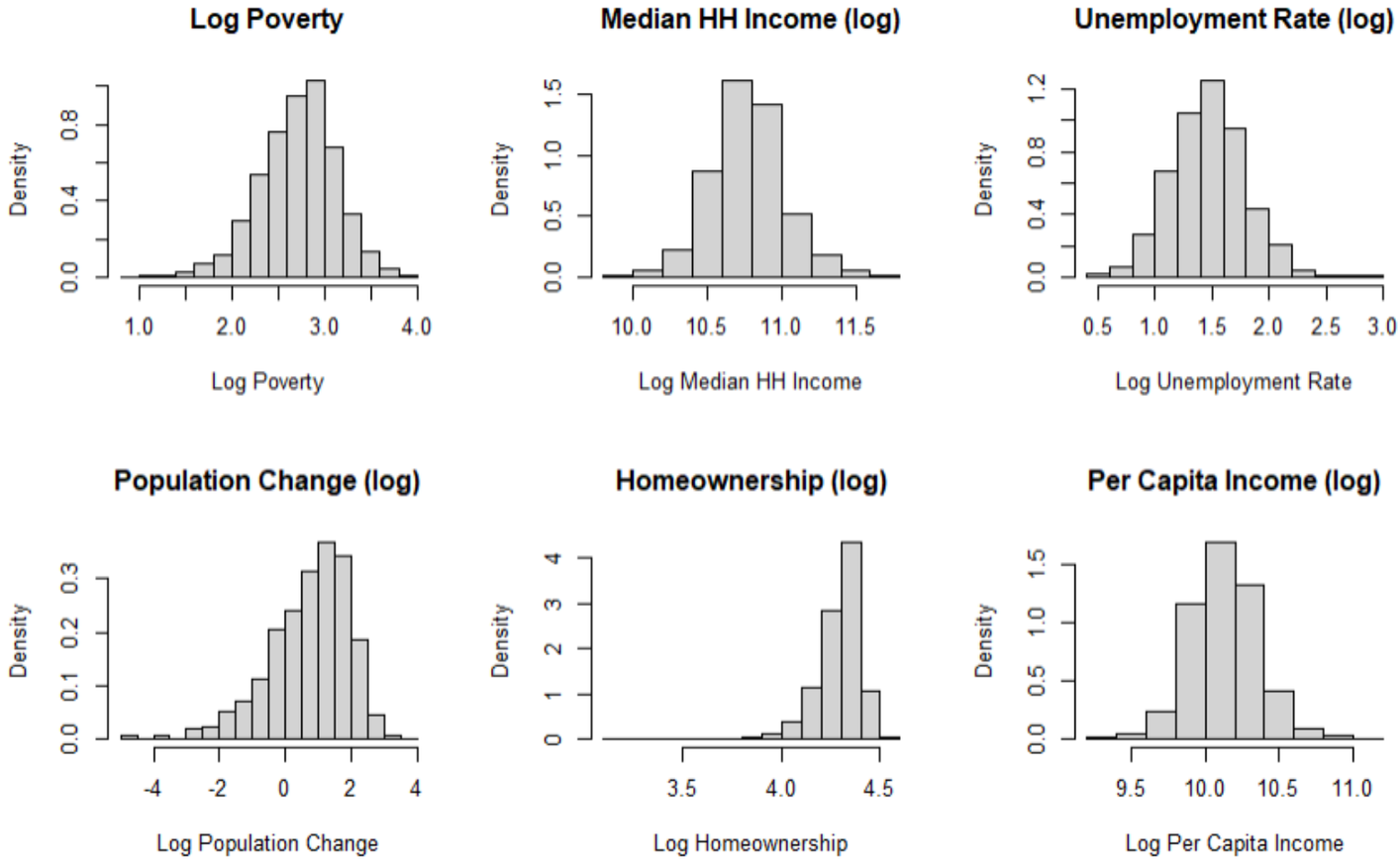
Median HH Income: Concentration of counties with higher log-transformed median household incomes, potentially indicating a more even distribution across different income levels. The histogram demonstrates symmetry, with an even distribution of data points on both sides.

Unemployment Rate: Elevated bars at lower log-transformed values suggest a concentration of counties with lower unemployment rates, indicating a skewed distribution towards lower unemployment levels. The histogram displays right skewness, with a higher density of data points on the right side.

Population Change: Higher bars at lower log-transformed values imply a concentration of counties experiencing population growth, suggesting a skewed distribution towards areas with positive population change. The histogram exhibits left skewness, indicating a greater concentration of data points on the left side.

Homeownership: Concentration of counties with higher log-transformed homeownership rates, potentially indicating a more even distribution across different homeownership levels. The histogram exhibits left skewness, indicating a greater concentration of data points on the left side.

