

Immigrant Residency and Happiness in New York City

Alison Tuiyott* Thomas Fisher* Karsten Maurer* Robert Garrett*
Ben Schweitzer* Lydia Carter*

Abstract

Exploring the quality of life of immigrants in New York City through housing and neighborhood conditions by creating a happiness metric. Utilizing data provided by the New York City Housing and Vacancy Survey (NYCHVS), the New York City Police Department, the New York City Department of Education, and the New York City Department of Health and Mental Hygiene and reports from Happy City and the New Economics Foundation, a happy score was assigned to each sub-borough in New York City. This happy score evaluated five main domains: work, place, community, education, and health.

Key Words: New York City Housing Vacancy Survey, Happy Score, Immigrant Residency, Data Expo

1. Introduction

Many people around the globe venture to the United States seeking the American Dream. In this analysis, we use the data provided by the New York City Housing and Vacancy Survey (NYCHVS). According to the New York City Department of Housing Preservation and Development (HPD):

the NYCHVS is a representative survey of the New York City housing stock and population. It is the longest running housing survey in the country and is statutorily required. The Census Bureau has conducted the survey for the City since 1965. HPD is the only non-federal agency that sponsors a Census product. The HVS is a triennial survey with data collected about every three years. Each decade, a representative sample of housing units is selected, which represents the core sample.

Our goal is to use this data to explore the quality of life of immigrants in New York City through housing and neighborhood conditions. Our results hope to help guide individuals understand more about immigrant households and how it relates to their quality of life.

To measure quality of life, we utilize a happiness metric. Happiness, according to Happy City and the New Economics Foundation, is a city's success in providing the conditions that create 'sustainable wellbeing'. Sustainable wellbeing is made up of five main domains: work, place, community, education, and health. According to the Happy City Index 2016 Report, the happiness metric,"aims to be a practical tool that can help local policymakers understand how well their city is doing in comparison to the other cities and prioritize key policy areas". Using the data from the housing surveys about immigrant residency, we attempt to find a connection to happiness.

*Miami University Department of Statistics, 501 E High St, Oxford, OH 45056

2. Data

In New York City, there are five boroughs: Brooklyn, Bronx, Manhattan, Queens, and Staten Island. These boroughs are smaller than cities, but larger than towns. Within each borough there are smaller subsections of land almost equivalent to a town that the data provided by the NYCHVS calls sub-boroughs. For the purpose of this paper, we focus on Manhattan and its sub-boroughs. However, the analysis completed for the 2019 Data Expo, was over all of the boroughs in New York City.

In order to compare immigrant residency to happiness, we extract the immigrant residency information from the data provided by the NYCHVS. To create the happiness metric, we need a measure for the five main domains: work, place, community, education, and health. For work and place, we utilize the NYCHVS data. The rest of the domains require data from external resources, including the New York City Police Department, the New York City Department of Education, and the New York City Department of Health and Mental Hygiene.

2.0.1 Immigrant Residency: Place of Birth

In the NYCHVS data, there is a field that identifies the place of the householder's birth. Using this field, we determine whether a given household is an immigrant household. We apply the sample weights provided to the number of immigrant households and calculate the percentage of immigrant households within each sub-borough. **Figure 1** is a map of New York City illustrating immigrant residency by sub-borough.

2.0.2 Work: Income

In the NYCHVS data, there is a field that identifies total household income. We assume this is a relatively decent measure of work. Using this field, we adjust for inflation and calculate the average total household income per sub-borough. **Figure 2** is a map of Manhattan illustrating average total household income by sub-borough.

2.0.3 Place: Rent

In the NYCHVS data, there is a field that identifies monthly contract rent. We assume this is a relatively decent measure of place. Using this field, we adjust for inflation and calculate the average monthly contract rent per sub-borough. **Figure 3** is a map of Manhattan illustrating average monthly contract rent by sub-borough.

2.0.4 Community: Crime

For a measure of the community, we choose crime data provided by the New York City Police Department. Our assumption is a community with high crime might negatively affect the happiness of the households in that area. This data is provided at the precinct level. A precinct is an area of a town as defined for police purposes. Every precinct in New York City is unique and may not be the same areas designated as sub-boroughs in the NYCHVS data. The crime data provided lists the total number of seven major felony offenses per precinct per year from 2000 to 2018. Since the NYCHVS data is collected every three years, we only selected years that overlapped with the NYCHVS data. **Figure 4** is a map of Manhattan illustrating total number of seven major felony offenses by precinct.

