Measuring Socio Economic Status in Jersey City Census Tracts

Selecting Indicators Using a Measurement Theory Approach

Overview:

We find strong evidence that five variables at the tract level — median household income, average educational attainment, percent of children living in two parent households, percent of households whose expenditures on rent or mortgage are manageable for their income, and percent of working age adults who are employed — are strong indicators of socioeconomic status in Jersey City, New Jersey.

An Issue of Measurement

Socio-economic status (SES) can be difficult to define. Is it yearly income? Yes. But not entirely. Does an individual with a law degree and a career in corporate law suddenly lose their SES because they become a district attorney and take a salary cut? No, not likely. Is it educational background? That could be part of it. But, an individual who does not attend college can start a successful company; education cannot be the sole measure. Maybe it is occupation or community involvement? School board members tend to be well-respected community members. But does being on the board automatically mean you have high SES?

We can think of all of these things as outcomes — or *symptoms* — of SES. In statistics, we call them *indicators*. Just like when you have a cough, you cannot automatically assume you have the flu, each component individually is not enough for us to draw conclusions about SES. But, when we look at them together, they give us more confidence about the underlying cause. When we find a set of strong indicators, we are able to say that as a person's SES increases, the likelihood of that person scoring more highly on each individual indicator also increases. If we can do this successfully, we can infer SES.

A Solution

This is the basis of measurement theory. When we try to measure a variable that we cannot directly observe, like SES, we need to find a group of indicators that (a) we have reason to believe might be related to SES and (b) have *shared variation. Shared variation* means that, as one indicator increases, the others tend to do so as well. There are statistical methods to assess variation. Using these methods, we can test whether or not it is reasonable to believe that a group of indicators are symptoms of the same underlying cause.

SES is Location Based

Generally, when we talk about SES, we speak in relative terms — one person's status is in contrast to other people in their community. This means factors related to socio-economic

status tend to change from place to place. For example, in a rural setting, owning significant amounts of land may be an indicator of higher SES. In contrast, in an urban setting, how close the land is to downtown may be more influential than how many vegetables can be grown on it. For this reason, we need to identify indicators of SES that are reliable within the context of Jersey City.

The Options

In this study, we explore three sets of indicators:

(1) The indicators that the City of Chicago uses to measure SES for their school admission.
(2) An adaptation of the Chicago model to better match the Jersey City context.
(3) Indicators selected using the National Center for Education Statistics definition of SES.

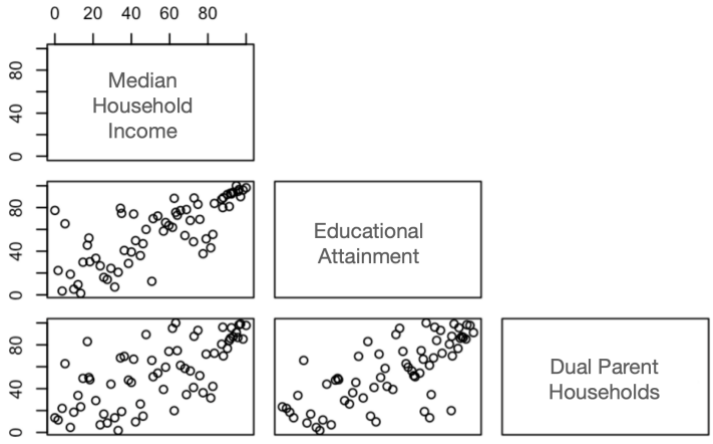
Note: All variables are reported at the tract level.