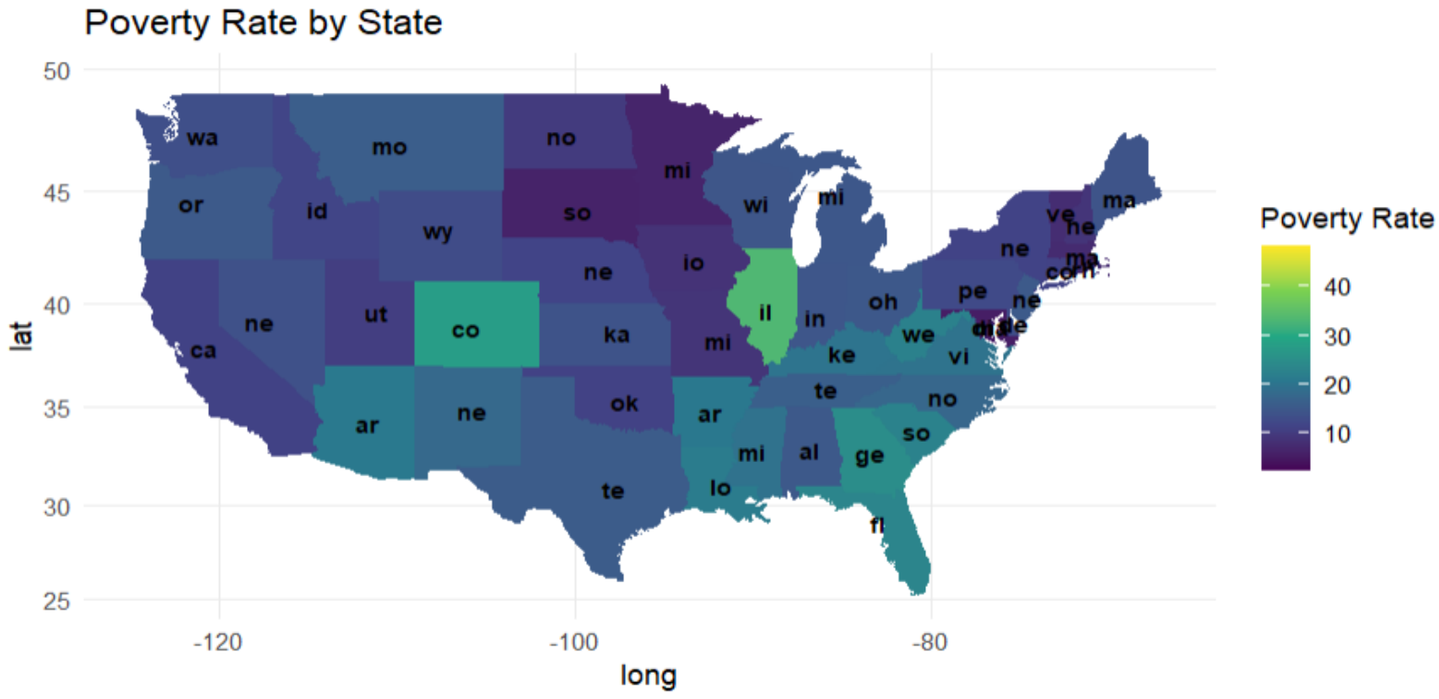
Per Capita Income: Elevated bars at higher log-transformed values suggest a concentration of counties with higher per capita incomes, indicating a skewed distribution towards higher income levels. The histogram demonstrates symmetry, with an even distribution of data points on both sides.



Graph 1.1 – Histograms for distribution across numerical variables

A Geographic Overview of Poverty Rates in US Counties:

The map for poverty plotted according to states shows variations in poverty levels across different states. Lighter colors on the map indicate higher poverty rates in those states, while darker colors indicate lower poverty rates. This color gradient provides a visual representation of the spatial distribution of poverty across the United States, allowing for easy identification of regions with relatively higher or lower poverty levels. By observing the map, viewers can quickly discern patterns and disparities in poverty rates among states, facilitating comparisons and highlighting areas where poverty is particularly prevalent or less prevalent. Overall, the map serves as a useful tool for visualizing and understanding the geographic distribution of poverty within the United States.

