Understanding the Effects of Diversity on Health Outcomes in New York City: Overdose and Gun Violence

Louis Pang
Adviser: Olga Troyanskaya

May 1st, 2023

Abstract

New York City is one of the most diverse cities in the entire United States. But how does this diversity affect health equity within the city? That's the question that my group and I have attempted to tackle over the semester. Prevailing data appears to show that Black and Hispanic New Yorkers experience higher rates of premature death due to chronic illnesses, childbirth, and violence. The goal of our project is to investigate these disparities and study how urban diversity affects many health outcomes for underrepresented communities across New York City's boroughs by pinpointing some of the factors, such as health care access, air quality, and race, which could affect healthcare outcomes. To accomplish this, we compiled public health data into a demonstrative visualization of health outcomes and health factors in New York City. Compiling these studies, we'll be left with a broad overview of how social factors affect health outcomes for everyday New Yorkers. Our hope is that a visualization similar to this one - lightweight, deployable, and built on free software - could be used to help inform future policy decisions. While the visualization is effective in some respects, it fails to provide useful correlative conclusions when not enough data is provided for a certain outcome or factor.

1. Introduction

Public health is a complex interaction of many real-world variables producing a number of real-world consequences. Despite these complexities, one thing remains certain - in the United States, minority groups experience disproportionately poorer health outcomes than their non-minority counterparts. But understanding and addressing this disparity is harder than it may appear. From healthcare access to air pollution, the list of factors to consider when attempting to address public health concerns is vast and never ending. Thus, it's important to carefully assess the possible effects of shifts in the public health system on the health of individuals before enacting any concrete change. We can measure these effects using the aforementioned health outcomes - a quantitative set of attributes that describe the health results of a single individual (e.g. life expectancy).

A comprehensive study of any and all possible factors influencing health outcomes in the United States is a tall task for any individual, so it's wise to narrow our scope. If you ask the average Princeton senior where they'll be moving after graduation, chances are they'll say "The City", by which they mean New York City (NYC). NYC is a hotbed of human activity. But the intrigue of NYC extends far beyond just the promise of a college student's future. It is also one of the most populous and diverse cities in the United States. It presents a profound challenge to public health, given the extreme population density (on average over two times that of Chicago) and remains an area of intense interest, with an abundance of data and prior study.

Situating this paper among the relevant literature, this study is one in a group of three, which aims to provide an overview of a wider swathe of factors (environmental, economic, and genetic) and their effects on a number of health outcomes. The related studies will focus on chronic illness and birth weight outcomes, while this one will focus primarily on drug overdose and gun death outcomes.

In 2021, nearly fifty thousand people died to injuries inflicted by firearms in the United States. More than double (106,000) died from drug-involved overdoses. In both categories, the United States is an outlier from the rest of the world, and in both categories, minority populations are disproportionately affected. Viewing NYC as a relevant metropolitan microcosm of American health, let's attempt to use the data at our disposal to rationalize the disparities between minority and non-minority communities in drug overdose and gun death outcomes - two statistically significant causes of premature mortality in NYC.

Our study aims to create a visualization tool to map health outcomes, as well as the factors influencing them, in order to allow researchers in public health to draw appropriate conclusions and formulate next steps in addressing health outcome disparities. Our hope is to create easily digestible resources, bridging the gap between data and policy in a way that standard correlation tables may not sufficiently address.

Precision health is a subset of healthcare which aims to provide customized health solutions and treatments to the individual. In practice, this has meant identifying individualized genetic risk factors and providing recommendations ahead of time in an attempt to improve health outcomes. In this model, environmental, economic, and lifestyle factors are treatments, often altered to mitigate the potential complications identified in the genetic component of precision health. However, we'd like to contribute to an ongoing shift: considering these mitigating factors as possible indicators of future health outcomes. With our visualization tool, we will consider a variety of factors, such as race, proximity to health care facilities, and air quality in order to determine their relevance in the study of health outcomes. So, our goals become two-fold: to create an accessible visualization, and to provide an evaluation of a variety of factors' contribution to our health outcomes.

2. Background and Related Work

2.1. Drug and Firearms Research in New York City

Drug and firearm outcomes have been a topic of interest in New York City for at least thirty years. One of the early papers that guided my research was published in 2002 and attempts to link the two health outcomes, exploring the role of drug use in eventual fatal firearm incidents. In a comprehensive overview of data spanning from 1990-1998, researchers tracked the percentage of firearm deaths in which victims or perpetrators tested positive for illicit substances of many types, including opiates. The study concluded that while the percentage of gun fatalities involving opiates had decreased over the 8-year cycle, the disparity with which gun fatalities affected men in minority groups (note the outdated definitions of African American and Latino groups), had diminished slightly, but remained ever-present. Furthermore, the slight decrease in rates in regards to both opiates and affected minority groups could be attributed to the fact that total firearm deaths per 100,000 people had nearly halved over the nine-year spell. While the study was well defined, it was clear to me from an early stage that there was an emphasis on race and individualized factors; that is to say, when attempting to ascertain the risk of drug-related firearm death for an individual, they would examine the individual's race, economic status, or drug use habits. This trend is indicative of the precision health model, in which health outcomes are determined by an individual's characteristics.

Subsequent studies departed from this individualistic consideration with substantial success. A 2003 paper surveyed the correlation between income distribution (as opposed to income) with drug overdose risk in NYC. The study found that when adjusted for demographics and neighborhood income, areas with more unequal income distribution had higher levels of unintentional drug overdose death. These findings were heavily influential in my research. How could we collect and evaluate the connections between environmental factors and health outcomes for a broader spectrum of factors in a way that would allow us to compare the individualistic and environmental approaches? It's important to note that while this study was promising, correlation between environmental factors and health outcomes does not indicate causation. So in an attempt to understand the correlation between potential factors in health outcomes, I sought out explanations for a number of individual and environmental factors.

I would be remiss not to mention the public health EPI briefs produced by the New York City Department of Health in my related work. Their exploration of Opioid and Fentanyl overdose by neighborhood helps to identify trends by grouping enough overdose data together to provide meaningful conclusions without generalizing by borough. It would eventually prove helpful in choosing how to group my data and provide meaningful conclusions.

