Education Attainment’s Impact on Median Income in Summit County, OH

Kevin Coleman

Final Research Project

Urban Data Analytics

2024, Fall Semester

EXECUTIVE SUMMARY:

The purpose of this study is to seek how the education attainment levels of high school only and bachelor degree holders in a given tract affect the tracts poverty and lower-middle class median incomes. Poverty is defined as a median income at or below $31,200 and lower-middle class is defined as a median income more than $31,200 and less than $60,000. The area of study is Summit County, OH home to the Akron, OH metropolitan area. The variables come from the ACS 2020 US Census Bureau.

Important findings include the statistical significance of the bachelor and high school holder variable on median income. A multivariate regression model and two bivariate choropleths were used to test the hypotheses. Both variables had a p-value less than .01 and the R^2 value accounted for half of the total variance. Additionally, Figures 4 and 5 revealed a key finding that many of the poverty median income tracts, centered around Akron, had both low levels of bachelors and high school holders. This suggests that the tracts likely had a high level of no high school holders.

TOPIC AND RESEARCH QUESTIONS:

The research question will seek to determine if and how the education attainment level of high school and bachelor degree holders in a tract affects median income. Specifically, this study observes how education attainment correlates with poverty and lower-middle class median incomes at the census tract for Summit County, OH. According to the US Department of Health and Human Services, a family unit of four that has a household income of $31,200 is at the 100% poverty level. [[1]](#footnote-1) Impoverished tracts will be defined as having a median income at or below $31,200. The lower-middle-class income will be defined as a tract with a median income of more than $31,200 but less than $60,000. According to the Pew Research income calculator, a 4-person household living in Akron, OH (Summit County area) making more than $60,000 per year is considered middle-class.[[2]](#footnote-2) Understanding how education affects both the lower and lower-middle-class can lead to more effective higher education policy and programs.

KEY CONCEPTS AND VARIABLES:

The dependent variable will be ‘MEDINC’ or median income for a given tract. Since the research question seeks to determine if education level is significant in affecting poverty and lower-middle-class median incomes for a given tract. A limitation of the dependent variable is the unit of measurement, it measures tracts rather than households or individual incomes. Measuring tract misses out on some of the local data or granularity of the data, but does provide a decent generalization of trends within a neighborhood or community which is represented by the census tract. Individual data would be harder to obtain due to privacy concerns. The dependent variable is not exploring or testing middle class and higher median income tracts. The focus of this study is on poverty and lower-middle median incomes. The predictor variables of high school graduates and bachelor’s degrees within a tract will represent the education level. The education variables are individual count measurements for individuals 25 years old and older. Limitations include individuals who are below 25 years old and individuals who do not have a high school degree or individuals who have degrees above a bachelor’s degree. However, this study assumes that tracts with high levels of bachelor’s degrees likely have higher levels of higher education compared to tracts with low levels of bachelor’s degrees. For this topic, the chosen education variables are valid for exploring income at the tract level. The US Census Bureau is known for collecting high quality and reliable data, there should not be any reliability issues with the data. Ecological inference may be an issue, as the data is usually collected at the individual or household level. The issue will be mitigated by analyzing general trends through the tract measurement which is a shared geographic measurement for each variable that aggregates the data from all individuals living within the tract. By focusing on the tract level, the general trends affecting income will be captured within Summit County.

HYPOTHESIS AND THEORETICAL RATIONALE:

For the primary hypothesis, tracts with lower median incomes (poverty and lower-middle class median incomes) are likely to have lower levels of educational attainment compared to other tracts. The secondary hypothesis asserts that tracts with predominately poverty level median incomes likely contain high levels of high school graduates only compared to bachelor’s degrees. The two hypotheses are based on a couple of assumptions. College education is associated with higher paying jobs and more stability, since higher paying jobs require critical thinking skills that are usually obtained through higher education. Additionally, communities with a low average income likely have worse public schools and fewer opportunities to pursue higher education.

METHODS AND DATA SOURCES:

The dataset used is from the 1-year ACS 2020 US Census Bureau for Summit County, OH.[[3]](#footnote-3) This dataset has the most available socio-economic predictors and provides data close to the current year. The US Census Bureau dataset was chosen due to the agency’s reliability, credibility, and varied access to many different kinds of free and publicly available data. To test the two hypotheses, this study will utilize a multiple regression analysis model. Regression analysis is used due to the nature of the dependent variable being a quantitative value. This kind of analysis provides a simple but effective method to observe the relationship between income and the various socio-economic predictors while controlling for other factors that may influence median income. The statistical significance of the model will be assessed using p-values, to determine if the null hypotheses are rejected or fail to reject. This approach allows for an evaluation of the hypothesized relationships to observe the general trends within each tract.