The Results:

(1) **City of Chicago Indicators:** Median Household Income, Educational Attainment, % Children in 2 Parent Households, % Households Who Own homes, % Households Who Speak English as the Only Language

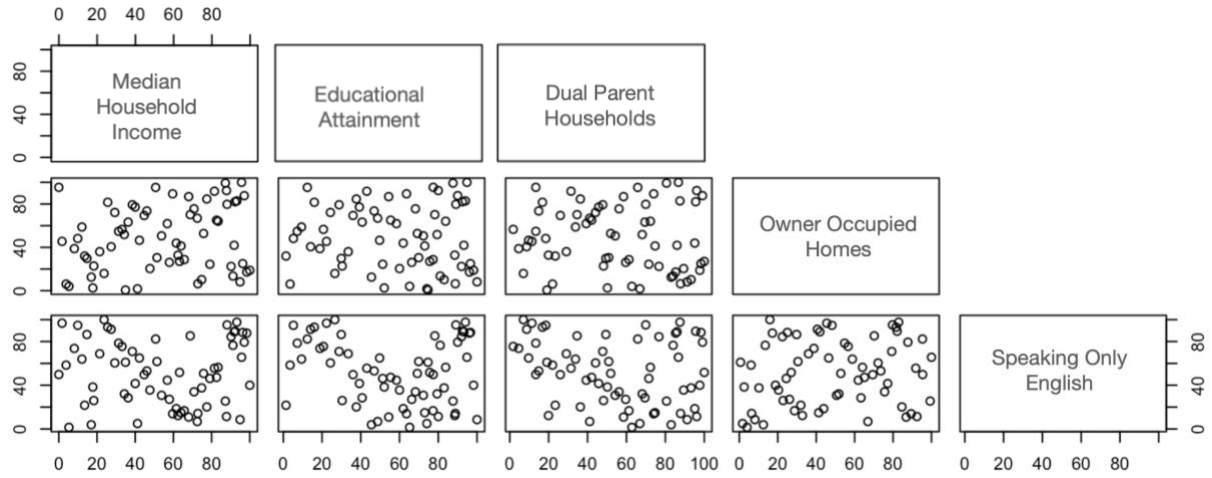
Based on our analyses, three indicators — Median Household Income, Educational Attainment, and % Children in 2 Parent Households — have significant shared variance. In other words, because they show similar patterns of increasing and decreasing, it may be reasonable to believe that they are all symptoms of the same thing — SES. Figure 1 shows this below. As you can see, the points in all three plots show trends running from the bottom left to the top right. This is a positive correlation. In these three factors, the correlation is strong.

Figure 1: Correlations Between Tract-level Median House Hold Income, Educational Attainment, and the Percentage of Students in 2 Parent homes.



Unfortunately, the other two indicators — % Households who own homes, and % Households Who Speak English as their Only Language — do not follow similar patterns. There is almost no relationship between Owner Occupied Homes and the other three variables. There is also a weak negative relationship between Speaking Only English and the three stronger variables. In other words, as Median Household Income, Educational Attainment, and Dual Parent Household Percent increase at the tract level, the Percent of Households who Speak Only English tends to decrease.

Figure 2: Tract Level Correlations



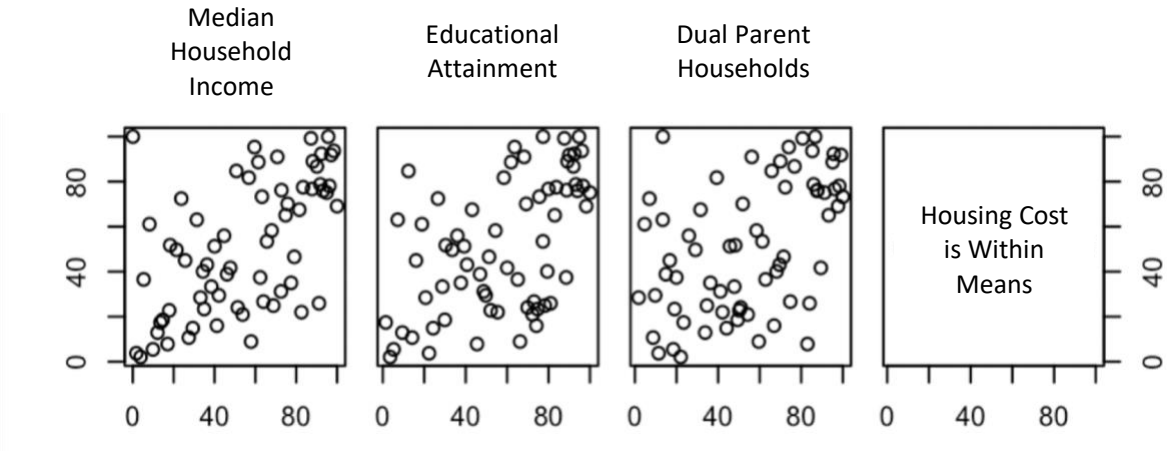
While Owner Occupied Homes and Speaking Only English may be indicators of higher SES in some tracts, the evidence they provide is not stable enough for us to use them confidently. Beyond examining the correlations, other procedures that we use in measurement theory (reliability and dimensionality) suggest that making an SES score based on these five variables would be neither advisable nor defensible.