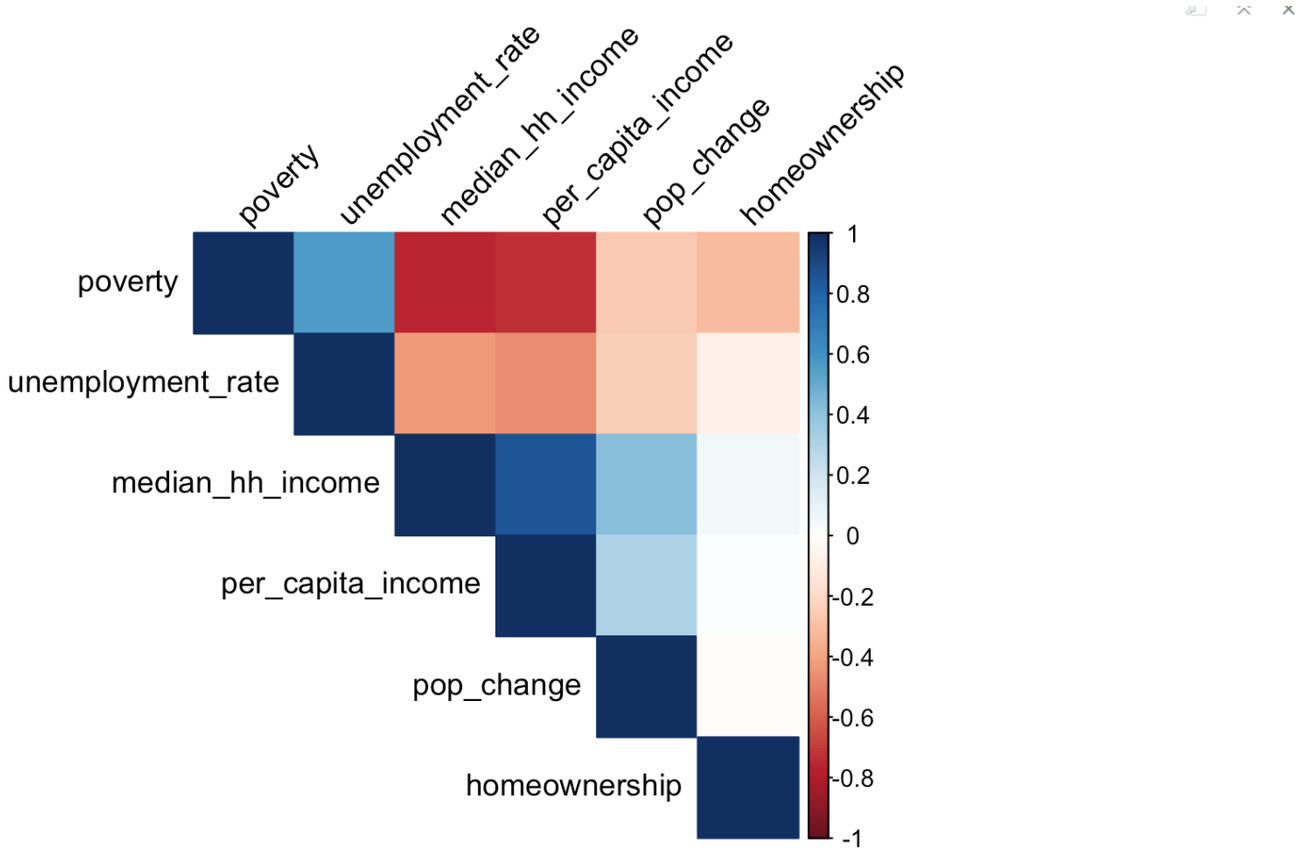


Graph 1.2 – Visualization of poverty rate across US counties

Correlation Matrix between the numeric variables:

The correlation matrix that is displayed in the below image offers a thorough examination of the connections between different variables that are available in the dataset. Important insights can be gained from the color-coded relationships of directionality and intensity. There is a clear negative correlation between poverty and both per capita and median household income, indicating that lower rates of poverty are linked to greater incomes. On the other hand, the unemployment rate demonstrates a negative link with per capita income and a positive correlation with poverty, suggesting that more unemployment is associated with more severe financial difficulties. While homeownership and population changes over time show more complex associations with the other

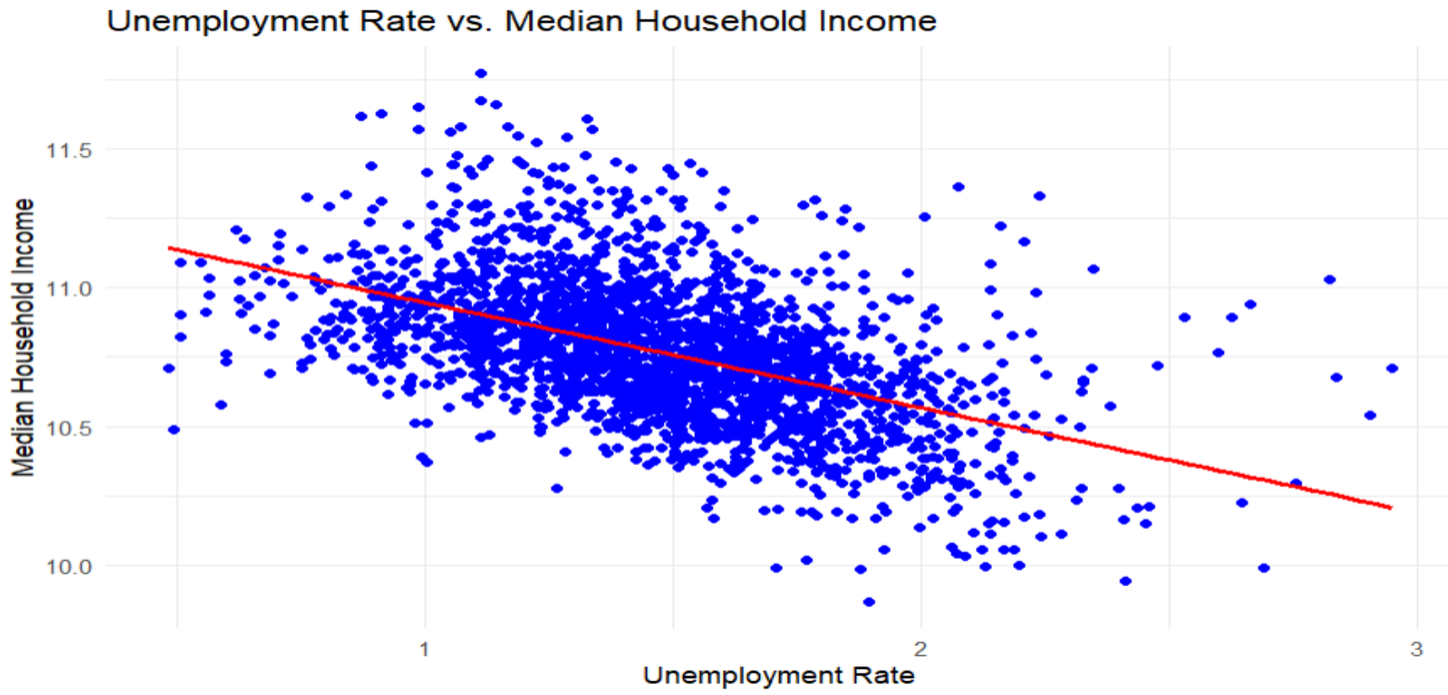
variables, median household income shows a positive link with per capita income. socioeconomic dynamics of a particular setting and guide the formulation of public policies meant to address problems like income inequality, unemployment, and poverty.