FINDINGS:

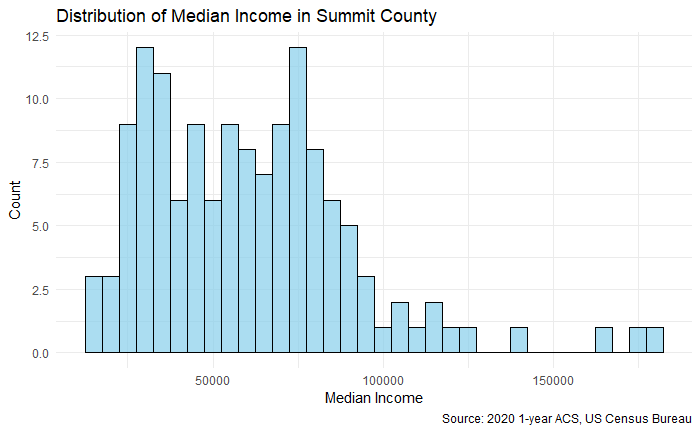
Table 1

Variable Descriptive Statistics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | Mean | Standard Deviation | Minimum | P25 | Median | P75 | Maximum |
| Median Income | $61,171 | $30,800 | $12,898 | $36,280 | $57,843 | $77,342 | $179,773 |
| High School | 746 | 388 | 83 | 457 | 748 | 1,002 | 2,038 |
| Bachelor’s Degree | 577 | 465 | 1 | 188 | 502 | 780 | 2,349 |

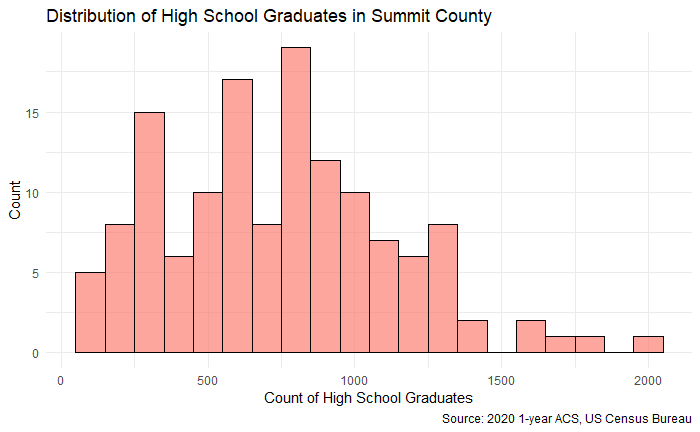
Table 1 shows the median income for Summit County, OH ranging from $12,898 to $179,773. The median income is less than the mean value indicating that higher income earners are affecting the average income in the county. This suggests that there are high-income tracts creating a positive skew in the distribution. Interestingly, the median income of $57,843 is less than the $60,000 mark used to identify lower-middle class incomes. Therefore, most tracts contain lower-middle class to poverty level median incomes compared to middle class and up median income tracts. The maximum and minimum differences between high school and bachelor degree graduates suggests that there are tracts with vast extremes of either bachelor degree and high school degree holders or possibly both. The relatively high standard deviations for both variables also suggest a large variation in tracts between high school and bachelor degree holders. For median income, values greater than $138,935 are considered outliers (77,342 + 1.5 \* 41,062). For high school holders, values greater than 1,819 are outliers (1,002 + 1.5 \* 545). For bachelor degree holders, values greater than 1,668 are outliers (780 + 1.5 \* 592). There are no lower bound outliers for any of the variables.

Figure 1



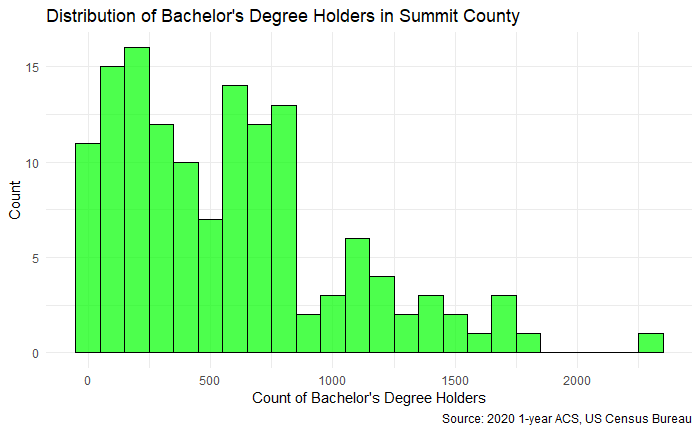
The histogram in Figure 1 confirms the positive skew assertion shown in Table 1. Most of the tracts fall within the 0-$60,000 range indicating that most tracts in Summit County, OH have a median income at the lower-middle class level. There are also a significant number of tracts that are below the federal guideline for poverty. Figure 1 shows that there are at least three high outliers. The right tail and high outliers indicate that a few specific tracts earn a median income much higher than the rest of the county.

