Comparative Analysis of New York City and Detroit Water Quality Through Population and Income

Ashley Grinstead & Ian Hinds
GEOG 176C: GIS Applications
9 June 2021

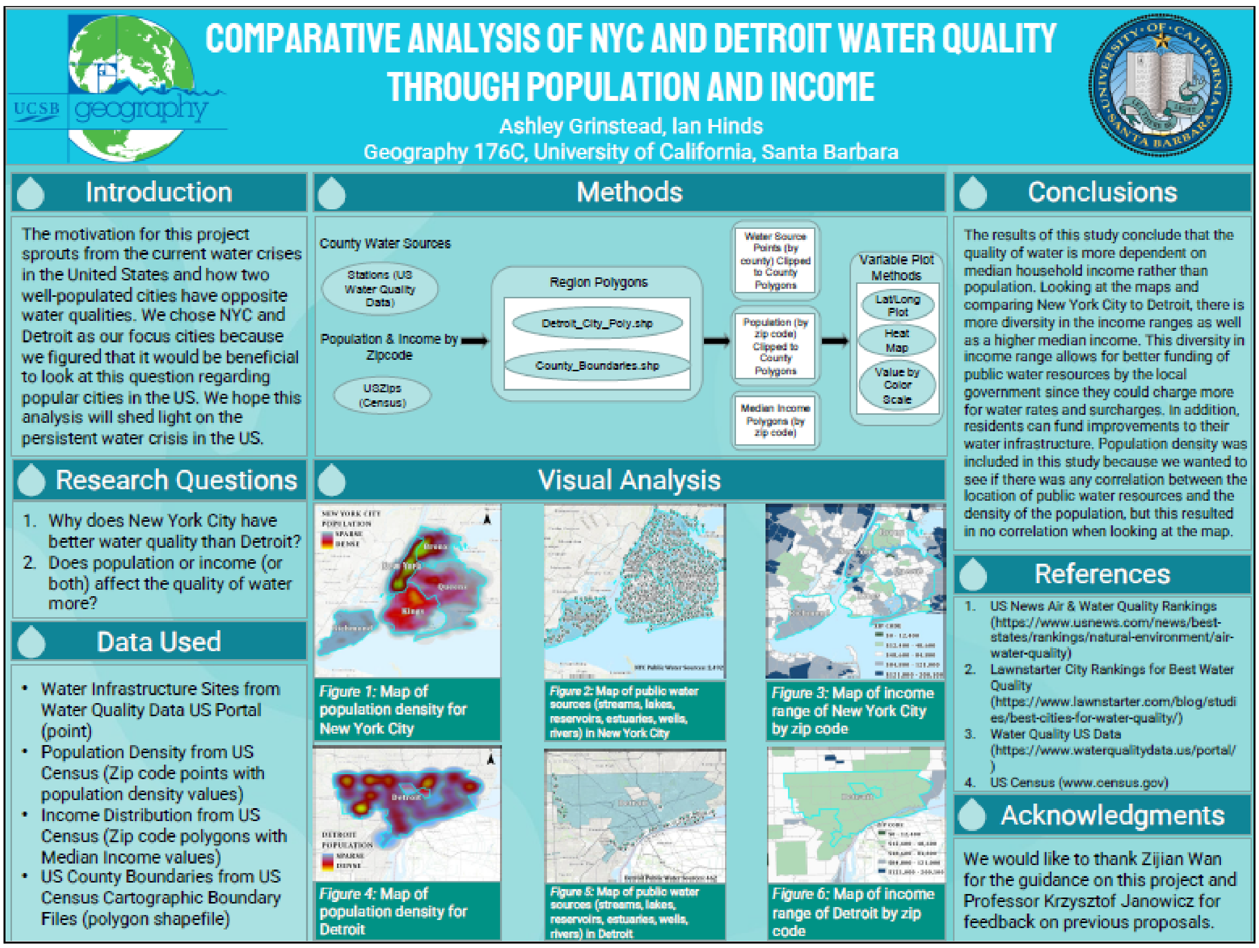


Figure 1: *Poster submitted for spatial@ucsb.global2021: Spatial Data Science for a Sustainable Future*

1. Research Questions

Why does New York City have better water quality than Detroit?

