How well does public transit serve the diversity of neighborhoods and people in New York City?

Jimmy Feng and Shih-Lung Shaw

2019-05-06

Abstract

The diversity of New York City is found in both its neighborhoods as well as its modes of transportation. In its densely-populated urban landscape, access and opportunity to different modes of transport are important in facilitating the city's functions as well as supporting human activities and interactions. This research analyzes the flows and trajectories of the city's public transit agency Metropolitan Transit Authority (MTA)'s Bus, Subway, and MetroNorth ridership. Neighborhood socioeconomic and demographic indicators from the American Community Survey, U.S. Census, and NYC agencies are integrated to provide a holistic picture, framed in the context of social equity, between the different types of public transit services and various neighborhoods and their respective populations in the city. Service-area network analyses reveal a disparate landscape in access to public transit stations as well as the provision of service, with residents in Manhattan enjoying the greatest service relative to those in the communities in the outer boroughs. However, linear regression finds that the relationship between minority populations and access to jobs and schools via public transit is indeed statistically significant but weak at best. This exemplifies the classical geographic concepts of scale, spatial heterogeneity, and the modifiable areal unit problem. Understanding the relationships between transportation services and neighborhood characteristics can shed light on the heterogeneous and disparate landscape of the vitality of both the city and its diverse population. This is important for future planning and policy making to foster a more equitable landscape that encourages and supports the diverse social fabric of New York City's communities.

Introduction

This paper explores the social, spatial, and temporal relationship of public transit as both a facilitator and prohibitor of educational and job accessibility for the diversity of population groups in New York City. A literature review provides the motivation and context to the following questions:

- What is the physical landscape of accessibility to public transit in New York City?
- What are the implications of unequal access? Why is it important to have equal access to public transit services? How does this manifest in terms of social equity and inequity?
- Which neighborhoods and population groups are particularly disadvantaged to educational (secondary education) and job opportunities as a result of unequal access to public transit?

Followed are a brief summary of the study area and a detailed explanation on the data and methods used as well as the results from the different analyses. Particularly, network and statistical analyses using GIS and the statistical programming language R are applied to public transit and demographic data to provide various quantifiable measures that allow for comparison of access between the different populations in New York City. A conclusion and some final thoughts for future directions ends the paper.

Literature Review

In 1987, the United Nations World Commission on Environment and Development published the Brundtland Report, or "Our Common Future", and popularized the notions of sustainable development and sustainability. "Sustainable development is the kind of development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (United Nations World Commission on Environment and Development 1987, 43). Broken down, sustainability is comprised of three fundamental dimensions of environmental protection, economic growth, and social equity.

Thus, social equity is intrinsically tied to the social dimension, and also economic and environmental dimensions, of sustainability. Sustainable development is greatly determined by social context and values which manifest differently across space (Wilbanks 1994). An equitable society is one without exclusionary or discriminatory practices that hinder individuals from economic, social, and political participation in society (Dempsey et al. 2011, 292) While the overarching notion of sustainability was proclaimed at a global scale, the local scale is also of critical importance especially in light of the common and every day experience of the built environment. Social exclusion and inequities manifest geographically as deprived areas with poorer living conditions and diminished access to different amenities, services, and facilities (Dempsey et al. 2011, 292). A sustainable world that provides equal opportunity for all peoples requires a global network consisting of sustainable local communities. This is only achieved when current economic and social systems are radically altered and equalized; as it is, poverty, homelessness, hunger, and violence remain rampant in all parts of the world. The world's social fabric is being threatened by these issues and continue to be exacerbated.

In many cities across the United States there exists an unequal landscape of opportunity as a result of extrinsic social forces. Notably, in 1965, John F. Kain theorized that housing market discrimination marginalized and restricted the residential geographies of African-Americans, also known as the *spatial mismatch hypothesis*. This in turn deprived them of quality jobs, education, and services (Meyer et al. 1965) which has significant implications for current and future success and quality of life. Later studies reported similar findings, with agencies and public authorities alike stressing the need for greater public transportation as part of the solution to eradicate the social inequalities arising from spatial mismatch (Kain 1992). Such problems are also shared by other minority and immigrant groups, including Hispanics (Hellerstein et al. 2009) and Asians (Easley 2018). Together, these studies emphasize the importance of accessibility to services in fostering equitable urban environments, and the many issues that continue to plague immigrant and minority groups. Accessibility to a wide range of both essential and non-essential amenities and services including jobs and education is a major characteristic of a socially equitable landscape and is largely facilitated and hindered by mobility and transportation. Understanding the landscape of social equitability via access to public transit is one way to understand the larger scope of social justice issues and ascertain whether a community, locale, or place is truly sustainable.

There is an abundance of literature on accessibility and equity in transportation geography. In their report *Evaluating Transportation Equity*, the Victoria Transport Policy Institute defined equity as the relative fair and appropriate distribution of costs and benefits and synonymous with *justice* and *fairness*. And in recognition of transportation planning decisions on equity, they suggested types of transportation equity: horizontal (equal distribution of impacts for people equal in ability and need) and vertical (distribution of impacts for people with differences in ability and need - i.e. by income or social class; and by mobility need and ability) (Litman 2002). This paper is concerned with vertical equity as it pertains to income and social class. Other research on related topics range from research on transit modelling improvements in GIS (JÃ¤rv et al. 2018; Manout et al. 2018), case studies on identification of facilitators and barriers to different modes of public transit (Dillahunt and Veinot, 2018), and methodological advancements to evaluate transport and transit investments and policies (Bocarejo and Oviedo 2011). Many in the past have devoted much effort to developing different ways of measuring and understanding flows and interactions in transportation. Notably, Hansen's gravity model to define accessibility as the potential for opportunities for interaction in 1959 (Hansen 1959) inspired a wide range of accessibility modelling in transportation. Similarly, Wilson established the framework for the family of spatial interaction models in locational analysis to understand spatial flows of people, goods, and/or information between areas (Wilson 1967), and from this emerged many other forms of spatial interaction modeling that are applied in a wide range of location-aware problems: from distribution, retail forecasting, discrete choice, to location-allocation. Of note include Evans' algorithm with the gravity model and linear programming for solving trip distribution and assignment problems (Evans 1973) and Fotheringham and O'Kelly's spatial choice modelling (Fotheringham and O'Kelly 1989), among many others. And the past few years have witnessed an influx in the use of information, communications, and technology-rich data such as cell-phone tracking, geotagged Flickr photos and Twitter tweets, and etcetera to better understand these varying concepts of accessibility and mobility. However, while methodology and using the appropriate methods and techniques are very important aspects, it is of even more importance in this research to understand how different groups of people are affected by transportation and their varying experiences resulting from this interplay between human and transport.

New York City presents as a unique place to study especially given its status as a global city and network, not only in its daily functions, but also in its populations. It is the most diverse area in the world, with people of different race, ethnicity, culture, age, ability, education, gender, socioeconomic status, religion, and etcetera all coexisting together. This dynamic interplay of unique characteristics makes it challenging to appropriately identify relationships. But it is of particular interest in this research to understand how racial minority groups are affected by public transit.

While literature in the broad field of accessibility and public transit is rich, there remains a gap in understanding how the New York City public transit system both facilitates and prohibits opportunity for the diversity of its population. There is a relative abundance of literature and research on the taxi and rideshare system in NYC, but less so on public transit. Qian and Ukkusuri used geographically weighted regression to understand the spatial variation of taxi ridership in the city and found higher educational achievement and subway accessibility to positively correlate with taxi ridership (Qian and Ukkusuri 2015), while Hochmair examined the spatiotemoral patterns of taxi ridership (Hochmair 2016). More related, Jin et al. analyzed the spatiotemporal relationship between the rideshare service Uber and public transit to measure the former's impact on urban transportation equity and found that income did not have much correlation with access and that Uber did not significantly improve transport equity (Jin et al. 2019). A report by Nate Silver and Reuben Fischer-Baum of FiveThirtyEight dove into a similar analysis by Qian and Ukkusuri, and found that where public transit infrastructure is relatively lacking correlates with lower availability of rideshare services (Silver and Fishcher-Baum 2015). An abundance of data, massive improvements in computational technologies, and significant advances in research have inspired an impressive amount of literature but there still remains a gap in understanding the dynamic spatiotemporal patterns of transportation (Shaw 2010, 137) let alone public transit as well as the relationship between the differences in service provision with the diversity of population groups in New York City.

