Immigrant Residency and Happiness in New York City

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

Abstract
We explore the quality of life of immigrants in New York City through housing and neighborhood conditions by creating a happiness metric to measure a households' quality of life. 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, the New York City Department of Health and Mental Hygiene, along with reports from Happy City and the New Economics Foundation (HCNEF) [1], a happiness score was assigned to each sub-borough in New York City. This happiness score evaluated five main domains: work, place, community, education, and health. As a result of this analysis, we discovered higher happiness scores were associated with lower percentages of immigrant households.

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

1. Introduction

Many people around the globe venture to the United States seeking the American Dream. Some immigrants travel with hopes of a better quality of life. In this analysis, we use the data provided by the New York City Housing and Vacancy Survey (NYCHVS) to create a happiness metric to measure quality of life. According to the New York City Department of Housing Preservation and Development (HPD) [5]:

> 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 help guide individuals to understand more about immigrant households and how being in an immigrant household relates to their quality of life.

To measure quality of life for New York City neighborhoods, we utilize a happiness metric. Happiness, according to HCNEF [1], 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" [3]. Using the data from the housing surveys about immigrant residency, we attempt to find a connection between happiness and immigrant residency.

*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. Each borough is further partitioned for local governance into roughly equivalent sized regions called sub-boroughs. 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 described by the HCNEF: work, place, community, education, and health. For work and place, we utilize the NYCHVS data. Community, education and health require data from external resources: the New York City Police Department, the New York City Department of Education, and the New York City Department of Health and Mental Hygiene, respectively.

2.1 Immigrant Residency: Place of Birth

In the NYCHVS data, there is a field that identifies the place of the householder's birth, and thus we are able to determine immigrant head of 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.

2.2 Work: Income

In the NYCHVS data, there is a field that identifies total household income. We utilize this as a measure of *work*. We use the total household income because higher household income can be associated with well paying jobs, which can be a sign of a thriving household. Using this field, we adjust for inflation and calculate the average total household income per sub-borough. **Figure 2** is a map of New York City illustrating the average total household income by sub-borough.

2.3 Place: Rent

In the NYCHVS data, there is a field that identifies monthly contract rent. Monthly contract rent is the total rent scheduled to be paid to the landlord [4]. We utilize this as a 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 New York City illustrating the average monthly contract rent by sub-borough.

2.4 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 the city as defined for police purposes. Every precinct region in New York City is unique and the boundaries do not match the sub-borough regions. 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 New York City illustrating the total number of seven major felony offenses by precinct.

2.5 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 region in New York City is defined by boundaries that do not match either the sub-borough or precinct regions. 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 utlize the number of high school graduates who achieved APM. This metric measures the quality of the education.

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 New York City illustrating the rate of the number of achieved high school graduates and the number of total graduates in the cohort by school district.

2.6 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 death age will negatively affect the happiness of the households in that area. This data is provided at the community district level. Every community district region is defined by boundaries that do not match the sub-borough, precinct or school district regions. 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 5 year age range per community district per year.

First, we calculate the average number of deaths by dividing the total number of deaths by two. For example, if there were 15 deaths in the 85 and older range, 20 deaths in the 80-84 age range, 5 deaths in the 75-79 age range, and no deaths in the remaining age ranges, there would be 40 total deaths. When we divide 40 in two, we see that our average number of deaths is 20.

Next, we start from the oldest age range (85 and older) and cumulatively sum up the total number of deaths. If we start from the oldest age range and cumulatively sum the total number of deaths, we would get: 15 cumulative deaths in 85 and older

range, 35 cumulative deaths in the 80-84 range, and 40 cumulative deaths in all of the subsequent ranges.

From here, we select the oldest age range that contains our average number of deaths. Going back to our example, we are looking for the age range that has more than or equal to 20 cumulative deaths. This means the 85 and older range is ruled out because it only has 15 cumulative deaths. The 80-84 range has 35 cumulative deaths and it is the oldest age range with more than or equal to 20 cumulative deaths. Even though the subsequent age ranges contain the 20 cumulative deaths, they are not the oldest age range. We repeat this process for each community district for each year.