Does population or income (or both) affect the quality of water more?

2. Motivation

The motivation for this project sprouted out of researching a new project idea since we were unable to execute our previous project idea. Our previous idea was trying to correlate how water quality affects spikes in COVID but we realized that there were too many other factors to consider and that we would not be able to come up with a clear conclusion. With that in mind, we were more interested in the current water crises in the United States than COVID so we decided to continue to research an idea that we were able to do for this project that related to water. We were inspired by the information we found through US News and Lawnstarter, which had articles that ranked US cities and states by water quality and water infrastructure. It was interesting to see how there is a drastic difference between water quality in two of the most popular cities in the United States, New York City and Detroit. In 2019, US News claimed that the state of New York had a drinking water quality rank of sixteen and Michigan had a drinking water quality rank of thirty-five. The metric that was used measured the number of violation points against the public water systems per one thousand residents served in each state. In another study done by Lawnstarter in 2021, Detroit had a higher water infrastructure vulnerability than New York City, meaning that water infrastructure in Detroit is in need of an upgrade. The infrastructure vulnerability was composed of the share of homes lacking plumbing or kitchen facilities and the share of homes with sewage disposal breakdowns in the last three months.

We decided that we would focus on these cities for our project, and analyzing water infrastructure, median household income, and population will hopefully tell us why these cities' overall water quality is ranked differently on these websites. We hope this analysis will shed light on the persistent water crisis in the United States, which will hopefully inspire people to fight for clean water for everyone.

3. Data

Water Infrastructure: EPA Drinking Water Quality Violations

To gain an initial perspective on how each city has maintained their water quality over a period of 30 years (1990-2020), data was observed through the EPA Consumer Confidence Report portal. Clearly organized by city, county, and state, these values were easy to locate. The values were observed by overall count by city within the described time period.

New York City had a reported total of 24 violations since 1990. Detroit had a reported total of 7 since 1990. This count seemed odd at first, but when later compared to the amount of physical infrastructure and population, these numbers exposed a troubling ratio. Overall, Detroit had more violations relative to their population than New York City. The offset ratio indicates that Detroit has less successful infrastructure in place, and matches the ranking diversity indicated by Lawnstarter.

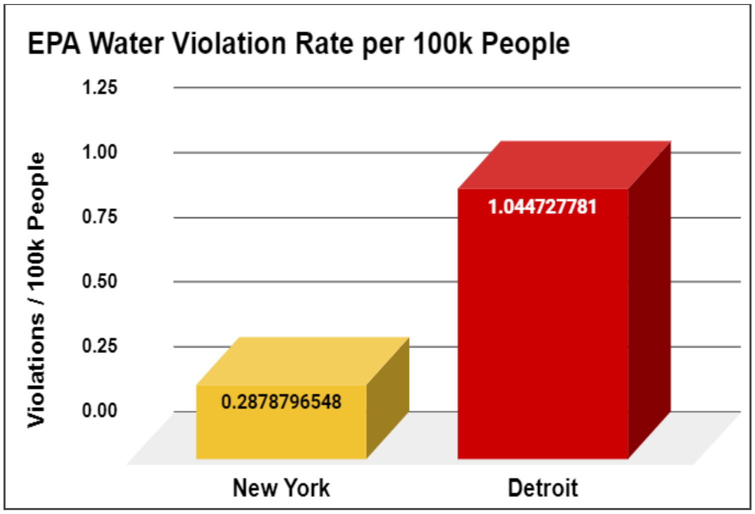


Figure 2: *Visual aid showing EPA's water violation rate per 100,000 people in New York and Detroit*

Water Infrastructure: Physical Resource Sites (points)

Data was downloaded from the US Water Quality Data Portal in the form of comma-separated value files organized by US County. In total, there were 6 counties to analyze for the project. Detroit City is located in Wayne County Michigan and its csv file required an additional data export within ArcGIS to focus specifically on Detroit and not the surrounding county. New York City includes 5 individual boroughs, also in the form of counties. These 5 counties included Richmond ("Staten Island" area), Kings ("Brooklyn" area), New York ("Manhattan" area), "Queens" (Queens area), "Bronx" (The Bronx area).

