# Hyper-Localized Analysis of New York City's Lead Poisoning Epidemic

## 205101

### Virtual NLC 2020-2021

## Introduction

Brooklyn's lead poisoning is (7) times higher than Manhattan's and 12 times higher than Staten Island's, showing the ineffectiveness of New York City's decades-long attempts to identify and solve factors relating to disproportionate lead-poisoning. New York City's lead poisoning management team needs comprehensive, localized data analysis that provides a geographical and personal profile for individuals who are at particular risk of lead poisoning. Our analysis seeks to magnify each New York City borough by analyzing which neighborhoods are disproportionately impacted, and identifying the root causes of such disproportionality. By specifically narrowing down high-risk neighborhoods and individuals, New York City can save time and resources while canvassing homes for lead poisoning more effectively. All in all, the holistic goal of this project is to facilitate New York City's campaign to eradicate lead poisoning among minors by identifying the most high-risk populations, based on several corroborating factors

## Purpose

Despite New York City's 86% overall drop in minor-related lead exposure from 2005. Exposure to these high levels of lead poisoning can lead to life-long implications such as permanent memory loss, tremors, and hallucinations. Over the past two (2) decades, New York City has dramatically decreased its lead poisoning epidemic; however contemporary studies show that the rate at which lead poisoning is decreasing has stagnated, severely affecting upwards of eleven-thousand children. As New York City's comptroller, Scott Stringer, has argued that "any lead poisoning of our children must be treated as a five-alarm fire, but the city isn't utilizing basic tools," New York City's lead poisoning management team has been unable to effectively extinguish lead poisoning, as in 2019 alone, 9,099 residences went unchecked for lead poisoning. A breadth of factors contributes to New York City's lead poisoning—old housing, industrial waste, and lead-ridden soil. To fight this multifactorial epidemic, the city has no room for overlooking critical residences in their campaign to eliminate lead poisoning. New York City's canvassing protocols can be vastly improved by clearly defining the most pertinent sections of the city. By analyzing New York City's children < 5 lead poisoning data with corroborating factors (income, poverty rate) and on a scale as small as neighborhoods, we seek to elucidate the high for correlations within specific boroughs. Additionally, external factors (income, poverty rate, water consumption, etc.) can be appended to analyze the damage of certain factors on elevated lead blood levels within children under the age of six.