2.0.5 Education: High School Achievement Rates

For a measure of the education, we choose high school graduation data provided by the New York City Department of Education. Our assumption is a community with many high school graduates might positively affect the happiness of the households in that area. This data is provided at the school district level. Every school district in New York City is unique and may not be the same areas designated as sub-boroughs in the NYCHVS data. Also, school districts may not align with the precincts either. The education data provided lists each school within a school district, their cohort, and the number of graduates in each cohort. A cohort is all students who first entered ninth grade in a given school year (the cohort of 2006 entered ninth grade in the 2006-2007 school year). Graduates are defined as those students earning either a Local or Regents diploma and exclude those earning either a special education (IEP) diploma or GED.

There is also a metric in the dataset that lists the number of graduates achieving Aspirational Performance Measure (APM). The New York State Education Department (NYSED) defined English/Math Aspirational Performance Measure (APM) as the percentage of students that after their fourth year in high school have met NYSED standards:

- Graduated by August with a Regents or Local diploma, AND
- Earned a 75 or higher on the English Regents, AND
- Earned an 80 or higher on one Math Regents

Instead of simply using the number of high school graduates as the metric to gauge the happiness of a school district, we utilize the number of high school graduates who achieved APM. This metric measures the number of graduates that are more likely to go on to college, obtain a job, and contribute to society.

The education data is only available for cohorts that graduated from 2010 to 2015. We extrapolate the data for 2017 using the 2015 data and aggregate the other years into 2011 and 2014. **Figure 5** is a map of Manhattan illustrating the rate of the number of achieved high school graduates and the number of total graduates in the cohort by school district.

2.0.6 Health: Mortality Age Range

For a measure of the health, we choose mortality data provided by the New York City Department of Health and Mental Hygiene. Our assumption is a community with a lower median death age range might negatively affect the happiness of the households in that area. This data is provided at the community district level. Every community district in New York City is unique and may not be the same areas designated as sub-boroughs in the NYCHVS data. Also, community districts may not align with the precincts or school districts either. The mortality data provided lists the total number of deaths in each age range per community district per year.

To calculate the mean death age, we calculate the average number of deaths. We start from the oldest age range (85 and older) and cumulatively sum up the totals in each category until we reach the mean death age. Whatever age range contains the mean death age becomes the median death age range.

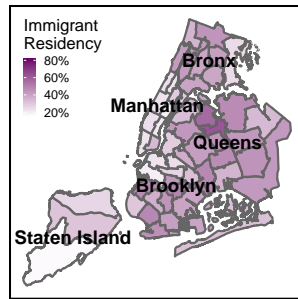


Figure 1

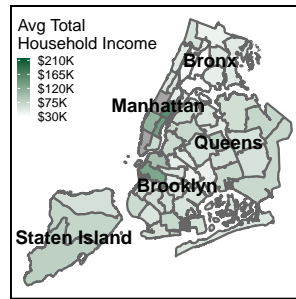


Figure 2

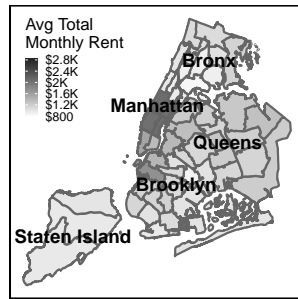


Figure 3

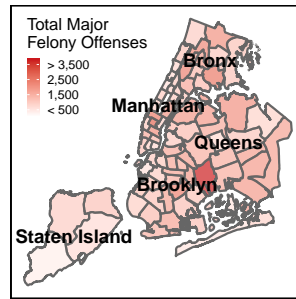


Figure 4

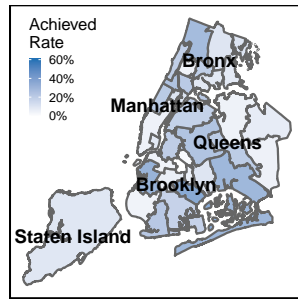


Figure 5

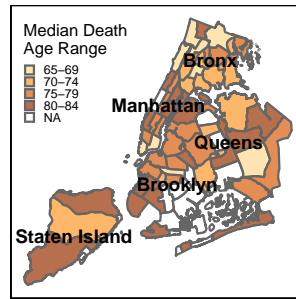


Figure 6

The data ranges from 1999 to 2014. Since the NYCHVS data is collected every three years, we only selected years that overlapped with the NYCHVS data. **Figure 6** is a map of Manhattan illustrating median death age range by community district.