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 New York City illustrating the median death age range by community district.

3. Methods

3.1 Aggregating the Data

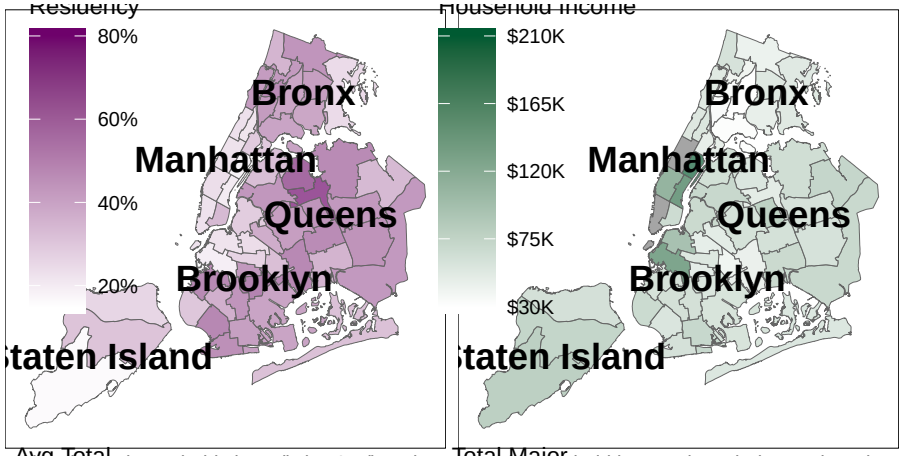
The shape file data is provided by the New York City Department of City Planning. This website provides the shape file data for the Census blocks or sub-borough regions which is what is used in the NYCHVS data, school district regions for education and community district regions for health. The precinct 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 [2]. 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 unique sub-boroughs. For the purposes of explaining this process in greater detail, we focus on Brooklyn. **Figure 7a** shows Brooklyn with its sub-boroughs outlined in red. We refer to these unique sub-borough regions as R_i, where i represents the region for the i^{th} sub-borough.

In order to aggregate all of the different regions (sub-boroughs, police precincts, 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 [7]. We use the `gIntersection` function from the same package to first intersect the sub-borough regions and the precinct regions. **Figure 7b** shows Brooklyn with its sub-boroughs outlined in red and precinct regions outlined in blue. We refer to these unique sub-borough and precinct regions as R_{ij}, where i represents the region for the i^{th} sub-borough and j represents the region for the j^{th} precinct.

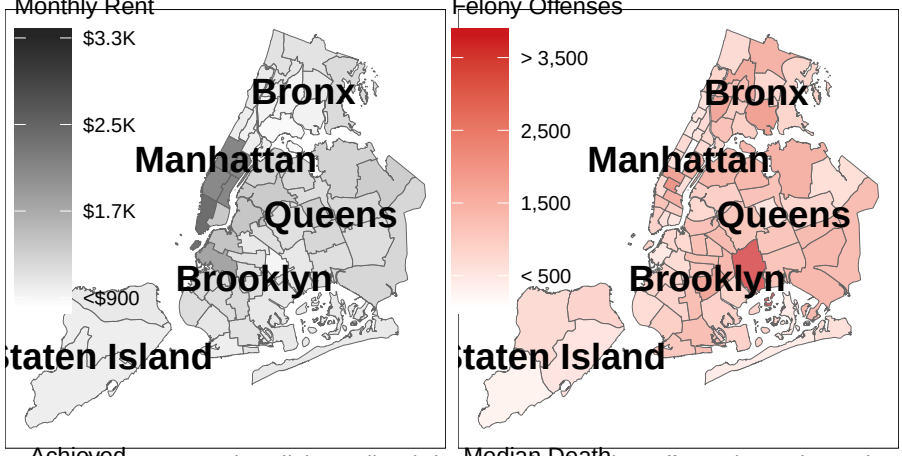
Building off this combination, we follow the same process to intersect the sub-borough and precinct regions and the school district regions. **Figure 7c** shows Brooklyn with its sub-boroughs outlined in red, precinct regions outlined in blue, and school district regions outlined in orange. We refer to these unique sub-borough, precinct and school district regions as R_{ijk}, where i represents the region for the i^{th} sub-borough, j represents the region for the j^{th} precinct, and k represents the region for the k^{th} school district.

