Urban Data: Final Project Milestone

Names & NetIDs:
Emma Condie - erc97
Sarang Pramode - sp872
Marie Demple - med287

Progress

As we have seen in the preliminary research we've done for this project, there have been many attempts to analyze the process of gentrification through a quantitative lens. Most of the research that we have encountered looks at basic demographic data such as income, housing costs, and education and how these metrics change over time in a given geographic area such as a census tract. Some studies add in other measures such as changes in the amount and type of 311 calls or Airbnb listings in a neighborhood. Our takeaway is that because there is no single definition of gentrification, there is room for us to experiment, expand upon, and hopefully contribute a new definition of this phenomenon to the existing body of research.

Some Definitions of Gentrification:

1. "The process by which central urban neighborhoods that have undergone disinvestments and economic decline experience a reversal, reinvestment, and the in-migration of a relatively well-off, middle and upper middle-class population." (*The Encyclopedia of Housing* (Smith 1998, 198)) (source)
2. "The replacement of low-income, inner-city working class residents by middle- or upper-class house-holds, either through the market for existing housing or demolition to make way for new upscale housing construction." (Hammeland Wyly (1996, 250)) (source)
3. "The process by which a neighborhood occupied by lower-income households undergoes revitalization or reinvestment through the arrival of upper-income households" (U.S.Department of Housing and Urban Development 1979, 4) (source)

In seeking to define the characteristics of a neighborhood undergoing gentrification, the following elements are consistent across many definitions:

1. The neighborhood is in an inner city (most literature on gentrification does not include suburban neighborhoods)
2. Historically, the neighborhood was populated by a majority of low-income households
3. The neighborhood has previously experienced disinvestment (e.g. due to "white flight" in the 1960's and 1970's)
4. The neighborhood experienced or is experiencing an influx of new residents who are affluent relative to the existing neighborhood population
5. The neighborhood itself experienced or is experiencing an increase in investment (real estate investment, new businesses and other amenities)

In addition to reading articles about how people have approached similar projects in the past, so far we have gathered data from several different sources to prepare for our analysis. We have also found a few existing 'gentrification score' formulas that we are using as a basis to try out different versions of our own formula. Our goal for this project is to create a formula that outputs a neighborhood's 'gentrification score' based on inputs such as demographic data from the U.S. Census (age, education, income, rent, house prices, etc.), building data (the age and quality of housing stock), and other sources such as the 311 database from OpenData NYC. Gentrification is a dynamic process, therefore it's necessary to look at changes in these characteristics across time.

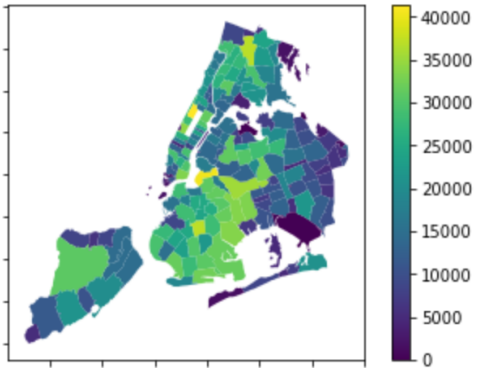
To quantify gentrification trends it is important to define the characteristics of the neighbourhood and information we are capturing. To this end we came up with the following definition of gentrification and the metrics which we felt best fit the trends after our research.

"The replacement of low-income, inner-city working class residents by middle- or upper-class house-holds, either through the market for existing housing or demolition to make way for new upscale housing construction"

- Median household income
- Home Value
- Rent
- Education
- Displacement of minor communities
- Age

Preliminary analysis included dataset exploration and familiarizing ourselves with census data. This included working with shapefiles and merging with census data to obtain maps to visualize trends occurring in neighbourhoods.

The plot below depicts the number of households for each zip code in NYC for 2019



Census data obtained from - data.census.gov - Table - S1901
Zip Code shapefile obtained from NYC Open Data

Data Sources

Explore Census Data
2010 Census Tracts | NYC Open Data
Housing New York Units by Building | NYC Open Data
311 Service Requests from 2010 to Present | NYC Open Data

Future Analyses

We will experiment with different versions of our formula and compare the outputs to see which formula produces scores that are most consistent with commonly held views on gentrification. For example, we know that certain neighborhoods in Brooklyn such as Bushwick, Bed-Stuy, and Crown Heights are widely regarded as gentrifying; our formula should produce relatively high scores for these neighborhoods. (Our goal is not to produce a formula that caters to preconceived notions; just to check to make sure we are on the right track). Once we have settled on the best way to "score" neighborhoods for gentrification, we plan to apply the final formula to NYC Census tracts and create maps and other data visualizations from this output.

Problems

Since gentrification is a process that happens over time, we cannot rely solely on data collected from a single year, but must define a range of years on which to base our analysis. Since we are working with Census data from the American Census Bureau, we will be looking at the data they have available from 2010 and 2019. Getting census data was more tricky than we anticipated! This timeframe conveniently cuts off just at the start of the Covid-19 pandemic, which definitely would affect the metrics we are including in our analysis. Additionally data collected at the zipcode level is only available for 2019. This made it difficult to quantity gentrification of a neighbourhood across time. To combat this we decided to use Census tracts as it provided more granular information and had data for our metrics between 2010 and 2019.