Each csv file included site names, types, latitude, longitude, and identification numbers. Although each of these values were important to organization and understanding, the lat and long values were used for our spatial analysis. Once plotted, the points had to be restricted to the respective city boundaries. To extract the relevant values, relevant site types and locations had to be filtered. A site was considered "relevant" if it was a freshwater source, or city-owned well within the perimeter of the city or county. These sources included lakes, rivers, streams, wells, and estuaries to mimic the natural relationship between human consumption and natural resources. Other site types included atmospheric and oceanic sources, but were not deemed relevant due to the absence of consistent freshwater capability. These other source types were also not featured in the data because they did not relate to city-owned infrastructure focused to provide clean drinking water.

Population Density from US Census (Zip code points with population density values)

Population density was used as a variable to better understand the differences of Detroit and New York City on a social scale. More water resources are needed to sustain a massive population such as New York City's. It was an immediate understanding that New York City needed more infrastructure than Detroit, but by plotting the densities by zip code, a spatial correlation existed between water resource sites and the core distribution of people depending on those resources.

Among each of the regions, New York county in New York City had the most densely populated area, followed by Detroit, Bronx, Kings, Queens, and Richmond. When combined, New York City dominates over Detroit by a ratio of 8.3 million to 670 thousand people.

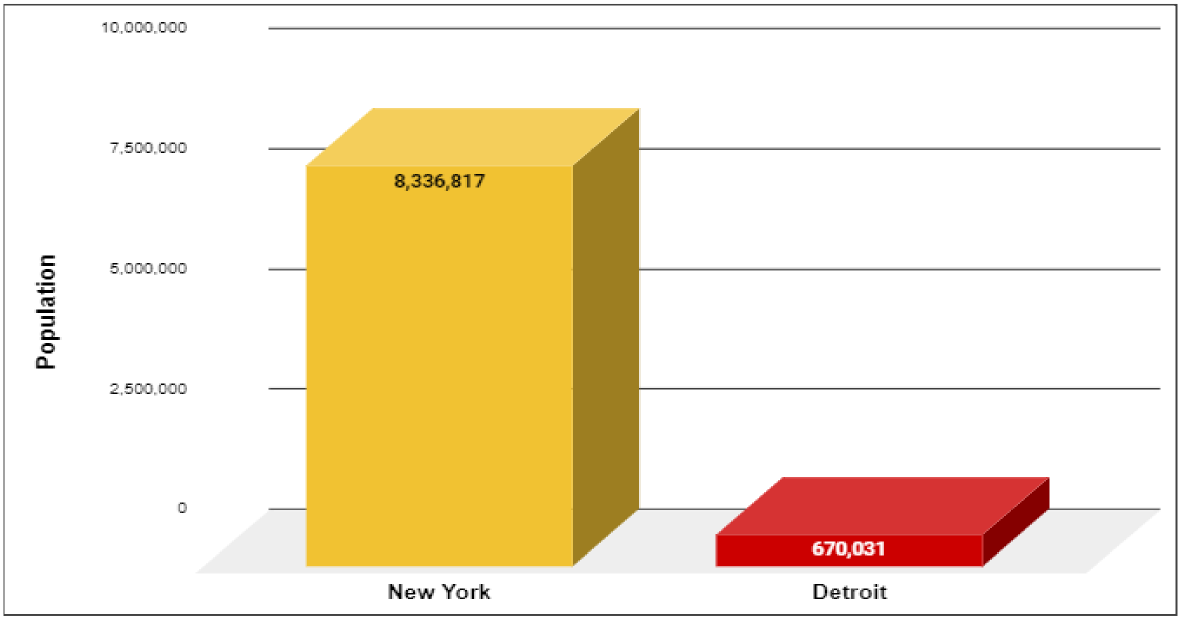


Figure 3: *Visualization of New York City and Detroit's population*

In order to plot the population density values, a csv file was obtained by the US Census, outlining US population by zip code. This csv file had to be filtered by the relevant study regions, meaning state, county, and city designations had to be narrowed down. Once filtered and districted properly, the values per zip code could be plotted using a density function. The densities provided an understanding of how population compares to physical infrastructure distribution spatially.