2.2. Individual Factors

As seen by the above studies, when examining an individual's characteristics, one of the first demographic variables invoked is income. Income can have substantial, long-term effects on health outcomes. For example, living with low income may give you poorer access to healthy food and adequate exercise facilities, leading to obesity. In addition low income can decrease your access to resources, leading to increased anxiety or emotional distress which may cause an individual to develop unhealthy drug or sleeping habits. While income can have an effect on health, it appears as though the nature of the relationship between low income and poor health outcomes is cyclical. Poor health can also inhibit people's access to education or lucrative job opportunities, and worsen the effects of poverty

Following income, the next most relevant demographic variable is often race or ethnicity. While ethnicity can be indicative of some genetic variability, it likely accounts for very little of the actual disparity in health outcomes. Likely, race is an intermediate variable, accounting for institutionalized racism, economic structures which negatively affect minority groups in the United States, and barriers to care enforced by social inequality. So while race is likely correlated with poorer health outcomes, we must be careful not to inherently attribute poorer health outcomes with race, and instead search for the causal factors behind this variable.

2.3. Environmental Factors

Air pollution has been linked with poor health outcomes for many years. The World Health Organization (WHO) posits that higher rates of particulates in ambient air lead to increased risk of diseases such as respiratory infections, heart disease and lung cancer. While air pollution in large American cities has improved vastly over the past 20 years, I still wanted to explore the effects of an environmental factor heavily linked to poorer health outcomes, especially in a city as densely populated and heavily trafficked as NYC.

Lastly, distance from a relevant care facility can be correlated with health outcomes. A study of NHS emergency calls found that a 10-km increase in straight-line distance from the nearest emergency facility led to a 1% increase in mortality rate. This is especially interesting in the case of a massive metropolitan area, where hospitals, chemical dependency facilities, and mental health facilities are so closely grouped. Does this pattern propagate on a smaller scale and how does access to public transportation affect health outcomes?

2.4. New York City Department of City Planning

The NYC Department of City Planning's "Capital Planning Explorer" tool, which is a web tool that allows a user to create a customizable map with all the New York facilities ranging from libraries to core infrastructure, had a notable influence on this project. It can feel daunting at first, providing a dizzying thirty-thousand possible markers, which blanket the map of New York City. After applying a filter, one can plainly observe the groupings and density of government services across the city. Extracting some of the geospatial data from the site, I was inspired to create a simpler visualization, which could be more user friendly and, when shown side-by-side with other demographic data, allow the user to draw relevant conclusions about facility proximity's effect on health outcomes.

3. Approach

3.1. Data Collection and Breadth

In an effort to determine the overall usefulness of environmental factors in predicting health outcomes, my research group chose to compare these factors with a broad swathe of outcomes. This will allow us to gain a deeper understanding of the effect of our factors on public health as a whole. We chose chronic illness, low birth weight, and external causes of premature mortality (overdose or firearm violence).

Following my survey of background material, I concluded that it would be beneficial to map both environmental and non-environmental (economic, racial/ethnic) factors in order to observe their respective correlations with health outcomes across NYC. We decided to select a standardized number of factors, at least one of which was linked with each of our areas of health outcomes. All of our data was publicly available.

Our study benefits from the breadth of factors and outcomes that we consider, allowing us to create a survey of possible factors affecting health outcomes in NYC. While the related works mention or demonstrate correlation between one or more outcomes and factors, there are none which draw from such a wide range of public health databases. Notably, our wide range of health outcomes is particularly helpful considering we are mapping such disparate factors. For instance, in the case of my paper, air quality may not be directly correlated with gun violence (although poor air quality can make you more susceptible to drug overdose), but it is closely related to outcomes of chronic illness. Similarly, the location of chemical dependency facilities are not particularly pertinent to low birth weight outcomes, but they are vitally connected to opiate overdose death rates.

3.2. Technology and Mapping

We chose to map over a single, contiguous geographic area in order for two reasons. The first is that it allowed us to map certain, related factors over one another, thus allowing us to quickly swap between them and more easily compare their effects. The second is that mapping over the same area of land (New York City) allows us to visually compare maps of our data. Many visualizations of drug and overdose data are presented at a state level, so our visualization is novel in its scope.

Lastly, we chose to use publicly available data and software in order to allow our visualization to be easily deployed and reproduced without refactoring of any kind. This is a public health project after all, and the intention of this visualization is to make it available to other researchers and policymakers who are seeking to draw conclusions which will help them narrow the disparity in health outcomes for minorities. Using our code as a boilerplate, one could easily map other factors and outcomes onto any geographical area and display them side by side for comparison

4. Implementation

4.1. Data Collection and Cleaning

In order to make our code simple and intuitive, we decided to use Python 3.10 and the well known data analysis library Pandas for our data collection and cleaning. For each map in our visualization, we required two key pieces of information: a table with the numerical data itself (e.g. Racial Makeup by Census Tract) imported as a pandas dataframe and a shapefile, which would associate our value table id's (e.g. Census Tract ID or County GeoID), with a geospatial area and location and outline on the map. We began by searching for tables first.

Our firearm mortality data came from the CDC Wide-ranging ONline Data for Epidemiologic Research (WONDER) Underlying Cause of Death system. By querying for all the ICD-10 causes of death codes associated with firearms, we were able to group total gun deaths and crude rates by county and racial makeup. But this is where we encountered our first problem. Data suppression and privacy is stringent in government health datasets. For groupings where total counts are below nine, data is suppressed, and for groupings where total counts are below 25, crude rates are considered unreliable. There are numerous reasons for this. The two largest are: (1)The CDC does not want researchers drawing conclusions about crude rate based on minimal data. (2) In cases where there are not many deaths in a certain geographical area, the CDC attempts to protect the privacy of the victims by not releasing demographic data. Because our data was grouped so specifically, and the majority of gun deaths in the United States are black males and white males, the data from all other groups was suppressed or unreliable. I made the decision to group data by county alone, thus losing demographic specificity, but gaining much more reliable data. It was downloaded as a text file and ingested using pandas' $.read_{c}sv()$, which is capable of reading text files if they are encoded in utf-8.