Continuing the previous combination, we follow the same process to intersect the sub-borough, precinct, and school district regions and the community district



rcentage of immigrant households by sub−borough region.

Figure 2: The average total household income by sub−borough region.



verage total monthly contract rent by sub−borough region.

Figure 4: The total number of major felony offenses by precinct region.

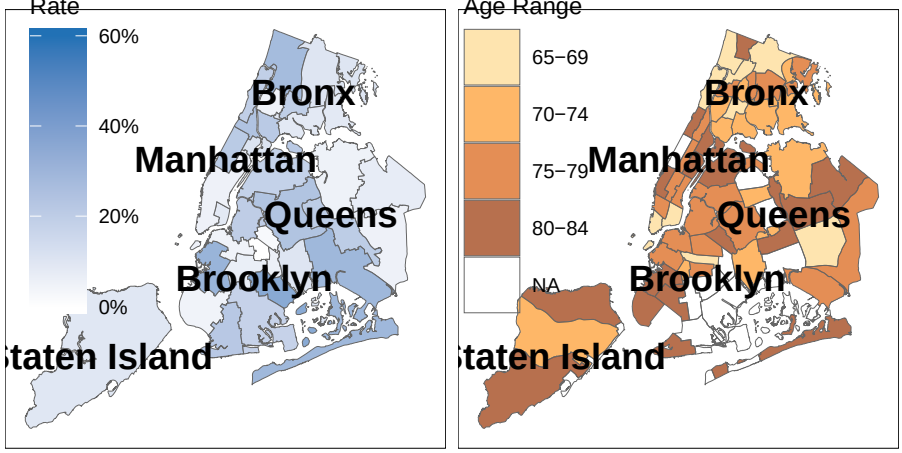


Figure 5: The rate of achievement of high school graduates by school district region.

Figure 6: The median death age range by community district region.

regions. **Figure 7d** shows Brooklyn with its sub-boroughs outlined in red, precinct regions outlined in blue, school district regions outlined in orange, and community districts outlined in purple. We refer to these unique sub-borough, precinct, school district and community district regions as R_{ijkl}, where i represents the region for the i^{th} sub-borough, j represents the region for the j^{th} precinct, k represents the region for the k^{th} school district, and l represents the region for the l^{th} community district.

After the intersections are complete, we project the data back to the geographic coordinate system of longitudes and latitudes. 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 (sub-boroughs, police precincts, etc.). We concatenate all four ID's to create a single ID (R_{ijkl}) 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 [6,8].

Figure 7e zooms in on a specific sub-borough in Brooklyn where we can notice all the different geographic regions that are necessary to calculate the happiness score index. Each region has a unique combination of sub-borough, precinct, school district and community district regions, R_{ijkl}.

3.2 Calculating Happiness Score Index

When calculating the happiness 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 Happiness Score Index.

4. Results

Figure 8 shows the happiness score index for each sub-borough in New York City. **Figure 9** is another visualization to better understand the association between immigrant residency and happiness scores. From **Figure 9**, we see that Manhattan and Staten Island tend to have lower percentages of immigrant households and higher happiness score indexes. The Bronx tends to have the reverse effect with higher percentages of immigrant households and lower happiness score indexes. Then, there are also those boroughs, Queens and Brooklyn, that are relatively average across.

5. 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 HCNEF to

Figure 7a: Region R_i of sub−borough regions in red.