Income Distribution from US Census (Zip code polygons with Median Income values)

The second main variable observed in the study was the distribution of median household income in the respective study regions. With population density representing the social scale, median income provides an economic scale to our study. With a greater number of people, New York City is able to generate more money into the city's infrastructure. Additionally, income distribution could potentially indicate where infrastructure is best due to local funding.

To plot this data, average median household income was calculated for each zip code in the US. To obtain the best visual, a pre-existing map of the data was provided by esri via the "living atlas" in ArcGIS. Once extracted, this data had to be restricted to the specific study regions to best understand the variable's influence on water infrastructure.

By creating a spatial comparison between known physical infrastructure points and income distribution, it was possible to draw conclusions on how this variable is more significant than that of population. Although income and water sites did not expose a direct geographical correlation, it was still extremely important to understand how the respective cities maintain their water infrastructure to provide clean and enjoyable drinking water to their populations.

US County Boundaries from US Census Cartographic Boundary Files (polygon shapefile)

The key factor to districting income, population, and water resource was the regional boundary data and polygons provided by the US Census. These boundaries were plotted before every other variable to clearly decipher and outline our spatial regions of study. Individual boundaries had to be obtained for each Detroit, Richmond, Kings, Queens, New York, and Bronx.

Initially, these boundaries were simple to extract, as they were available by county. However, Detroit is the only area not listed as a county, and it's boundary had to be extracted from the city's open data platform. From this point onward, each forementioned variable was restricted to the available boundaries to provide accurate visualization and understanding of our data.

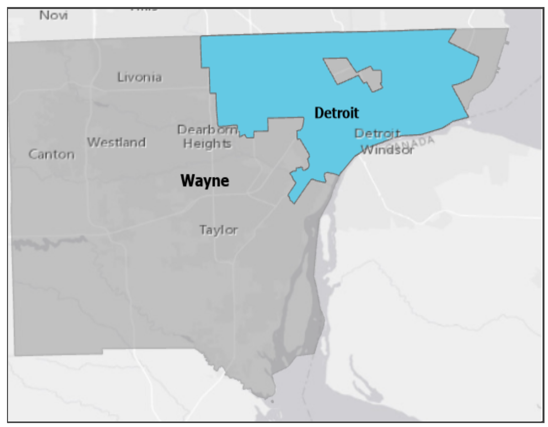


Figure 4 (top left): *Map of New York City's five county boundaries; New York, Bronx, Queens, Kings, and Richmond*
Figure 5 (top right): *Map of Detroit's county boundary*