The next dataset was our overdose data, which was found on the New York Department of Health's EpiQuery database. Here, in order to observe trends on a smaller scale, the total counts and crude rates of overdose death over the entirety of NYC are grouped by UHF neighborhood, 42 zones designated by the United Hospital Fund, each involving multiple zip codes. As I explored more epidemiological and public health databases containing data from NYC, I began to realize how commonly these UHF neighborhoods were used to classify and group data. Where possible, I began to seek out neighborhood-grouped data so that factors and outcomes could be compared one-to-one. The overdose data was downloaded as a **.csv**, and easily ingested as well. After stumbling upon a dataset with a total count of opiate-related hospitalizations by county, I downloaded and cleaned the data in order to compare it to the fatal overdoses data.

I found a facilities database (FacDB) on the New York City Department of City Planning's website, which allowed me to search for thousands of facilities across NYC. After downloading all health related facilities as a **GeoJSON** file, I selected and grouped facilities from just three categories - hospital or clinic, chemical dependency facility, and mental health facility.

For air quality data, I sought out the New York City Environment and Health Data Portal, which provided multiple years of data on several pollutants in a single **csv** file. Examining the data year over year, it was clear that there had been a nearly linear decline in all pollutants over time. Thus, I decided to select data from only the most recent year and instead focus on the type of particulate matter. After examining the data available on all particulates, I decided to focus on three: nitrogen dioxide, ozone, and black carbon. These had the most variation over NYC (concentrated in different UHF neighborhoods) and I wanted to explore which health outcomes would be correlated with which particulate type.

Poverty rates and uninsured rates were also grouped by UHF neighborhoods on the EpiQuery database. Grouped under a single community health survey and downloaded as an excel file, I extracted the necessary columns in pandas.

Finally for racial makeup, I decided to consult the census level data. While information was available about racial makeup for UHF neighborhoods, I decided to graph by census tract, a relatively small zone containing between 1,200 and 8,000 people, so that I could more easily ascertain the density of racial/ethnic demographics. As a particularly divisive factor, I wanted as much information as possible when drawing conclusions based on racial makeup.

Now that all of my data was collected and imported as dataframes in pandas, I went about cleaning and normalizing it. I left null values in the tables, with a plan to graph them identically to zero values. For the few cells where data was suppressed, I replaced the string value with a numeric value of the mean between zero and the maximum suppressed value. For example, for firearm data, where all counts below nine were suppressed, I replaced values with 4.5. While this may not be accurate, as the values may not be normally distributed or even centered around the mean, I'm willing to replace them for the ease of graphing, as they will not be used to calculate an R or R^2 value of any kind. Additionally, because the suppressed values are so small when graphed, they will appear nearly identical to the null color value or zero value. For values given in total counts, not percentages, I calculated crude rates in order to control for population size. Crude rates in this case were on the scale of number of incidents per 100,000 people.

In regards to the geographic specificity of data, it may seem counter-intuitive to gather data which is grouped by different geographic metrics (borough, UHF neighborhood, census tract, single point). However, due the scarcity and sensitivity of public health data, it was unlikely that I would have been able to find both factors and outcomes at the same geographic scale. In addition, some variables are more sensitive to geographic scale and have larger variances over the same area. For these variables, I want to observe them with as much detail as possible. To keep my geographic bounds constant between variables, I would have been forced to downsample the data (calculating averages) until all of it was grouped by the largest geographic area (Borough). At such a large scale, I don't think this would have been very indicative of any trends in health outcomes. Given that I am evaluating my data visually, and not performing any linear regression on it, I was satisfied to gather data at different geographic scales.

The only downside of this approach is that for each geographic breakdown of NYC, I had to find and merge different shapefiles to the data as they had different codes and geometry. Thankfully, this was relatively simple as the New York City Open Data website provides shapefiles for NYC broken down by borough, UHF neighborhood, and census tract.

4.2. Identifying Relevant Factors

Now that our data was gathered and imported, I had to decide how to group our data in order to graph it. The logical conclusion was to graph every factor against every outcome, grouping related data or multiple years of the same data in singular graphs. I settled upon the following hierarchy. Each bullet is a graph and each sub-list are the layers within it.

4.2.1. Factors

- Air Quality
  1. Nitrogen Dioxide Levels
  2. Ozone Levels
  3. Black Carbon Levels

- Facilities
  1. Mental Health Facilities
  2. Hospitals and Clinics
  3. Chemical Dependency Facilities
- Poverty and Uninsured
  1. Poverty Rate
  2. Uninsured Rate
- Racial Makeup
  1. Hispanic %
  2. Black %
  3. AAPI %
  4. White %

4.2.2. Outcomes

- Drug Overdose Fatalities
  1. 2014
  2. 2015
  3. 2016
  4. 2017
- Opioid Overdose ER Visits
- Gun Violence Deaths
  1. 2018
  2. 2019
  3. 2020
  4. 2021

```python
uninsured.explore(
    m=factor_map, # Factor map to add layer to
    column="Poverty %", # make choropleth based on "Percentage" column
    tooltip= ["Name", "Poverty %"], # Information shown on hover
    popup=True, # Information tagged on click
    legend = False, # Display Legend
    style_kwds=dict(color="black", weight=0.2, opacity=0.4), # Outline style
    cmap="RdPu", # Set matplotlib colormap
    name = "Uninsured", # Name Layer
    missing_kwds=dict(color="white"), # Color of value null
    tiles="CartoDB positron", # Specify Background Tiles
    show=False # Show by default when opened
    )
```

Figure 1: Default settings used to create graphs

4.3. Folium Plots

Now that hierarchy had been decided and I knew the contents of each graph, I imported **geopandas**, a pandas extension for working with geographic data, and **folium**, a JavaScript based mapping library to superimpose and position our dataframes on the interactive world map generated by **LeafLet**, another open-source JavaScript library. For each graph I merged the relevant shapefile to the data by the geographic id, and created a Folium factor map by calling geopanda's $.explore()$ method, using a number of constant parameters, such as tiles and default popups. Some parameters were chosen graph-by-graph, varying based on what provided the most information without cluttering the map. Figure 1 shows the default mapping parameters for one call to $.explore()$. Note that I used the "CartoDB positron" tile settings, in order to create contrast between the maps and graphed colors. I repeatedly called the $.explore()$ function on all layers under each graph hierarchy until all of them were on the factor map. Finally, I added layer control, allowing the viewer to toggle between maps, and saved the maps as **.html** files.

