Michelle Rice

DSC680-T301 Applied Data Science

Project 1

Income Level as a Predictor of Academic Achievement

9/18/2021

**Executive Summary**

According to Wikipedia, “Educational inequality refers to unequal access to [education](https://en.wikipedia.org/wiki/Education), and the unequal outcomes for students that result. The disparities present in academic access among students in the United States are the result of several factors including: government policies, [school choice](https://en.wikipedia.org/wiki/School_choice), family wealth, [parenting style](https://en.wikipedia.org/wiki/Parenting_styles), [implicit bias](https://en.wikipedia.org/wiki/Implicit_bias) towards the [race](https://en.wikipedia.org/wiki/Race_(human_categorization)) or [ethnicity](https://en.wikipedia.org/wiki/Ethnic_group) of the student, and the resources available to the student and their school.”

Since family wealth has been identified as a factor that results in educational disparities, it is important to identify and consider variances in income and how they are impacting the success of students. One we see where the gaps in income are, we can compare that to the educational achievement in those areas and determine how dependent, if at all, educational achievement is on income level.

Understanding these gaps could be very valuable in helping us make changes and implement strategies to focus on improving achievement where it is currently lower. This could guide where and how funding should be spent to provide more resources where they are needed.

**Abstract**

This project is to study the correlation between income level and educational achievement. The data that I will be using comes from the USDA Economic Research Service website. The dataset contains very detailed information, at a county level, about the population of the county, demographics, income levels, jobs, etc. The information that I will specifically be focusing on is the median household income, poverty percentages, and education levels. The data breaks the educational achievement down as follows: less than high school diploma, high school diploma, some college, associates degree, or 4 years of college or more. I will focus on less than high school, high school only, and 4 year degree or more. The data also gives each county a rural/urban continuum code which assigns the county a code based on population. I will also evaluate that code as it relates to education level to see if living in a rural or urban area has any correlation to outcomes.

**Background**

The median household income in the United States in 2019 was $62,843. The poverty rate for children age 0 – 17 in the US was 18.52% and the percent of children living in deep poverty (a household with total cash income below 50% of it's poverty threshold) was 8.16%. The hypothesis being tested is that family wealth is a factor in determining educational achievement. If this is in fact true, we would expect to see low graduation and low college rates among those 18% for sure, and would also expect to see some sort of linear relationship between income levels and education levels.

In this study, I would like to look at how strong this correlation is, and if we can identify any income levels where we see trends shift. This could help determine where resources should be spent and where we might want to make changes.

**Problem Statement**

The problem statement is: can we predict educational outcomes based on income level? If there is correlation, can we find any specific income levels where outcomes noticeably shift? Also does being in an urban or rural community have any correlation?

**Initial Data Exploration**

The first thing that I did was to look at some high level values for the entire US. As mentioned previously, the median income was $62,843. Figure 1 shows a plot of the percentage of people in each education level in the United States. While 32% of Americans have a 4 year degree or higher, 12% have not finished high school. Just over half of Americans have at least a high school diploma and possibly some college or an associate degree. Figure 2 shows the distribution of income. We can see that there is a slightly heavier to the left of the median income, or more lower income families than higher income. I then plotted different education levels against median income (Figure 3-7).

**Methods**

In order to evaluate the correlation between income and educational achievement, I ran a linear regression model. In order to evaluate the correlation between the rural/urban continuum code and education, I first ran a linear with the continuum code as a predictor of education level (specifically college degree level), then ran a multiple regression model with both the continuum code and income level as predictors of a college degree.

**Results**

Looking at the plots of income and educational achievement, we can see that in the first plot, which is less than high school, the distribution is relatively low across all income levels, which coincides with our data and graph showing that only 12 percent of Americans fall into this category. We see heavy distribution around that median income level, with a bit heavier in lower incomes levels. The plot which shows high school diploma again shows a high concentration of data around that median income with only a slightly heaving distribution in the lower incomes. The plots that show some college and associates degree are fairly evenly distributed along the lower percentages, but we do start to see a bit more distribution to the right of median income. Finally, the plot which shows 4 year degree or more, shows a distribution in the higher percentages and also possibly a distribution slightly more to the higher incomes.

