**Impact of Socioeconomic Status on Academic Achievement Across the United States**

DS2500 Team Members:

* Isabella Caruz: [caruz.i@northeastern.edu](mailto:caruz.i@northeastern.edu)
* Elda Gobena: [gobena.e@northeastern.edu](mailto:gobena.e@northeastern.edu)
* Abigail Sandquist: [sandquist.a@northeastern.edu](mailto:sandquist.a@northeastern.edu)

**Problem Statement/Background**

Socioeconomic status (SES) plays a critical role in shaping academic achievement across the United States. By analyzing factors such as personal income, access to educational resources, poverty levels, and geographical disparities, this study seeks to uncover how SES influences educational outcomes. This topic was chosen because of the growing recognition of the importance of educational equity in breaking cycles of poverty and promoting social mobility. The findings have the potential to pave the way for greater educational equity, addressing disparities that perpetuate cycles of poverty and hinder social mobility. As academic success often serves as a predictor of future economic and societal opportunities, tackling these inequities can contribute to broader societal progress. Examining regional differences further emphasizes the need for targeted interventions to ensure fairness and opportunity for all students.

**Introduction to Data**

The data for this analysis is sourced from multiple reputable organizations. For the Urban vs. Rural Comparison and Advanced Degrees vs. Personal Income analysis, data is obtained from the USDA, U.S. Census Bureau, and the Bureau of Economic Analysis, covering years 1929–2023. The focus is on the highest degree obtained, the urban-rural continuum code, and personal income per capita. Educational Resources data is sourced from the National Center for Education Statistics (NCES) and includes student resource usage from 2005–2010 across all U.S. districts. Poverty and Education data, specific to Massachusetts (2018–2022), are sourced from the HDPulse portal, provided by the National Institute on Minority Health and Health Disparities. These datasets include county-level metrics on poverty rates, education attainment, and rankings, with no privacy or ethical concerns identified.