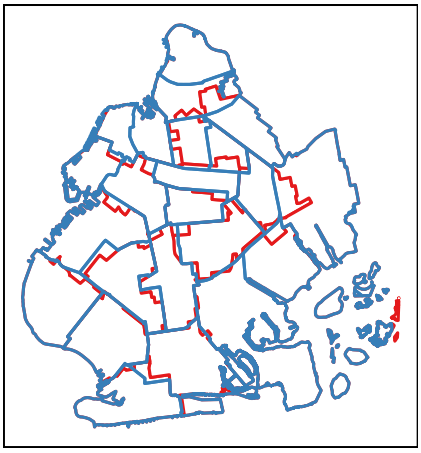


Figure 7b: Region R_{ij} of sub−borough and precinct regions in red and blue, respectively.

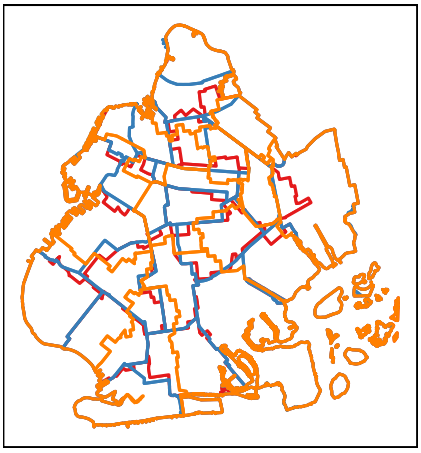


Figure 7c: Region R_{ijk} of sub−borough, precinct and school, district regions in red, blue and orange, respectively.

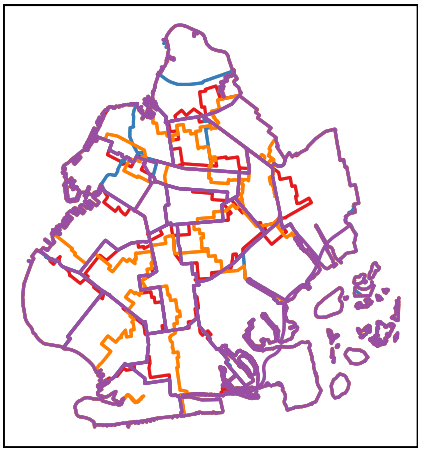


Figure 7d: Region R_{ijkl} of sub−borough, precinct, school district, and, community district regions in red, blue, orange, and purple, respectively.

create a metric for each sub-borough in New York City. From the happiness 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 the Bronx. It would be interesting to see if over time, these happiness scores change within a given sub-borough. When analyzing the scatter plot visualization with all of the sub-boroughs, it would be interesting to delve deeper into the original countries of some of these immigrant households. Also, if length of residency in the United States is associated with higher happiness scores. 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

1) "Happy City Index." What Works Centre for Wellbeing, 27 Oct. 2014, whatworkswellbeing.org/case-study/happy-city-index/.
2) Kyle Walker (2018). tigris: Load Census TIGER/Line Shapefiles. R package version 0.7. https://CRAN.R-project.org/package=tigris
3) 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.

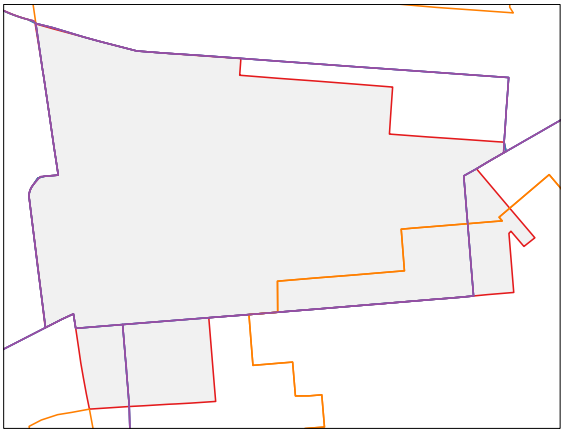


Figure 7e: Region $R_{1,01}$ of sub–borough region 1 in red, precinct region in blue, school district region in orange, and community district region in purple.