4.4. Dash

Dash is an open-source framework for building visualizations in python. Built on top of **Plotly** and **Javascript**, it provides a lightweight framework for building dashboards without writing any **Javascript** whatsoever. Once our graphs have been saved as HTML files, it is relatively simple to embed them into an iframe, an HTML wrapper which allows us to insert HTML elements (like our maps) into other HTML code. We can do this by building a dash library with a hierarchy consisting of a main app file ($dash_app.py$) and navigation bar (*navbar.py*) which at runtime, are displayed with a default homepage. This default homepage is located in a "pages" folder which contains subsequent pages, each containing two maps. These pages are accessed through a dropdown menu in the navigation bar. There is one for each combination of factor graph and health outcome. The hierarchy appears as follows:

- dash_app.py
- navbar.py
- pages
  - home.py
  - airquality_overdose.py
  - facilities_overdose.py
  - povertyuninsured_overdose.py
  - racialmakeup_overdose.py
  - airquality_firearm.py
  - facilities_firearm.py
  - povertyuninsured_firearm.py
  - racialmakeup_firearm.py

5. Evaluation

In evaluating my visualization, I attempted to ascertain how helpful my graphs would be at drawing conclusions about the correlation and/or the effect of environmental factors on health outcomes. Let's walk through a series of graphs, beginning with Figure 2, which superimposes chemical dependency facilities on a choropleth map of overdose rates per 100,000 across UHF neighborhoods. Note that among the neighborhoods with the highest rates of overdose death (Ridgewood, Port Richmond, and Throgs Neck), only Port Richmond has any chemical dependency facilities whatsoever. Perhaps Ridgewood and Throgs Neck have extremely high rates of opioid overdose

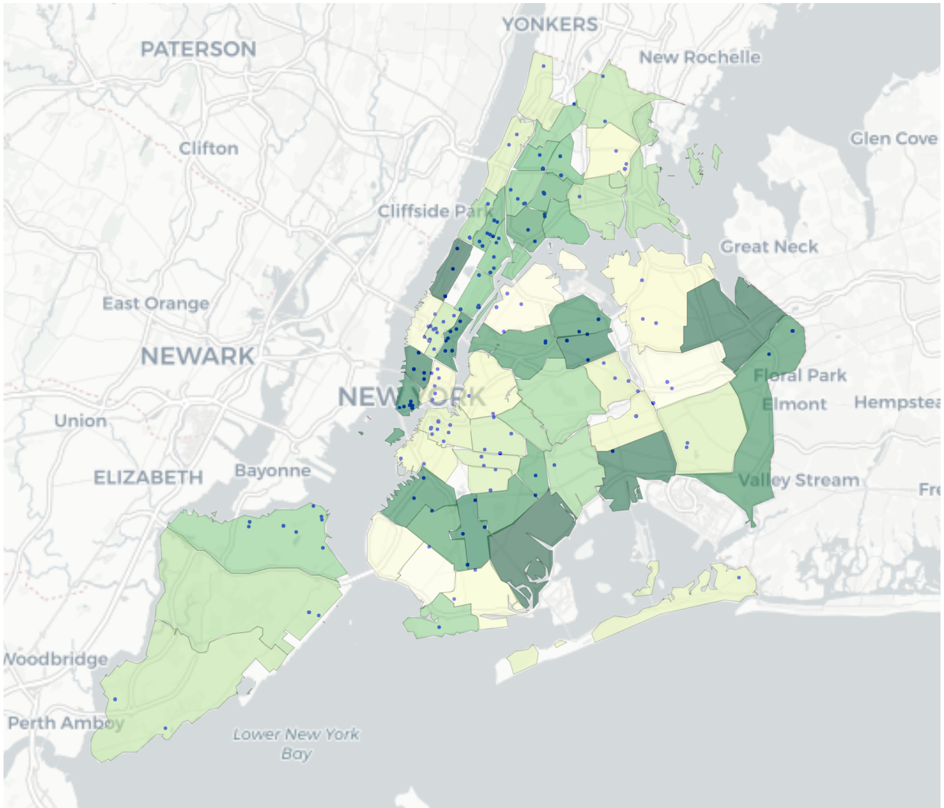


Figure 2: Chemical dependency facilities plotted over 2017 overdose death rates. (Higher yellow saturation indicates higher death rates)

deaths because their residents don't have access to any nearby chemical dependency facilities. For a public health researcher, this may indicate an unfulfilled need in the community. Perhaps this disparity in access is due to a socioeconomic gap between Port Richmond and Ridgewood/Throgs Neck. A lack of funding could account for a lack of services. However, examining Figure 3, we can clearly see that the percentage of people living under the poverty line in Port Richmond is actually higher than that of either Ridgewood or Throgs Neck.

Let's look at another graph, Figure 4. This is an overlay of hospital and clinic locations on a map of UHF neighborhoods where the darker the orange, the higher the percentage of uninsured individuals. Noticeably, areas with higher percentages of uninsured people have a higher density of hospitals and clinics than areas with lower percentages of uninsured people. Areas like the Upper West Side in Manhattan and South Beach in Staten Island have a very low percentage of uninsured individuals, 5% and 4% respectively, and only a few hospitals or clinics scattered throughout,