(2) **Option 2: A More Appropriate Measure of Housing Cost:** Median Household Income, Educational Attainment, % Children in 2 Parent Households, % Households Whose Expenditures on Rent or Mortgage are Manageable for their Income

We saw that the first three indicators — Median Household Income, Educational Attainment, and % Children in 2 Parent Households — were strongly related to the same measure of SES. But if we want to incorporate a measure of housing affordability, we need to choose something more appropriate for Jersey City. In a space where renting a luxury apartment may be more an indicator of SES than home ownership, we tested a metric of rent burden at the tract level. The Department of Housing and Urban Development defines an individual as rent burdened if they spend more than 30% of their monthly income on rent. We use this metric to estimate the percent of households who can afford their housing.

As you can see in the charts below, the population who can afford housing is positively correlated with the other three economic variables. Other statistical tests of reliability show these four indicators to be stably related to each other and to the same underlying cause, although with some caveats that show room for improvement.

Figure 3: Correlations Between Tract-Level Housing Cost and Other Factors



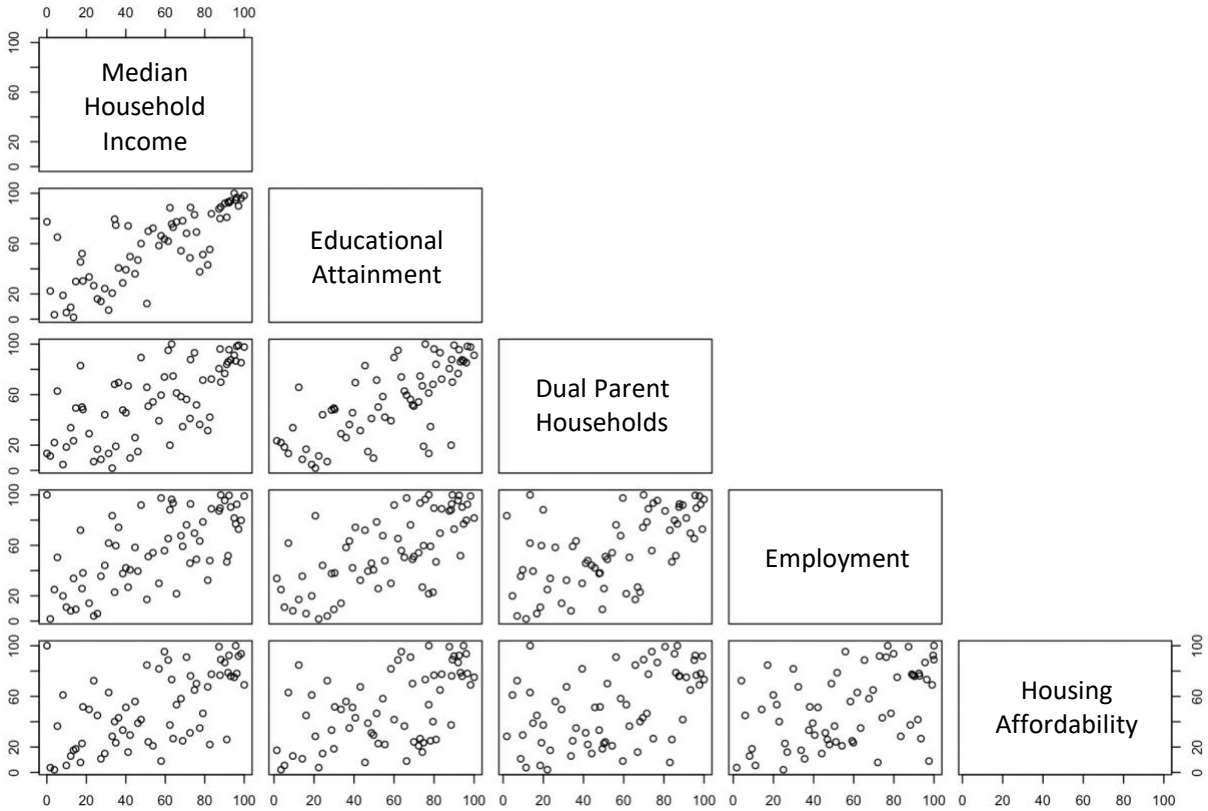
(3) **Option 3: Incorporating NCES Considerations:** Median Household Income, Educational Attainment, % Children in 2 Parent Households, % Households Whose Expenditures on Rent or Mortgage are Manageable for their Income, & % Working Age Employed