2.1 Methods

2.1.1 Aggregating the Data

The shape file data is provided by the New York City Department of City Planning. This website provides the shape file data for the Census blocks which is what is used in the NYCHVS data, school districts for education and community districts for health. The precinct shape files for the crime data are provided by New York City OpenData.

Starting with the NYCHVS data, we use the `pumas` function in the `tigris` package in R to extract New York specific data. We use the geoid's provided by the NYCHVS data to extract the necessary geoid's from the Public Use Microdata Area (PUMA) data. The data is read in and projected to the Long Island State Plane coordinate reference system. From there we can create an outline of New York City and the small sub-boroughs.

In order to aggregate all of the different regions (Census blocks, police precincts,

Brooklyn Aggregate Data

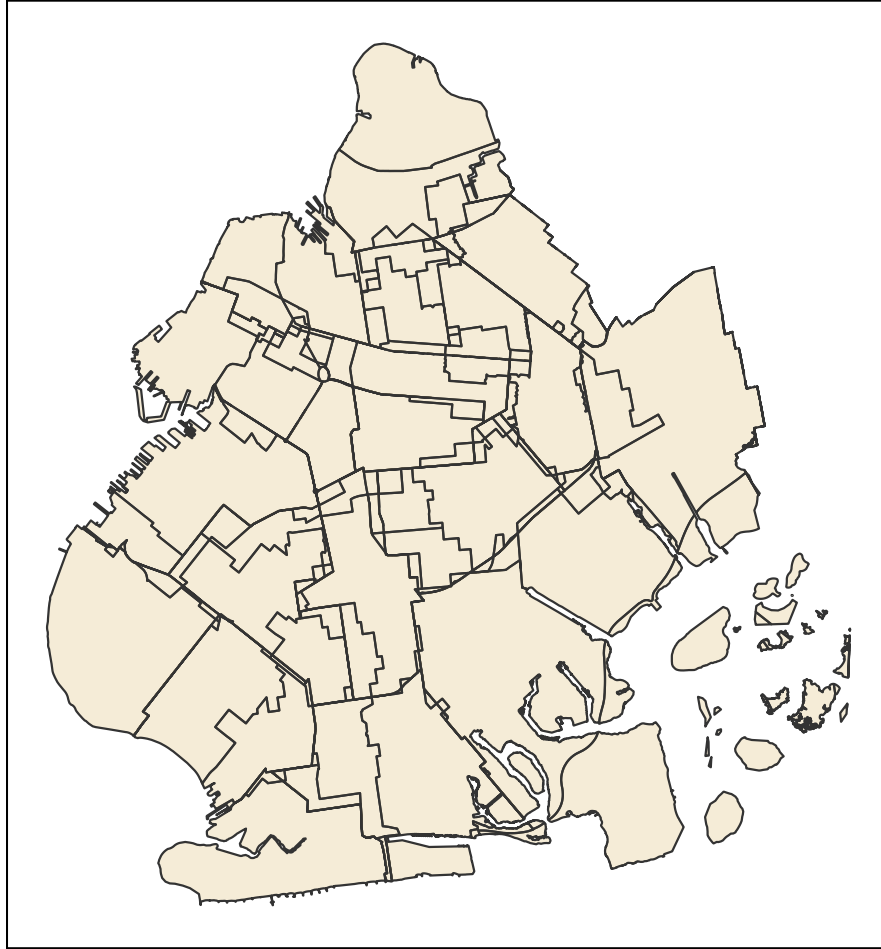


Figure 7

school districts, and community districts), we first remove all self intersections. To do this, we use the `gBuffer` function from the `rgeos` package in R to make the widths equal to zero. We use the `gIntersection` function from the same package to intersect the four different regions. At each intersection, we only keep the polygons we need. After the intersection is complete, we project the data back to the geographic coordinate system of longitude and latitudes.

After the intersections are complete, we return to the original four regions and extract the ID's for each dataset type by using the `sapply` function from base R. Every ID is unique for each polygon in the different regions (Census blocks, police precincts, etc.). We concatenate all four ID's to create a single ID for each unique polygon in the intersected data. We add this ID to the spatial polygon data frame. Using these ID's we can utilize the `SpatialPolygons` function from the `sp` package in R to left join all of the necessary data from the different regions to their respective polygons.