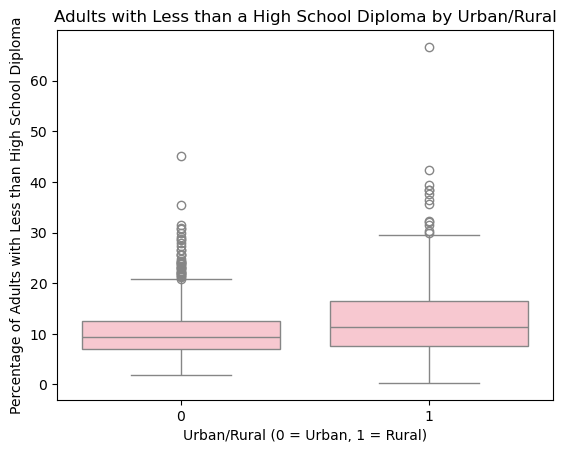
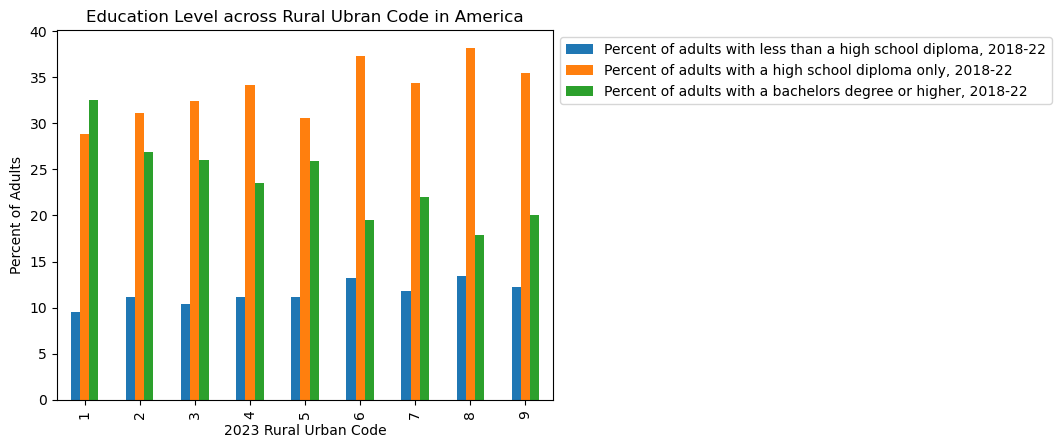
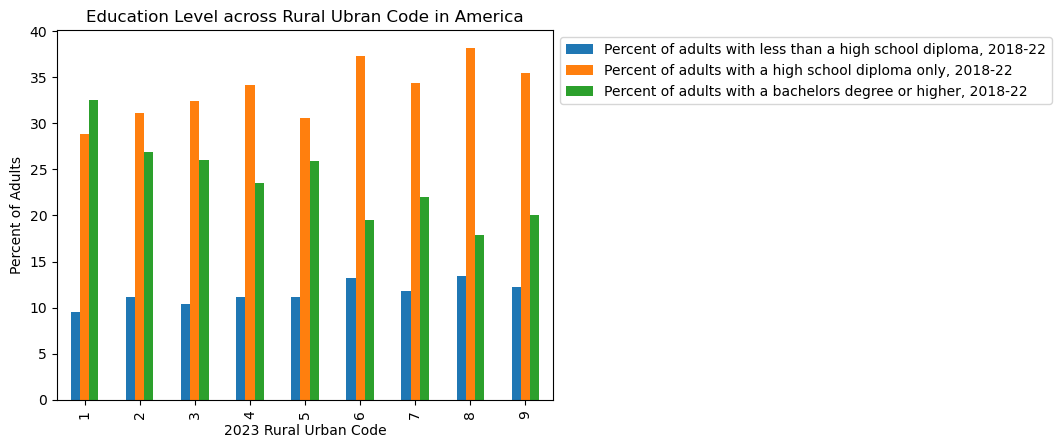
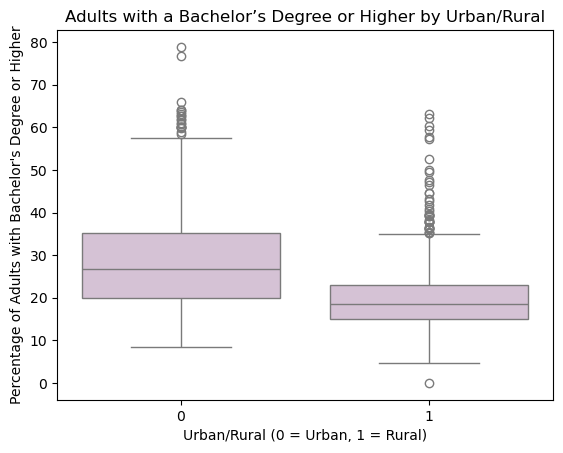
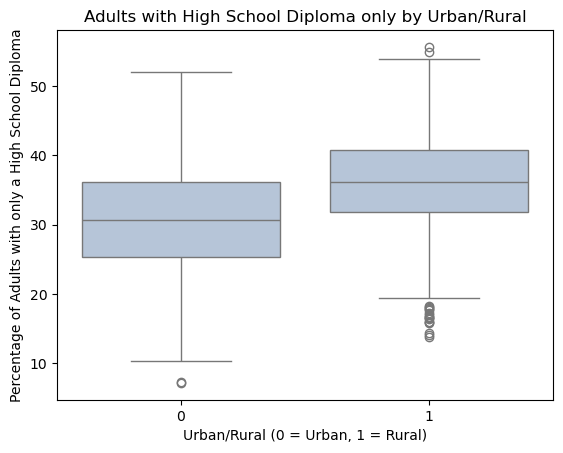
**Data Science Approach**

For Urban and Rural analysis, we initially made a bar graph to visualize the data and see if there are any obvious differences between the groups. However, after we will group the area codes into urban or rural areas and create a box plot to compare each academic level. We will first be using Pearson Correlation to see if there is a relationship between living in a rural or urban area and academic achievement, which in this case will be the highest degree attained. After testing correlation, we will also be running a z test between the 3 different groups to confirm if the differences seen are statistically significant.

