STEM Fellowship High School Big Data and AI Challenge 2024-2025

Investigating the Influence of Poverty on High School Graduation Rates: An Analysis of the Impact of Socioeconomic Inequities on Adolescents

Arnav Sinha[1], Harper Yang[1], Justin Nguyen[1], Lucien Lin[1], and Richard Gao[1]

[1]Webber Academy

January 11, 2025

Abstract

As part of the United Nations' Sustainable Development Goals, education is a significant part of how our world functions. It teaches to youth not only essential skills for personal development but also knowledge to be applied further on in life. However, the accessibility and quality of education for adolescents are being hindered by a variety of socioeconomic disparities, especially poverty. Without reaching equity in learning opportunities, access to essential academic resources is limited while cycles of poverty are perpetuated. Despite various legislative measures in the United States and other countries, these inequities persist globally and call for more targeted measures. This study examines the relationship between adolescent poverty and high school graduation rates from 2021 to 2024 in counties within the United States, utilizing data science tools to clean, analyze, and visualize datasets, including data from sources such as the US Census Bureau. Unlike most prior research that addresses these variables independently, or without a connection to potential steps to combat the inequality, this analysis correlates high school educational outcomes with poverty rates and proposes ways to decrease the detrimental connection between regional adolescent poverty and high school dropout rates. Through identifying correlation between multiple variables, the study uses analytics to draw conclusions. The findings provide policy recommendations aimed at mitigating the effects of poverty on quality of education, narrowing the gap in opportunities between students from different socioeconomic standpoints, and advancing the goal of equitable education for all.

Keywords
Socioeconomic Disparity, Education, Poverty, High School Graduation, Adolescents

1 Introduction

Poverty has long been a barrier to adolescents' academic success worldwide, particularly in completing their education. Goal 4 of the United Nations' 17 Sustainable Development Goals proves globally equitable education to be vital in lifelong opportunities and quality of life [1]. With the increasing importance of equality and equity within modern developed countries, allowing all youth to have the same chances of success in the future is necessary. Educational attainment plays a large factor in the success of adults. Unfortunately, disparities in educational access are deeply rooted in socioeconomic factors such as income, social status, and healthcare [2]. In 2022, with regards to people in the United States aged 25 to 34, those who have graduated high school had a median income $3,890 to $14,250 higher than those who had not [3]. Additionally, high school graduation is generally required in order to attend higher levels of education, each allowing for higher median income in the future. In order to ensure future economic success, youth should be able to have equal ability to graduate high school.

In the United States, over 11.4 million adolescents live below the poverty line of approximately 30,900 USD for a family of four [4]. Despite policies such as the Every Student Succeeds Act (ESSA) and Title I funding, limited access to schooling still impedes a variety of adolescents [5] [6]. These challenges, fueled by poverty, are not limited to the US but are evident globally.

Understanding the specific mechanisms through which poverty influences educational outcomes is essential to designing effective interventions.

This study will begin by investigating the impact of child poverty on high school graduation rates, employing statistical techniques to analyze trends and correlations across four years. After verifying this relationship, methods that can be used to reduce such educational and opportunistic inequities were studied using data analytics and targeted strategies that delve into fields such as economics and sociology, aiming to identify a path forward toward an equitable future for all students.

2 Materials & Methods

2.1 Data Source

The analysis for uses data from government and educational databases from the United States, including regional poverty and graduation rates between 2021 and 2024. This specific time period was chosen because it is recent and relevant to the present day. Adolescent poverty data points were found through the Small Area Income and Poverty Estimates (SAIPE) Datasets from the United States Census Bureau, and graduation rates were measured using the United States Census Bureau Public Elementary-Secondary Education data [7] [8]. The rate of adolescent poverty in each county was linked with the high school dropout rates in that same region. Data about Public Elementary-Secondary Education Finance Data released by the United States Census Bureau was also utilized for the comparison of incomes with and without a high school diploma. Furthermore, along with data on educational fiscal policies, the data on the dropout rates of American adolescents living in households where the income was within the lowest 25th percentile of American households was derived from the National Center for Education Statistics. This data ranged from 2002 to 2016 due to limitations in the data, although there were no identifiable outliers.