Since the scatter plots are a bit difficult to interpret with the number of data points, I created the plots with the smooth function and see some interesting curves in the lines of data. We can see that the red line, which represents less than a high school diploma, is highest at the lowest income levels and then goes down as income goes up. The black line, which represents a 4 year degree or more, has the opposite trend, it is lowest at lower income levels and increases as income increases. The blue line, which is a high school diploma only, is interesting as it increases briefly at a low income level, but then steadily decreases as income goes up. A key point to note is that at about $100,000, the blue and black lines meet, indicating that at that income level, there is about an equal rate of college. Before that level, the percentage of college degrees is lower than high school diplomas, but above that level the amount of 4 year degrees is higher than high diplomas only. This would initially indicate that the hypothesis that achievement is correlated to income would be true. The green line, representing some college, goes up slightly as income level rises at a lower range, but then levels out across the remaining income levels. The yellow line representing associates degree is steady around the 9% range for all income levels.

The regressison models all returned a very low p value, again indicating that these variables do have correlation to educational completion. I then looked at the correlation coefficients to see how strong the correlation is for each group. The results are below:

Correlation based on 10,000 change in income

* Less than High School: -0.76760
* High School Only: -0.7578
* Some College: 0.04101
* Associates Degree: 0.06095
* College Degree and Above: 1.42348

As the charts showed, we can see a negative correlation with income and less than high school or high school diploma only. The correlation then shifts to positive as we look at college levels. The highest correlation is for the college and above group.

The regression models using the Rural Urban Continuum showed that there was a

-0.41047 correlation coefficient for the continuum, however, when adding income to the regression model, that coefficient dropped to -0.1640, so it appears that the relationship to the setting is not as strong as it it to income.

**Discussion/conclusion**

I feel that the data collected and analyzed provides us with enough information to say that income level can be used as a predictor for educational achievement. It also appears that $100,000 is an income level where students are more likely to achieve a 4 year degree. It also appears that somewhere around $40,000 is where the number of people who do not finish high school seems to level out. This would indicate that attention should be given in particular to families with income levels below that $40,000 mark.

**References**

Wikipedia. Educational Inequality in the United States. Retrieved from: <https://en.wikipedia.org/wiki/Educational_inequality_in_the_United_States>

UCDavis. What is Deep Poverty. Retrieved from: <https://poverty.ucdavis.edu/faq/what-deep-poverty#:~:text=The%20U.S.%20Census%20Bureau%20defines,percent%20of%20its%20poverty%20threshold.&text=While%20poverty%20thresholds%20vary%20by,income%20below%20%246%2C243%20in%202016>.

Dataset retrieved from: <https://www.ers.usda.gov/data-products/atlas-of-rural-and-small-town-america/download-the-data/>

