

# Introduction

Both studies and simply walking around a city will tell you that not all neighborhoods are equal. At least in New York City, there is large variability among neighborhoods in wealth and, thereby, access to amenities and necessities, including access to healthcare. I hypothesize that certain neighborhoods have many more locations to receive medical care than do others. This decreased access to care can lead to poor outcomes as it makes it more difficult to be seen by a healthcare provider.

The fundamental question I attempt to answer is this: Which neighborhoods in NYC are underserved by healthcare locations?

Answers to this question would allow a city to invest in healthcare access in an equitable way, instead of simply building medical centers where convenient. Thus, city planners, social scientists, public health officials, and healthcare workers will find this work of interest.

## Data

Conveniently, Foursquare has a venue category called "Medical Center" which encompasses the following venue categories pertinent to this project:

- Acupuncturist
- Alternative Healer
- Chiropractor
- Dentist's Office
- Doctor's Office
- Emergency Room
- Eye Doctor
- Hospital
- Maternity Clinic
- Medical Lab
- Mental Health Office
- Nutritionist
- Physical Therapist
- Rehab Center
- Urgent Care Center

The "Medical Center" category also includes veterinarians, but I will be excluding them from this analysis.

Neighborhoods are of all different shapes and sizes, so we cannot simply look at the counts of medical venues as this will increase the count for neighborhoods that are simply larger or more dense populated. We can normalize by comparing the number of medical venues to the number of a different venue type. For this project, I will use the number of stores selling alcohol as this normalizing factor and compute the medical center to alcohol store ratio for each neighborhood (hereafter, MC:AS). Foursquare has the following venue categories for stores selling alcohol:

- Liquor Store

- Wine Shop
- Beer Store

To obtain basic information, I scraped data from Wikipedia in order to define the neighborhoods of NYC.

## Methodology

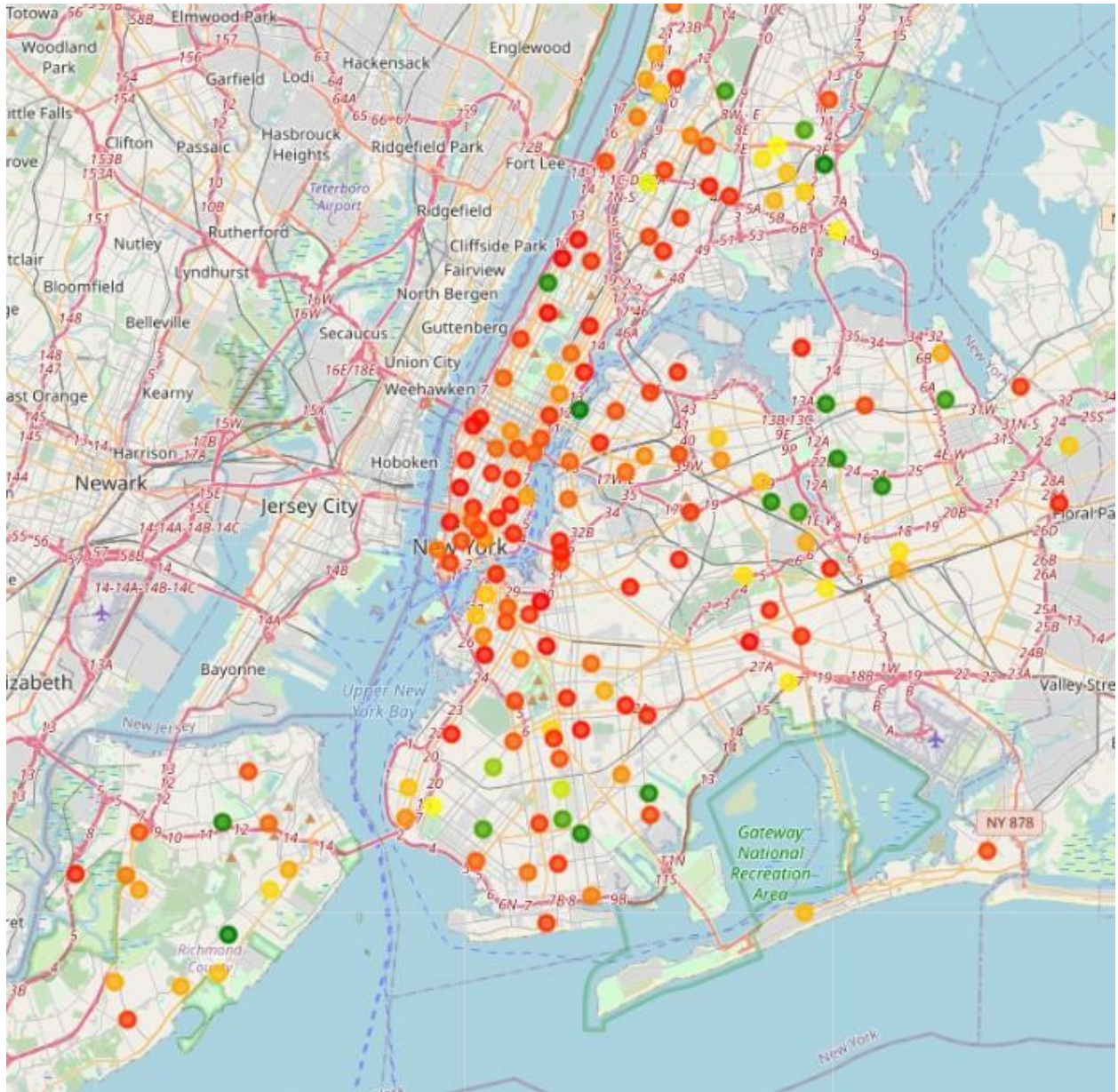
After acquiring the data, I noticed that many neighborhoods had only a handful of venues, so made the decision to exclude neighborhoods with fewer than 10 venues returned by the Foursquare API.