Figure 7 overlays all the different geographic regions that are necessary to calculate the happy score index. Each region has a unique combination of sub-borough, precinct, school district and community district.

Happiness Score Index

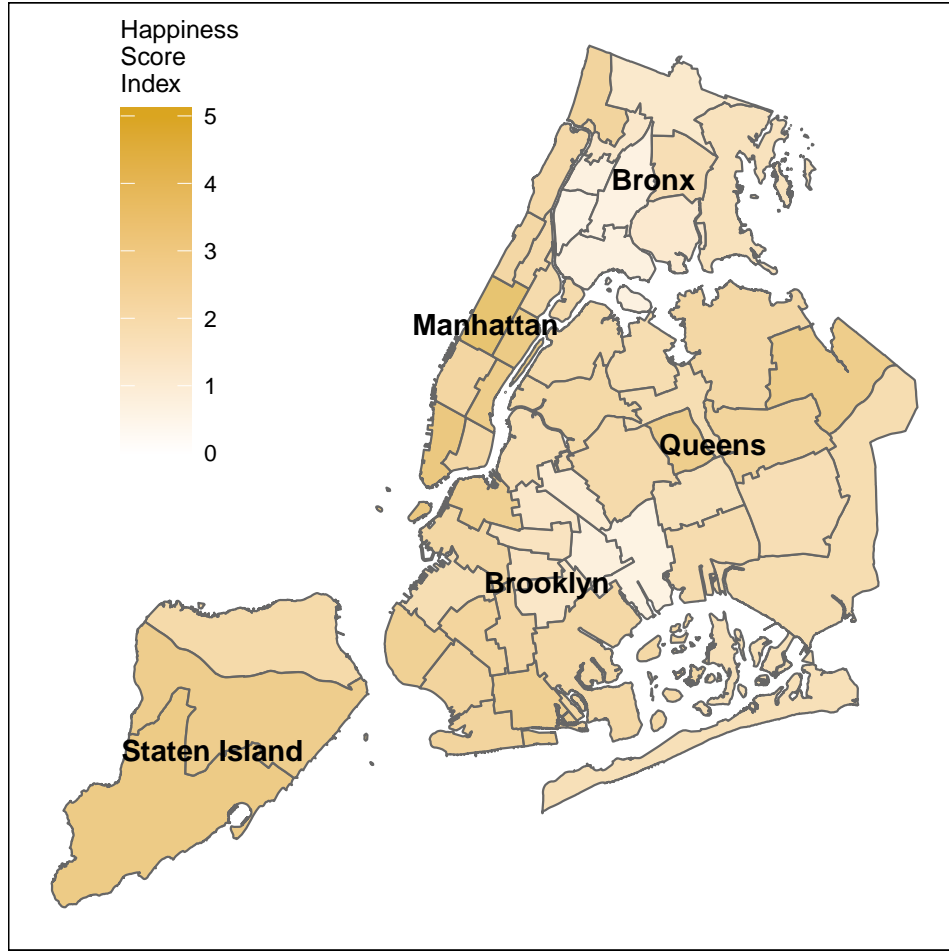


Figure 8

2.1.2 Calculating Happy Score Index

When calculating the happy score index, we first take all of the data from the intersected regions and left join them with the data from precincts, school districts and community districts. We extract the area of each unique intersecting region, group by the sub-borough and take the weighted average of the different measures of happiness. (Note: For the housing measure of rent, we take the weighted average of the proportion of average monthly contract rent to average total household income. For the health measure of median death age range, we take the weighted average of the lowest value in the range.). We scale each happiness measure from 0 (Worst) to 1 (Best). We, then, take a sum of all five scaled measures to create the Happy Score Index.

3. Results

Figure 8 shows the happy score index and the immigrant residency for each sub-borough.

Figure 9 is another visualization to better understand the association between immigrant residency and happy scores.