Graph 1.3 – Correlation matrix between the variables

Scatter plot between unemployment rate and median household income:

In simple terms, this statement suggests that when unemployment rates go up, median household income tends to go down. In other words, there is a relationship between higher levels of unemployment and lower household incomes. This could be because when more people are out of work, there may be fewer job opportunities and less income coming into households, leading to a decrease in median household income. It reflects the economic challenges that arise when unemployment rises, impacting the financial well-being of families and communities.



Graph 1.4 – Scatter plot between unemployment rate vs median household income

Calculating the average population change from 2010 to 2017:

This means that from 2010 to 2017, the population of the areas studied increased by an average of 58.4%. This indicates a significant growth in the number of people living in those areas over that period. Such a substantial increase in population suggests factors such as migration, births, and possibly economic development contributing to the growth of these areas. Overall, it highlights the trend of population expansion within the studied regions during that seven-year period.

Convert 'poverty' variable to numeric:

```
county$poverty <- as.numeric(as.character(county$poverty))
```

Filter counties with poverty rate greater than 12%:

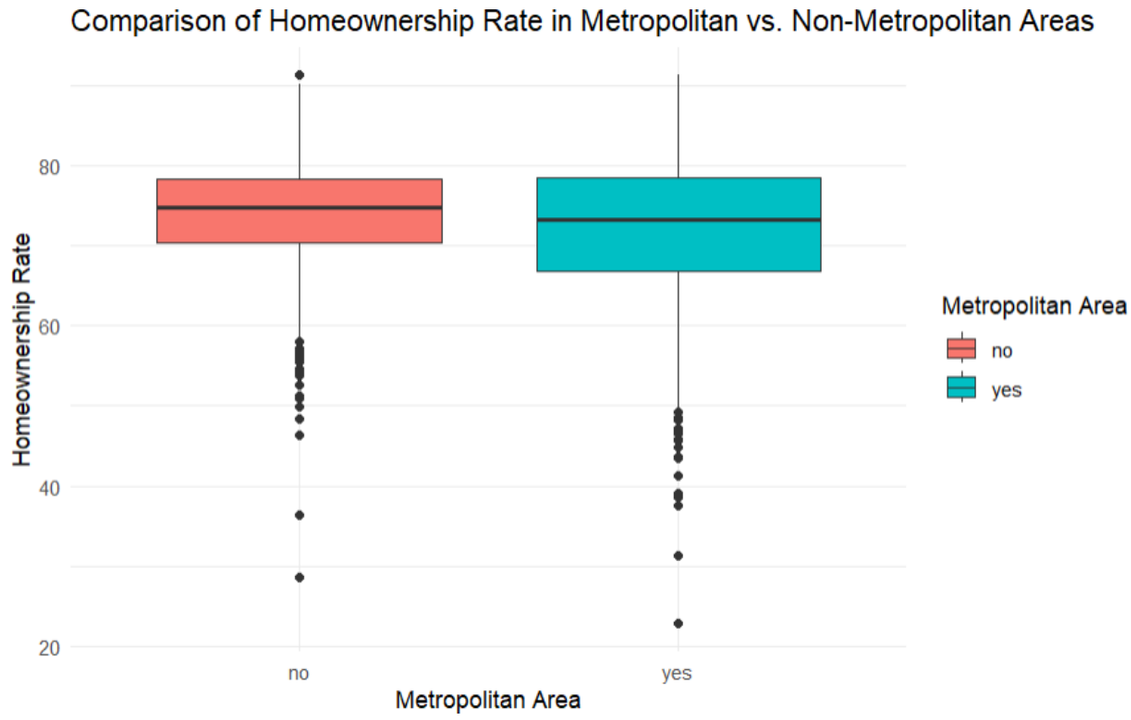
```
high_poverty_counties <- county %>% filter (poverty > 12)
```

In simple terms, this instruction tells us to identify counties where the poverty rate is greater than 12% and assign them to a group called "high_poverty_counties." Essentially, we're looking for counties where more than 12 out of every 100 people are living in poverty. Once we identify these counties based on their poverty rates, we'll group them together under the label "high_poverty_counties" for further analysis or consideration. This helps us focus on areas where poverty is more prevalent and understand the specific challenges or needs, they may face.

Comparison of Homeownership Rate in Metropolitan vs. Non-Metropolitan Areas:

The homeownership rates in each of the dataset's counties are contrasted between metropolitan and non-metropolitan areas using a boxplot visualization. The two categories, "Metropolitan" and "Non-Metropolitan," on the x-axis, show whether a county is home to a metropolitan region. The percentage of households in each county that own their homes is represented by the homeownership rate on the y-axis.

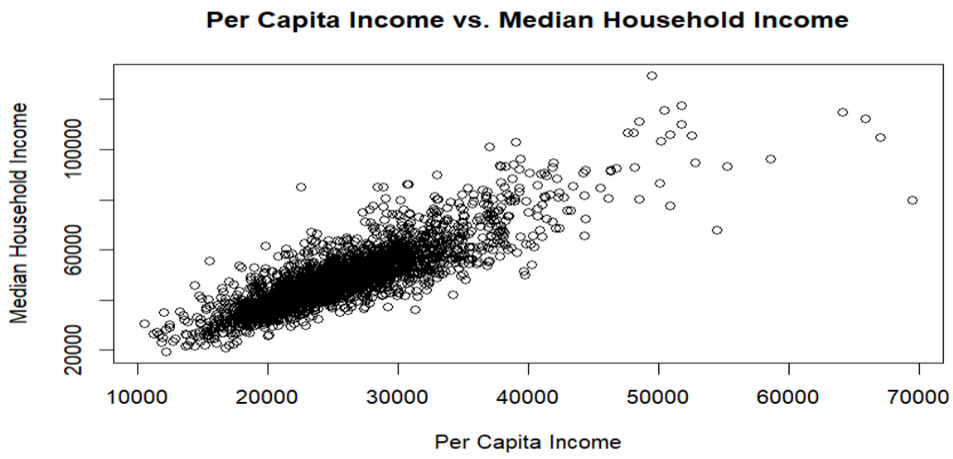
The distribution of homeownership rates within each group is visible in the boxplot. In comparison to non-metropolitan areas, the boxplot for metropolitan areas indicates a marginally lower median homeownership rate. Furthermore, compared to non-metropolitan areas, the interquartile range (the box) appears to be shorter in metropolitan areas, suggesting lower fluctuation in homeownership rates. There are some outliers in both categories, but the whiskers reach the lowest and highest values within 1.5 times the interquartile range from the lower and upper quartiles, respectively.