2.2 Data Cleaning

Outliers, if left unaddressed, could disproportionately influence the regression model and distort the observed relationships. Outliers were defined, for this project, as data points that stray unusually far from the average or are seemingly impossible to accomplish. In addition to careful selection of data, this formula of the inter-quartile range (IQR) was used:

$$\text{IQR} = Q_3 - Q_1$$

along with the 1.5 inter-quartile range rule, a widely accepted method for identifying outliers in statistical datasets:

$$\text{Lower Bound} = Q_1 - 1.5 \cdot \text{IQR}$$

$$\text{Upper Bound} = Q_3 + 1.5 \cdot \text{IQR}$$

The method was chosen due to its simplicity and effectiveness in handling uneven data distributions, making it suited to the socioeconomic data analyzed in this study. [9]

Two outlier data points were set aside in the data. For the county of Kennedy, Texas, the high school graduation rate was well below the lower limit given by our 1.5 IQR calculations. Similarly, Issaquena, Mississippi, had an unusually high adolescent poverty rate of 89 percent in 2024. The two outliers could disproportionately influence our results. By using the IQR method, this ensured a more accurate and reliable analysis.

2.3 Statistical Techniques

This study investigated the relations between a variety of factors using graphs and forms of supervised machine learning to analyze and visualize data. The graphs summarized key trends in factors such as household income and graduation rates, making it easier to identify patterns and disparities in different regions. Consequently, correlations were able to be identified between the subjects that were being investigated.

2.4 Mathematical Modeling

Mathematical modeling, specifically forms of linear regression, was used to quantify the relationship between child poverty and graduation rates. The linear regression model employed was effective in finding relationships within the dataset. [10]

2.4.1 Potential Directions for Mathematical Modeling

Future research could use multivariate regression models to include variables such as parental education, school funding, and mental health resources. Advanced machine learning techniques, such as ensemble models, could look for more complex interactions that linear regression cannot. Geo-spatial analysis could also help identify regional disparities and create more targeted interventions. [11]

2.5 Visualization Tools

Graphs such as scatter plots, regression lines, and time series visualizations were generated using Python libraries Matplotlib and Seaborn. Through scripts utilizing these tools, the relationships and trends identified in the data were clearly illustrated.

3 Results

3.1 Residual Analysis

Residual analysis was used to evaluate the model's accuracy and reliability in predicting high school graduation rates.

Residuals The residuals in our data were calculated using the formula:

$$\text{Residual} = Y_i - \hat{Y}_i$$

where Y_i is the actual graduation rate, and $\hat{Y}_i$ is the predicted rate. This analysis helped identify outliers and assess the model's fit. Residual analysis was essential because it highlighted discrepancies between observed and predicted values, helping us understand where the model might be underperforming or missing key factors. From the residual plot, it was determined that linear regression was a valid form of data analysis.

3.2 Comparing Local Adolescent Poverty Rates and Household Income to High School Graduation Completion Rates

To begin, it was important to identify whether there exists a correlation between adolescent poverty rates in the included counties and their respective high school graduation rates. Figure 1 shows a residual plot that was created to ensure that linear regression would be a valid form of data analysis. This residual plot showed an average spread of points along the predicted regression line. There were no heteroscedasticity nor any other patterns present that could make a linear regression model invalid. Knowing this, utilizing the linear regression model shown in Figure 2 would be valid for comparing child poverty with high school graduation. Figure 2 displays data from County Health Rankings & Roadmaps on the correlation between child poverty rates in each county and the respective high school graduation rate in the years 2021 - 2024, compiling our analysis from each year [12]. A correlation coefficient of approximately -0.425 was found, demonstrating a moderate negative correlation. The slope was close to -0.46, meaning that for every 10 percent increase in child poverty, graduation rates are predicted to decrease by around 4.6 percent.