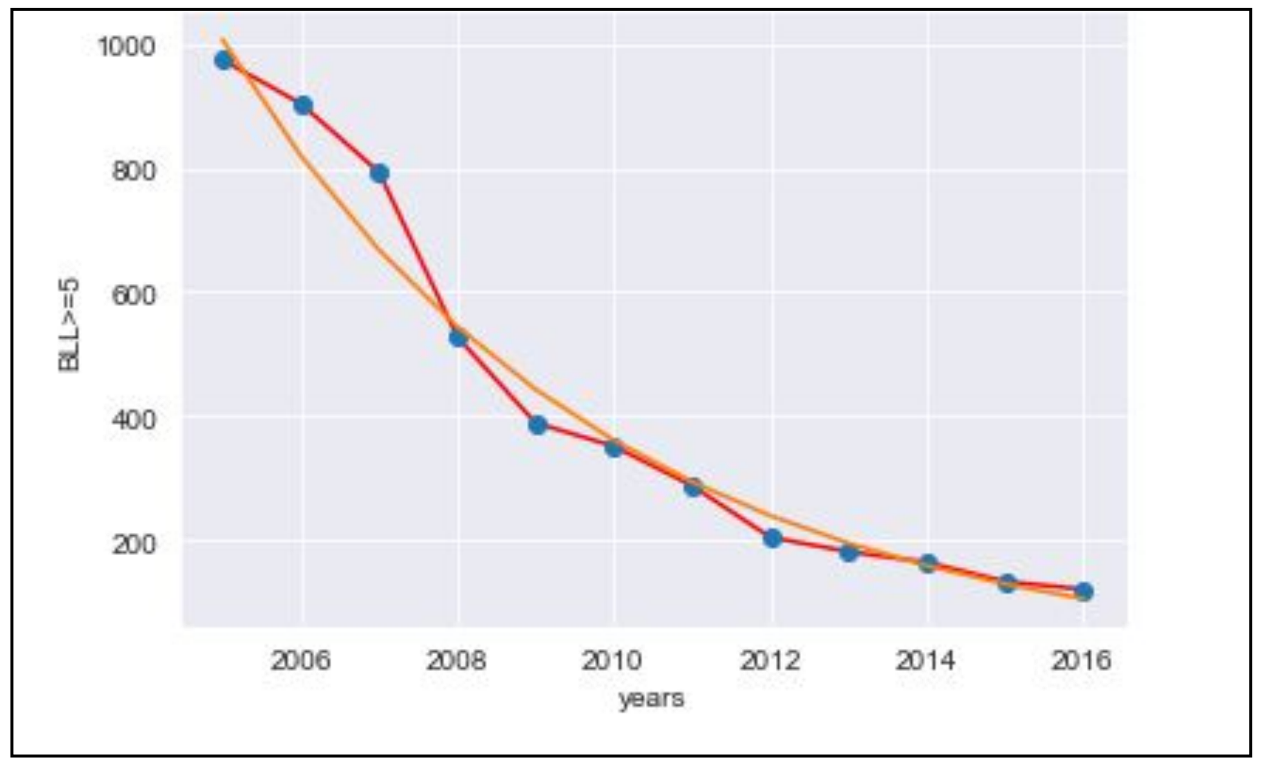
## Methods

1. NYCOpenData: We utilized New York City's open-source database to sift through datasets concerning lead poisoning among children in New York City. By conditioning datasets for relevance and the amount of redacted data, we utilized datasets that were recent (within the past decade) and also accounted for many different variables that may create outliers within the data.
2. Point2Homes: Point2Homes provides small-scale data on several neighborhoods throughout New York City, making it convenient for finding statistics on poverty rate, average income, and median income for the respective neighborhood.
3. NYC Open Data Maps: NYCOpenData provides numerous shapefiles and GeoJSON mappings of New York City's layout. To construct geographical maps, we imported NYC shapefiles that contain information about the city's neighborhoods and their polygonal bounds respective to each other.

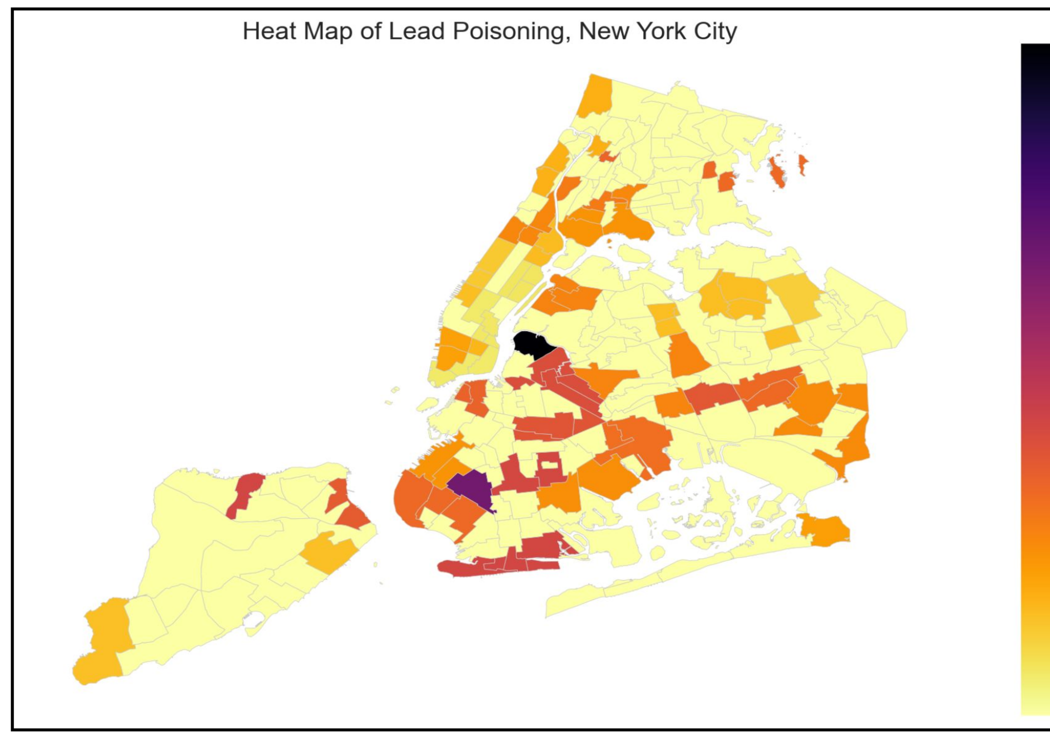
- Principal Component Analysis: Principal Component Analysis is used to reduce a data set with many variables into one with only 2 components. First, it uses elimination to standardize the data to two different axes. Then, it scales and shapes that data according to a matrix to create covariances in the data.
- Heatmap: Heatmaps are one of the best ways to visualize data on a map in a way that allows viewers to easily understand how the data connects to a geographical context.
- Scatter Plots: Scatter plots do a good job of demonstrating data of two numeric values and with a lot of variance in data points. It makes it easy to read and understand what the data is showing and the correlation between the two factors.
- Line Plots: Line plots were used in this analysis to apply regression algorithms and observe future trends. The most notable example is the graph displaying New York City's lead poisoning cases over time, with an accompanying exponential regression.
- Distribution Plots: A distribution plot is used to show how the data is spread across a certain value. This two-part graph demonstrates how close a certain value actually is to the hypothesized line. For our project, this analyzed the distribution of lead poisoning in the boroughs/neighborhoods in New York City and how it related to the poverty rate using a bell curved line and outliers in an interval.