Graph 1.5 – Box plot of homeownership in metro vs non-metro areas

Capita Income vs. Median Household Income

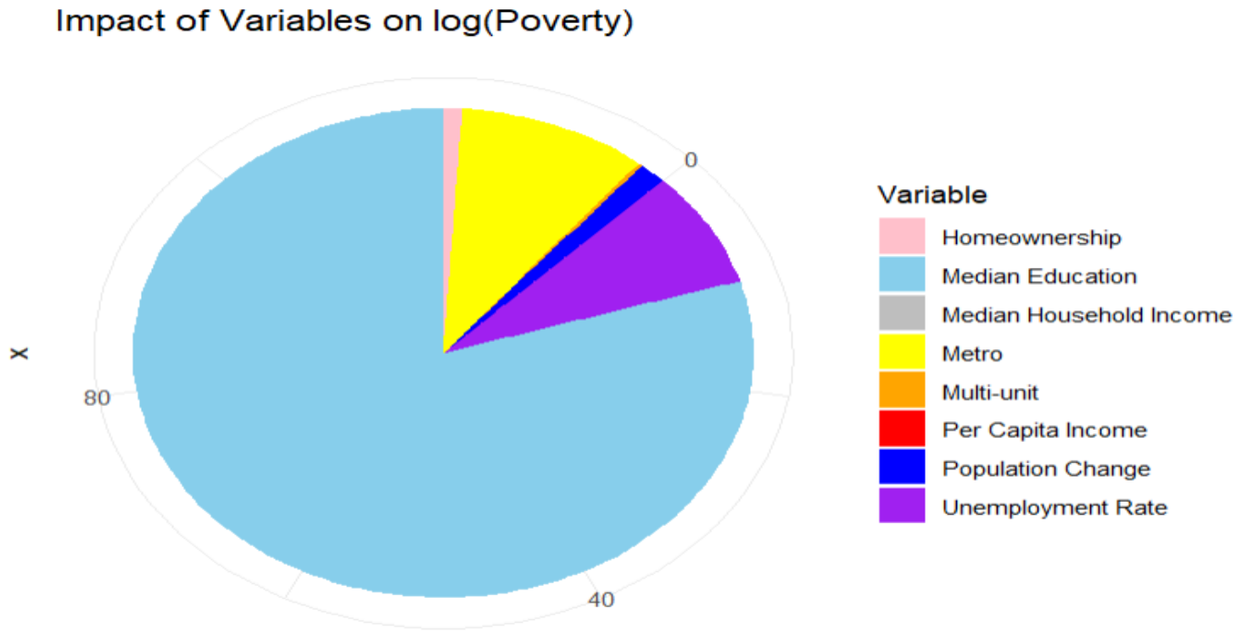
The scatter plot illustrates the correlation between the median household income and per capita income for each county in the dataset. The scatter plot shows that the median household income and per capita income have a positive correlation. The rising trend of the plot's points indicates that counties with higher per capita income also typically have higher median household incomes. This implies that people generally have greater household incomes when living in counties with higher per capita income levels.



Graph 1.6 – Scatter plot of per capita income vs median household income

Data Visualization and Variable Impact:

Given the focus on understanding socioeconomic factors, poverty serves as a comprehensive measure encapsulating various aspects of economic hardship and deprivation within a population. Additionally, poverty is inherently tied to the other variables in the dataset, such as population change, unemployment rate, median household income, homeownership rate, metro, and per capita income.



Graph 1.7 – Pie chart of variables that shows the impact on the target variable

The proportionate impact of several predictors on the logarithm of poverty (log (Poverty)) is displayed in a pie chart. In terms of explaining the relative importance of variables such as metropolitan status, education level, household income, unemployment rate, population change, homeownership rate, multi-unit housing, and per capita income in explaining poverty levels within the dataset, each segment outlines the percentage contribution of a predictor variable to the overall variance in log (Poverty). Particularly, the unemployment rate, metropolitan status, and median education stand out as having the most effects on poverty levels.