Study Area

The area of focus in this research is New York City, divided into 55 unique communities that adhere to the public use microdata area (PUMA) classification as defined by the United States Census. These are areas that contain at least one-hundred thousand residents and are the equivalent of the conventional communities, such as Park Slope, Upper East Side, Midtown, SoHo, which are commonly understood by the layman. New York City is characterized by a significant metropolitan population with nearly nine-million residents. Its history as the gateway to America and a place conducive to the pursuit of the 'American Dream' continues to permeate modern culture and capture the allure of foreign-born peoples. The city's immigrant population constitutes a third of its overall population and this reflects the diversity of the city's functions and vibrancy. Immigrants and many other minority groups especially rely on the city's public transit system

Data and Methodology

The data retrieved for the various analyses applied in this research were gathered from different public government agencies and non-profit and educational institutions. These datasets and their respective origins are illustrated in the following table:

DATA	SOURCE
NYC Streets Centerline	NYC Department of Information Technology & Telecommunications
GTFS: MTA Subway	NYC Metropolitan Transit Authority
GTFS: MTA Bus	NYC Metropolitan Transit Authority
GTFS: MTA Metro North	NYC Metropolitan Transit Authority
2010 Public Use Microdata Areas	NYC Department of City Planning
2010 Census Tracts	NYC Department of City Planning
High Schools & Performance	NYC Deparment of Education

DATA	SOURCE
Demographic Indicators	US Census Bureau & American Community Survey
Rental Market Indicators	New York University's Furman Center
Origins/Destinations of Jobs	US Census Bureau - Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics

Preprocessing Data

To answer the different questions about job and educational accessibility posed earlier, various network analyses were applied in this research. All analyses were processed with the city's street centerline network as the network layer to model travel via public transit and also account for non-transit segments in commutes (i.e. walking from an origin to a public transit station.) General Transit Feed Specification (GTFS) data, the worldwide format for transit data were acquired from New York City's Metropolitan Transit Authority. Specifically, data was obtained for three modes of public transit offered by the agency: bus, subway, and Metro North. Despite part of its network located in the city and also being operated by the MTA, the Long Island Railroad was omitted from this research because it is considered to be a suburban rail system and not an intra-city transit option. Each mode contained various files providing different information about that particular network, including run-time schedules for all lines and routes, trips, and stop locations. For the sake of analysis, geometric centroids of each PUMA were computed with the 'Find Centroids' tool in ArcGIS, and also assessed with the 'st_centroid' function in the 'sf' package in R, to represent the start and end points in network analyses.

Determining Physical Landscape of Access to Public Transit

Assessment of the first question, what is the physical landscape of accessibility to public transit in New York City, was done with a service-area network analysis performed in Esri's ArcMap. In particular, service-areas or isochrones were defined at five and ten minute walking distances, the equivalent of one-fourth and one-half miles. Existing literature have often used these metrics as the baseline for distances people are comfortable walking to get to a public transit station. Together with a shapefile of all fifty-five PUMAs, the 'Tabulate Intersection' tool computed the percentage of total area within each PUMA that could reach at least one public transit station within ten minutes walking distance.

Determining Average Hourly Frequency of Public Transit Service

Many of the additional analyses relied on a series of tools for analyzing public transit systems developed by Melinda Morang of Esri, available at https://github.com/Esri/public-transit-tools. This required the conversion and pre-processing of the GTFS MTA data into a SQL database. Doing this allowed for querying and sub-selection of trips that met specified parameters (i.e. time of day, specific route, etcetera) with various Python scripts. To determine the frequency of public transit service across the five boroughs, the 'Count Trips in Polygon Buffers' tool was run for a given weekday and weekend, Wednesday and Saturday, based on high-resolution schedule GTFS data downloaded from MTA's data portal on February 27, 2019. Polygon service areas at a threshold of ten-minutes walking distance (one-half mile) were generated around all public transit stops, and the number of unique trips at each stop at every minute of every hour in each day was determined and summed. These total counts of trips at each stop were divided by twenty-four hours to provide a measure of the average hourly frequency of transit service.

Determining Job Accessibility

Determining job accessibility first required pre-processing of employment data at each census tract. This data was acquired from the United States' Census Bureau Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES). The version, state, type, and most recent year of the LODES data used in this research are, respectively, LODES7, New York, Workplace Area Characteristics (WAC), and 2015. Because the data provided by LODES contains information for the entirety of New York State and at the scale of census tracts, pre-processing was necessary to (1) extract job statistics only in New

York City, (2) aggregate them for analysis at the scale of public use microdata areas, and (3) transfer these data as a polygon shapefile for further analysis in a GIS. This was done with the R programming language and the 'readr', 'dplyr', and 'sf' packages.

Afterwards, the 'Calculate Accessibility Matrix' tool was run in ArcGIS to determine the number of reachable jobs by each PUMA. Broadly summarized, this tool solves an origin-destination cost matrix weighted by the jobs at each destination while also accounting for public transit trips from the provided GTFS schedule data. A one-hour travel time threshold was specified; and this was determined to be the cutoff because the average commute time for residents in all of New York City is at least 40 minutes (New York City Department of City Planning 2019) and it is an easily understood measurement of time. Essentially, all PUMAs that are reachable within one-hour travel time via public transit are identified for each PUMA origin; and this includes the PUMA itself (it would reason that a resident should be and is able to reach the jobs that are in their PUMA.) The number of jobs at each reachable PUMA destination are then aggregated to provide the total reachable number of jobs for each PUMA. This analysis was performed at each individual hour; and within each hour, every minute. Analysis at this precision of time enabled for the calculation of reachable jobs at various percentages of start times (i.e. how many jobs are reachable within 30, 50, and 90% of all possible trips between 8 and 9 AM?). While it is relatively more computationally and time intensive, this is necessary to capture a more accurate landscape of access to jobs as service varies temporally; the level of service is not the same at every minute and a bus that runs every other half an hour means lower frequency of service than a subway that comes every five minutes. Similarly, a person who arrives to a station thirty seconds after a bus has left may have a longer total commute time than another person who was able to catch the bus and make their appropriate time-sensitive transfers to other bus and/or subway lines.

In total, this tool was iterated forty-eight times (every hour for a weekday and a weekend) with ArcGIS' ModelBuilder. These data were then integrated into R for further regression analyses to discover potential correlations between job access and various socioeconomic and demographic characteristics for each PUMA. The 'tidycensus' and 'tidyr' packages were used to retrieve and process the United States Census 5-Year American Community Survey 2016 indicator variables.

Determining Educational Accessibility

A similar workflow to determine job accessibility was used in assessing educational accessibility. There were a few changes: 'quality' high schools replaced the destinations in origin-destination cost matrix analyses and the number of quality schools replaced the number of jobs as the weighted factor in the calculations. In addition, the average start time of schools was computed to be 08:31 with the 'lubridate' package in R and, because the travel threshold was one-hour, the analysis was only run at a start-time of 07:30. Further, this research is not interested in access to schools during non-opening hours and only when students can arrive to school at a reasonable time without being late.

In total, there are a total of 435 public high schools. Private schools were intentionally disregarded because they mostly cater to students of higher socioeconomic class and are not equally accessible due to their cost. However, it would be remiss to only analyze the number of high schools that are reachable within a one-hour window. The New York City Department of Education provides statistics for various metrics of student population and performance, with the most recent data of average SAT scores available for the school year of 2015 and graduation rate and racial composition available for the school year of 2018. These variables were processed together in R for the development of a quality index. Ultimately, the three aforementioned variables were integrated in the index: average SAT, graduation rate, and racial diversity; and the formula for this is

$$Quality_{HS} = RacialDiversity_{HS} + AverageSAT_{HS} + GraduationRate_{HS}$$

$$RacialDiversity_{HS}, AverageSAT_{HS}, GraduationRate_{HS} \in [0, 1]$$

. For this research, racial diversity is defined as the probability of randomly identifying students of different race, and this was calculated with the formula:

$$RacialDiversity = 1 - (P_{White}^2 + P_{Black}^2 + P_{Hispanic}^2 + P_{Asian}^2)$$

All three variables were scaled to values between 0 to 1 and then added to provide a score reflecting the quality of the school, with a score of 3 indicating highest quality and 0 indicating lowest quality. These scores were then grouped by percentile. Of interest in this research was accessibility to high schools with a quality score above the 50th percentile or the upper-half of all schools in the cleaned dataset, and only these high schools were input as the destination locations in processing the accessibility matrix in ArcGIS.

Other potential variables such as college enrollment rate were eliminated due to issues of multicollinearity (i.e. college enrollment rate was determined to be highly correlated rate with graduation rate with $R = 0.848$) or because many of the other variables (i.e. various quality review metrics) were assessed in different years for different schools and also lacked much data. On the topic of missing values, with the 'MICE' package, it was revealed that 80 of the 435 high schools were missing data in either one or more variables of use in the quality index and it was intentionally decided not to impute them. One reason is because many of these schools were only recently established and thus had no students take SATs nor graduate yet and another reason is based on the author's personal understanding, having been born and raised in New York City, that these schools are also of lower quality; the results of the analysis would not be influenced with imputed values and given the nature of this topic, it did not seem appropriate to impute values. These schools would not be considered in the origin-destination cost matrix analysis given that the research is concerned with accessibility to 'high' quality schools, defined as the upper-half of all public schools in New York City.

Miscellanious

Principal Components Analysis (PCA) was used as an exploratory method to understand the abundance of potential explanatory variables for a particular hour of job access at the census tract level. The first principal component explained 24.83% of the proportion of variance and subsequent principal components explained less of the variation in the sample dataset. It was decided not to discuss these results in depth because of the difficulty in interpreting transformed and scaled data and because the first principal component correlated to less than a fourth of the variation in the dataset and it was felt that this was too low of a value; even considering the first, second, and third principal components cumulated in less than half of all variation.

Mixed effects models were also used for analyzing job and educational access and their relations to the minority populations and other socioeconomic, housing, and demographic variables at the PUMA scale but this resulted in no variable being statistically significant which raised considerable doubt to the validity of the model. For this reason, it was also decided not to pursue this method further.

Results & Discussion

Areas Accessible to Public Transit

Service-area analyses with the pedestrian network of NYC determined the physical landscape of accessibility and identified all areas that were within five and ten minute walking distance of a station of public transit. Both thresholds are the equivalent of 1/4th and 1/2th miles.