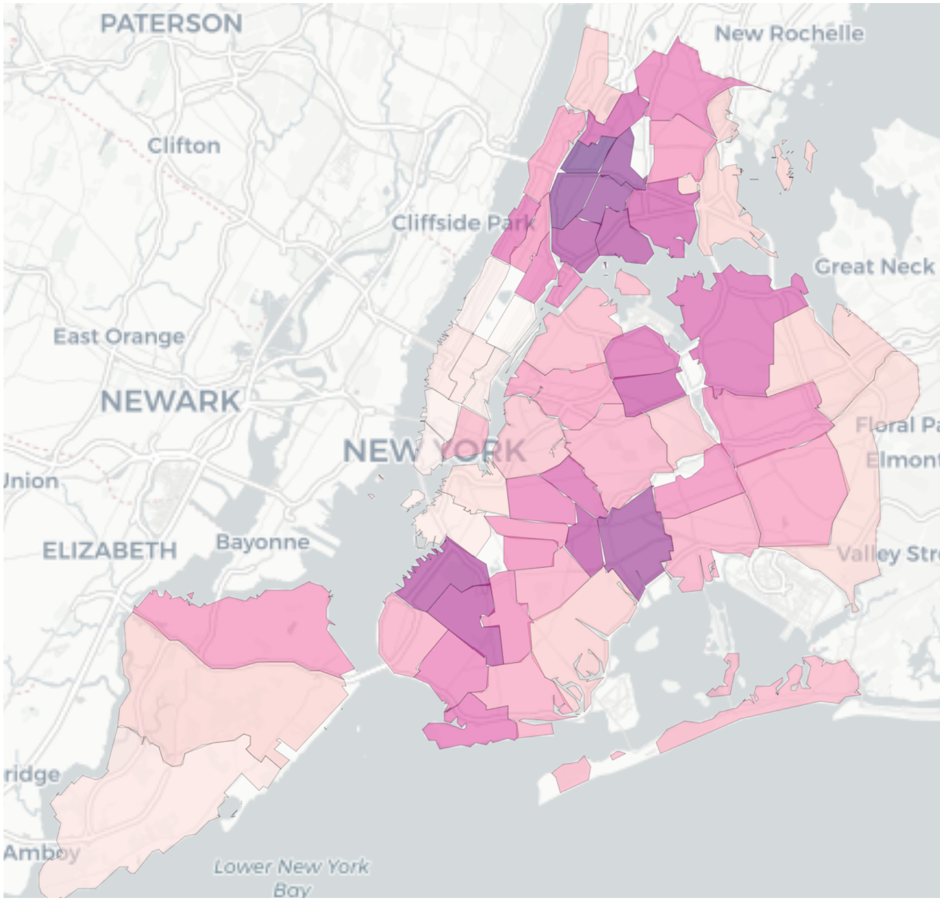


Figure 3: Percentage of people living under the poverty line (lighter color means higher percentage)

whereas neighborhoods like Sunset Park in Brooklyn and Mott Haven in the Bronx have much higher percentages, 22% and 14%, respectively, and have dense clusters of hospitals and clinics. Perhaps with uninsured patients, for profit hospitals and clinics are more likely to survive. Another explanation could be that for neighborhoods with lower uninsured rates, and likely more wealth, inhabitants are willing to travel further to go to the hospital, as the entirety of New York City is rather small and dense, geographically. Richer inhabitants may not view living extremely close to a hospital as favorable, and instead are content to travel a short distance to one. One final explanation is that these richer neighborhoods could contain private practices that aren't listed in the facilities dataset. However, that is unlikely because as far as I know, the dataset contains all registered Hospitals and clinics in New York City.

However, we can note one more similarity using our uninsured graph. Regarding Figure 5, the high rates of uninsured individuals appear to be closely correlated with the percentage of people of

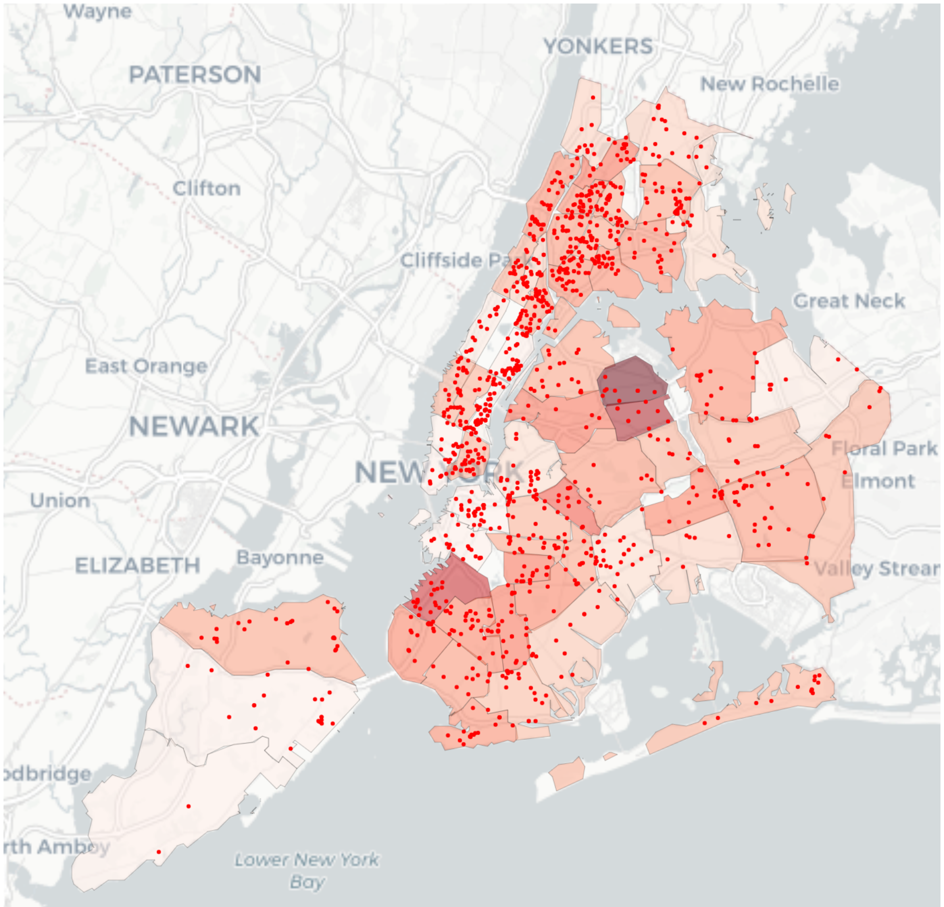


Figure 4: Hospital and clinic locations graphed over rates of uninsured individuals (Darker orange indicated higher uninsured rates

hispanic origin living in the neighborhood. This tracks with the realization that people of Hispanic and Latino origin have among the highest uninsured rate in the United States, at around 18%. What public health measures can we take in order to decrease this extremely high rate? Perhaps this graph could spur public policy leaders into action.