4) "NYCHVS Glossary." US Census Bureau, New York City Housing and Vacancy Survey, 2008, www2.census.gov/programs-surveys/nychvs/about/glossary/gloss08.pdf.
5) "NYCHVS in the ASA Data Challenge Expo." NYCHVS in the ASA Data Challenge Expo, New York City Housing and Vacancy Survey, 2019, www1.nyc.gov/site/hpd/about/nychvs-asa-data-challenge-expo.page.
6) Pebesma, E.J., R.S. Bivand, 2005. Classes and methods for spatial data in R. R News 5 (2), https://cran.r-project.org/doc/Rnews/.
7) Roger Bivand and Colin Rundel (2018). rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.4-2. https://CRAN.R-project.org/package=rgeos
8) Roger S. Bivand, Edzer Pebesma, Virgilio Gomez-Rubio, 2013. Applied spatial data analysis with R, Second edition. Springer, NY. http://www.asdar-book.org/

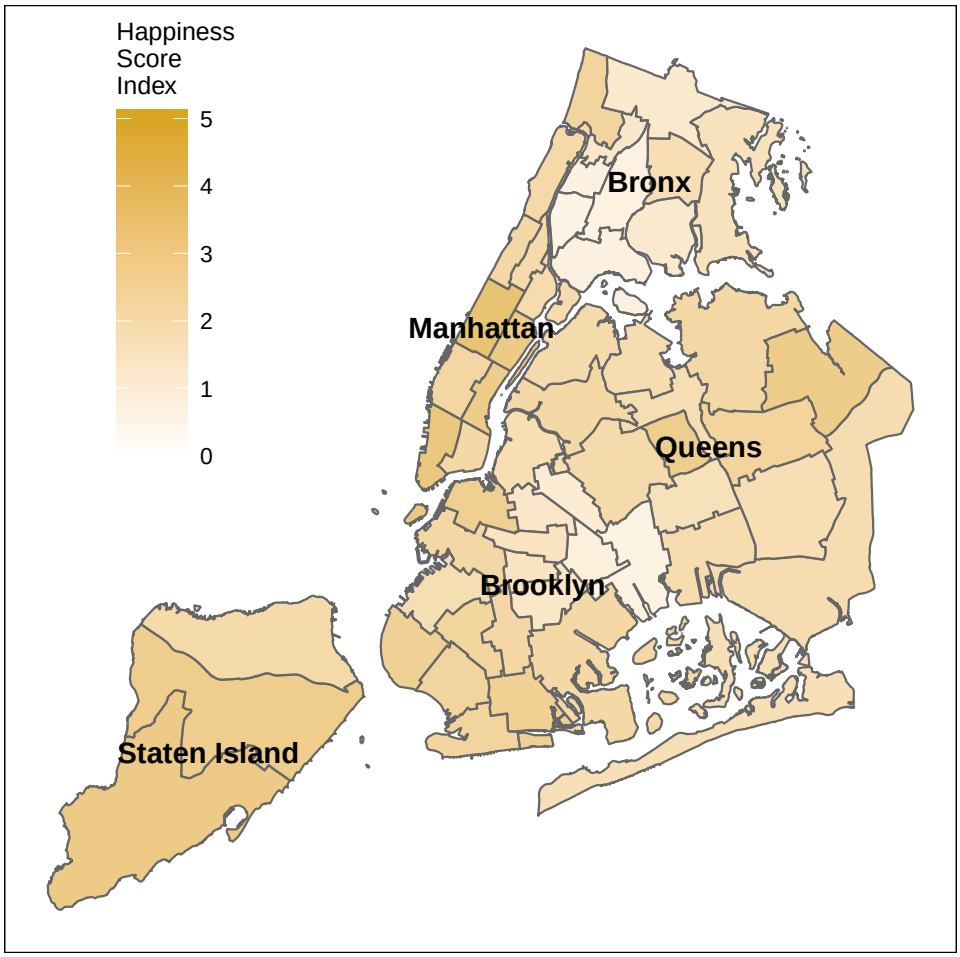

Happiness Score Index
Happiness Score Index
5
4
3
2
1
0
Bronx
Manhattan
Queens
Brooklyn
Staten Island