Figure 2



The histogram in Figure 2 appears to have a more normal distribution than Figure 2. There is a slight right tail indicating that specific tracts have high amounts of just high school degree holders. However, most tracts appear to hold 500-1,000 high school degree holders. Figure 2 shows at least 1 high outlier, this is due to the tract containing a value greater than 1,819. However, the median and mean are almost the same value indicating a normal distribution. Overall, the high school graduate distribution suggests that most tracts within Summit County, OH will contain low, medium, and high levels of high school graduates.

Figure 3



The histogram in Figure 3 has a significant right tail or positive skew. Most tracts fall within the range of 0-800 bachelor degree holders. Figure 3 shows that there are about 2-3 high outliers, this is due to these tracts containing values greater than 1,668. The presence of high outliers, and the distribution having a higher mean value compared to its median indicates a positive skew. Overall, the bachelor degree univariate analysis suggests that most tracts within Summit County will contain low to medium levels of bachelor degree holders.

Figure 4

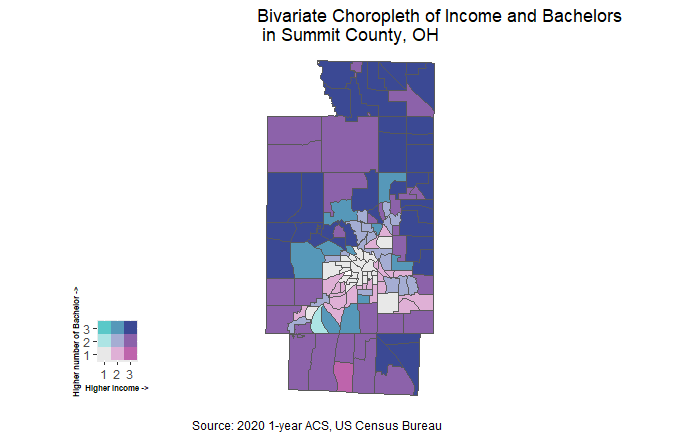


Figure 4 is a bivariate choropleth showing the relationship between median income and bachelor degree holders within each tract. The darker color variations represent tracts with middle class and up median incomes. The medium shaded color variations represent tracts with lower-middle class median incomes. The lightly colored variations represent tracts with poverty median incomes. The blue color signifies tracts with higher bachelor holders, while the lighter colors represent tracts with lower levels of bachelor holders. For example, figure 4 has no teal-colored tracts, this indicates that there are no tracts with poverty level median incomes and a high level of bachelor holders. Figure 4 reinforces the primary hypothesis that tracts with lower median incomes are likely to have lower levels of education attainment. There are more tracts with middle class and up median incomes with a higher level of bachelor degrees compared to tracts with middle class and up median incomes with a lower level of bachelor degrees. There is only one tract with a middle class and higher median income with a low number of bachelor degree holders. Comparatively, there are at least over a dozen tracts with middle class and up median incomes with a higher level of bachelor degrees. Figure 4 supports a strong relationship between middle class median income and bachelor holders.

Figure 5

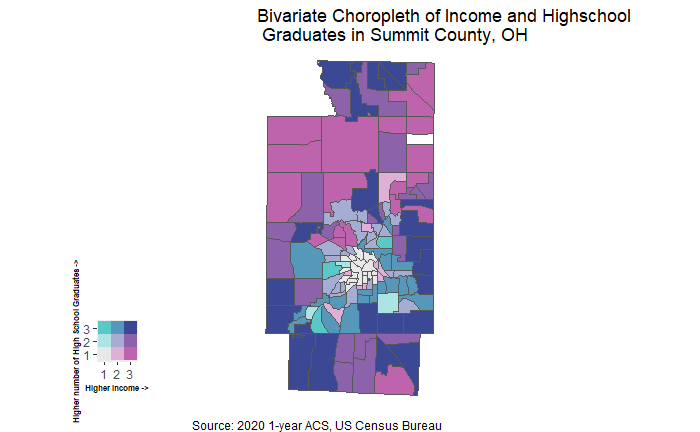


Figure 5 is a bivariate choropleth showing the relationship between median income and high school holders within each tract. The color definitions are the same as in Figure 4, except this choropleth is measuring high school holders rather than bachelor holders. Figure 5 somewhat supports the secondary hypothesis that tracts with predominantly poverty level median incomes likely contain high levels of high school degree holders. Many of the dark blue colored tracts from Figure 4 are pink in Figure 5, suggesting that the middle class and up median income tracts are tied with higher education level. However, both Figure 4 and 5 contain tracts with middle class and up median incomes that have higher levels of high school and bachelor degree holders. Overall, Figure 5 has a higher amount of poverty level median income tracts with higher levels of high school holders compared to Figure 4 that had no tracts with poverty level median incomes and a higher level of bachelor degree holders. Figure 4 and 5 contain white colored tracts around the Akron metro area. This indicates that both bachelor degree and high school degree holder levels are not influencing poverty median income levels there. It is likely that the tracts centered in the Akron metro area contain high levels of no high school degree holders.