A final note is that while we were able to graph firearm death rates, it's extremely difficult to draw any conclusions based on borough-level data, when the rest of our data is more specific. Note that Figure 6 attempts to graph firearm death rates with Hospital and Clinic locations to very little result. The best correlation that we can produce is with the hospitalization rate per 100,000 due to opioids by borough. Looking at Figure 6 and 7 side-by-side, it's clear that in both, the Bronx has high rates. However, looking at the next-highest borough, Staten Island has more opioid related ER visits, while Brooklyn has a higher rate of firearm death.

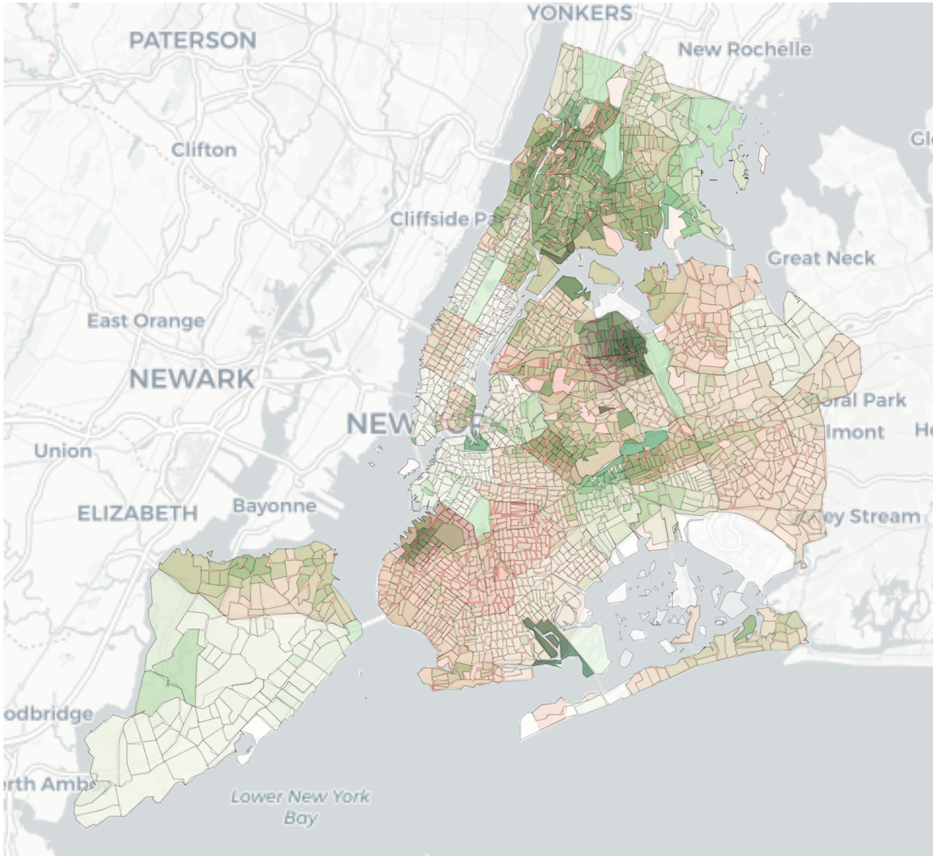


Figure 5: People of Hispanic/Latinx ethnic origin graphed over uninsured rates in Orange (Darker green implies higher percentage)

6. Conclusions and Future Work

Reflecting on my evaluations, I feel as though I was capable of drawing relevant conclusions from my visualizations. In this respect, I feel as though my research was successful. The dash implementation is free, accessible, and available through the github repository, which can be found here: *(https://github.com/lpang2143/COS398IW)*. These are early development images, but with User Interface, and graphs side-by-side, the web-page looks roughly like Figure 8, with a dropdown menu to toggle between graphs. However, I don't feel as though I was particularly successful in achieving my second goal, providing an evaluation of a variety of factors' contribution to our health outcomes. While one can draw conclusions about correlation between many of the factors and drug overdose death rates, as mentioned above, I feel as though the Borough-wide data on premature gun deaths prevents us from drawing any real conclusions about the effect of our graphed factors

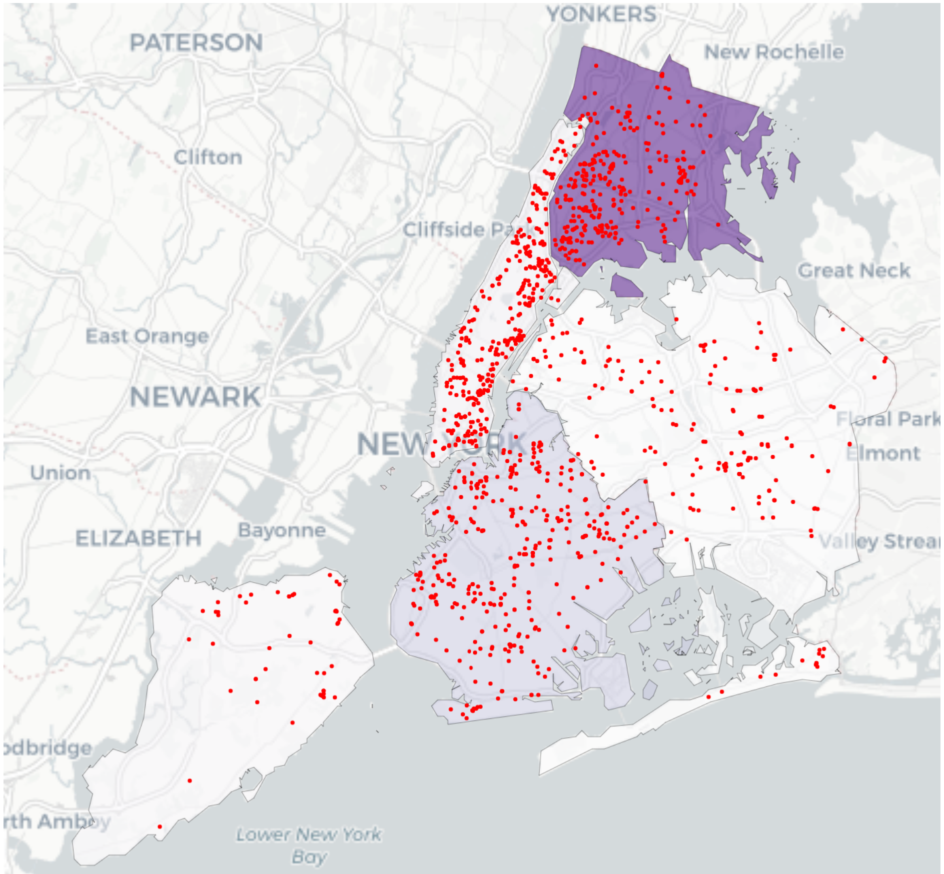


Figure 6: Gun violence death rates graphed over Hospital and clinic locations (note that darker purple indicates a higher rate of premature firearm deaths per 100,000).

on premature gun deaths. In order to draw concrete conclusions about correlation and possible causation, we would need more data on a neighborhood or even census tract scale.