Figure 8

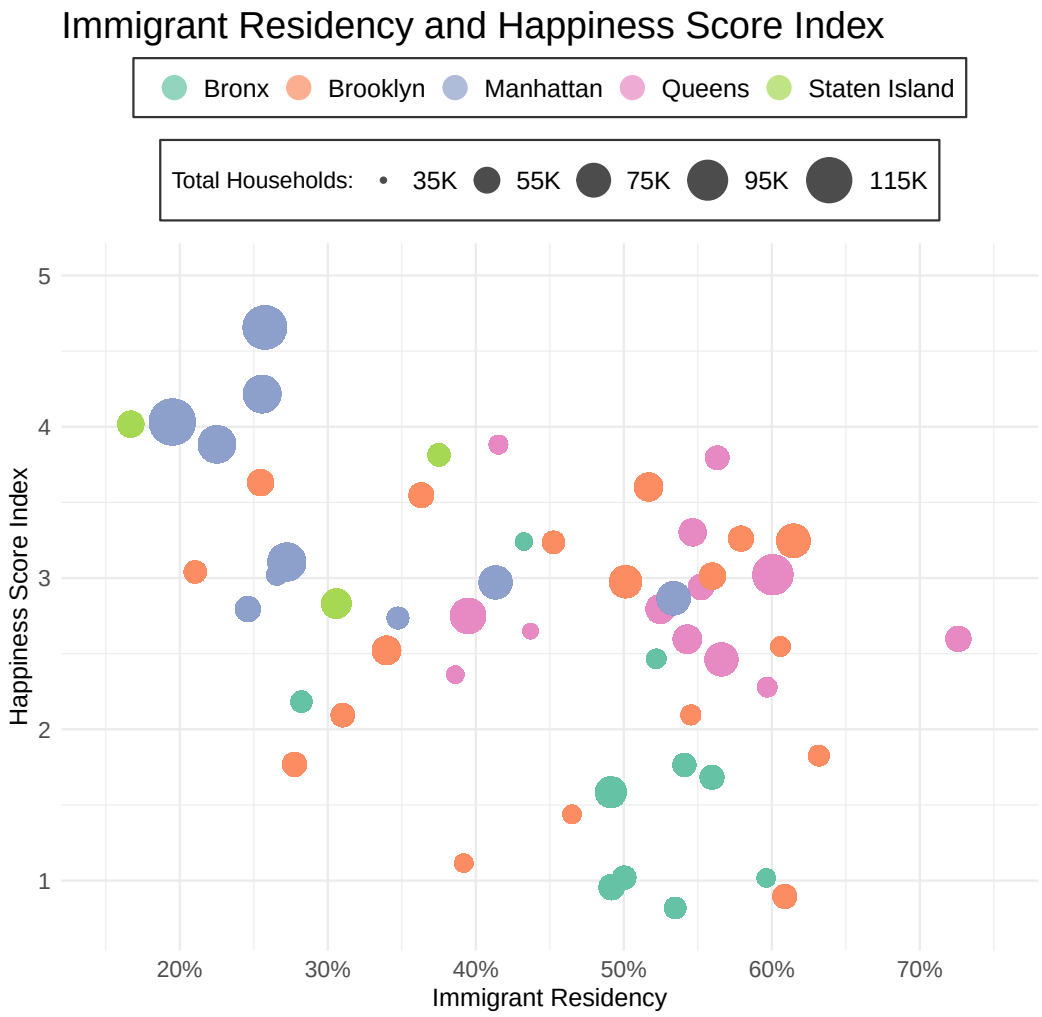

Immigrant Residency and Happiness Score Index
Bronx
Brooklyn
Manhattan
Queens
Staten Island
Total Households:
35K
55K
75K
95K
115K
Happiness Score Index
5
4
3
2
1
20%
30%
40%
50%
60%
70%
Immigrant Residency

Figure 9