Table 2

Multivariate Regression Model Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Standard Error | P- value | Adjusted R^2 |
| Highschool | -14.92 | 4.94 | 0.003 | .49 |
| Bachelor | 47.24 | 4.12 | ~0.0 |  |

Both the predictor variables in the multivariate regression model were very significant at the 1% level. This is due to both predictors having a p-value less than .01. The high school variable had a negative coefficient suggesting that tracts with high levels of high school degree only holders have a lower median income than other tracts. For each additional individual with a high school diploma in a tract, the median income in that tract decreased by $14.92. The bachelor degree variable had a positive relationship with median income. For each additional individual with a bachelor degree in a tract, the median income in that tract increased by $47.24. Additionally, the adjusted R^2 value of nearly .5 indicates that the model accounts for nearly half of the total variance affecting median income at the tract level.

DISCUSSION:

Based on the multivariate regression model and the findings from the two bivariate choropleths, the primary and secondary hypotheses support the alternative hypotheses. Both predictor variables were very statistically significant in predicting median income at the tract level. The visualizations supported the statistical significance of the model as well. Generally, tracts observed with middle class and higher median incomes yielded higher levels of bachelor degree holders. However, there was some overlap in middle class and up median income tracts with higher levels of both bachelor and high school holders. It is likely that the bachelor degree holders within the tract raised the median income to middle class level based on the findings from the multivariate regression model. Conversely, tracts with poverty level median incomes generally had higher levels of high school only holders and lower levels of bachelor degree holders. Combined, the bivariate choropleths highlighted a new finding that the study didn’t account for. Tracts with a higher level of no high school degree holders. The white shaded tracts in both choropleths represent the Akron, OH metropolitan. The white colored tracts represented tracts with a poverty median income and a low level of high school or bachelor degree holders. Since both choropleths had the same white colored tracts, this likely means that these tracts contained a high level of no high school degree holders. Assuming that these tracts do not have a high level of master degree or higher holders. To confirm this new finding, future modeling and testing can use a no high school degree variable to confirm or deny if the Akron metro area has a high level of no high school degree holders. Further, it would be interesting to see how master degrees and PhDs affect median incomes per tract. This is a relatively straight forward process that is easily achievable, it was not included in this analysis due to length requirements. The model captured almost half of the total variance, adding in additional predictors may lead to a greater capture of the total variance. There are other socio-economic variables that were not tested including race, marital status, price of housing, rent costs, etc. These socio-economic variables may capture more insights into how they affect median income at the tract level. The future hypothesis would claim that tracts with poverty median incomes have a higher level of no high school degree holders and have a higher black population. The US Census contains variables related to no high school degree holders and has a black race variable. The future study would conduct a multivariate regression model using median income as the dependent variable again with bachelor, high school, black, and no high school as predictor variables. The additional choropleth visualizations and new multivariate model would likely capture more of the variance and test the new hypothesis.

APPENDICES:

Ohio Department of Children and Youth. (2024). *2024 federal poverty guidelines*. Education Ohio. <https://education.ohio.gov/getattachment/Topics/Early-Learning/Early-Childhood-Education-Grant/Early-Childhood-Education-Grants-for-Administrator/2024-Federal-poverty-Guidelines.pdf.aspx?lang=en-US>

Fry, R. (2024, September 16). Are you in the american middle class? find out with our income calculator. Pew Research Center. <https://www.pewresearch.org/short-reads/2024/09/16/are-you-in-the-american-middle-class>

Bureau, U. S. C. (2020). Summit County, Ohio. Explore census data. <https://data.census.gov/profile/Summit_County,_Ohio?g=050XX00US39153>

1. Ohio Department of Children and Youth. (2024). *2024 federal poverty guidelines*. Education Ohio. https://education.ohio.gov/getattachment/Topics/Early-Learning/Early-Childhood-Education-Grant/Early-Childhood-Education-Grants-for-Administrator/2024-Federal-poverty-Guidelines.pdf.aspx?lang=en-US [↑](#footnote-ref-1)
2. Fry, R. (2024, September 16). Are you in the american middle class? find out with our income calculator. Pew Research Center. https://www.pewresearch.org/short-reads/2024/09/16/are-you-in-the-american-middle-class [↑](#footnote-ref-2)
3. Bureau, U. S. C. (2020). Summit County, Ohio. Explore census data. https://data.census.gov/profile/Summit\_County,\_Ohio?g=050XX00US39153 [↑](#footnote-ref-3)