## Results

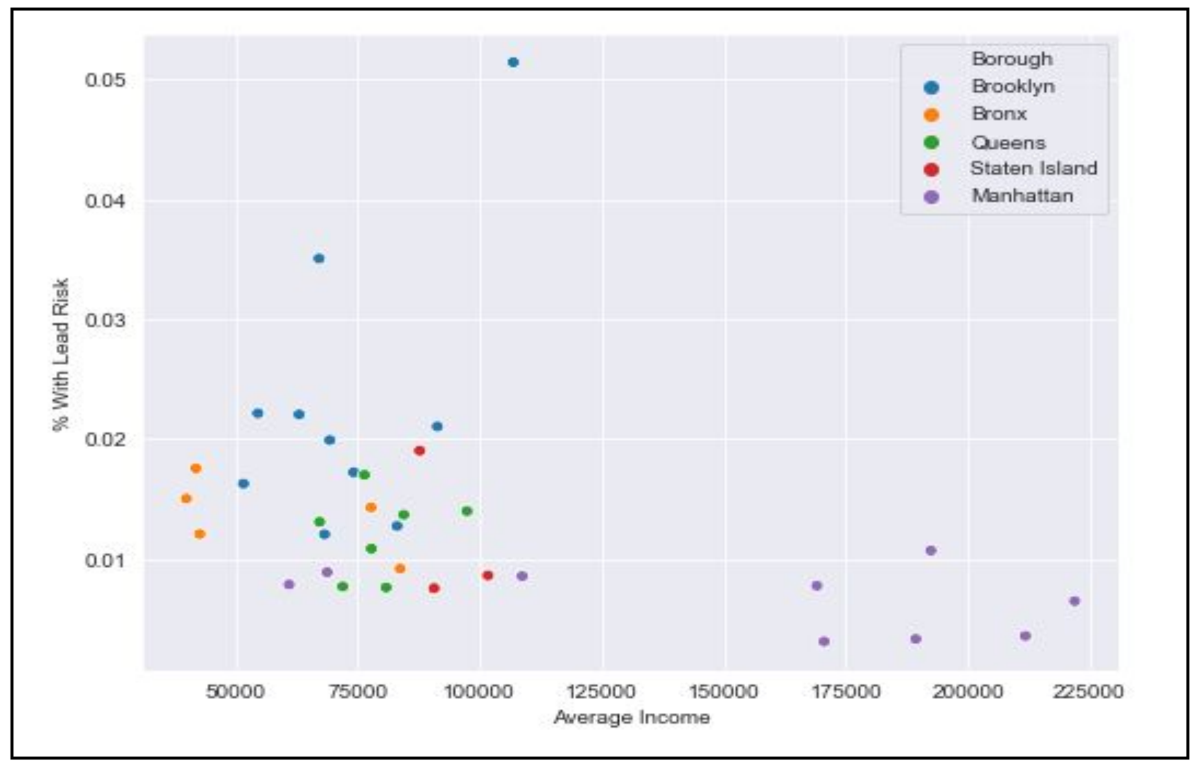
We are only interested in children with a BLL >= 5. As such, we graphed the children with a BLL >= 5 over the course of 2005 - 2016. The data indicates that NYC has done a tremendous job handling the lead epidemic from 2005-2012. However, their progress has significantly slowed in recent years. After fitting the data onto an exponential curve, the flattening of the decrease in lead poisoning cases becomes more evident. The flattened curve between 2014-2016 marks that NYC's lead poisoning epidemic is now flattening, but not dying.