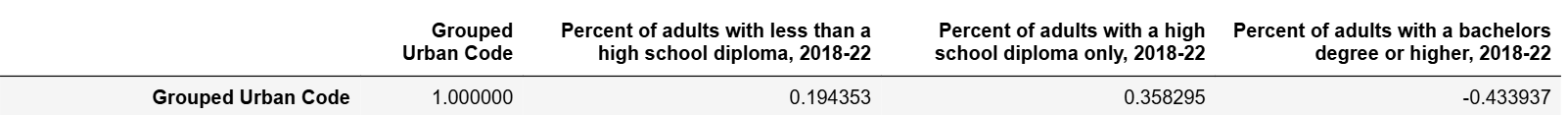
To analyze the effect of personal income on education, regression graphs will be made to examine the correlation between the percentage of adults with advanced degrees and personal income per capita across states. After assessing the correlation, time series graphs will be created for personal income and advanced degrees to compare trends over multiple years in three states. Given the strong relationship between these variables, we will construct a linear regression model to predict the percentage of adults with advanced degrees based on personal income. Finally, two more regression plots will be made to determine whether the number of students utilizing educational resources in 2005-2006 correlates more strongly with advanced degrees or with adults lacking a high school diploma by 2018-2022, ensuring generational consistency in the analysis.For the poverty and education analysis, the data will first be cleaned to remove irrelevant entries, focusing exclusively on individual counties. A correlation analysis will be performed between high school dropout rates and poverty levels using Pearson Correlation to explore the relationship between these variables. A scatter plot will be created to visually inspect the correlation. Following this, a linear regression model will be employed to assess how well poverty levels predict high school dropout rates across Massachusetts counties. Additionally, any statistically significant differences in dropout rates between counties with varying poverty levels will be examined.

**Results**

Urban vs. Rural Analysis

The Rural-urban Continuum Code is a system designed to categorize counties along a continuum that shows the transition from rural to urban. It classifies counties based on their population size and degree of urbanization. The Rural-urban Continuum Code ranges from 1-9 and each county is assigned a number. When grouped 1-3 signifies counties in metropolitan areas (from large to small metropolitan counties). 4-6 is non metropolitan counties (with varying levels of rurality and proximity to metro areas). Lastly 7-9 represents more isolated rural counties with little or noI grouped counties from 1-3 as urban and 7-9 as rural to run some analysis. The box plot above shows the comparison of urban and rural environments on the 3 different education levels. 

To test if the differences observed between each group are statistically significant I decided to run a paired z test. The null hypothesis is that there is no difference between the urban and rural group for each of the 3 categories. From the z-test the P values for all of them were less than 0.05. This signifies that the differences observed are statistically significant and we can reject the null hypothesis.

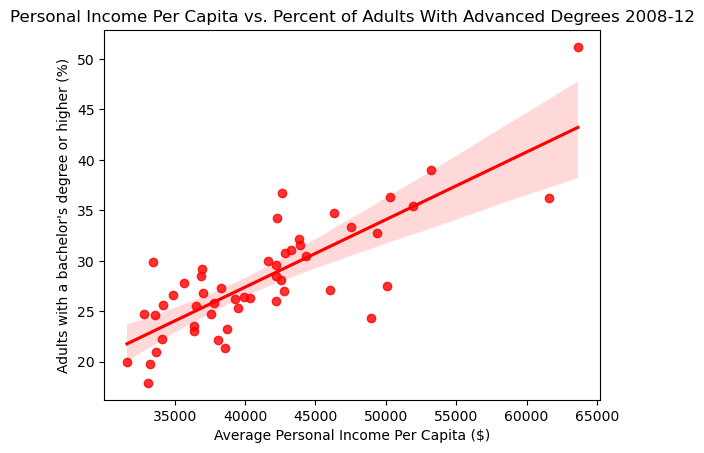
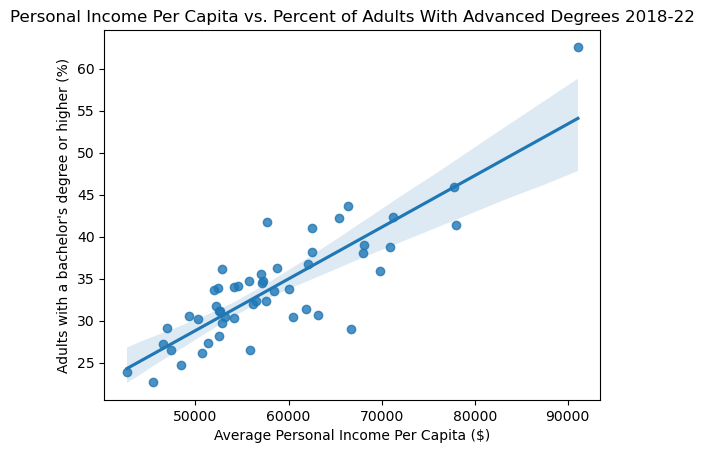