Appendix

Figure 1 Percentage of Americans at each education level

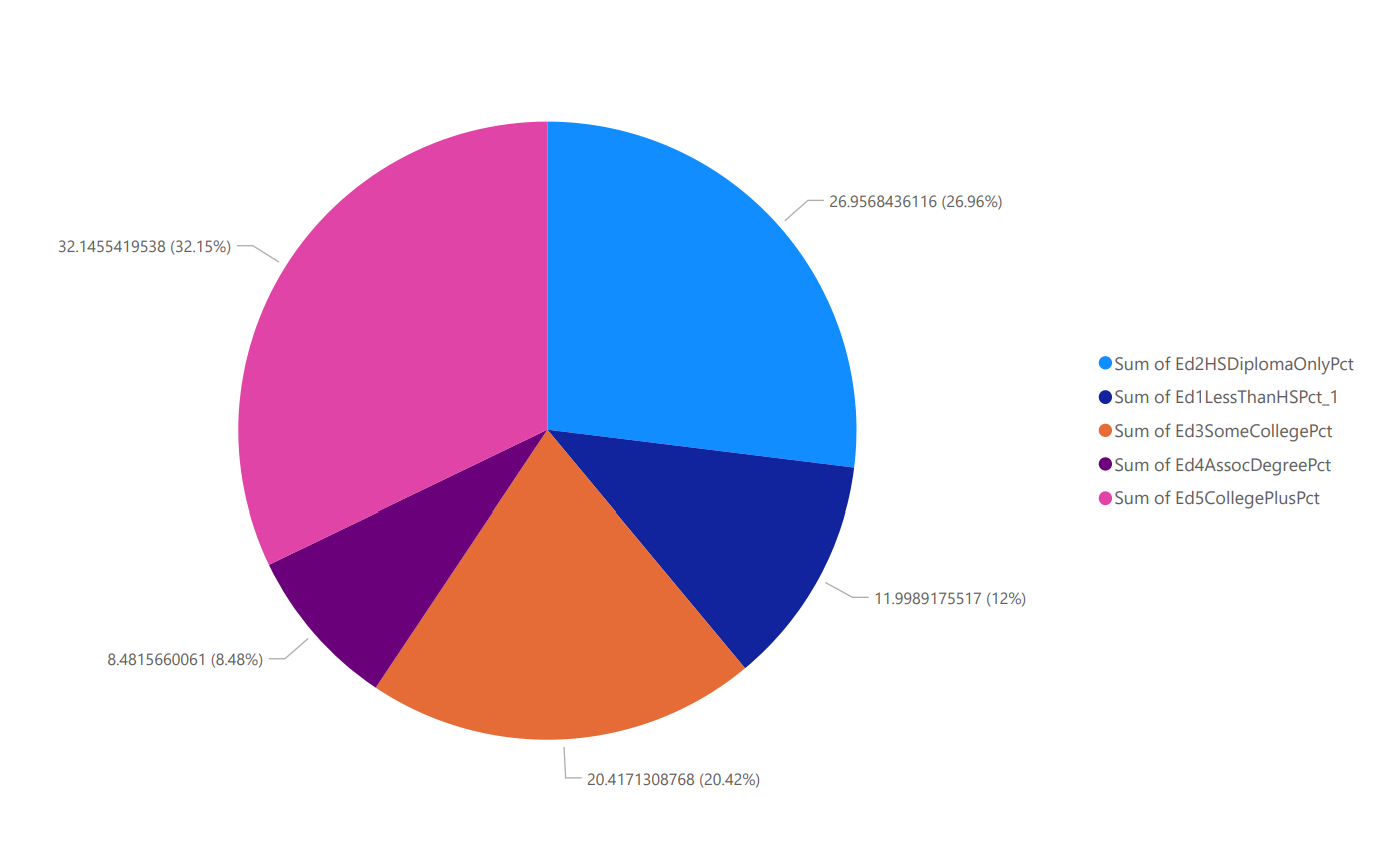


Figure 2 Distribution of income in the United States