4. Methods

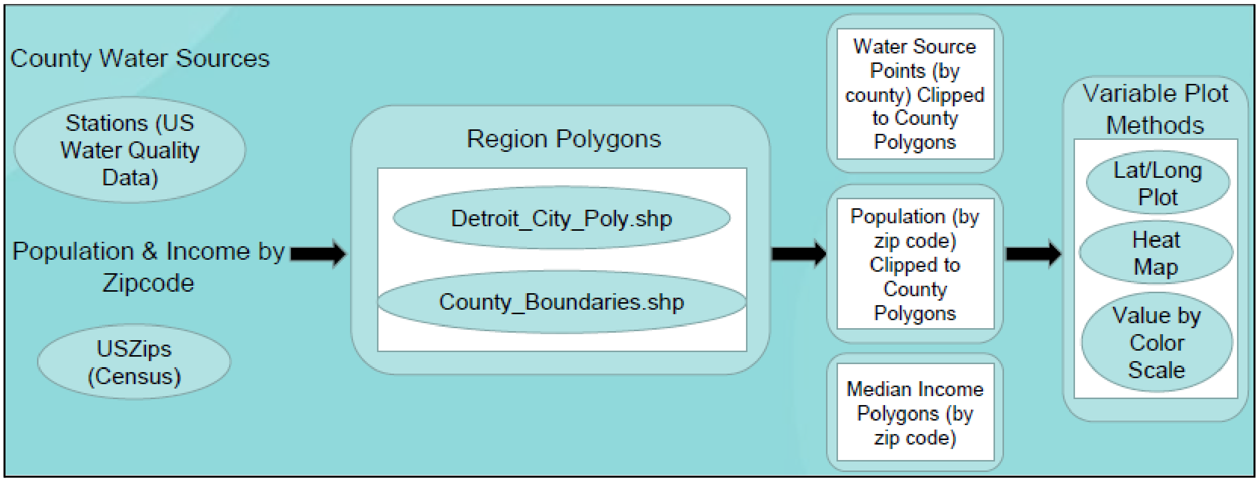


Figure 6: *Workflow chart for the methods used to produce the maps of New York City and Detroit*

County Water Sources:

The csv files provided latitude and longitude coordinates for each water resource site. To visualize these points in ArcGIS, the coordinates were geo-referenced to the WGS 1984 coordinate system, then simply plotted onto the map layout. Although these point features were simple and easy to understand, they had to be altered to best fit the regions of study and perspective of the viewer (of the poster).

Using Symbology, each point was given the same symbol and font size, to best fit the parameters of the map. The esri teardrop symbol was selected as it narrows at its base to indicate specific location. The teardrop was chosen over the dot feature symbol, because the dots grouped together too densely and did not allow the viewer to see the topography surrounding the location.

Once symbolized to preference, the regional polygons were layered below the points, to help visualize the distribution of resources throughout the city's governance. However, simple layering was not enough to match the realistic parameters of the study. To be realistic to our study, the irrelevant resources were removed (i.e. atmospheric and oceanic) as they did not represent the city's immediate water infrastructure of freshwater. Once the relevant features were summarized, the clip tool was used to extract only the resources that lay within the boundary of the city and county. Outlying points were removed from the map, and all water resource points were simplified and easy to understand when compared to the other variables of our study.

Population by Zip Code:

______The csv files provided population values by zip code and state. To organize the data, the values were first plotted as point features with WGS 1984 and ArGIS location finder as references for US zip codes. Once plotted, these points were then clipped to fit the county or city boundary polygon, similar to the water resource points.

Instead of leaving the points as dots, which did not represent corresponding population values, the symbology function was used in ArcGIS to aggregate the point features via heat map.

The heat map used was reliant on population values, and indicated relative density compared to nearby zip codes. The color scheme used was strategically selected for visual aesthetic. The yellow extreme in the visuals are the higher density values, whereas the blue color extreme are the lower density values.

Once simplified, the density values also had to be viewed at a specific scale to represent spatial distribution across the entire region of study. Detroit's map scale was 1:400,000 and New York City was 1:520,000, these scales varied because of geometric area differences. Detroit has a smaller area than New York City, and required a smaller scale to best visualize the population density variable. New York City's larger scale encompassed all 5 of the boroughs within the city limits, to best understand the spatial variations across the entire region.