If this data were available, I think that future research could consist of a linear regression model to predict these health outcomes. With a larger set of health outcome data, labeled with demographics (e.g. grouped by race or income level) we would be able to predict the likelihood of a set of health outcomes (premature gun/drug deaths) for an individual in New York City based on a wide variety of social factors. This will provide a useful, centralized reference for those trying to understand how societal factors such as, healthcare access, transportation access, and other equity factors affect the average New Yorker. It could be used as a tool by policymakers to inform data-driven decisions on how to uplift minority populations and equitably allocate healthcare resources within New York City.

A final note is that I feel as though my visualization would have benefited greatly from an

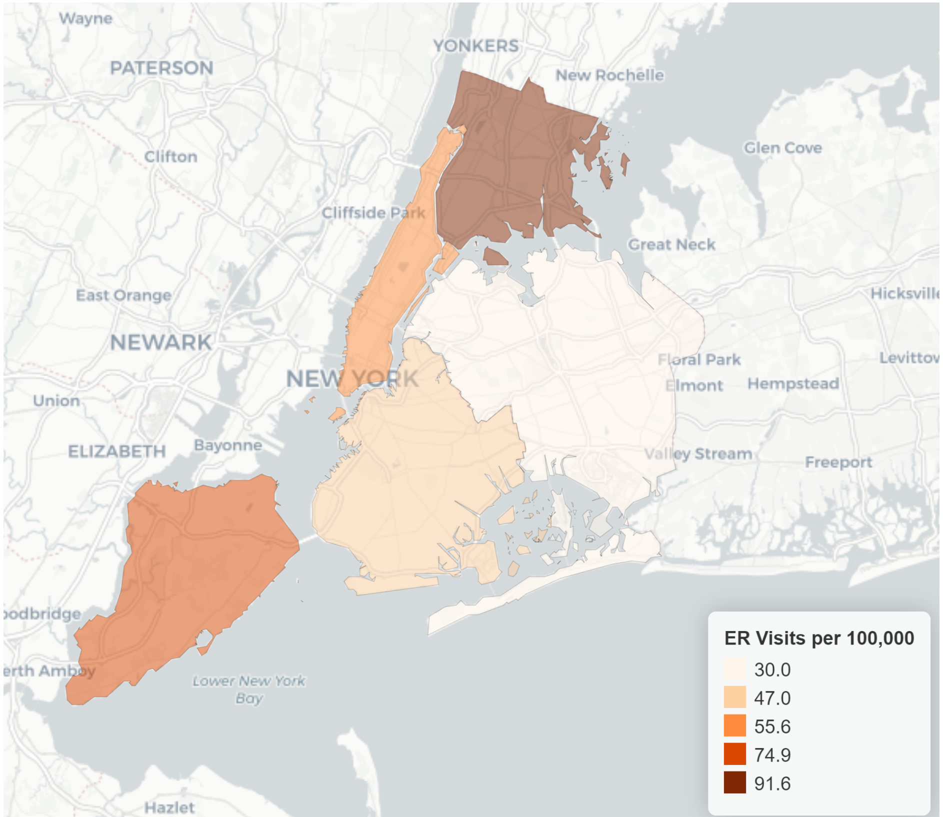


Figure 7: Hospitalization rates involving opioid overdose per 100,000.

additional layer containing population density. This would have allowed us to rationalize the density of facilities. For instance, if there are relatively few hospitals in South Beach, this could be due to the fact that South Beach has a much smaller population than places like the Upper West Side.

7. References

"2018 Community Health Profiles - NYC Health." Accessed May 1, 2023. https://www.nyc.gov/site/doh/data/data-publications/profiles-2018-community-health-profiles.page.

"2020 Census -DCP." Accessed May 1, 2023. https://www.nyc.gov/site/planning/planning-level/nyc-population/2020-census.page.

"Borough Boundaries | NYC Open Data." Accessed May 1, 2023. https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm.

Branch, Breauna, and Douglas Conway. "Health Insurance Coverage by Race and Hispanic Origin: 2021," 2021.

Environment Health Data Portal. "Air Quality Data for NYC." Accessed May 1, 2023. https://a816-dohbesp.nyc.gov/IndicatorPublic/beta/data-explorer/air-quality/.

Environment Health Data Portal. "Explore Data on How Environments Shape Health in New York City's Neighborhoods." Accessed May 1, 2023. https://a816-dohbesp.nyc.gov/IndicatorPublic/beta/.

"GIS Data Files For Download." Accessed May 1, 2023. https://www.nyc.gov/site/doh/data/data-sets/maps-gis-data-files-for-download.page.

Gramlich, John. "What the Data Says about Gun Deaths in the U.S." Pew Research Center (blog). Accessed May 1, 2023. https://www.pewresearch.org/short-reads/2023/04/26/what-the-data-says-about-gun-deaths-in-the-u-s/.

"Health Consequences of Air Pollution on Populations." Accessed May 1, 2023. https://www.who.int/news/item/15-11-2019-what-are-health-consequences-of-air-pollution-on-populations.

"How to Create a Multipage Dash App | by Michael McManus | Medium." Accessed May 1, 2023. https://medium.com
/@mcmanus_data_works/how-to-create-a-multipage-dash-app-261a8699ac3f.

Institute for Health Metrics and Evaluation. "On Gun Violence, the United States Is an Outlier," March 25, 2021. https://www.healthdata.org/acting-data/gun-violence-united-states-outlier.

Kyaagba, Shachia. "Integrating Folium with Dash." Medium, September 8, 2018. https://medium.com/

@shachiakyaagba_41915/integrating-folium-with-dash-5338604e7c56.