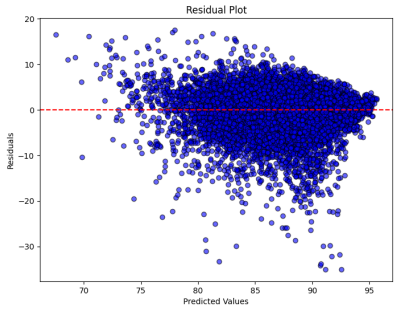

Figure 1: A residual plot showing the relationship between percent adolescent poverty rates in every US county, shown on the x-axis, and high school graduation rates per county on the y-axis for the year 2021-2024.

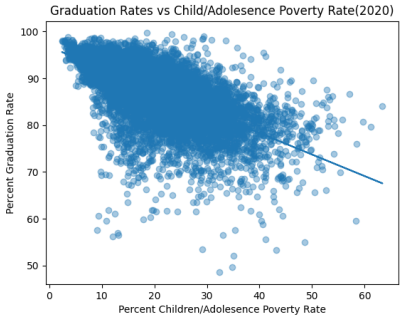

Figure 2: This graph correlates adolescent poverty rates for every US state, shown on the x-axis, and the high school graduation rate, shown on the y-axis for every county in the contiguous United States from 2024-2021.

3.3 Comparing the Median Salary of Individuals With a High School Diploma Versus Those Without

Figure 2 is a line graph from the National Center for Education Statistics depicting the difference in median salary in the United States between individuals who did not graduate from high school and individuals who did [13]. There is a steady increase in the median salary, likely due to natural economic inflation. Most importantly, the median income for individuals who have completed high school is approximately 33% more than those who have not—40,000 and 30,000 dollars, respectively.

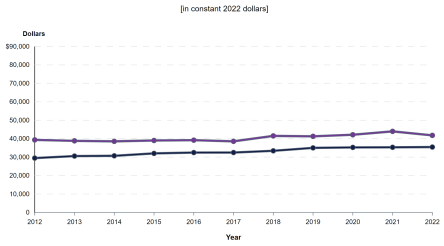

Figure 3: This double line graph shows the median income of individuals in the United States who have completed high school, shown in purple, and the median income of individuals who do not have a high school diploma, shown in black.

3.4 Comparing Adolescent Dropout Rates in the US Lowest-Income Quartile with the Average Government Funding per Individual over the years 2002-2016

To connect the analyzed data with a possible factor and solution, data on government educational funding in the United States was correlated to the previously concerned high school dropout rates. A residual graph was created to validate the use of linear regression. This residual graph had an average spread of data along the predicted regression line, with no patterns that could invalidate the use of linear regression. Figure 3 relates the adolescent dropout rates in the US lowest-income quartile and the average amount of government spending per pupil. The correlation coefficient was around -0.843, showing a strong negative correlation. The slope was about 2.4e-3, meaning for every 1,000 dollars spent per individual, there was a predicted decrease in dropout rates by 2.4 percent.

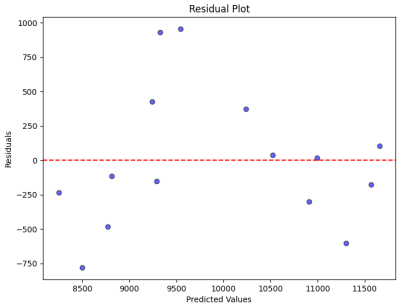

Figure 4: This residual plot relates government spending per pupil, shown on the x-axis, and dropout rates of adolescents in the lowest earning quartile of households in the US from 2002 to 2016.

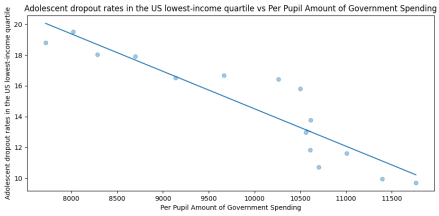

Figure 5: This graph relates government spending per pupil, shown on the x-axis, and dropout rates of adolescents in the lowest earning quartile of households in the US on the y-axis, in 2002-2016.