The resulting map illustrates that most of the city is well within walking distance of at least one station of public transit. However, not all PUMAs are equally as connected to the public transit system. Nearly the entirety of the Manhattan borough enjoys relatively greater access and connectivity and this finding does not come as a surprise as many of the city's major financial institutions, government agencies, amenities, and other prominent organizations are centrally located within the area. And it is not just mere access; someone located in Manhattan can easily find them self within a five-minute walking distance to some form of public transit while many residents in the other four boroughs, especially those in the outskirts and more suburban parts of the city, have to travel farther to even reach a station. Furthermore, significant portions of residents throughout Staten Island, east-Queens, east-Brooklyn, and the north-Bronx require more than ten-minutes to reach even one station.

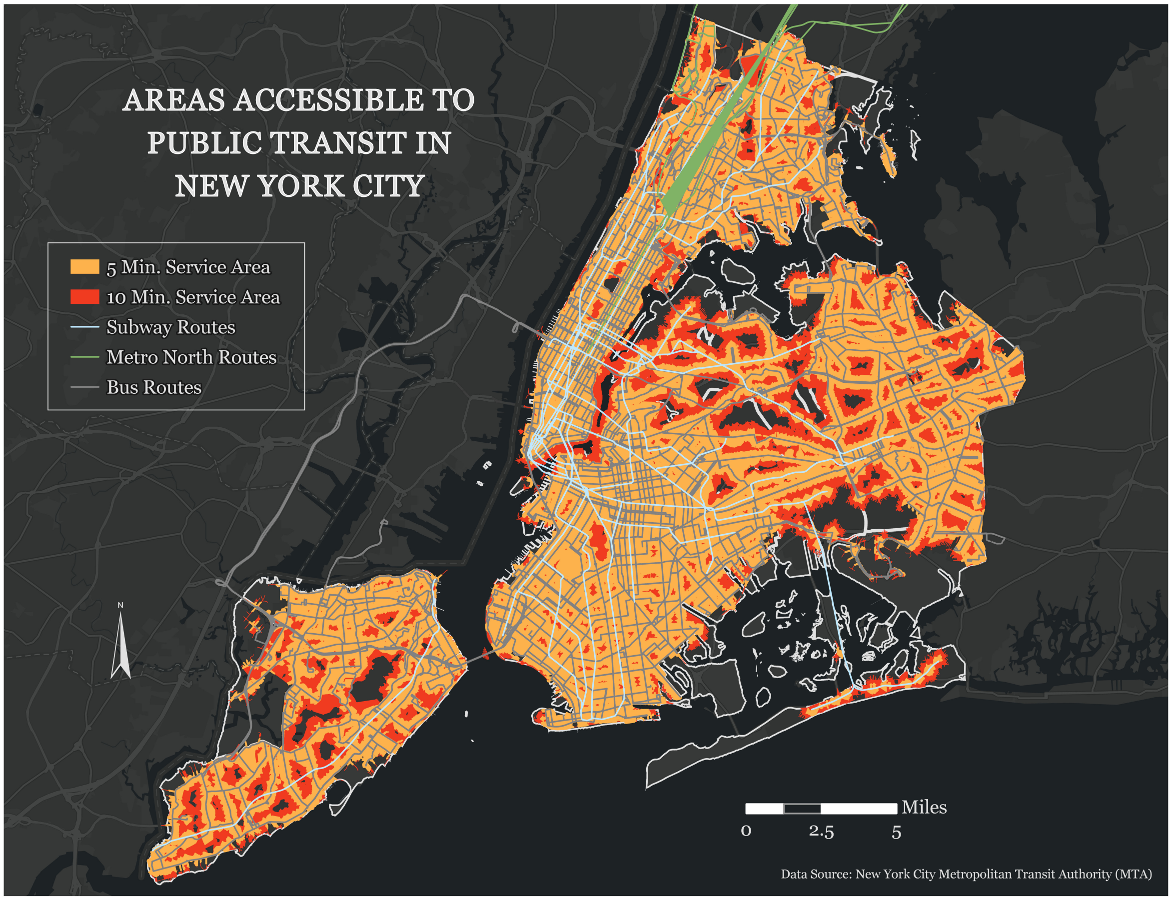
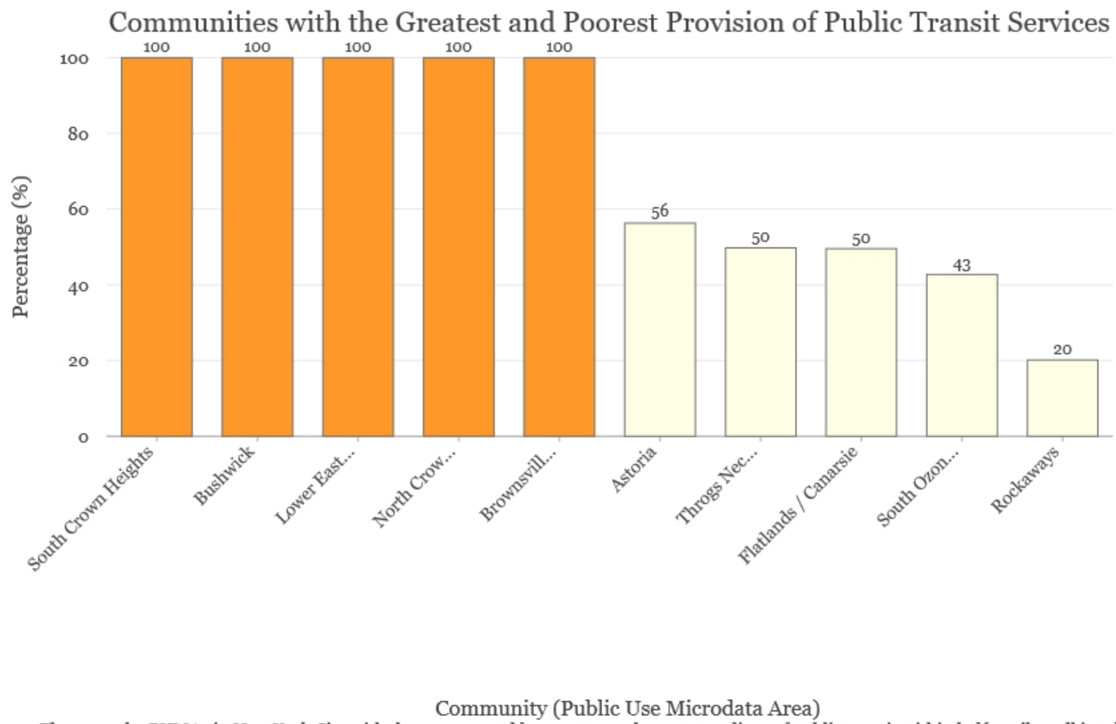


Figure 1: New York City Public Transit Stations, Routes, and Half-Mile Service Area



These are the PUMAs in New York City with the greatest and least coverage by some medium of public transit within half a mile walking distance.

Overlaying the shapes of the PUMAs on the cumulative service-areas and calculating the percentage of service area within each individual PUMA reveals that the communities with the greatest connection to public transit are located in Brooklyn (South Crown Heights, Bushwick, North Crown Heights, and Brownsville/Ocean Hill) and Manhattan (Lower East Side.) On the opposite end of the spectrum, those with the poorest connection are located in Queens (Astoria, South Ozone, and Rockaways), Brooklyn (Astoria and Flatlands/Canarsie), and the Bronx (Throgs Neck.) This also reflects intra-borough differences as there are many communities in Brooklyn that have great connection to public transit, but there are also many others that have significantly lower access. In several of the least-connected PUMAs, half or less of their areas are even reachable to a station. Only a fifth of the Rockaways can even access public transit within ten-minutes walking. This exploration of basic access and connectivity indicates that there is a disparity in public transit provision and infrastructure that manifests across New York City.

However, service area analysis only captures a part of the entire picture of service provision and as discussed later, job and school accessibility. It is important to note that not all modes of public transit are created equally as subway services are more frequent than bus services. Even more, the MetroNorth network, most prominent in the northern Bronx is a significantly more expensive mode of transportation than subway and bus, and is more often than not, used for commuting between the areas north of the city and midtown. Beyond comparing between the different services, service is also fundamentally tied to the temporal nature, and in reality, there are often disruptions, unscheduled maintenances, and more critically, the frequency of service varies spatially and temporally.