0.194 = weak positive correlation between urban/rural living and percent of adults with less than a high school diploma

0.358 = weak positive correlation between urban / rural living and percent of adults with only a high school diploma

-0.433 = moderate negative correlation between urban/rural living and the percent of adults with a bachelor degree or higher

Personal Income and Higher Education

Residual Graphs and Correlation - Personal Income

2008-12: correlation = 0.8047 = strong positive correlation between the percent of adults with advanced degrees and personal income per capita.

2018-22: correlation = 0.8483 = strong positive correlation between the percent of adults with advanced degrees and personal income per capita.

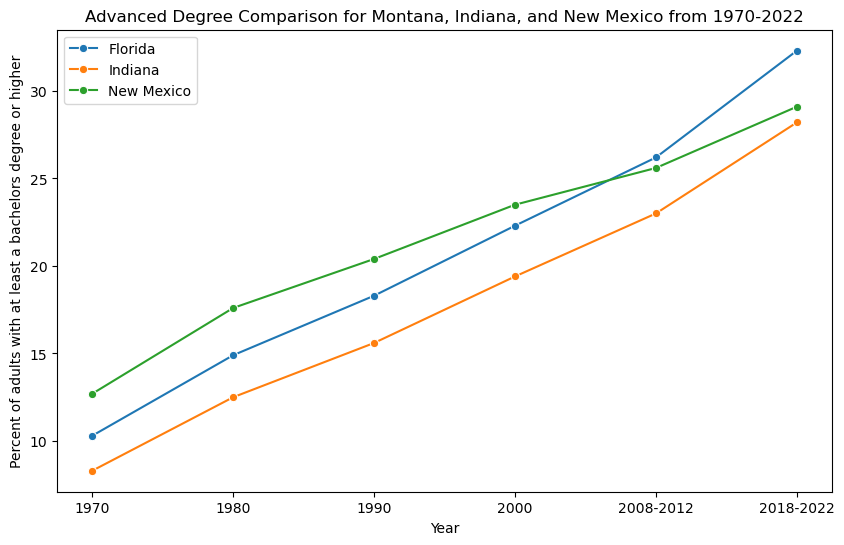
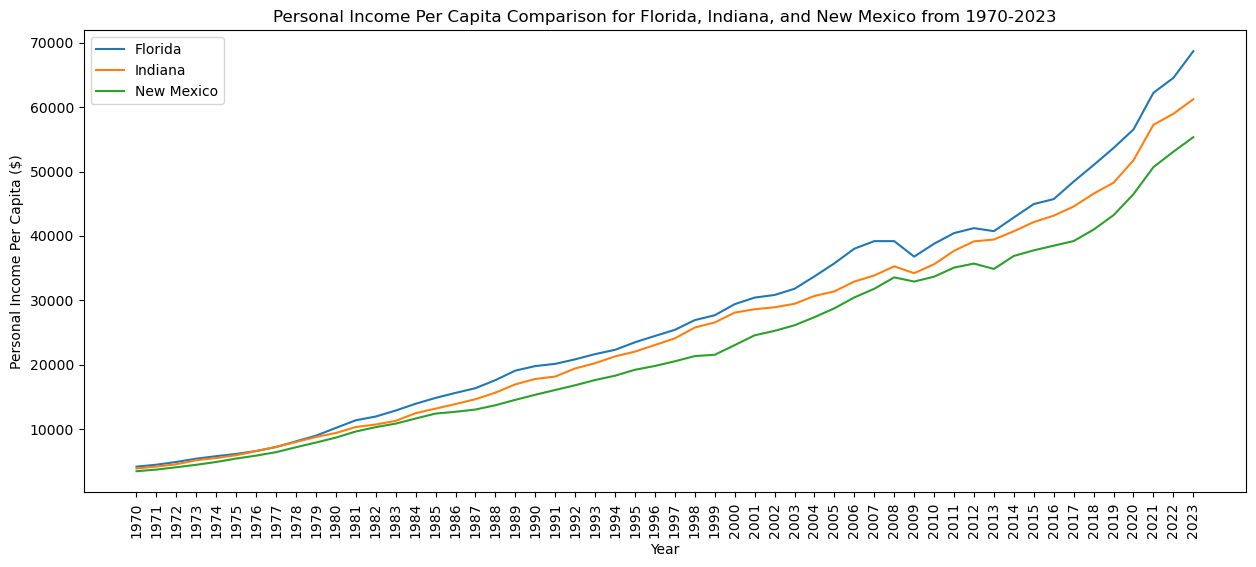
I found the correlation to be strong in both half decades which shows a significant relationship between a state’s personal income and the percent of adults in that state that have advanced degrees.