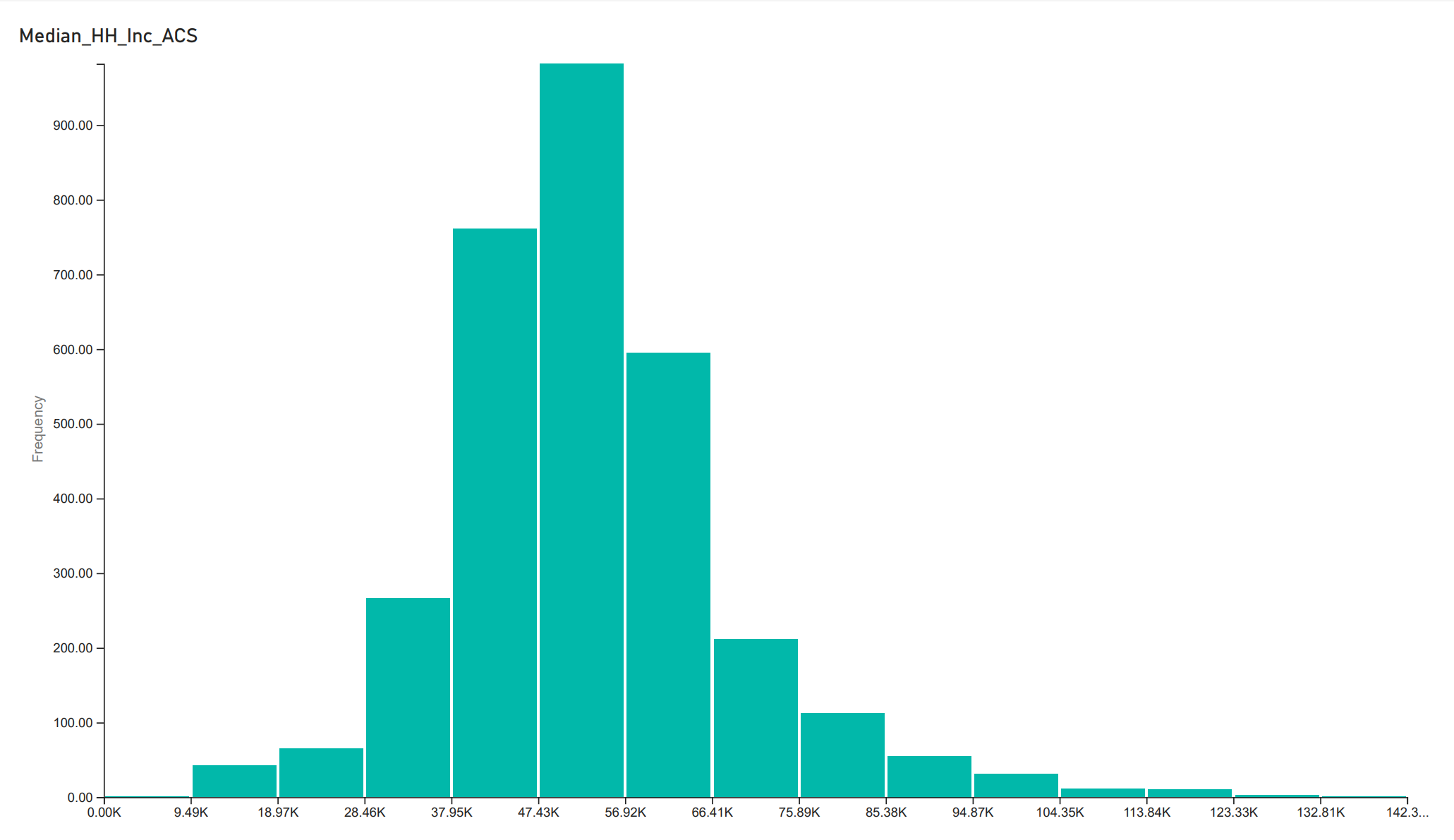


Figure 3 Percentage of individuals with less than a high school per income level