TIME MATTERS

These two maps illustrating average hourly frequency of service indicate that public transit runs more often during the weekday versus the weekend. While this finding is nothing new and makes logical sense as many industries and businesses only operate on the weekdays, it is important to discuss how reduction in service is not uniformly implemented across the entire city. In addition, while many people are in the fortunate situation to not have to work on the weekends, the single parent supporting their elderly parents and kids or the immigrant working multiple jobs to send remittances back home relies on public transit services equally as much on the weekend as the weekdays. There remains a large population that rely heavily on the public transit system across all days of the week and not just during conventional working periods. While analysis was also performed by the hour, regression analysis accounting for service differences with particular time

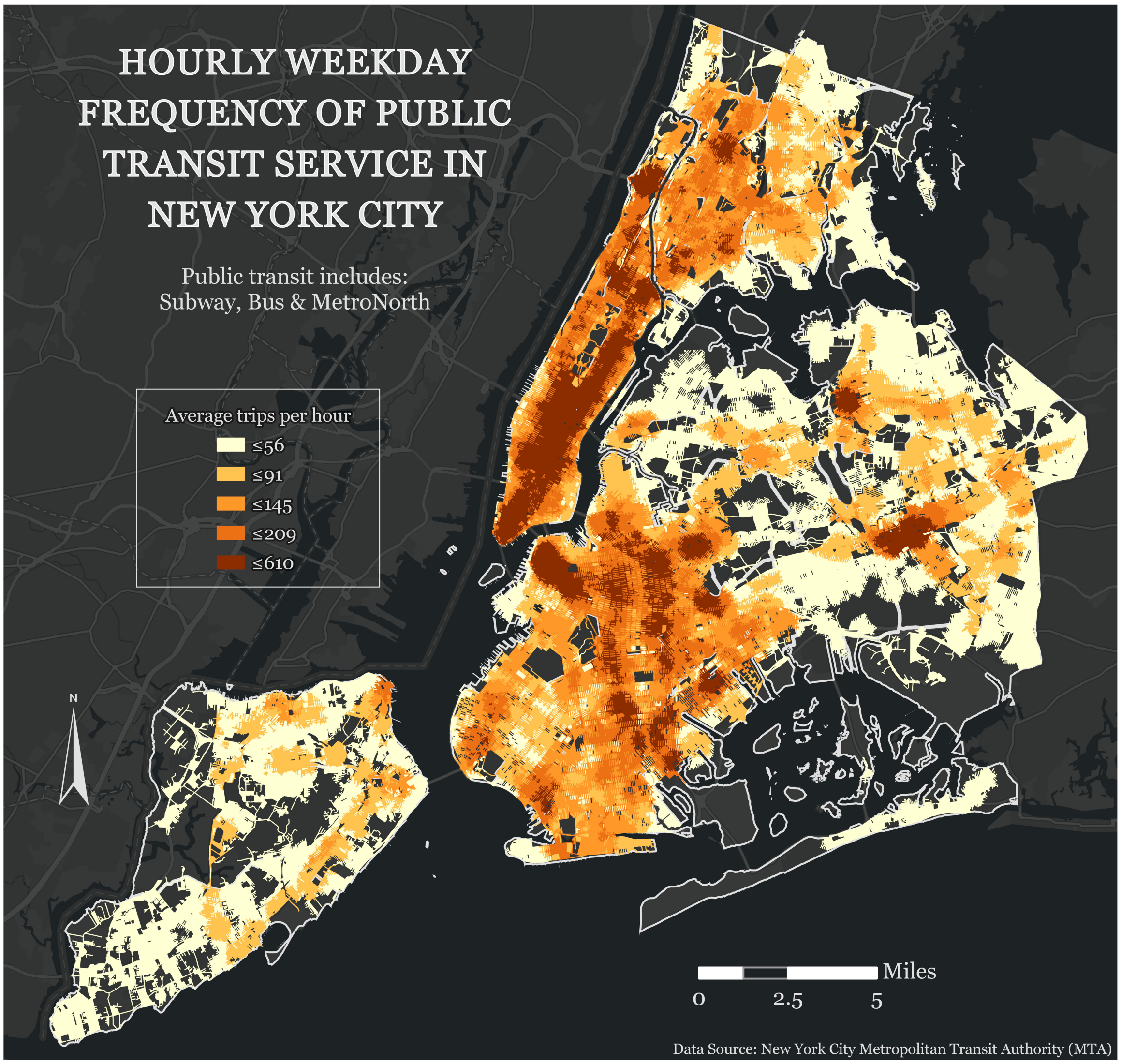


Figure 2: Average Weekday Public Transit Service

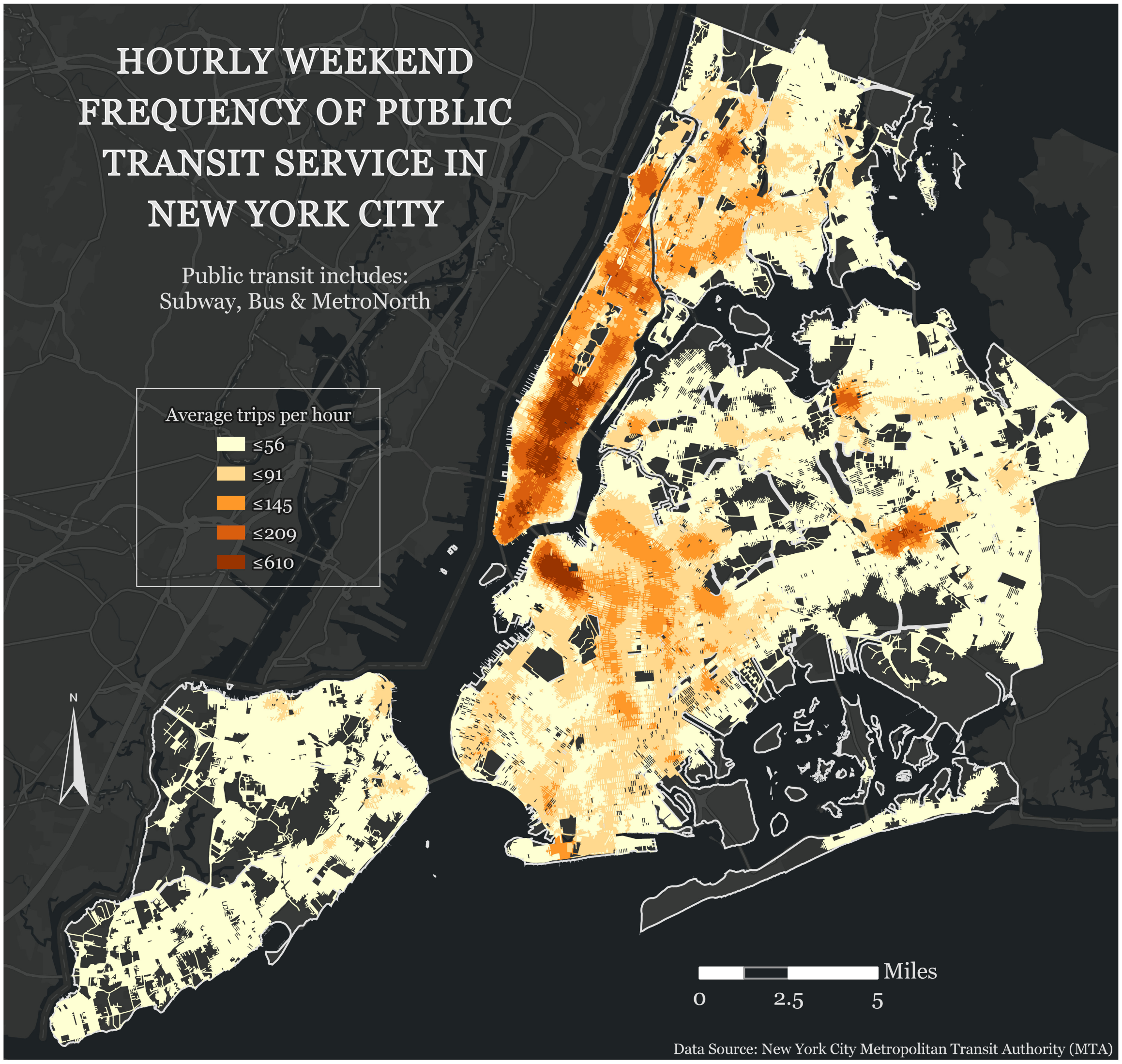


Figure 3: Average Weekend Public Transit Service

chunks (i.e. rush hour vs. late evening and early morning hours vs. off-peak vs. mid-day hours) did not reveal any significant differences in service provision between PUMAs. Service did fluctuate and change by the hour but because it aligns with basic principles of supply-and-demand; service increases during peak times (i.e. when people are getting to work) and decreases when there are fewer users (i.e. in the early morning when people are sleeping.)

Overall frequency is diminished, but comparatively more so in the outer boroughs where service is relatively lacking to begin with. It is also important to add that frequency does not have a linear relationship with accessibility and connectivity, as lower frequency is less forgiving when it comes to making time-sensitive transfers to other subway or bus lines. There is thus a disparity in service provision that manifests over space. If a bus only runs every half-an-hour and a person misses a bus by a minute, they will have to wait another twenty-nine minutes for the next bus to come. Even despite what the schedule dictates are the run-times of each bus and subway, there are unaccountable factors as traffic congestion is problematic in the city and many routes are placed in major corridors and roads that are heavily trafficked. Bus drivers may also take longer breaks especially as retaliation for low pay and dissatisfaction with contracts which contributes to further delays. Unions are also very influential and while they will not go on full strike, they may pressure subway conductors and bus drivers to delay service.

Maximum waiting time was also analyzed but the spatial distribution of low, medium, and high waiting times was comparatively more heterogeneous all throughout New York City. This is defined as the longest amount of time between one bus, subway, or Metro North train and the one thereafter approaching a given stop.

Where are the schools and jobs in New York City?