4 Discussion

4.1 Insights and Interpretation

Despite the strong correlation shown by the Pearson R Coefficient, many counties deviated from the regression line. This suggests the presence of additional factors that may have a considerable influence on high school graduation rates. Nonetheless, our analysis revealed a correlation that higher youth poverty rates are linked to a decrease in graduation rates, supporting our hypothesis that adolescents' socioeconomic status affects their equity regarding opportunities for success in the future. Furthermore, the significance behind the slope of the regression line in Figure 2 showed a P-value of 0:

$$df = 12764 - 2$$

$$df = 12762$$

$$t = \frac{b - \beta 0}{SEb}$$

$$t = \frac{-0.46 - 0}{0.019889114071150526}$$

$$t = -23.128229762$$

$$P\text{-value} = P(T \geq 23.128|H_0) + P(T \leq -23.128|H_0)$$

$$P\text{-value} = 0$$

P-value confirmed to the 10th decimal place.

Let $\alpha = 0.05$.

Because the P-value $< \alpha$, this shows the results are significant, meaning that the correlation was not merely coincidental.

4.2 Additional Influencing Factors

Beyond socioeconomic status, several other factors affect high school graduation rates. Parental involvement improves academic performance and motivation. Limited access to technology hinders students' ability to complete assignments and develop online skills. School campus, including access to libraries and laboratories, influences learning. Health and nutrition,

especially in low-income areas, affect cognitive development and energy levels. Future studies should include these variables for a more detailed and accurate analysis. [14]

4.3 Relation to Hypothesis

The findings largely support our hypothesis that lower household income correlates with reduced graduation rates. Nonetheless, additional influencing factors, such as mental health and school location, suggest that income is not the only determinant of graduation outcomes. However, the presence of poverty in a region can itself drastically influence some identified other factors, including parental decisions and community availability of resources.

4.4 Correlation Analysis

A supervised machine learning model was used to analyze data, utilizing the least squares regression formula within a linear regression framework.

P-Value Calculations: This P-Value equation was calculated using the formula:

$$p = P(T \geq t|H_0) + P(T \leq -t|H_0)$$

where T is the test statistic, t is the observed value, and H_0 is the null hypothesis. The p-value confirmed the correlation between adolescent poverty and graduation rates, ensuring adding basis behind the hypothesis and validating the purpose of the study.

Linear Regression: The linear regression model was originally calculated using the formula:

$$Y = \beta_0 + \beta_1 \cdot X$$

where Y represents the predicted graduation rate, X represents household income, β_0 is the intercept (the expected value of Y when $X = 0$), and β_1 is the slope (the change in Y for a one-unit increase in X).

Slope Calculation (β_1): The slope β_1 was found using the formula:

$$\beta_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

In this formula, X_i and Y_i represent individual data points for household income and graduation rates, respectively, while $\bar{X}$ and $\bar{Y}$ denote their mean values. The slope β_1 quantifies how much the graduation rate changes for each unit increase in household income. This formula was essential because it allowed us to measure the direction and strength of the relationship between income and graduation rates, providing a clear numerical representation of this relationship.

Intercept Calculation (β_0): The intercept β_0 was calculated using the formula:

$$\beta_0 = \bar{Y} - \beta_1 \cdot \bar{X}$$

Here, β_0 represents the baseline graduation rate when household income is at its average value. This intercept provides context for understanding how income levels relate to graduation outcomes when income is near the mean. The intercept was crucial for anchoring the regression line at a meaningful starting point, ensuring the accuracy of our predictions when income is average.

After excluding outliers, the updated regression model was calculated using the formula:

$$Y = 91.5 - 0.43X$$

This shows that for every 1% increase in child poverty, the graduation rate decreases by approximately 0.43% in the respective region. The intercept of 91.5 represents the expected graduation rate when there is no poverty.