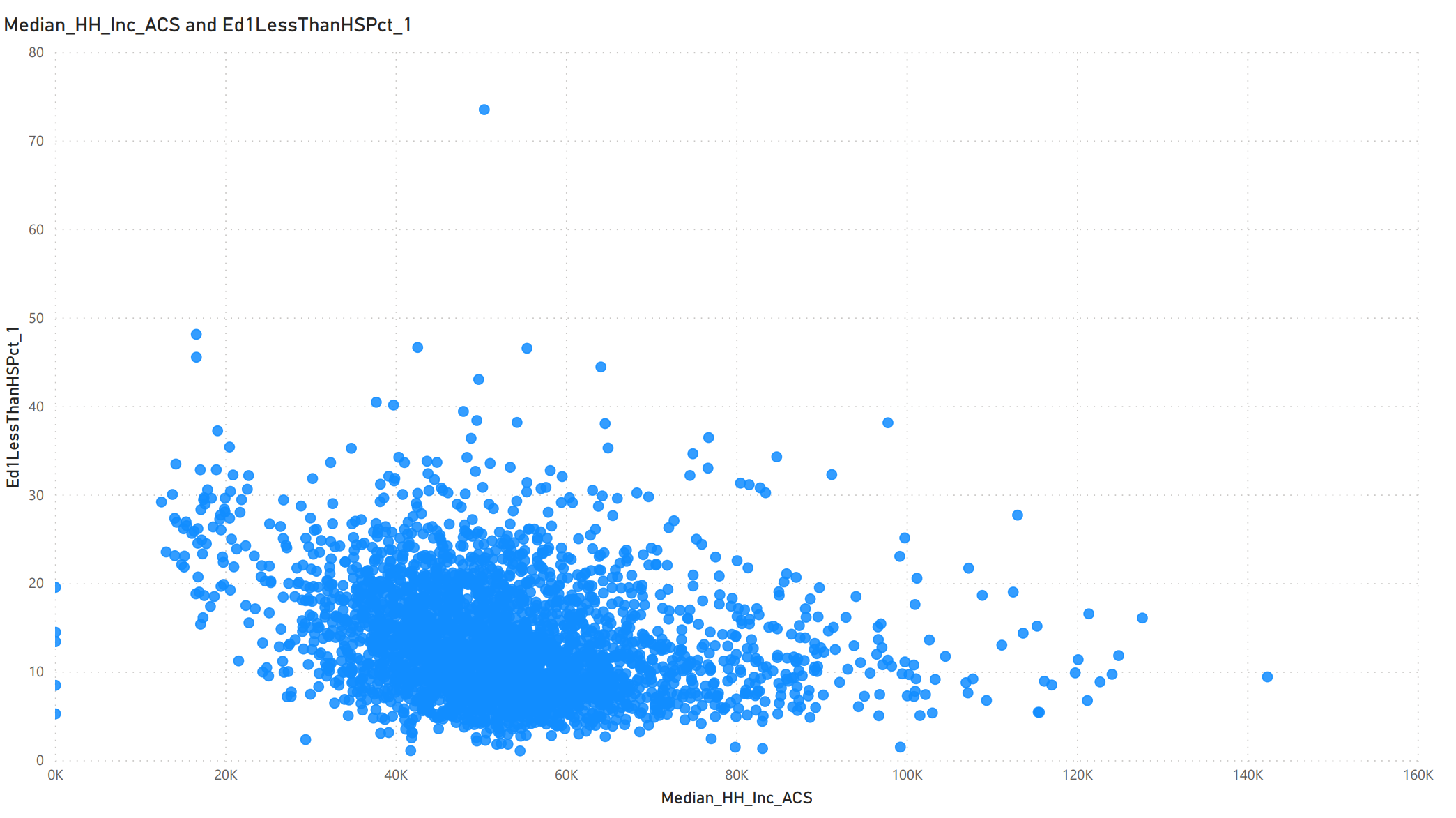


Figure 4 Percentage of individuals with a high school diploma per income level

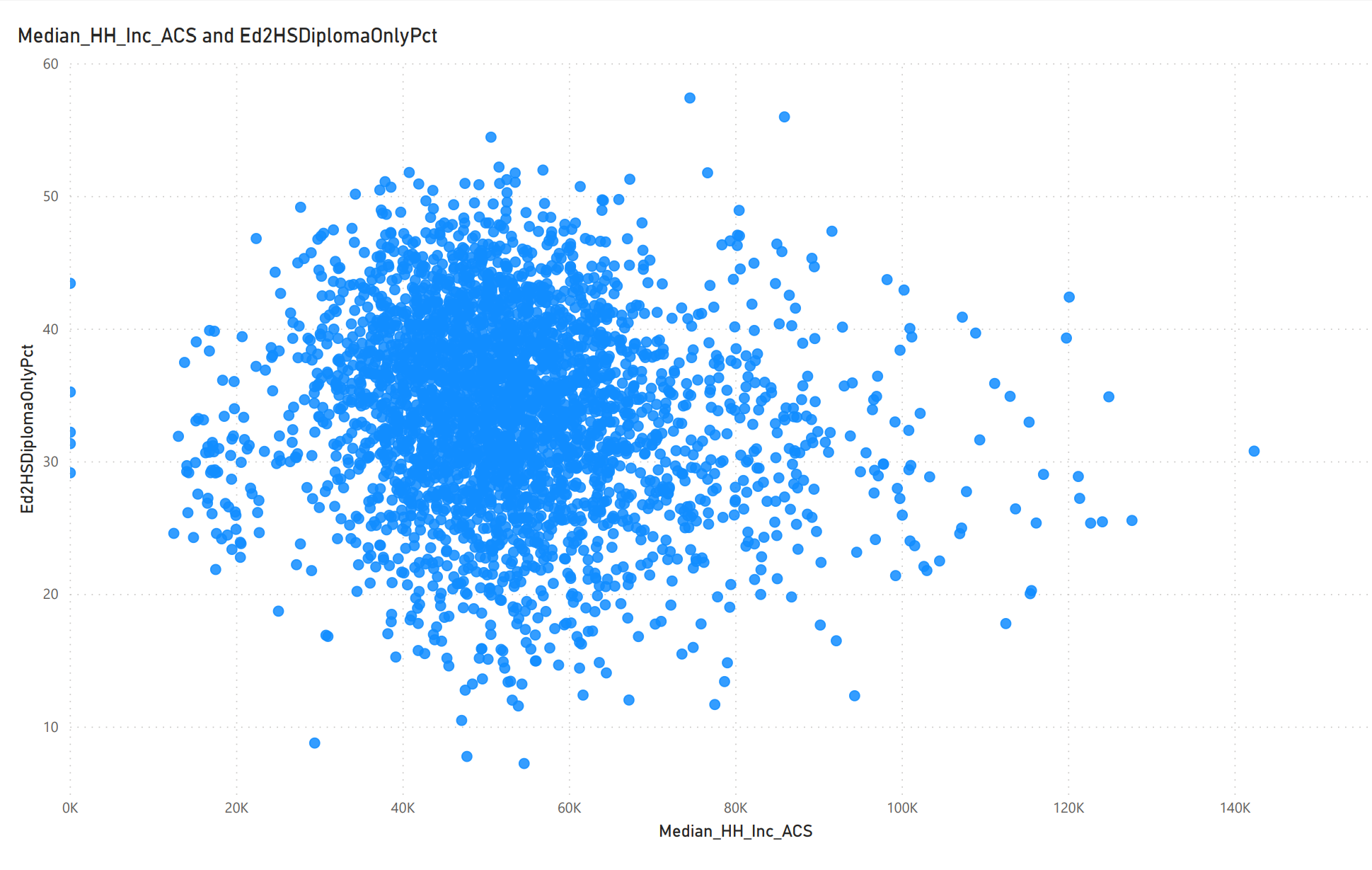