All of these temporal and spatial variations in public transit service affect job access between neighborhoods. Greater density of stations and public transit results in greater connectivity to other areas of the city, and thus improved job access. Employment data from the Longitudinal Origin Destination Employment Survey dataset by the Census, aggregated first at the census tract and then the PUMA level, returned a total of 4.2 million jobs in the city with the greatest concentration (standardized by the number of people) in the lower half of Manhattan and west-Brooklyn. The distribution of quality schools is similar, with greater concentration throughout Manhattan, west-Brooklyn, and parts of Queens. Jamaica, an anomaly in the eastern half of the city, is determined to have relatively higher job access as it is where John F. Kennedy International Airport is located and this transport hub provides a significant amount of jobs. As Figure 5 illustrates, job accessibility is greatest throughout Manhattan and west Brooklyn, including the affluent communities of Greenwich Village, Upper West Side, Upper East Side, Financial District and SoHo.

Minority Population and Communities with Least Job Accessibility

These communities also contain the lowest non-minority and immigrant populations, as shown in a comparison with Figures 6 and 7. Conversely, the least accessible communities, delineated with blue outlines in Figure 6 are also areas that contain significant minority populations. In Flatlands/Canarsie, which has very poor access with only 50% of its total area reachable to a public transit station within a reasonable walking distance, there are many African-Americans and West-Indian immigrants; in Jamaica, many Asians and Blacks, and in the area of the northern Bronx, both Blacks and Hispanics.

However, when individual and aggregate minority populations (percentage of each minority racial group in total population) are regressed on job accessibility, the same story was not told, despite the demonstrable differences in access between communities on either end of the spectrum.

```
##
## Calls:
## Nonwhite: lm(formula = JobsPerPerson ~ Minority, data = pumaJobs)
## Asian: lm(formula = JobsPerPerson ~ Asian, data = pumaJobs)
## Black: lm(formula = JobsPerPerson ~ Black, data = pumaJobs)
## Hispanic: lm(formula = JobsPerPerson ~ Hispanic, data = pumaJobs)
##
## ==================================================================
```