Methodology

Model Selection:

1. Multiple linear regression with Interaction:

Model 1 (Baseline): $\widehat{Poverty} = \hat{\beta}0 + \hat{\beta}1$Median_edu

Model 2: $\widehat{Poverty} = \hat{\beta}0 + \hat{\beta}1$Median_edu+ $\hat{\beta}2$Median_hh_income

Model 3: $\widehat{Poverty} = \hat{\beta}0$ + β^1Median_edu + $\hat{\beta}2$Median_hh_income + $\hat{\beta}3$unemployment_rate

Model 4: $\widehat{Poverty} = \hat{\beta}0$ + β^1Median_edu + $\hat{\beta}2$Median_hh_income + $\hat{\beta}3$unemployment_rate + $\hat{\beta}4$metro

Model 5: $\widehat{Poverty} = \hat{\beta}0 + \widehat{\beta}1$Median_edu + $\hat{\beta}2$Median_hh_income + $\hat{\beta}3$unemployment_rate + $\hat{\beta}4$metro + $\hat{\beta}5$per_capita_income

	R_squ_adjusted	MSE
Model_1	0.2398	0.13
Model_2	0.6604	0.058
Model_3	0.6999	0.051
Model_4	0.7171	0.049
Model_5	0.7281	0.047

Table 1.2 – Summary of multiple linear model

The models progressively improve in explanatory power and accuracy, with Model_5 demonstrating the highest adjusted R-squared and lowest mean squared error (MSE), indicating superior overall performance among the presented models.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	4.375	6.186e-02	70.721	< 2e-16
Median_edubelow_hs	-7.266	3.752e+00	-1.937	0.00529
median_eduhs_diploma	-5.48E-01	4.29E-02	-12.786	< 1.8E-16
median_edusome_college	-5.64E-01	3.99E-02	-14.116	< 8.69E-09
median_hh_income	-2.01E-05	6.65E-07	-30.208	< 3.6E-16
unemployment_rate	4.63E-02	2.93E-03	15.809	< 2.00E-16
metroyes	1.16E-01	9.79E-03	11.849	< 2.00E-16
per_capita_income	-1.51E-05	1.48E-06	-10.23	< 2.03E-16
Residual standard error: 0.2159 on 2552 degrees of freedom				
Multiple R-squared: 0.7289			Adjusted R-squared: 0.7281	
F-statistic: 980.2 on 7 and 2552 DF			p-value: < 2.2e-16	

Table 1.3 - Summary statistics for model 5

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
median_edu	3	105.639	35.213	755.31	< 2.20E-16
median_hh_income	1	184.396	184.396	3955.27	< 2.30E-16
unemployment_rate	1	17.36	17.36	372.38	< 1.20E-16
metro	1	7.597	7.597	162.96	< 2.20E-16
per_capita_income	1	4.879	4.879	104.65	< 2.20E-16
Residuals	2552	118.975	0.047		

Table 1.4 – ANOVA statistics for model 5

We build our final model using unemployment_rate, per_capita_income, median_hh_income, metro and median_edu. The regression equation comes out as:

Log(poverty)_hat = 4.375 + 0.04631* unemployment_rate – 0.00001511* per_capita_income – 0.00002008* median_hh_income + 0.116*metroyes – 0.6422*median_edu_below_hs – 0.5479*median_edu_hs_diploma – 0.5636*median_edu_some_college

The p-value for entire model is less than 2.2×10^{-16}, which is strong evidence that these predictors together are contributing to the response variable. The R-squared value of the model is 0.7289, indicating that the model accounts 72.89% of the variance in poverty rate. The adjusted R – square value of the model is 0.7281, which ensures that our model was not overfit, as it is still very high.

The interpretation of the coefficients for the provided linear regression model is as follows:

- **Intercept:** The intercept is approximately 4.375. This represents the expected value of the log(poverty) when all other predictor variables are set to zero.
- **Median Education Below High School:** For every unit decrease in median education level below high school, the log(poverty) is expected to decrease by approximately 0.642.
- **Median Education High School Diploma:** For every unit decrease in median education level with a high school diploma, the log(poverty) is expected to decrease by approximately 0.547.
- **Median Education Some College:** For every unit decrease in median education level with some college, the log(poverty) is expected to decrease by approximately 0.564.
- **Median Household Income**: For every unit increase in median household income, the log(poverty) is expected to decrease by approximately 0.00002008.
- **Unemployment Rate**: For every unit increase in the unemployment rate, the log(poverty) is expected to increase by approximately 0.04631.
- **Metropolitan Area (Yes/No)**: Being in a metropolitan area (Yes) is associated with an increase in log(poverty) by approximately 0.116.
- **Per Capita Income:** For every unit increase in per capita income, the log(poverty) is expected to decrease by approximately 0.00001511.

Hypothesis Testing:

At a significance level of $\alpha = 0.05$, our null and alternative hypothesis is:

Null Hypothesis (H0): $\beta\hat{5} = 0$

Alternative Hypothesis (HA): $\beta\hat{5} > 0$

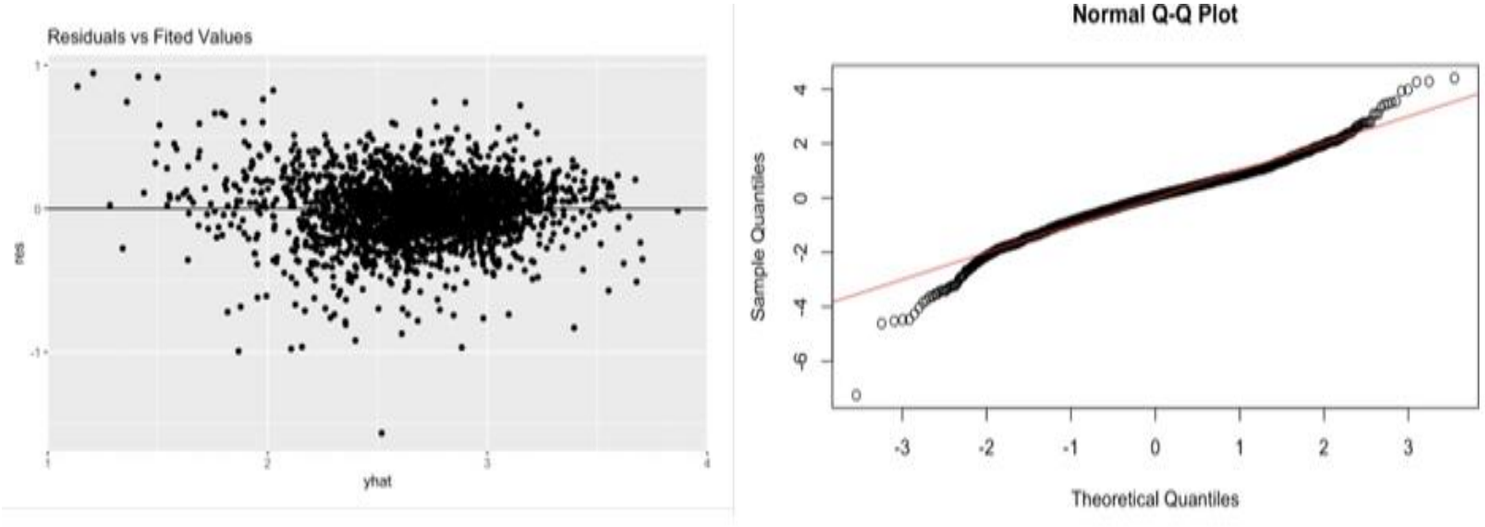
In this case, the p-value (Pr(>F)) for the overall model is less than 0.05 (< 2.20E-16), suggesting that we reject the null hypothesis. Therefore, we conclude that the regression model, which includes median_edu, median_hh_income, unemployment_rate, metro, and per_capita_income as predictor variables, significantly explains the variability in the outcome variable.

Residuals vs Fitted Values graph:

This graph shows the residuals (the differences between the actual values and the predicted values) plotted against the fitted values (the predicted values). This plot is used to check for patterns or trends in the residuals, which can indicate violations of the assumptions of the statistical model being used. In this case, the plot shows a scattered distribution of points, with no clear pattern or trend, suggesting that the assumptions of the model are likely met.

QQ Plot:

QQ Plot used to assess whether the residuals (or the data) follow a normal distribution. The theoretical quantiles (the expected values if the data were normally distributed) are plotted on the x-axis, and the sample quantiles (the actual values) are plotted on the y-axis. The points should fall on a straight line if the data has a normal distribution. The fact that the points in this instance nearly resemble the straight line suggests that the residuals, or the data, are roughly normally distributed.



Graph 1.8 – Residuals vs fitted and Normal QQ Plot

2. Logistic Regression:

Prediction Accuracy Using Confusion Matrix:

The confusion matrix is like a summary that shows how well a logistic regression model performs based on certain predictors. By setting a threshold value of 12 percent, anything above it is classified as representing poverty (labeled as 0), and anything below represents wealth (labeled as 1). This method, using data from the United Nations, allows for a clear distinction between poor and wealthy categories, making it easier to analyze and make decisions. We transformed numeric poverty data into categorical data for logistic regression, using reference from the US census data, which stated that the average poverty rate in the US in 2017 was between 11.7% and 12.3%.

As measured by the percentage of correctly categorized examples (True Positives and True Negatives) relative to all instances, the accuracy of the model is roughly 82.42%. Although accuracy gives an overview of the model's performance, other metrics like precision, recall, and F1-score must also be taken into account to assess the model's performance more thoroughly, particularly when dealing with imbalanced datasets.

In conclusion, the accuracy measure and confusion matrix offer insightful information about how well the logistic regression model predicts high and low poverty rates depending on the chosen factors. Additional performance measures can be analyzed and evaluated to get a more comprehensive picture of the model's efficacy and possible areas for development.

Actual/Predicted	***Predicted 0***	***Predicted 1***
Actual 0	TN = 229	FP = 31
Actual 1	FN = 59	TP = 193

Table 1.5 – Confusion Matrix after doing Logistic regression.

True Negatives (TN = 229): The counties accurately identified as having low rates of poverty. The model correctly identified these counties as not having high rates of poverty.

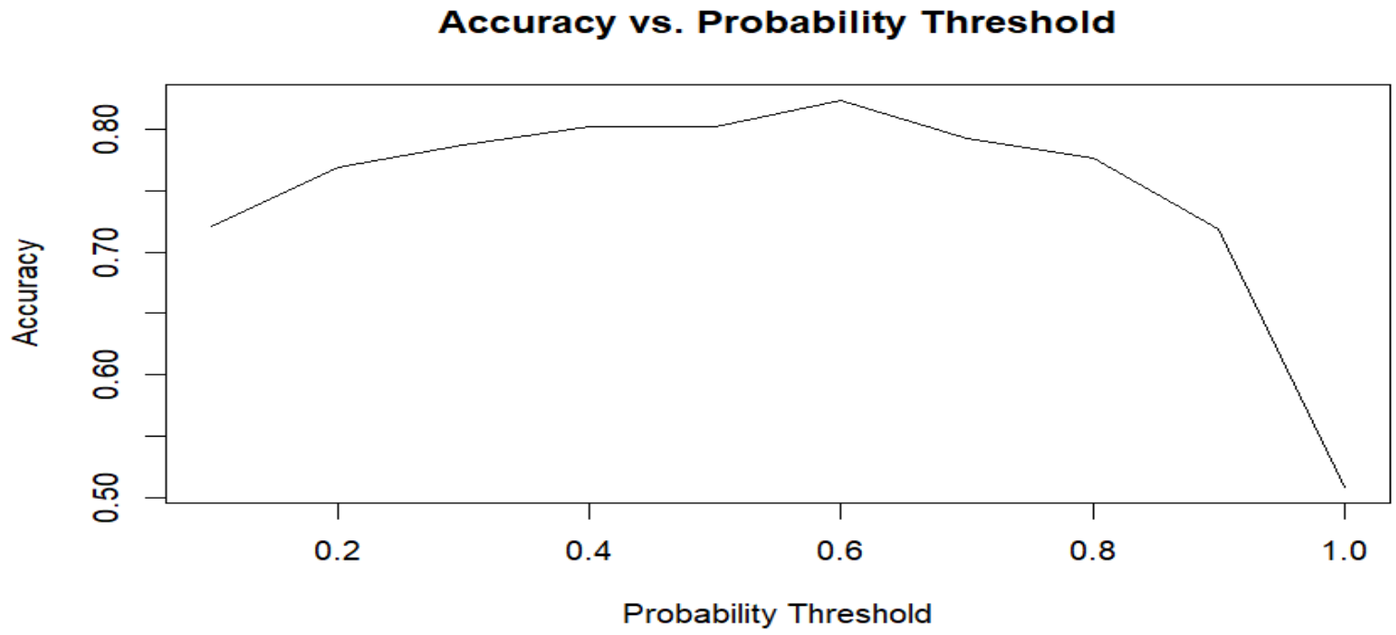
True Positives (TP = 193): The counties are accurately categorized as having high rates of poverty. The model correctly identified these counties as having high rates of poverty.

