Understanding how urban diversity affects health outcomes for minority communities in New York City: Birth outcomes

Aria Nagai

Class: COS 398 09 | Adviser: Olga Troyanskaya and Tavis Reed

Motivation and Goal

As one of the most diverse cities in the world, New York City is home to millions of people from various racial, ethnic, and financial backgrounds. While variations in backgrounds may differ, everyone should have the opportunity and resources they need to stay healthy and access healthcare. However, Black and Hispanic New Yorkers are less likely to have adequate access to care and are more likely to experience complications with childbirth, chronic health conditions, and premature death.[1] The goal of my project is to highlight these disparities in New York City's health equity amongst minority populations and pinpoint possible correlation between certain factors such as income level, access to public transportation, and race, with health outcomes, in particular birth outcomes, through a regression model. My project is part of a larger group where each of us analyze a specific health outcome, with the others being chronic diseases and violence/drug related deaths, which we will then compile to generate an overview of current health disparities in New York.

Problem Background and Related Work

There have been a number of studies that examine the relationship between infant mortality and income inequality or social polarization over the years in an attempt to highlight disparities in health access and outcomes. In a study by Nancy Sohler, Peter Arno, Chee Chang, Jing Fang, and Clyde Schechter in 2003, it was found that in New York City, an "increase of one standard deviation in income inequality was associated with an increase of 0.80 deaths/1000 live births", indicating that there was a correlation between the two variables[2]. Similarly, another study done in 2018 by Juynh, Spasojeciv, Li, Maduro, Van Wye, Waterman, and Krieger assessed the relationship between social polarization and infant mortality and preterm births found that "women who lived in areas with the least privilege were more likely to have preterm birth or infant mortality as compared to women living in areas with the most privilege"[3]. As evident in these two studies, there are correlations between income/privilege with certain birth outcomes; however, few studies create models based on more than one or two factors that could contribute to a health outcome. My project will essentially combine these studies to look at more factors that could contribute to certain health outcomes and build a model to have a more precise understanding of the effects of these factors.

---

[1] https://www.nyc.gov/assets/doh/downloads/pdf/episrv/disparitiesone.pdf
[2] https://link.springer.com/article/10.1093/jurban/jtg071
[3] https://journals.sagepub.com/doi/pdf/10.1177/1403494817701566

Approach

I will be creating a regression prediction model in order to model the birth outcomes of minority populations in New York City based on income, poverty levels, healthcare funding, public transportation in their area, and social vulnerability. Birth outcomes include premature births, quality of prenatal care, and infant mortality. This approach is novel not only because it encompasses more factors than previous work, but because this project is part of a larger project that takes into account a wider array of health outcomes (chronic disease and violence/drug related deaths) in a concentrated geographic area. This will provide a more in depth look at the health disparities found in New York City.

Plan

I plan to create a tool that can be used by policymakers to make data-driven decisions on how to best allocate healthcare resources in New York City. By using datasets from NYC Open Data, EquityNYC, and New York Department of Health on infant mortality, prenatal care access, and preterm births, I will create a regression model that takes in social inputs about an individual to predict their health outcome in regards to childbirth. After acquiring the datasets that pertain to childbirth outcomes for different income levels, races, ethnicities, and boroughs, I will clean up the data in a way that will allow me to better train and test the model. Moreover, combining my model with the two other models will allow us to test for different health outcomes and as a stretch goal we hope to implement a visualization that contextualizes the different health outcomes each of us analyzed in one view.

Evaluation

Because I am creating a regression model, evaluation will be based on mean square error and test accuracy. I hope to be able to find datasets of patient data that includes the different factors I analyzed and their health outcomes and thus compare the results to my predicted results. By minimizing the mean square error and maximizing both training and testing accuracy, I will be able to measure the success of my model.