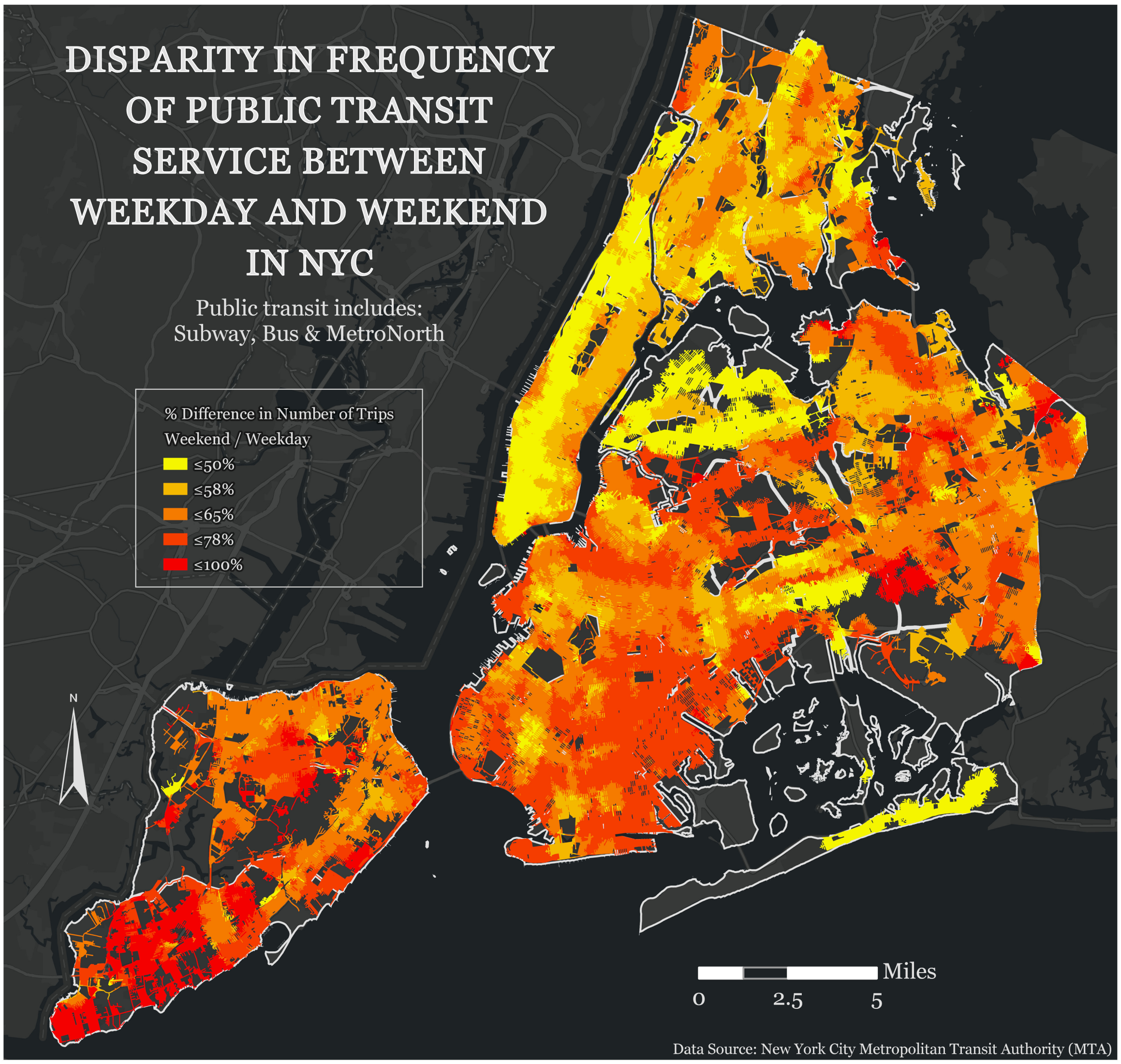


Figure 4: Percent Difference in Service between Weekend and Weekday

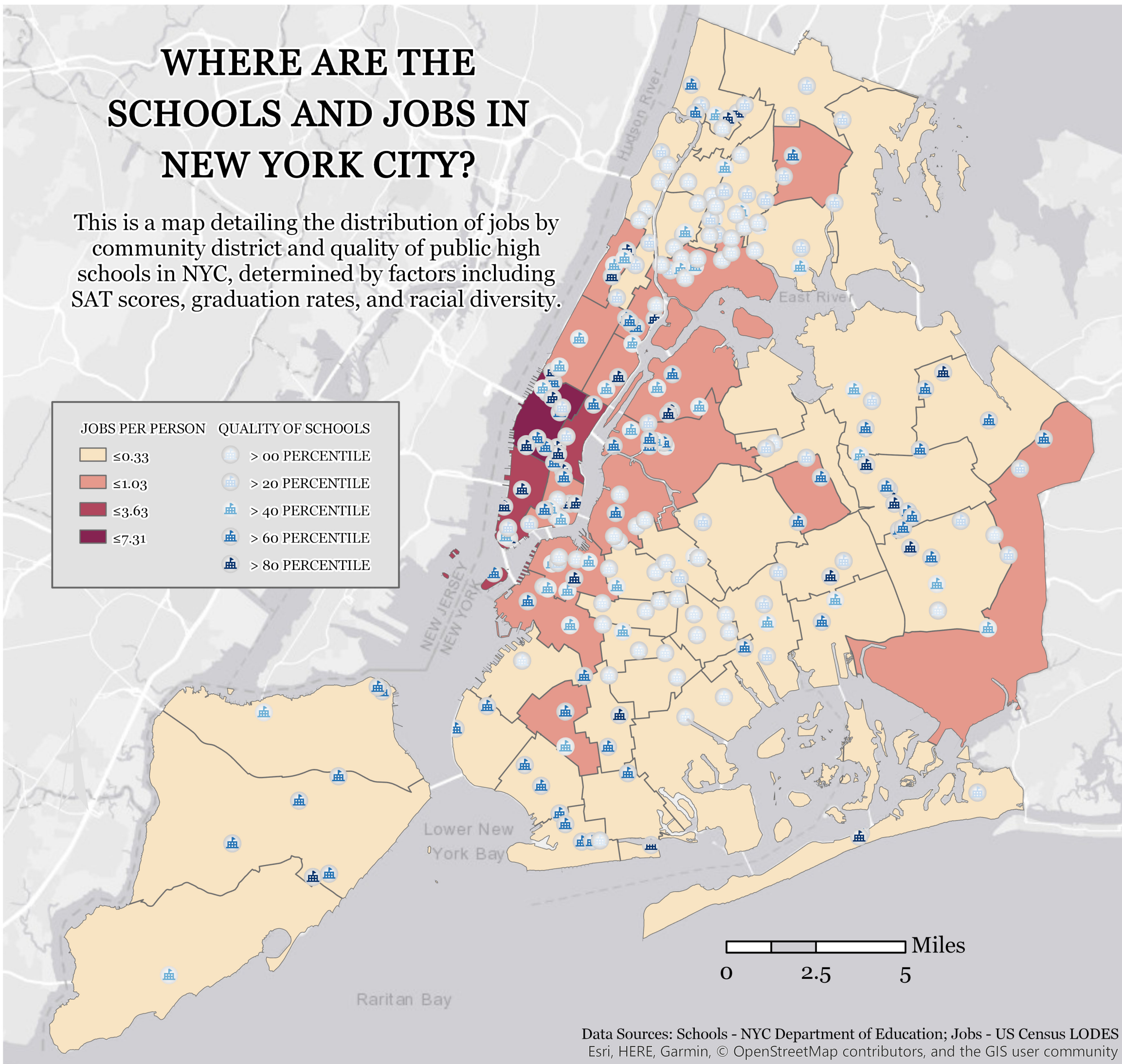


Figure 5: Schools and Jobs Distribution by PUMA in NYC

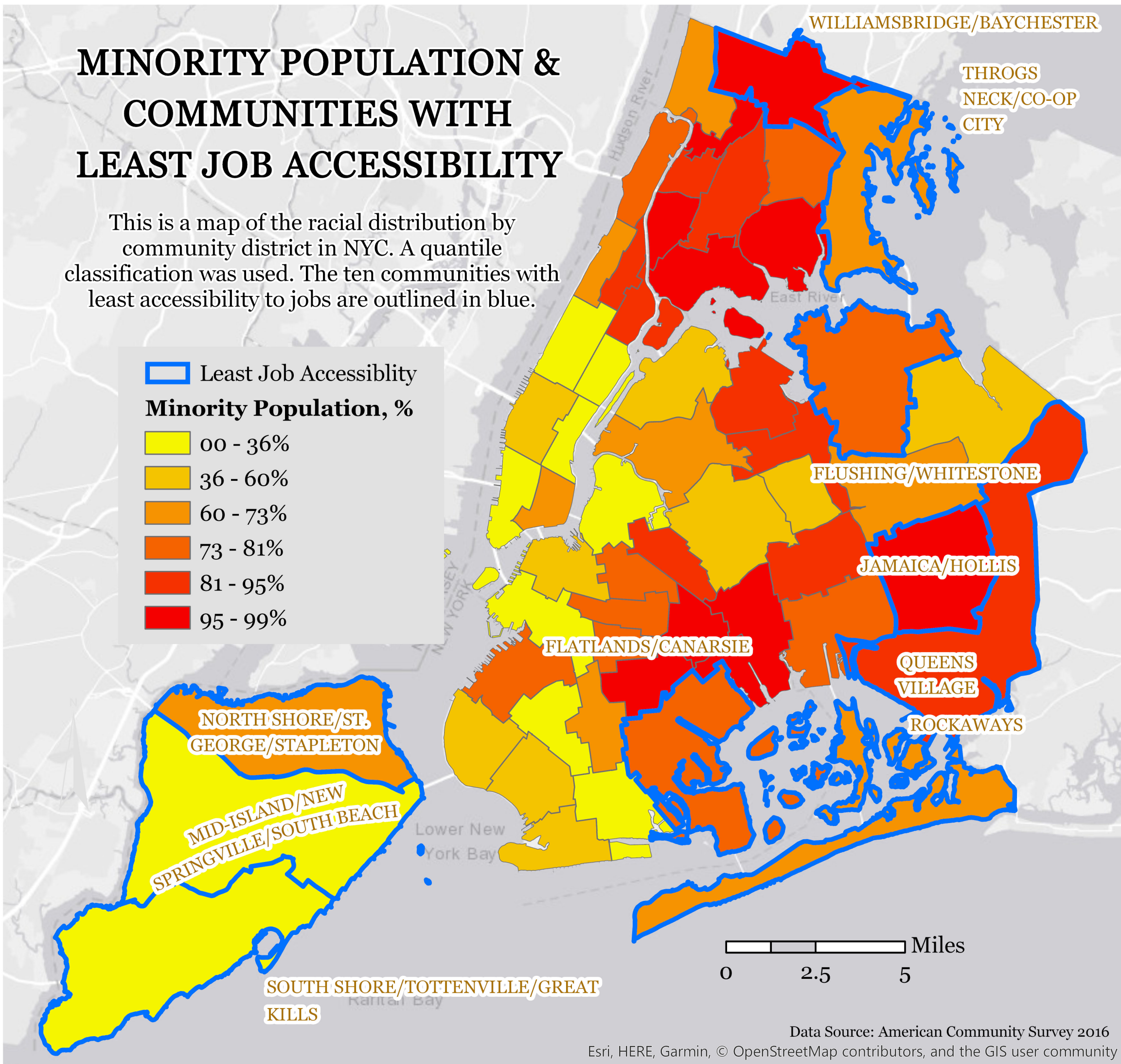


Figure 6: Minority Population and Job Access

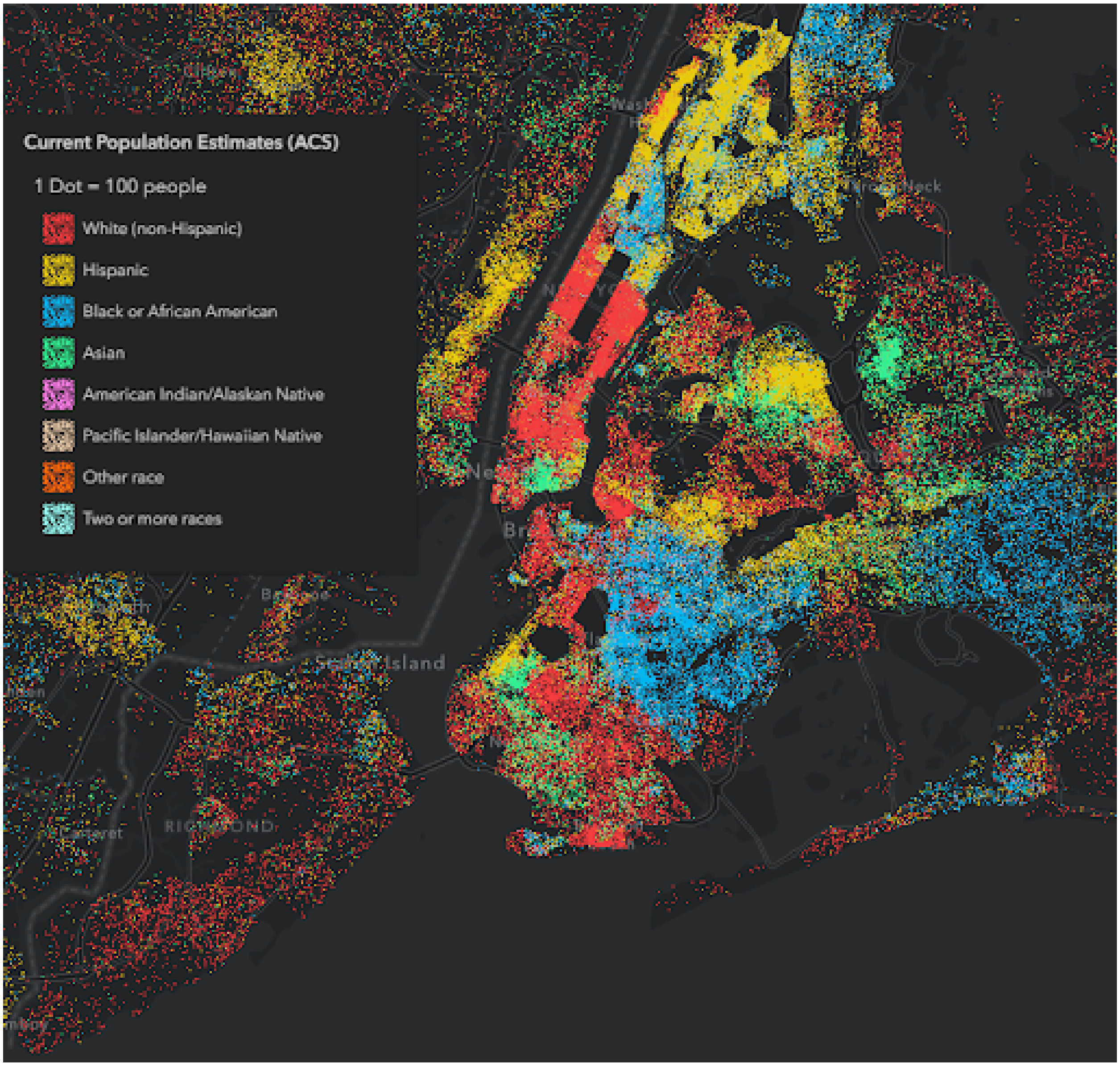


Figure 7: Racial Breakdown in New York City by Kristian Ekenes of Esri

```
##                Nonwhite      Asian       Black      Hispanic
## ------------------------------------------------------------------
##   (Intercept)    16.199***    14.791***    15.109***    11.851***
##                  (0.507)      (0.246)      (0.234)      (0.297)
##   Minority       -3.304***
##                  (0.703)
##   Asian                       -6.316***
##                               (1.353)
##   Black                                    -5.069***
##                                            (0.715)
##   Hispanic                                               7.210***
##                                                         (0.844)
## ------------------------------------------------------------------
##   sigma           8.624        8.625        8.579        8.543
##   R-squared       0.008        0.008        0.019        0.027
##   F              22.092       21.802       50.225       73.031
##   p               0.000        0.000        0.000        0.000
##   N            2640         2640         2640         2640
## ==================================================================
##   Significance: *** = p < 0.001; ** = p < 0.01; * = p < 0.05
```

Regressing the independent variable of percentage of minorities and each individual racial group is indeed statistically significant and negatively associated with job access, but the resulting R-Squared values are all less than 0.1. This reflects upon the issue of using a large aggregate unit of measurement in understanding accessibility as it is related to issues of social equity and inequity, and analyzing at the PUMA level is a prime example of the importance of scale in geography. The heterogeneity of the population throughout New York City and within each community makes it difficult to quantitatively assess the relationships between race and access as the city is so unique in its population structures with incredible diversity and the coexistence of many different groups of people. Similarly, using Census racial groups is not entirely illustrative of the minority population because different ethnic groups with varying socioeconomic characteristics are all lumped together (i.e. an immigrant Jamaican enclave has a different spatial distribution and socioeconomic characteristics than a African-American enclave but they are assumed to all be one entity.) PUMAs themselves are units with more than 100,000 people, essentially individual small cities, and the distribution of different minority groups throughout New York City is more scattered and does not conform to the administrative PUMA and census block boundaries demarcated by the US Census. Similarly, each of the 2177 census tracts that comprise the 55 PUMAs contain an average of four-thousand residents which are significant populations in of themselves. Beyond the issue of unit of measurement, linear regression may not be the best method for understanding the relationship between minority communities and job access.