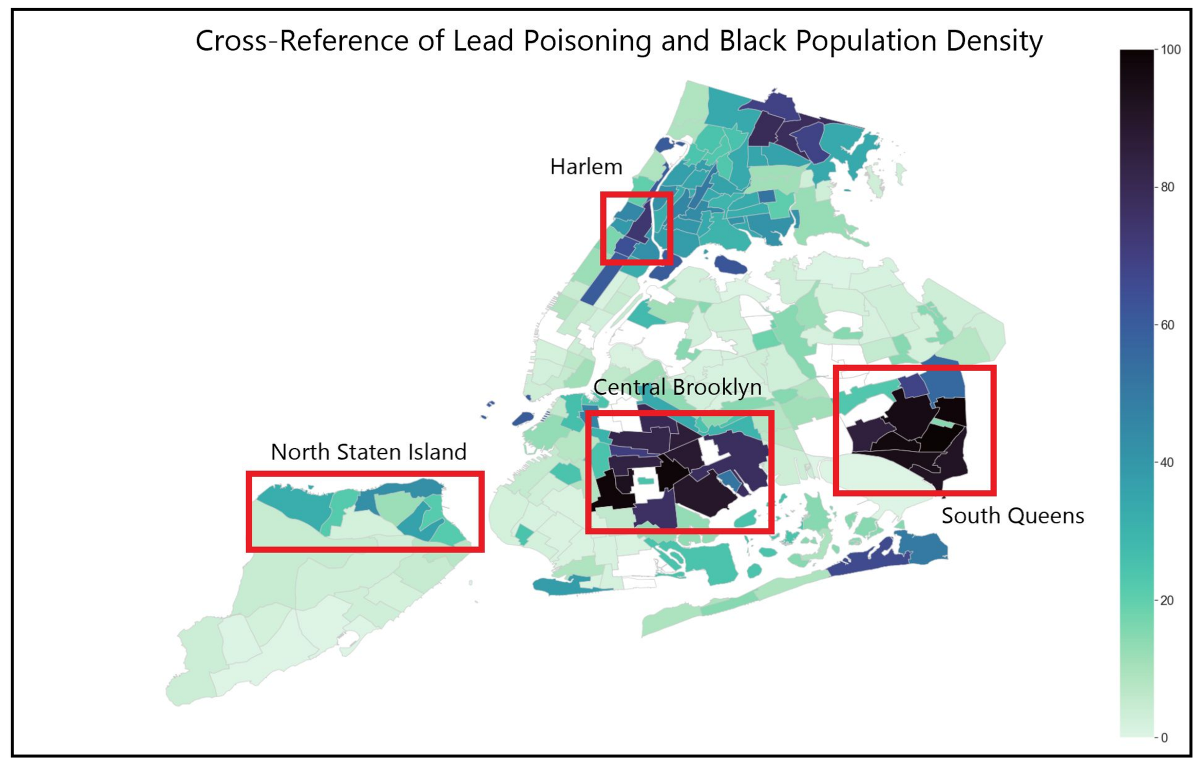



## Results



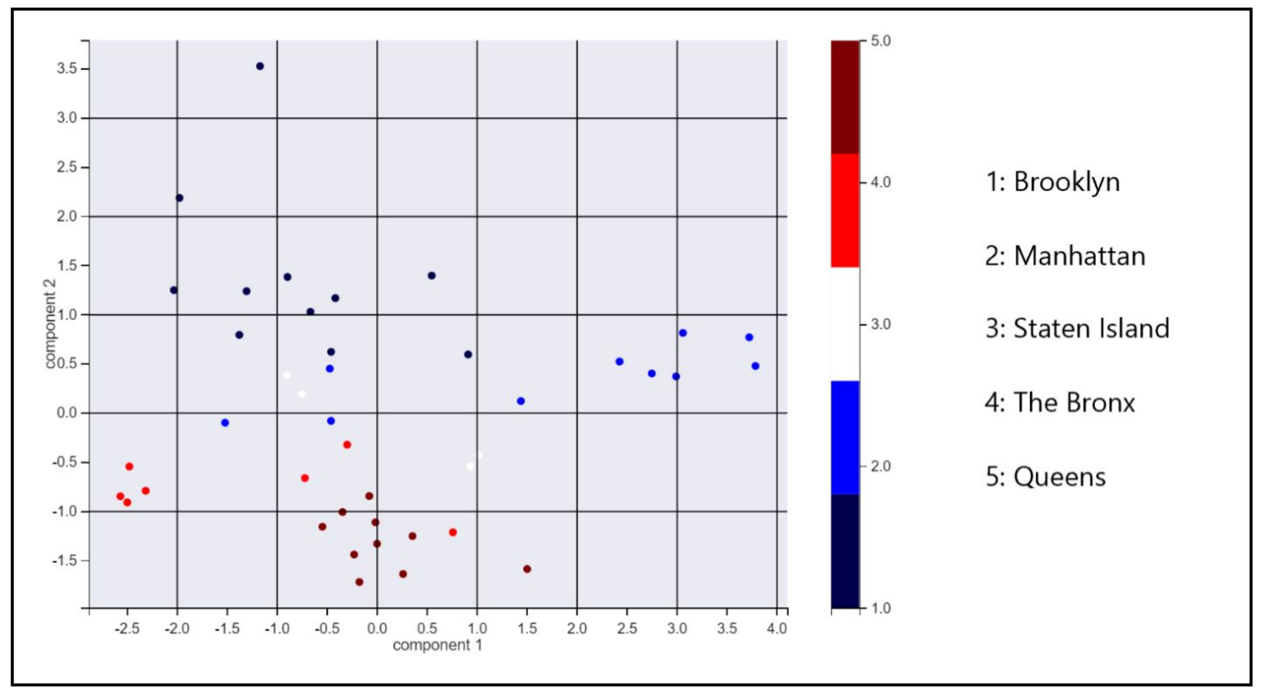

This graph is the same as the prior one, with the exception that it is now colored coded according to borough. There are a few distinct clusters within the graph, most notably the leftenmost cluster that is confirmed as belonging to Manhattan. There is also a cluster of blue points representing Brooklyn that demonstrate low average income and relatively higher lead risk. From this graph, we can clearly identify two clusters belonging to Brooklyn and Manhattan that observe opposite characteristics. Between these two clusters, there is also a patch of green points above the 0.01 threshold, representing Southeast Queens and its middle-class yet lead-struck neighborhoods.




Compare this heatmap with the heatmap concerning lead poisoning in New York City. Specific areas such as Brooklyn, Southeast Queens, Harlem, and Northern Staten Island on the demographics heatmap are consistent with hotspots on the lead poisoning heatmap.

After graphing the PCA, we applied a numeric-color code to the data points with each number signifying a borough. The legend on the left indicates which color represents which borough. As depicted by the graph, the clusters differentiating each borough stayed when all factors were applied to the principal component, lead risk. We acknowledge an area in the middle of the graph where some of the data points merge together. However, the majority of data points within Queens, Brooklyn, and Manhattan fit into distinct clusters. Hence, this analysis confirms poverty rate, average income, median income, black population contribute to the disproportion of lead poisoning cases in New York City's boroughs. For NYC to develop a system in which they can profile individuals and neighborhoods for lead-risk, they must consider a broader range of factors beyond housing age to combat lead poisoning.