While the four indicators presented in **Option 2** are strong indicators of SES, adding a fifth indicator could stabilize the measurement. The National Center for Education Statistics[1] identifies three core components of socio-economic status—family income, parental education, and parental occupational status. From those dimensions, we already include indicators of family income/assets (household income and housing cost) and indicators that measure the resources available to a student to support and augment their learning (Community Educational Attainment and Having Two Parents in the Home). Adding employment data, specifically the percent of working age individuals who are employed, would be a reasonable addition to round out the model.

As shown in Figure 4, below, all of the variables exhibit strong correlations with each other. As each variable increases, so do all of the others. Further statistical analyses also suggest that the shared variance between Employment and Housing Affordability helps bridge Housing Affordability and the other three variables. This means that, if these five indicators were combined into a single measure, we could be confident that as the composite measure increases, so would all of the component indicators.

[1] https://nces.ed.gov/nationsreportcard/pdf/researchcenter/socioeconomic_factors.pdf

Figure 4: Pairwise Correlations of Best Five Indicators



This five indicator model has a few other advantageous statistical properties.

First, it has high reliability (also known as internal consistency). Reliability is a measure of how much the indicators vary individually in comparison to how much they vary together. In other words, high reliability means that if a census tract experiences a change in SES, we can be confident that its indicators will change as well and that, using our model, we will be able to detect that change. We measure Reliability with a statistic called alpha. Generally, in social science, alpha values above .7 are defensible, above .8 are rare but very good, and above .9 are exceptional. This scale has an alpha value of .88, which is very strong.

Second, the scale is strongly unidimensional. Dimensionality assesses how many underlying factors may be causing the indicators observed. For example, an article of clothing, like pants has two dimension to it — waist and length. Conducting a dimensionality assessment on indicators such as the height, weight, and BMI of male students wearing pants would likely show that height is associated with a length dimension whereas BMI is associated with a separate a waist dimension. Because weight correlates to both pant length and pant waist size, the assessment would show that some variance in weight loads onto the same factor as student height, whereas some variance in weight loads onto the same factor as BMI. The point is, examining the data using a dimensionality assessment will reveal whether or not the

indicators are all generated from one or from multiple underlying dimensions. In our case, the dimensionality assessment suggests that all of the indicators arise from the same, shared factor. The two factor and three factor models tested did not sufficiently explain the variance as well as the one factor model.

Given these properties, we can conclude that this five-indicator model is a defensible method to measure SES in Jersey City.