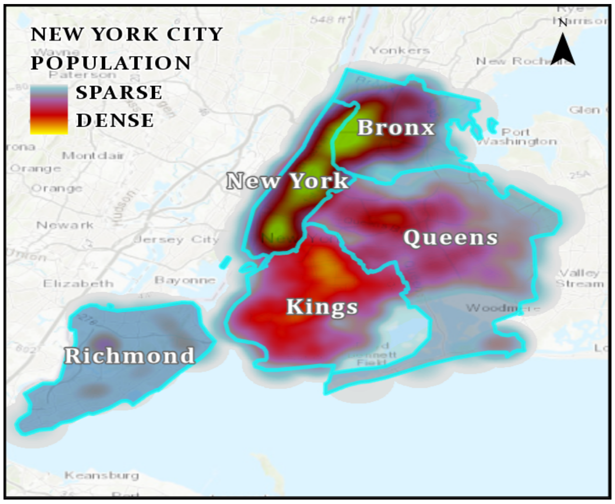
Median Household Income by Zip Code:

Straying away from the original csv, and pointing to polygon practice of the previous two variables, the living atlas was used in ArcGIS to explore a professional geospatial product and apply it to our study. The 2020 US Median Household Income map provided by esri clearly organized a map depicting income distribution by state, county, city, and zip-code.

As a fixed shapefile, and not editable in ArcGIS, a different visualization focus was required to see the impacts of the income distribution variable. The layer itself was overlaid by the respective regional polygons. The polygons were then altered to 100% transparency, and selected by attribute, being the name of the city or county. The select by attribute tool helped delineate the boundaries of the study regions throughout each map, but was especially important for this variable plot as the median household incomes of surrounding zip codes were initially blending with the specific regions we needed to understand.

Utilizing the living atlas (the last source of data and tools discovered in the length of our project), provided a realistic application of professional products to best answer research questions and find conclusions. Although it would've been more advantageous to plot these values as aesthetically as ESRI, it is important to understand the resources available to the team to achieve the goals of study.

5. Results



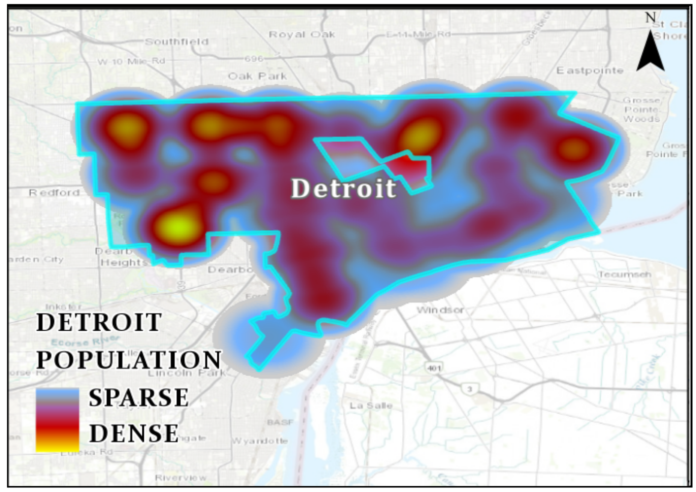
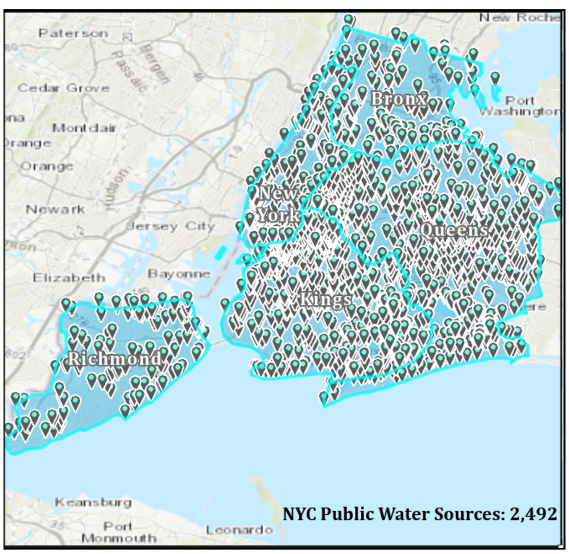