4.5 Solutions

To address this issue concerning graduation rates among children living in poverty, very specific interventions are required to break the cycle of poverty and improve education rates. For example, local governments can create programs that provide free meals, stable housing assistance, and access to tutoring, which have been shown to directly support students in need [15]. Investment in school funding, particularly in lower-income districts, can ensure access to quality teaching, technology, and enrichment opportunities. Additionally, community partnerships and mentorship programs can provide emotional and academic support. Long-term strategies like creating policies that center around affordable childcare, parental job training, and livable wages could also focus on reducing poverty. The findings of this study offer important insights for elected officials and educators. To improve the equity regarding high school graduation rates, increasing funding for schools in economically impoverished areas and creating more government-funded programs would be a forward step [16] Addressing immediate and long-term educational barriers can help break the cycle of poverty and improve graduation outcomes.

4.6 Limitations

This study's reliance on aggregated regional data limits its ability to find individual-level nuances. Furthermore, the absence of longitudinal tracking prevents causal-level inferences. Future research should use more granular data and control additional variables such as parental involvement and school resources. However, the study was still able to make relevant conclusions based on reliable data sources and illustrate the degree to which some factors contribute to a major global issue.

4.7 Relevance to Existing Literature

Our findings yield similar results to previous research. Studies have shown that poverty restricts access to essential academic resources, diminishing students' opportunities for success. For example, Duncan et al. explore how childhood poverty negatively impacts both short-term academic performance and long-term educational prospects. Their work centralizes on how economic hardship worsens disparities in educational achievement that happen throughout life. This supports our findings that higher adolescent poverty rates are linked with lower high school graduation rates. [17]

The points further away from our line of regression additionally align with findings in earlier literature. It was previously suggested that additional factors—such as parental education, school quality, and community resources—play a significant role in impacting graduation outcomes. For example, research by Barbra Ritter discusses the many influences, including school quality and community factors, on high school graduation rates. [18]

By confirming and adding on to these earlier findings, our study strengthens the evidence relating to how socioeconomic inequities, such as adolescent poverty, negatively impact high school graduation rates.

4.8 Future Research Directions

Building on the current findings, future research could: use longitudinal datasets to establish causative links between poverty and graduation outcomes rather than correlation [19], use additional variables, such as parental education and access to mental health resources, to widen the scope of analysis and better find out how much each factor is contributing to lower graduation rates [20] and assess the effectiveness of specific interventions, including programs which directly support those with increased financial needs, and increased school funding [21], through experimental or quasi-experimental designs [22]. Looking more deeply into what impacts increasing graduation rates more within the per pupil spending done by the government could also prove to be interesting in the future. This would not only provide deeper insight into the needs of adolescents in terms of their education but also would enable the most efficient use of resources in solving this issue.

5 Conclusions

This study demonstrates the relationship between child poverty and high school graduation rates, showing a decent correlation between these variables. Targeted interventions that address these issues are essential to providing all adolescents equal access to education and life opportunities. Policymakers must prioritize investments in education and community resources to mitigate the detrimental effects of poverty and create a more equitable society. Future research should focus on longitudinal studies and policy evaluations to refine and expand these findings. Compared to schools in districts with lower poverty rates, those in areas with higher poverty levels do not receive an equal or adequate amount of funding from the government. Even with reforms that aim to distribute funds for education, a considerable portion (approximately half) of the funding for schools is provided by the district or state governments. This comes from local taxes, meaning that areas with higher poverty rates receive less money to ensure equal educational opportunities. Given the profound impact of adolescent poverty on graduation rates and future success, as revealed through our research and analysis, districts with higher estimated poverty rates must receive adequate support in the form of subsidies from federal or state governments to create more equitable educational opportunities. After investigating the impact of an increase in government per-pupil spending, the impact was clearly visible. With more resources being provided to education, whether it be higher wages for teachers who can provide higher quality education to students or more money to invest in resources for students, it is evident that graduation rates are increasing. The support provided by the government will be a forward step in providing students in all situations, geographically and socioeconomically, an equal opportunity to graduate high school and allow for greater success in their futures.