Residual Graphs and Correlation - Educational Resources:

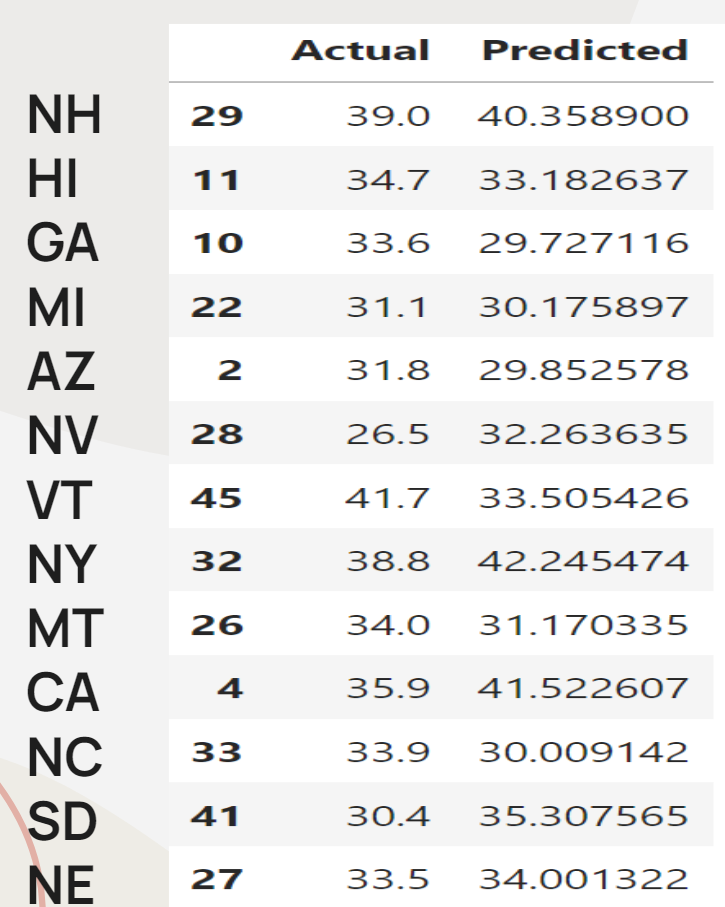
Advanced Degrees: correlation = 0.04339 = weak positive correlation between number of students using educational resources and the percent of adults who have advanced degrees.

Less the High School Diploma: correlation = 0.5464 = moderate positive correlation between the number of students using educational resources and the percent of adults who have less than a high school diploma.