Additionally, not all neighborhoods had alcohol stores, and my metric of MC:AS is undefined when the denominator is zero. Therefore, for the purposes of calculating MC:AS, I set the number of alcohol stores equal to the minimum of 1 and the actual number of alcohol stores. I then calculated the MC:AS for each neighborhood.

I then turned my attention to clustering the neighborhoods. I transformed the categorical data via one-hot encoding then computed the proportion of venues in each neighborhood that fell into each venue category. To cluster the neighborhoods, I used the *k-means* algorithm, with  $k = 3$ , as I was interested in seeing if the clusters would roughly follow the distribution of MC:AS via high-, medium-, and low-access clusters. I selected *k-means* because I wanted an unsupervised clustering algorithm.

## Results

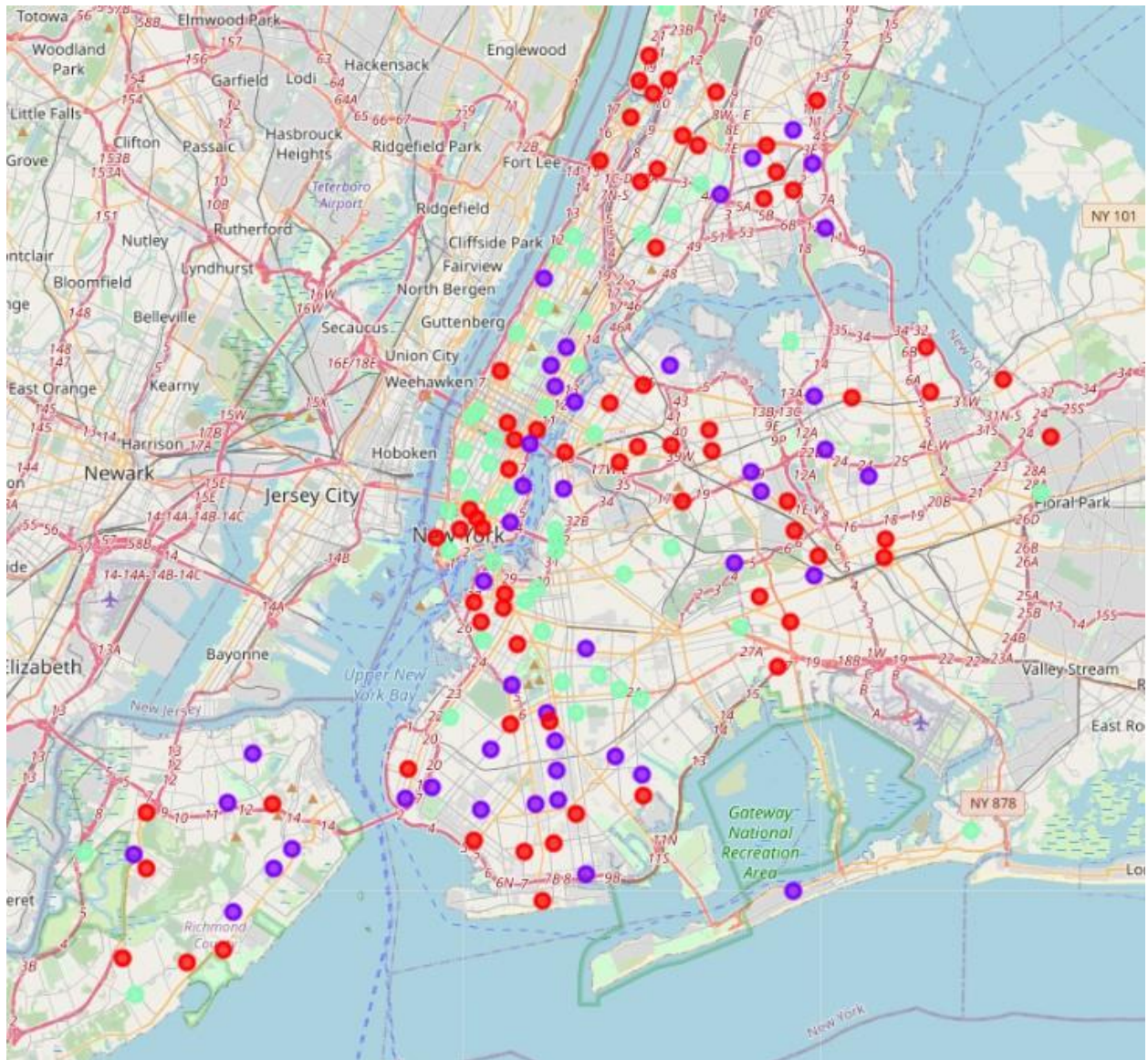
My MC:AS ratio ranged from 0.45 at the lowest, up to 30 at the highest, indicating large disparities in access to healthcare in NYC. The minimum value occurred in Manhattanville, Manhattan and the maximum value occurred in Pelham Bay, Bronx; Roosevelt Island, Manhattan; and New Dorp, Staten Island.



MC:AS tended to be lowest in Manhattan and highest in the outlying suburbs, roughly correlating with population and building density.

Clustering the neighborhoods according to the medical center and alcohol store venues gave mixed results, with the clusters relatively even spread throughout the boroughs. There may be some correlation between clusters and MC:AS, though it is difficult to tell from visual inspection.





## Discussion

There is a wide range of densities of medical centers in neighborhoods in New York City. The results presented above demonstrate clear spatial variability in those densities, both in terms of my MC:AS ratio and in terms of neighborhood clustering according to venue counts. The MC:AS ratio appears to anticorrelate somewhat with population and building density, though this does not explain the entirety of the disparities between neighborhoods. Further work would be valuable in exploring the possible correlations between neighborhood racial make-up, average neighborhood wealth, and these results.

These data should be examined by public health officials in NYC to determine which areas are being underserved by medical centers. Developers could be encouraged to open new medical center venues in these underserved areas.

## Conclusion

We know that ZIP code plays a huge role in a person's lifelong health, and easy access to healthcare surely influences that health. These data support the idea that healthcare in NYC is not evenly distributed and can hopefully help improve access for those populations that need it the most.