Figure 7 *(top left)*: *Map of the population density in New York City. Blue means that the population is more sparse and red and yellow represents a more densely populated area.*

Figure 8 *(top right)*: *Map of the population density in Detroit. Blue means that the population is more sparse and red and yellow represents a more densely populated area.*



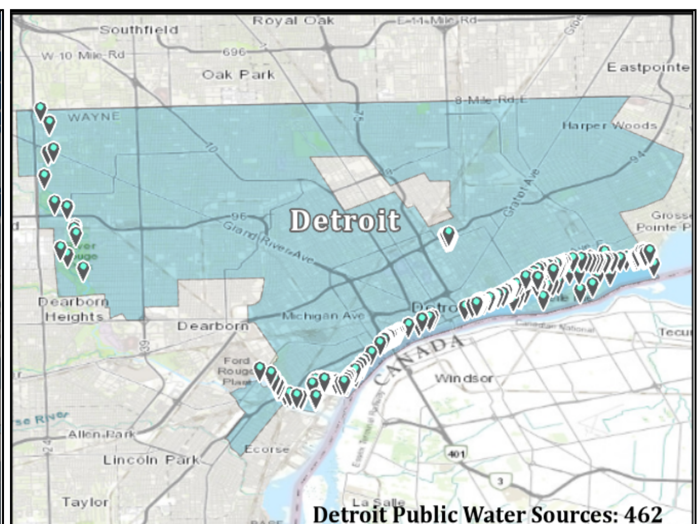
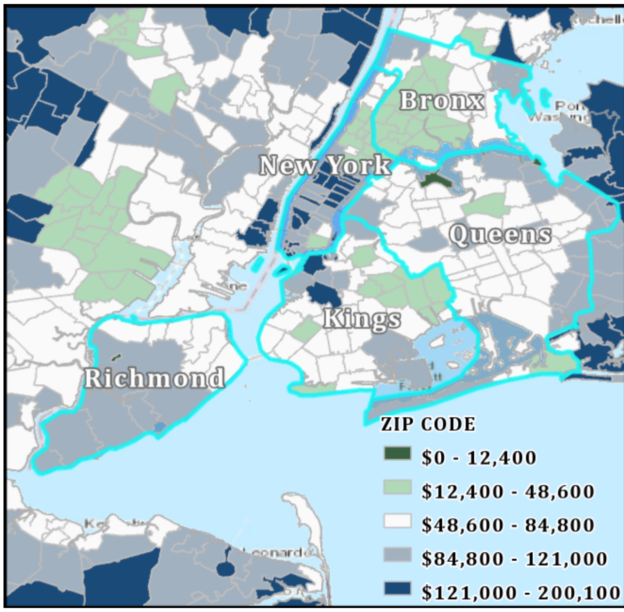


Figure 9 *(top left)*: *Map of the public water sources in New York City. The public water sources that were mapped include streams, lakes, reservoirs, estuaries, wells, rivers.*

Figure 10 *(top right): Map of the public water sources in Detroit. The public water sources that were mapped include streams, lakes, reservoirs, estuaries, wells, rivers.*