From the plot above, we see that Manhattan and Staten Island tend to have lower

Immigrant Residency and Happy Score Index

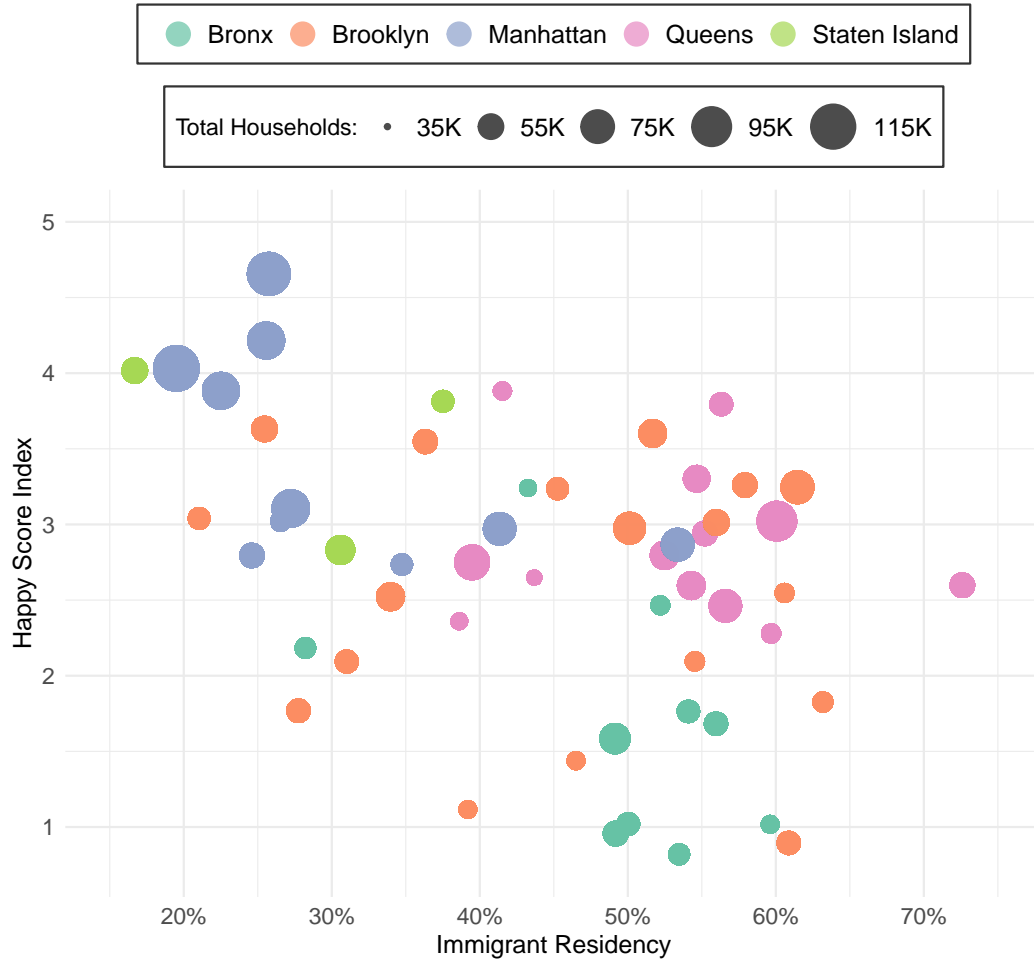


Figure 9

percentages of immigrant households and higher happy score indexes. The Bronx tends to have the reverse effect with higher percentages of immigrant households and lower happy score indexes. Then, there are also those boroughs, Queens and Brooklyn, that are relatively average across.

4. Discussion

In this analysis, we utilized the NYCHVS data and other external resources to create a happiness metric to evaluate the quality of life for immigrant households in New York City. We took advantage of the happiness score index created by Happy City and the New Economics Foundation to create a metric for each sub-borough in New York City. From the happy score index map visualization, it is clear that most of Manhattan and Staten Island tend to be happier than some areas of Brooklyn/Queens and Bronx. It would be interesting to see if over time, these happy scores change within a given sub-borough. When analyzing the scatter plot visualization with all of the sub-boroughs, it would be interesting to figure out exactly which sub-boroughs were different compared to the borough averages. Overall, we know that happiness, in our calculations, is heavily skewed in favor of wealth. However, we know that being wealthier may not equate to being happier,

but more wealth can alleviate other real life problems.

REFERENCES

- "Happy City Index." What Works Centre for Wellbeing, 27 Oct. 2014, whatworkswellbeing.org/case-study/happy-city-index/.
- Lewis, Sam Wren, and Saamah Abdallah. Happy City Index 2016 Report. Happy City Measurement and Policy Team, 2016, Happy City Index 2016 Report, www.happycity.org.uk/wp-content/uploads/2016/10/Happy-City-Index-2016-Report-FINAL.pdf.