This finding is further reinforced in a multivariate linear model that integrates different variables related to race, income, income diversity, education, rental housing characteristics, homeownership, immigrants, and period of transit service.

```
##
## Call:
## lm(formula = JobsPerPerson ~ . - PUMA - Minority, data = DFPUMAWEDANDSAT)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6395  -2.3771   0.3234   2.8561  10.6978
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                48.2385     5.9173   8.152 5.48e-16 ***
## IncDiv                      0.8997     0.1026   8.769  < 2e-16 ***
```

```
## MedInc                      -8.1059     1.6706  -4.852 1.29e-06 ***
## Unemploy                   -36.4614     6.5398  -5.575 2.72e-08 ***
## LaborForce                   1.0125     3.2911   0.308   0.7584
## Immigrant                   -6.9340     1.3818  -5.018 5.57e-07 ***
## White                       -5.3181     4.9261  -1.080   0.2804
## Black                       -7.0811     4.9186  -1.440   0.1501
## Asian                       10.9796     5.3106   2.067   0.0388 *
## Hispanic                     1.4746     4.9793   0.296   0.7671
## Senior                     -92.9555     4.6054 -20.184  < 2e-16 ***
## Bachelors                   20.7218     2.6055   7.953 2.68e-15 ***
## NotFinHS25                 -33.2875     2.9741 -11.192  < 2e-16 ***
## Homeownership               -7.3372     1.4877  -4.932 8.64e-07 ***
## PropertySales              -12.6031     0.5062 -24.899  < 2e-16 ***
## RacialDiv                  -14.1900     1.0669 -13.300  < 2e-16 ***
## ServiceOff Peak             -2.7891     0.2532 -11.014  < 2e-16 ***
## ServicePeaks                 0.2210     0.2233   0.990   0.3225
## ServicePre and Post Peaks  -0.3655     0.2532  -1.443   0.1491
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.199 on 2621 degrees of freedom
## Multiple R-squared:  0.7664, Adjusted R-squared:  0.7648
## F-statistic: 477.7 on 18 and 2621 DF,  p-value: < 2.2e-16
```

The all-inclusive model identifies explanatory variables of higher rates of immigrant (*Immigrant*), minority (*Minority*), and senior (*Senior*) populations, people 25 or older who have not graduated high school (*NotFinHS25*), homeownership (*Homeownership*), racial diversity (*RacialDiv*) and property sales (*PropertySales*) to all be negatively correlated with job access; meanwhile, greater income diversity (*IncDiv*) and higher average educational attainment (*Bachelors*) are positively correlated with job access. Of the four racial groupings, only Asian is statistically significant at $p < 0.05$; with otherwise identical populations (in respect to the other variables in the model), a 10% increase in the Asian population corresponds to approximately an increase in accessible jobs by one. But this ignores the differences between different Asian ethnic groups (i.e. while east-Asians including the Chinese, Japanese, and Korean, are relatively high-earners and achievers in the United States, many southeast-Asian groups such as the Burmese and Laotian live in abject poverty.) This particular model should be viewed with great caution despite the relatively-high adjusted R-squared value of 0.7648 and normally-distributed residuals as there may be issues arising from similarities and potential redundancies within the data. The dataset contains forty-eight entries for each of the 55 PUMAs, as it includes the temporally-sensitive measure of job access across all twenty-four hours for a weekday and weekend. Put another way, each hour is associated with one hundred and ten observations. Modelling these complexities is a difficult task; differences between PUMAs, hours, and day of the week all have to be accounted for. And these issues exasperate the problem of using an aggregate unit of measurement as discussed in the previous paragraph, thereby building errors upon errors and potentially masking the underlying relationships between not only race but also other socioeconomic and demographic factors with job access.

```
##
## Call:
## lm(formula = JobsPerPerson ~ Homeownership + Bachelors + MedInc,
##     data = pumaJobs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.7176  -3.7302   0.4732   2.6948  13.7031
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.4188     0.2796  62.310  < 2e-16 ***
## Homeownership -29.7810     0.8011 -37.177  < 2e-16 ***
## Bachelors      25.0674     1.3063  19.190  < 2e-16 ***
## MedInc         -4.2993     1.1085  -3.879 0.000108 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.162 on 2636 degrees of freedom
## Multiple R-squared:  0.6449, Adjusted R-squared:  0.6445
## F-statistic:  1596 on 3 and 2636 DF,  p-value: < 2.2e-16
```

A much simpler model with the highest F-statistic of all models explored finds homeownership, educational attainment, and median income (which was transformed by dividing by 100,000 and then scaling to a mean of 0) to be the most influential explanatory variables; with two otherwise identical PUMAs, the one with 10% greater homeownership correlates to approximately 3 fewer jobs. Similarly, a PUMA with a comparatively 10% increase in overall population with at least a Bachelor's degree correlates to an increase in more than 2 jobs per person. This is to no surprise, as greater educational achievement not only means more potential job opportunities, but also greater access to quality ones. Homeownership is negatively correlated with job access and this is likely attributed to the concentration of jobs in the most expensive parts of the city, primarily in Manhattan and west Brooklyn where the cost of land is extremely costly, so most people living there are renters. It may also be said that with continued urbanization and the exorbitant rising cost of land in New York, more people will be renters and it will be interesting to study this dichotomy and their relationship with mobility and accessibility as time progresses. Those with higher income are more likely to be able to afford homes in the suburban areas of the city. And more money also means more flexibility in residential location and choosing places that offer more amenities and services. This particular association between higher median income and lower job access is also corroborated by a study by Pathak et al. that found census tracts in Atlanta with comparatively better access to public bus transportation also have a higher proportion of low-income residents (Pathak et al. 2017).

Both *Homeownership* and *MedInc* were then processed into a multivariate linear model with and without interactions with the *Minority* variable. The resulting F-statistics suggest that the model without interactions is a superior choice. However, both models fall short of the linear model with *Homeownership*, *Bachelors*, and *MedInc* as the sole explanatory variables. But putting that aside momentarily, there are notable differences in the values obtained between the two models accounting for the minority population. In the model with interactions, the coefficients for *Minority* and *Homeownership* are positive but not statistically significant. However, all combinations of interactions are statistically significant with $p < 0.01$; and they suggest that a minority population with greater rates of (1) median income is positively correlated with greater job access, (2) homeownership is negatively correlated with job access, and (3) both median income and homeownership are positively correlated with job access. This suggests that minority populations of higher socioeconomic class alone as well as those who also own their homes enjoy greater job access. But if median income is disregarded, minorities who own their own homes have poorer job access. This may reflect on the housing stock and to a certain extent the economic landscape in the areas that minorities reside; minorities own houses of lower-quality in places that are also disadvantaged in terms of access to public transit and thus jobs. This should be taken with a grain of salt as the data on homeownership and median income are for the entirety of PUMAs and not just the minority populations within them.

```
##
## Calls:
## With Interactions: lm(formula = JobsPerPerson ~ Minority * MedInc * Homeownership,
##     data = pumaJobs[, c(3, 7, 11, 15)])
## Without Interactions: lm(formula = JobsPerPerson ~ Minority + MedInc + Homeownership,
##     data = pumaJobs[, c(3, 7, 11, 15)])
##
## ==============================================================================
```

```
##                                      With Interactions  Without Interactions
## ----------------------------------------------------------------------------
##   (Intercept)                             10.505**           24.337***
##                                           (3.542)            (0.786)
##   Minority                                 4.781             -5.890***
##                                           (3.646)            (0.651)
##   MedInc                                  17.463***          10.720***
##                                           (3.850)            (0.681)
##   Homeownership                           21.686            -39.707***
##                                          (11.174)            (0.632)
##   Minority x MedInc                       12.118**
##                                           (4.558)
##   Minority x Homeownership               -94.803***
##                                          (12.996)
##   MedInc x Homeownership                 -57.953***
##                                          (12.431)
##   Minority x MedInc x Homeownership       78.823***
##                                          (14.013)
## ----------------------------------------------------------------------------
##   sigma                                    5.293              5.427
##   R-squared                                0.627              0.608
##   F                                      633.041           1360.228
##   p                                        0.000              0.000
##   N                                     2640               2640
## ============================================================================
##   Significance: *** = p < 0.001; ** = p < 0.01; * = p < 0.05
```

Educational Accessibility

These findings are also descriptive and similar to that of the landscape of educational accessibility. The same areas that enjoyed the greatest job access are also those with the greatest access to quality schools. This is not unexpected given that most of the quality schools are concentrated in Manhattan and west Brooklyn, where jobs are also relatively greater in number. As the map of schools and jobs distribution illustrated, and when compared with the dot density map of race, there is a dearth of schools in the outer boroughs. It was also found that access to quality schools is significantly correlated with educational achievement, measured as the percentage of the population with a Bachelor's degree or higher. This suggests that having access to quality education and a number of choices to schools results in improved educational achievement outcomes, and all of this is tied to the racial distribution and demographics of communities.