The correlations show that the more students in each state who use additional educational resources tend to have less than a high school degree as an adult and less likely to have an advanced degree as an adult. This may seem counterintuitive, but could reflect that while resources are being allocated, they may not be effectively addressing the root causes of low educational attainment and factors such as socioeconomic status might be playing a more significant role. The Graphs are in the presentation slides (slides 14 and 15) for this.

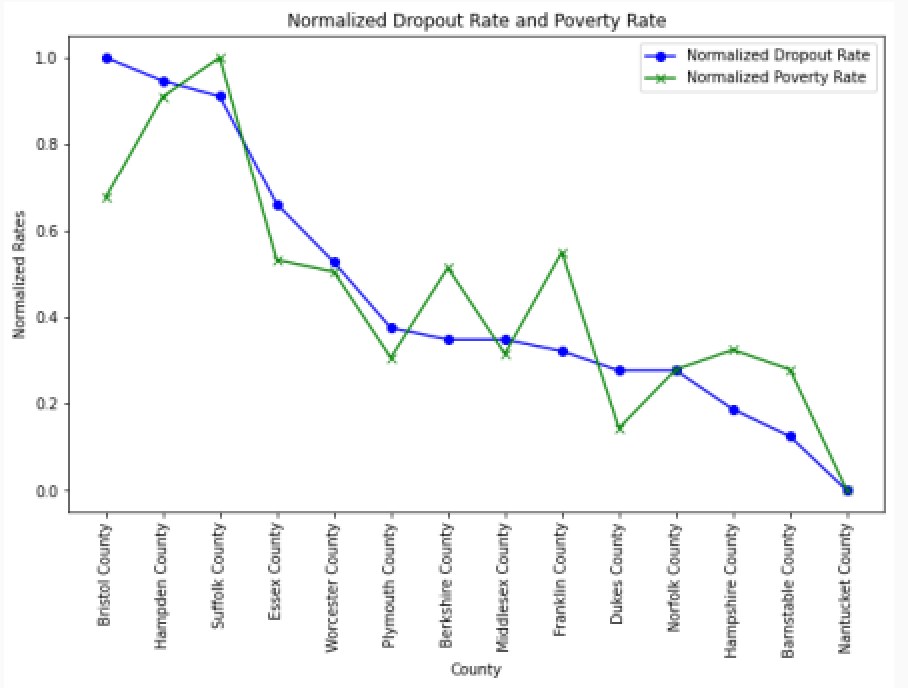
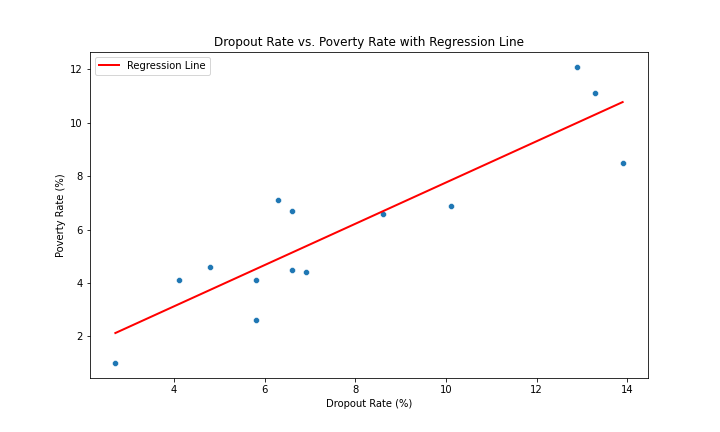
Time Series- Personal Income

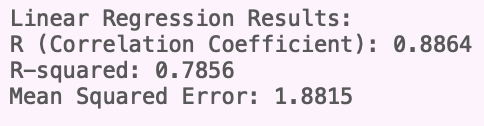
It is clear that the graphs both steadily increase at relatively similar rates, further justifying that personal income per capita and percentage of adults with advanced degrees are at least somewhat related.



Linear Regression Model- Personal Income

This prediction is really close(less than 1%) for some states but also over or underestimates by 4% or 5% for other states showing that personal income is a strong predictor for advanced degrees, but including other socioeconomic factors would make this predictor stronger.