Figure 5 Percentage of individuals with come college per income level

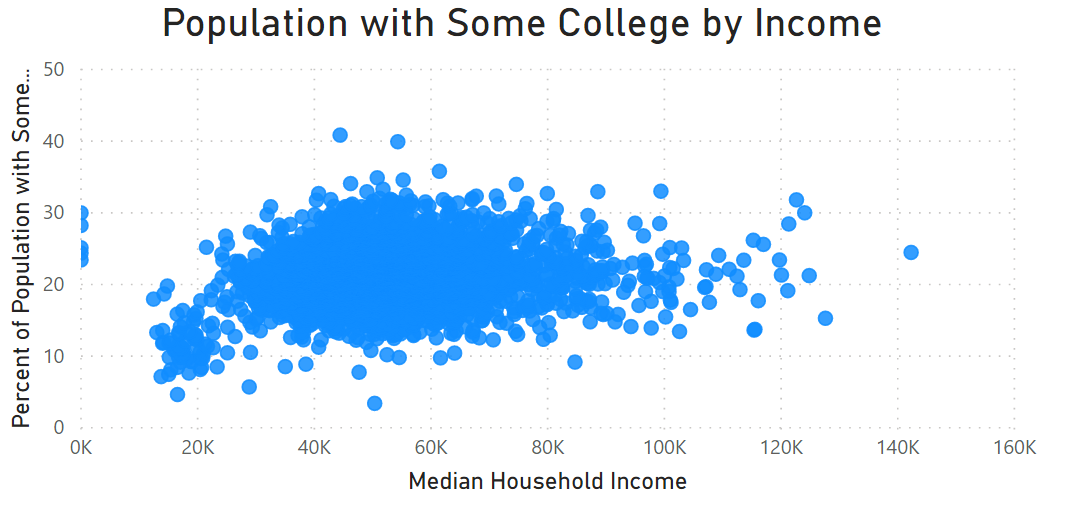