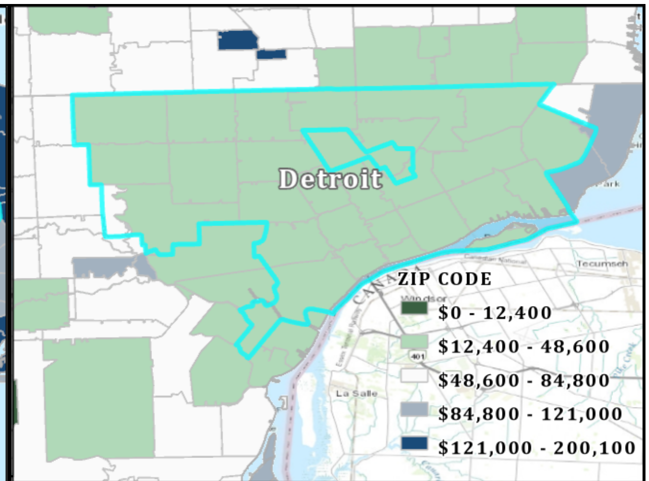


Figure 11 *(top left): Map of income range of New York City by zip code.*

Figure 12 *(top right): Map of income range of Detroit by zip code.*

The results of this study conclude that the quality of water is more dependent on diversities of household income rather than population. We came to this conclusion after performing a visual analysis using the maps that we created above. When comparing Figure 11 to Figure 12, there is more diversity in the income ranges in New York City, meaning that there is a higher median income as well. This diversity in income range allows for better funding of public water resources by the local government since they could charge more for water rates and surcharges. In addition, residents can fund improvements to their water infrastructure. When comparing the population densities of both cities, we found that there was no correlation between the location of public water resources and the density of the population. However, when looking at Figure 8 and Figure 10, we found it interesting that Detroit has most of its water resources located on the opposite end of where the population is most dense. We are unsure of why this is, but it may relate to geographic features that were not originally variables for the study. For example, the majority of points in Detroit line the Detroit River and Lake St. Clair, meaning that the two major water bodies act as the key water resource infrastructure for the people of the city.

6. Challenges

Because we only had ten weeks to complete this project, we came across a multitude of challenges. In the beginning of our project, we had a difficult time coming up with an idea that we both found interesting and would motivate us to keep up with the project. We spent a lot of time brainstorming and not enough time to narrow down what exactly we wanted to do. When we finally came up with the idea of finding a relationship between COVID and poor water infrastructure, we had trouble trying to make sense of how much data we needed to gather as well as how we would present a conclusion, since there were many other factors that needed to be considered. After that idea proved to be unsuccessful, we had to shift our project idea entirely and focus on something else, which meant we needed to collect new data, come up with new research questions, and basically redo our entire proposal.

Additionally, we were only able to conduct a visual analysis of our results because of the limited amount of time that we had to complete this project. We were planning on using a linear regression to see the statistical significance and correlation between population density and water resources and median household income and water resources, but we were unable to research how to conduct a linear regression and get the help that we needed. There were many suggestions during our poster session that referred back to doing a linear regression, but with other classes and finals coming up, we just had no time to do it.

7. Appendix

8. Bibliography

2020 USA Median Household Income Distribution by Zip. arcgis.com. (n.d.). https://www.arcgis.com/home/item.html?id=20a60423d37c49ba9253526859ba93e1

2021's Best Cities for Water Quality. Lawnstarter. (2021, May 13). https://www.lawnstarter.com/blog/studies/best-cities-for-water-quality

Bureau, U. S. C. (2020, May 18). *Cartographic Boundary Files - Shapefile*. The United States Census Bureau. http://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html

Environmental Protection Agency. (n.d.). *EPA Drinking Water Quality Violations*. EPA. https://ofmpub.epa.gov/apex/safewater/f?p=136%3A102%3A%3A%3A%3A%3A%3A

New York City *Median Income Distribution*. Median Incomes. (n.d.). https://data.cccnewyork.org/data/map/66/median-incomes#66/39/6/107/62/a/4

New York City Open Data. (n.d.). https://data.cityofnewyork.us/browse?Dataset-Information_Agency=Department%2Bof%2BEnvironmental%2BProtection%2B%28DEP%29

U.S. Census Bureau QuickFacts: Detroit city, Michigan. (n.d.). https://www.census.gov/quickfacts/detroitcitymichigan

U.S. Census Bureau QuickFacts: New York city, New York. (n.d.). https://www.census.gov/quickfacts/newyorkcitynewyork

U.S. News & World Report. (n.d.). *States With the Best Air and Water Quality in the U.S.* U.S. News & World Report. https://www.usnews.com/news/best-states/rankings/natural-environment/air-water-quality

Water Quality Data. (n.d.). http://www.waterqualitydata.us/portal/