National Institute on Drug Abuse. “Drug Overdose Death Rates,” February 9, 2023. https://nida.nih.gov /research-topics/trends-statistics/overdose-death-rates.

“New York State Opioid Data Dashboard.” Accessed May 1, 2023. https://webbi1.health.ny.gov/SASStoredProcess/g

Nicholl, Jon, James West, Steve Goodacre, and Janette Turner. “The Relationship between Distance to Hospital and Patient Mortality in Emergencies: An Observational Study.” //Emergency Medicine Journal: EMJ 24, no. 9 (September 2007): 665–68. https://doi.org/10.1136/emj.2007.047654.

NYC Capital Planning Explorer. “

NYC Capital Planning Explorer.” Accessed May 1, 2023. https://capitalplanning.nyc.gov.

NYC Open Data. “Borough Boundaries.” Accessed May 1, 2023. https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm.

Tomar, Anmol. “Dash for Beginners: Create Interactive // Python Dashboards.” Medium, November 10, 2022. https://towardsdatascience.com/dash-for-beginners-create-interactive-python-dashboards-338bfcb6ffa4.

“Underlying Cause of Death, 2018-2021, Single Race R//equest Form.” Accessed May 1, 2023. https://wonder.cdc.gov/controller/datarequest/D158.

Unwin, Antony. “Why Is Data Visualization Important? What Is Important in Data Visualization?” Harvard Data Science Review 2, no. 1 (January 31, 2020). https://doi.org/10.1162/99608f92.8ae4d525.

“U.S. Census Bureau QuickFacts: Chicago City, Illinois; United States.” Accessed May 1, 2023. https://www.census.gov/quickfacts/fact/table/chicagocityillinois,US/POP060220.

“U.S. Census Bureau QuickFacts: New York City, New York.” Accessed May 1, 2023. https://www.census.gov/quick

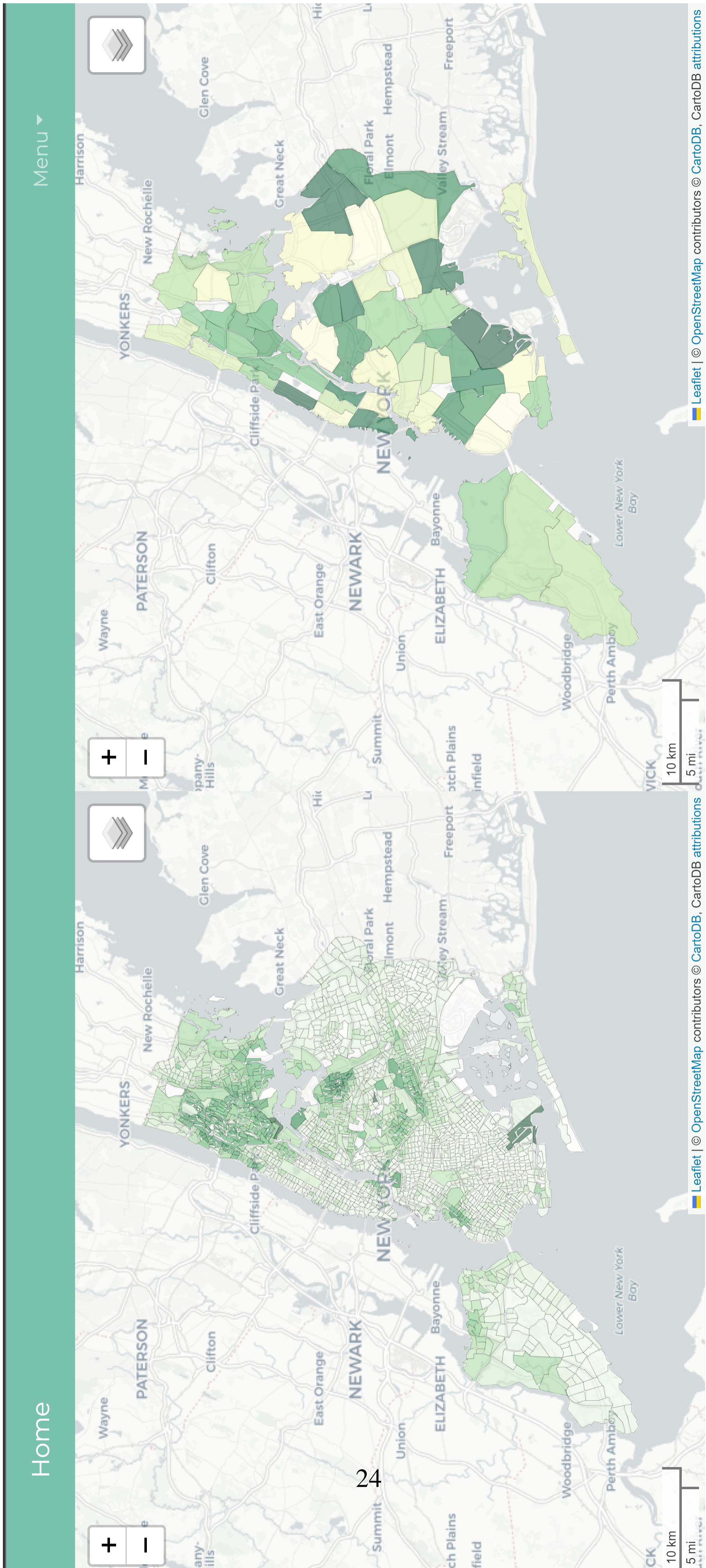
8. Acknowledgements

A special thank you to my advisers, Professor Troyanskaya and Tavis, as well as my team Aria and Herve. Without you all, I never would have created this project.

9. Appendices

Feel free to evaluation the visualization yourself. I'm including an HTML graph in my zip file which contains all the layers I've graphed.

Figure 7 below. It would not fit on any other page.


Home
Menu
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions
10 km
5 mi
YONKERS
PATERSON
NEWARK
NEW YORK
ELIZABETH
New Rochelle
Great Neck
Glen Cove
Hempstead
Freeport
Valley Stream
Floral Park
Elmont
Cliffside Park
Clifton
East Orange
Bayonne
Union
Summit
Woodbridge
Perth Amboy
Wayne
Harrison
Lower New York Bay