Acknowledgments

We would like to thank Webber Academy, Mr. Niven, Aaron Zhang and Owen Kuang for their guidance on statistics.

References

[1] United Nations. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. *United Nations Department of Economic and Social Affairs*, 2012.

[2] Jo Blanden, Matthias Doepke, and Jan Stuhler. Education inequality. *Centre for Economic Performance*, April 2022.

[3] National Centre for Education Statistics. Annual earnings by educational attainment. *U.S. Department of Education, Institute of Education Sciences*, May 2024.

[4] John Creamer and Emily Shrider. Poverty in the united states: 2022. *United States Census Bureau*, September 2023.

[5] Jennifer B Ayscue, Lance D. Fusarelli, and Elizabeth M Uzzell. Equity and early implementation of the every student succeeds act in state-designed plans during COVID. *Educ. Policy (Los Altos Calif.)*, 37(7):1917–1949, November 2023.

[6] Ray Domanico. Title i is a clunky, overbroad failure. low-income students deserve better. *Manhattan Institute*, February 2024.

[7] US Census Bureau. Small area income and poverty estimates (saipe) program.

[8] US Census Bureau. Census data tables.

[9] Sergio Hleap. The interquartile range method (iqr) to unmask outliers. *Procogia*, January 2024.

[10] Stuart W Grant, Graeme L Hickey, and Stuart J Head. Statistical primer: multivariable regression considerations and pitfalls†. *European Journal of Cardio-Thoracic Surgery*, 55(2):179–185, December 2018.

[11] Korryn Bodner, Chris Brimacombe, Emily S Chenery, Ariel Greiner, Anne M McLeod, Stephanie R Penk, and Juan S Vargas Soto. Ten simple rules for tackling your first mathematical models: A guide for graduate students by graduate students. *PLoS Comput Biol*, 17(1):e1008539, January 2021.

[12] County Health Rankings & Roadmaps. Data & documentation. *University of Wisconsin Population Health Institute School of Medicine and Public Health*, 2024.

[13] National Centre for Education Statistics. Digest of education statistics. *U.S. Department of Education, Institute of Education Sciences*, 2022.

[14] Ronald Stolberg. Factors that lead to higher graduation rates and less drop-out. *Eye on Psi Chi Magazine*, 28(4):20–21, 2024.

[15] Marc Frenette. Why are youth from lower-income families less likely to attend university? evidence from academic abilities, parental influences, and financial constraints. *Statistics Canada*, February 2007.

[16] Sheldon Solanki. Graduation rates vs. neighborhood income: How does the economic welfare of a community affect the graduation rates of schools within those communities? *ArcGIS StoryMaps*, January 2021.

[17] Hb Ferguson, S Bovaird, and Mp Mueller. The impact of poverty on educational outcomes for children. *Paediatr Child Health*, 12(8):701–706, October 2007.

[18] Barbara Ritter. Factors influencing high school graduation. *Washington Student Achievement Council*, 2015.

[19] Abhishek Sheetal, Zhou Jiang, and Lee Di Milia. Using machine learning to analyze longitudinal data: A tutorial guide and best-practice recommendations for social science researchers. *Applied Psychology*, 72(3):1339–1364, July 2023.

[20] Andrew Lanners. The impact of p the impact of poverty on high school gr ty on high school graduation rates aduation rates. *Bethel University*, 2022.

[21] Amberley Ruetz, Gabrielle Edwards, and Flora Zhang. The economic rationale for investing in school meal programs for canada: multi-sectoral impacts from comparable high-income countries. *Ruetz Consulting*, February 2024.

[22] Donald Campbell and Julian Stanley. Experimental and quasi-experimental designs for research. *Houghton Mifflin Company*, 1959.