Figure 6 Percentage of individuals with Associates degree per income level

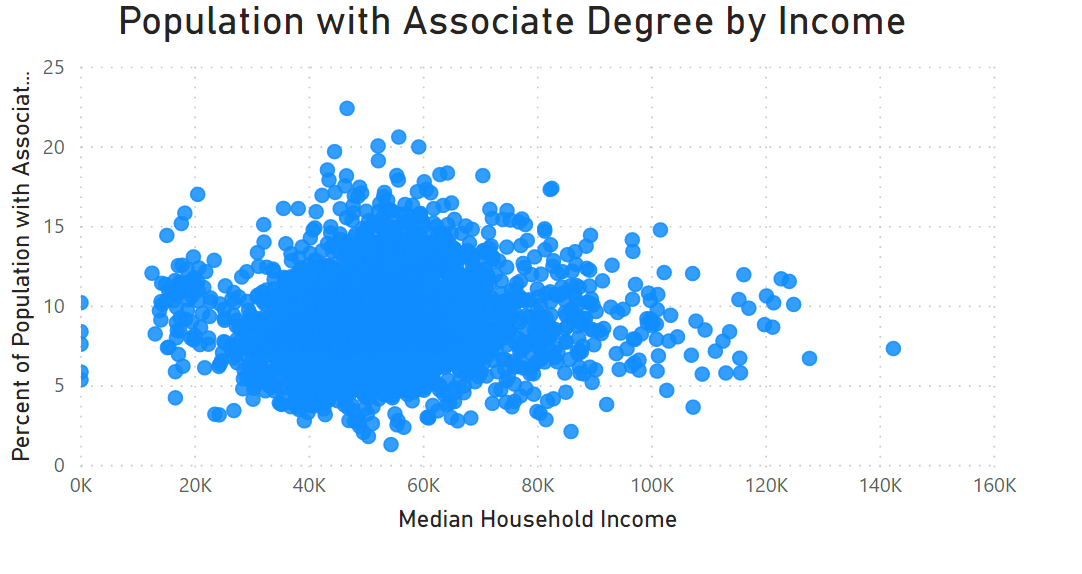


Figure 7 Percentage of people with a 4 