Conclusion

While the quantifiable results did not reveal a meaningful relationship with the minority populations and access in New York City, it is still the author's strong belief that there are issues of social inequities that manifest because of a disparate physical and social landscape of public transit service as it both facilitates and hinders job and educational accessibility. Increasing distance and travel time from the main areas of job provision has significant implications on the socioeconomic conditions of a community and contribute to greater unemployment as well as poverty. Service varies temporally from hour to hour and from weekday to weekend and spatially between more affluent areas and more impoverished ones and between more racially heterogeneous and more homogeneous communities.

Analysis at the PUMA level is a classical example of scale in geography and is not able to fully capture the relationship between accessibility and minority communities. In the geography discipline, this is commonly understood as the modifiable areal unit problem (MAUP) and this analysis exhibits both the aggregation and zonal issues. As PUMAs are aggregates of census tracts which are aggregates of census blocks which are aggregates of individual survey data, relationships may be masked and/or obscured. Similarly, the intentional choice of grouping results in relationships that may or may not be observed with different and other groupings. The spatial heterogeneity of populations may be understated because of such issues. As ethnic minority and immigrant communities do not exactly align to the administrative boundaries established by governmental agencies, potential relationships between minorities and job and educational access may be entirely disregarded. Ethnic, immigrant, and refugee enclaves are all spatially organized in areas smaller than any PUMA. This is further complicated with the incredible dynamism in all aspects of New York City's landscape as many different groups of people coexist in the same PUMA (i.e. the author who is ethnically Chinese was raised in a predominantly African-American and West-Indian immigrant community in east Brooklyn.) With the greatest concentration of diversity in all the world, it may be difficult to represent this diversity especially as data and models are oversimplifications of reality. And even despite the abundance of data for the city of New York, not all are fit for this analysis as different agencies use different systems for delineating the different parts of the city (i.e. the New York City Department of Sanitation and New York City Policy Department runs their analyses on community districts, which are similar to PUMAs but not entirely.) But this does not mean that the analyses from this research are not useful nor that they are wrong. It simply means that analysis at the scale of PUMAs is not the best for understanding the particular question of public transit service as it relates to educational and job opportunities for minority communities.

However, in light of the questions posed at the start of the research, the analyses and results suggest that while there may be a disparate landscape of access structured by time and across space, public transit service provision as a conduit of job and educational access is not inherently disadvantageous to any particular minority nor immigrant subpopulation. Yes, job and educational access may be impacted by public transit service and social disparities in the landscape may arise because of these impacts but there is a lack of conclusive evidence that suggests this in the meantime.

Future Directions

This research is not even close to being done, nor that it will ever be â€˜complete' in a more abstract sense given the dynamic interplay between cultural and urban processes and policies, but the ambition is to leverage this knowledge and further advance the current understanding which will lead to concrete and feasible change, whether through policy or infrastructural changes. Some thoughts are to analyze at the census tract level, and while this provides more precision, it is significantly more difficult to communicate this and make these results more actionable as there are a total of 2177 compared to just 55 PUMAs. This raises the issue of the computational capacity necessary to run these various tools for analysis. Parsing the high-resolution GTFS data altogether with the 2177 locations in origin-destination cost matrices in ArcGIS is computationally intensive. Future steps will look to improve these methods and reduce the necessary time and resources to run these functions so that this workflow can also be applied to understanding accessibility with public transit data in other cities and places. Another direction is to integrate the shadow transit network which is essentially individual systems of dollar vans (similar to jitneys) run by different ethnic enclaves to understand how they improve accessibility and/or complement existing public transit services. Because data on this is lacking and because it is technically an illegal mode of transit service (although there are active measures being taken to legally integrate them by the city government), fieldwork would be necessary and this potentially includes riding along the entirety of the different networks, precisely mapping all stops, interviewing users, recording user demographics, and recording the different schedules. A completely (but not entirely unrelated) direction is to overlay parcel data at the resolution of individual buildings with information about zoning, home values, and vacant lots to try to understand how housing and planning ordinances may affect service provision.

References

Bocarejo S., J. P., & Oviedo H., D. R. (2012). Transport accessibility and social inequities: a tool for identification of mobility needs and evaluation of transport investments. Journal of Transport Geography, 24, 142-154. https://doi.org/10.1016/j.jtrangeo.2011.12.004

Dempsey, N., Bramley, G., Power, S., & Brown, C. (2011). The social dimension of sustainable development: Defining urban social sustainability. Sustainable Development, 19(5), 289-300. https://doi.org/10.1002/sd.417

Dillahunt, T. R., & Veinot, T. C. (2018). Getting There: Barriers and Facilitators to Transportation Access in Underserved Communities. ACM Transactions on Computer-Human Interaction, 25(5), 1-39. https://doi.org/10.1145/3233985

Easley, J. (2018). Spatial mismatch beyond black and white: Levels and determinants of job access among Asian and Hispanic subpopulations. Urban Studies, 55(8), 1800-1820. https://doi.org/10.1177/0042098017696254

Ekenes, Kristian. (2019). Population density by race/ethnicity in New York City. Retrieved April 02, 2019, from Github website: https://ekenes.github.io/esri-js-samples/4/visualization/dot-density/population-race.html

Evans, S. P. (1973). A relationship between the gravity model for trip distribution and the transportation problem in linear programming. Transportation Research, 7(1), 39-61. https://doi.org/10.1016/0041-1647(73)90005-1

Hansen, W. G. (1959). How Accessibility Shapes Land Use. Journal of the American Institute of Planners, 25(2), 73-76. https://doi.org/10.1080/01944365908978307

Hellerstein, J. K., McInerney, M., & Neumark, D. (2009). Spatial Mismatch, Immigrant Networks, and Hispanic Employment in the United States (Working Paper No. 15398). https://doi.org/10.3386/w15398

Hochmair, H. H. (2016). Spatiotemporal Pattern Analysis of Taxi Trips in New York City. Transportation Research Record: Journal of the Transportation Research Board, 2542(1), 45-56. https://doi.org/10.3141/2542-06

Jarv, O., Tenkanen, H., Salonen, M., Ahas, R., & Toivonen, T. (2018). Dynamic cities: Location-based accessibility modelling as a function of time. Applied Geography, 95. https://doi.org/10.1016/j.apgeog.2018.04.009

Jin, S. T., Kong, H., & Sui, D. Z. (2019). Uber, Public Transit, and Urban Transportation Equity: A Case Study in New York City. The Professional Geographer, 71(2), 315-330. https://doi.org/10.1080/00330124.2018.1531038

Kain, J. F. (1992). The Spatial Mismatch Hypothesis: Three Decades Later. Housing Policy Debate, 3(2), 371-460. https://doi.org/10.1080/10511482.1992.9521100

Litman, T. (2002). Evaluating Transportation Equity. World Transport Policy & Practice, 8(2), 50-65.

Manout, O., Bonnel, P., & Bouzouina, L. (2018). Transit accessibility: A new definition of transit connectors. Transportation Research Part A: Policy and Practice, 113, 88-100. https://doi.org/10.1016/j.tra.2018.03.028

Meyer, J. R., Kain, J. F., & Martin Wohl. (1965). The Urban Transportation Problem. Cambridge: Harvard University Press.

New York City Department of City Planning. (2019, April 27). NYC Population Facts. Retrieved April 27, 2019, from New York City Population website: https://www1.nyc.gov/site/planning/data-maps/nyc-population/population-facts.page

Pathak, R., Wyczalkowski, C. K., & Huang, X. (2017). Public transit access and the changing spatial distribution of poverty. Regional Science and Urban Economics, 66, 198-212. https://doi.org/10.1016/j.regsciurbeco.2017.07.002

Qian, X., & Ukkusuri, S. V. (2015). Spatial variation of the urban taxi ridership using GPS data. Applied Geography, 59, 31-42. https://doi.org/10.1016/j.apgeog.2015.02.011

Shaw, S.-L. (2010). Geographic information systems for transportation: from a static past to a dynamic future. Annals of GIS, 16(3), 129-140. https://doi.org/10.1080/19475683.2010.513155

Silver, N., & Fischer-Baum, R. (2015, August 28). Public Transit Should Be Uberâ€™s New Best Friend. Retrieved April 22, 2019, from FiveThirtyEight website: https://fivethirtyeight.com/features/public-transit-should-be-ubers-new-best-friend/

United Nations World Commission on Environment and Development. (1987). Report of the World Commission on Environment and Development: Our Common Future. Retrieved from Oxford University Press website: http://www.un-documents.net/our-common-future.pdf

Wilbanks, T. J. (1994). Presidential Address: "Sustainable Development" in Geographic Perspective. Annals of the Association of American Geographers, 84(4), 541-556.

Wilson, A. G. (1967). A statistical theory of spatial distribution models. Transportation Research, 1(3), 253-269. https://doi.org/10.1016/0041-1647(67)90035-4