Poverty and Education



The two visualizations provide complementary insights into the relationship between high school dropout rates and poverty levels in Massachusetts counties. The linear regression graph reveals a strong positive correlation, with an R-squared value of 0.7856 and a Pearson correlation coefficient (r) of 0.886, indicating a very strong linear relationship between the two variables. This suggests that 78.56% of the variation in dropout rates can be explained by poverty levels. The model’s Mean Squared Error (MSE) further supports its predictive accuracy, emphasizing the socioeconomic impact of education. The normalized rates graph provides a more granular perspective, highlighting how dropout and poverty rates vary across specific counties. For example, Bristol and Hampden counties exhibit high normalized dropout and poverty rates, reflecting significant challenges, while Nantucket and Barnstable display low values for both metrics. The alignment of the two lines in the second graph underscores the close connection between educational attainment and economic hardship, while slight deviations highlight localized factors that may influence this relationship. Together, these analyses underscore the critical need to address educational disparities to reduce poverty.

**Future Work**

To continue the work of our project, we will focus on expanding the analysis in several key areas. First, the geographic scope will be broadened to include additional states and counties, allowing for comparisons of socioeconomic status (SES) impacts on dropout rates and academic achievement across diverse regions. Second, a longitudinal analysis will be conducted to track changes in personal income and educational policies over time, providing insight into their influence on advanced degree attainment. Finally, disparities between urban and rural areas will be explored by examining differences in educational resources, teacher quality, and their cumulative impact on academic outcomes. These directions will help provide a more comprehensive understanding of how socioeconomic factors shape educational achievement.

**Conclusion**

Socioeconomic factors such as urban versus rural location, income levels, access to educational resources, and poverty rates play a pivotal role in shaping educational outcomes and long-term success in the United States. Key findings reveal a statistically significant relationship between urban and rural living conditions and degree attainment, with lower income levels strongly correlating with a reduced likelihood of adults achieving higher education degrees. In Massachusetts counties, high poverty rates are closely tied to elevated high school dropout rates, underscoring the profound impact of economic hardship on educational attainment. These disparities emphasize the urgency of addressing systemic inequities to promote equal opportunities, break cycles of poverty, and build a more inclusive and equitable society where every individual has the chance to succeed. By addressing these disparities through targeted interventions and policies, we can work toward an education system that is not only more accessible but also more effective in fostering social and economic mobility.

Works Cited

“Common Core of Data (CCD).” *Table 3.–Number of Students Receiving Educational Services,by State and District Charter Status: School Years 2005–06 through 2009–10*, nces.ed.gov/ccd/tables/table\_3\_State\_and\_District\_Charter.asp. Accessed 12 Nov. 2024.

“Download ZIP Files.” *BEA*, apps.bea.gov/regional/downloadzip.htm. Accessed 12 Nov. 2024.

*Educational Attainment for Adults Age 25 and Older for the United States, States, and Counties, 1970–2022*, 20 June 2024, ers.usda.gov/webdocs/DataFiles/48747/Education.xlsx?v=7207.9. Accessed 12 Nov.2024.

National Institute on Minority Health and Health Disparities. “Education Data by Social and Demographic Characteristics: Massachusetts.” *HDPulse Data Portal*,<https://hdpulse.nimhd.nih.gov/data-portal/social/table?socialtopic=020&socialtopic_options=social_6&demo=00005&demo_options=education_3&race=00&race_options=race_7&sex=0&sex_options=sex_3&age=081&age_options=age25_1&statefips=25&statefips_options=area_states>. Accessed 3 Dec. 2024.

National Institute on Minority Health and Health Disparities. “Poverty Data by Social and Demographic Characteristics: Massachusetts.” *HDPulse Data Portal*,<https://hdpulse.nimhd.nih.gov/data-portal/social/table?age=001&age_options=ageall_1&demo=00007&demo_options=poverty_3&race=00&race_options=race_7&sex=0&sex_options=sexboth_1&socialtopic=080&socialtopic_options=social_6&statefips=25&statefips_options=area_states>. Accessed 3 Dec. 2024.