year degree or more by income level

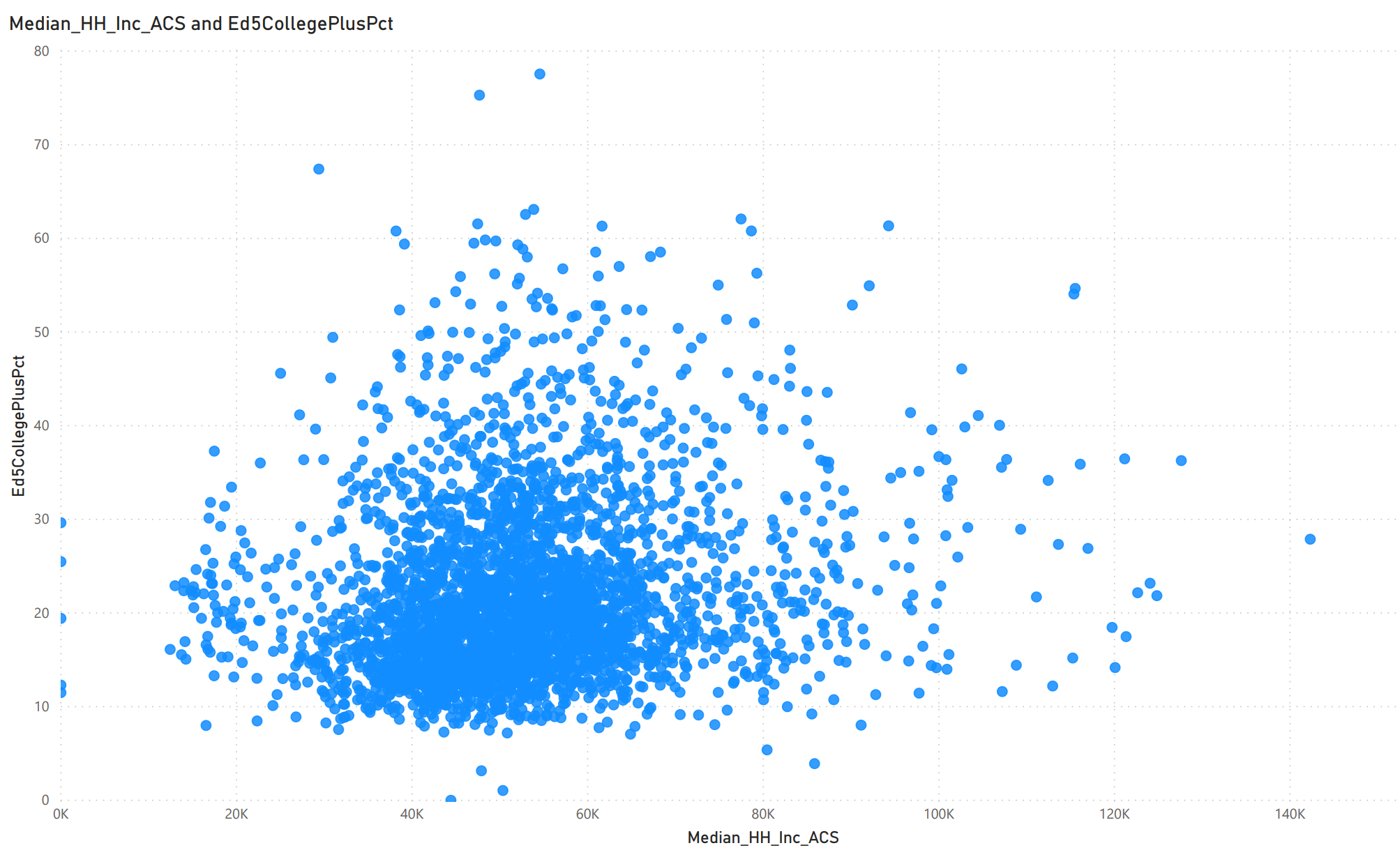


Figure 8 Adding a smooth line to more easily compare these charts