False Negatives (FN = 59): The counties that are falsely classed as having low rates of poverty when, in fact, they have high rates. The model incorrectly classified these counties as not having high rates of poverty.

False Positives (FP = 31): The counties that were mistakenly labeled as having high rates of poverty while they have low rates. The model incorrectly identified these counties as having high rates of poverty.

Accuracy vs. Probability Threshold:

The graph illustrates the relationship between accuracy and probability thresholds in a logistic regression model. As the probability threshold increases from 0.1 to 1, the model's accuracy generally improves, indicating that higher thresholds lead to more conservative predictions with lower chances of misclassification. However, variations or plateaus in accuracy at specific thresholds may occur due to unique dataset properties and the balance between sensitivity and specificity in classification tasks. Additionally, determining the optimal probability threshold for maximizing accuracy depends on the application's context and the relative costs associated with false positives and false negatives.



Graph 1.9 – Accuracy vs Probability Threshold graph

Predicted Probability	Actual Probability
more than 10%	0.7207031= 72.07%
more than 20%	0.7695312= 76.95%
more than 30%	0.7871094= 78.71%
more than 40%	0.8027344= 80.27%
more than 50%	0.8027344= 80.27%
more than 60%	0.8242188 = 82.42%
more than 70%	0.7929688= 79.29%
more than 80%	0.7773438 = 77.73%
more than 90%	0.71875 = 71.87%

Table 1.6 – Predicted vs Actual probability.

All things considered, the graph sheds light on how the logistic regression model performs across various probability thresholds, assisting in the selection of a suitable threshold that strikes a balance between classification accuracy and reliability. In this instance, the 0.6 threshold, at which accuracy is assessed, seems to produce comparatively high accuracy, indicating that it is a good cutoff point for forecasting.

Conclusion

Our project aimed to uncover the key factors influencing the poverty rate across the various counties in the US. We tried fitting 2 models for our data. From the exploratory data analysis, we have done we got the variables that were affecting the poverty rate. We have created models using multiple linear regression with interaction and logistic regression.

In the multiple linear regression model, the adjusted R_square value came around 72.81%. By the model equation, the poverty rate has a positive correlation with Unemployment_rate and median_hh_income, and has a negative correlation between median_edu, metro, and per_capita_income. The poverty rate varied across the different counties concerning all the independent variables by using this model. However, pop_change and homeownership has a lesser impact on poverty so couldn't consider them as independent variables for the model.

The logistic regression model, trained on the dataset, estimates the probability of whether counties exhibit high or low poverty rates based on various socio-economic factors. Achieving an accuracy of around 80%, the model demonstrates respectable predictive ability. Examination of the confusion matrix suggests a balanced classification, with comparable counts of true positives and true negatives. Additionally, the plotted accuracy versus probability threshold graph illustrates the model's performance sensitivity to different threshold levels. While the model offers valuable insights into poverty determinants, continued refinement and validation could bolster its predictive capacity, thereby facilitating more effective policy recommendations and interventions targeted at alleviating poverty across American counties.

Data Sources

1. National Poverty in America Awareness Month: January 2024 : https://www.census.gov/newsroom/stories/poverty-awareness-month.html
2. Income and Poverty in United States in 2017 https://www.census.gov/content/dam/Census/library/publications/2018/demo/p60-263.pdf
3. Unemployment rate in United States: https://rpubs.com/steveye/846257
4. A profile of working poor in 2017 https://www.bls.gov/opub/reports/working-poor/2017/home.htm#:~:text=About%2039.7%20million%20people%2C%20or,to%20the%20U.S.%20Census%20Bureau.&text=(See%20the%20technical%20notes%20section%20for%20examples%20of%20poverty%20levels.)
5. Social Capital https://www.socialcapital.org/?dimension=EconomicConnectednessIndividual&geoLevel=county&selectedId=&dim1=EconomicConnectednessIndividual&dim2=CohesivenessClustering&dim3=CivicEngagementVolunteeringRates&bigModalSection=&bigModalChart=scatterplot&showOutliers=false&colorBy=
6. Unites State Counties: https://www.openintro.org/data/index.php?data=county