## Conclusion

As confirmed by the Principal Component Analysis, disproportion exists between neighborhoods and boroughs in New York City. Lead poisoning in NYC disproportionately affects areas with high minority populations and lower income. Accordingly, areas such as Central Brooklyn, Northern Staten Island, Southeast Queens, and Harlem all display the highest lead poisoning risks. Our analysis can be used in several different ways and scopes. Foremost, NYC's lead poisoning management team can use this analysis to observe the worst affected individual neighborhood and focus their efforts there. Additionally, since median income, average income, poverty rate, and minority population are confirmed to influence child lead poisoning risk, New York City can use this new knowledge to design algorithms that can geographically and personally develop profiles to target high-risk families and/or neighborhoods. In the scope of boroughs, NYC officials also have further evidence confirming that neighborhoods such as Brooklyn suffer a significantly higher proportion of child lead poisoning than Manhattan. In essence, our analysis can be viewed in several dimensions, and in every aspect provide a unique perspective on NYC's child lead poisoning crisis.

## Next Steps (TPACK)

Hundreds of neighborhoods have no data on lead poisoning. New York City solely relies on the city's comp controller, Scott Stringer, to collect lead poisoning data, leading to huge data gaps. As New York City develops definitive tests to collect lead poisoning data on a holistic scale, we will be able to analyze the data more deeply with a machine-learning algorithm. More data means that our algorithm will better utilize factors such as age, geographic location, income, and race to quantify an individual's risk of lead poisoning. Moreover, to help New York City's lead poisoning team better evaluate hotspot areas, we seek to implement minority-specific data to find lead poisoning trends in minority groups.

## References

Please refer to our logbook for references