

Poverty, Income, and COVID-19 in New York City

Irene Zhang*

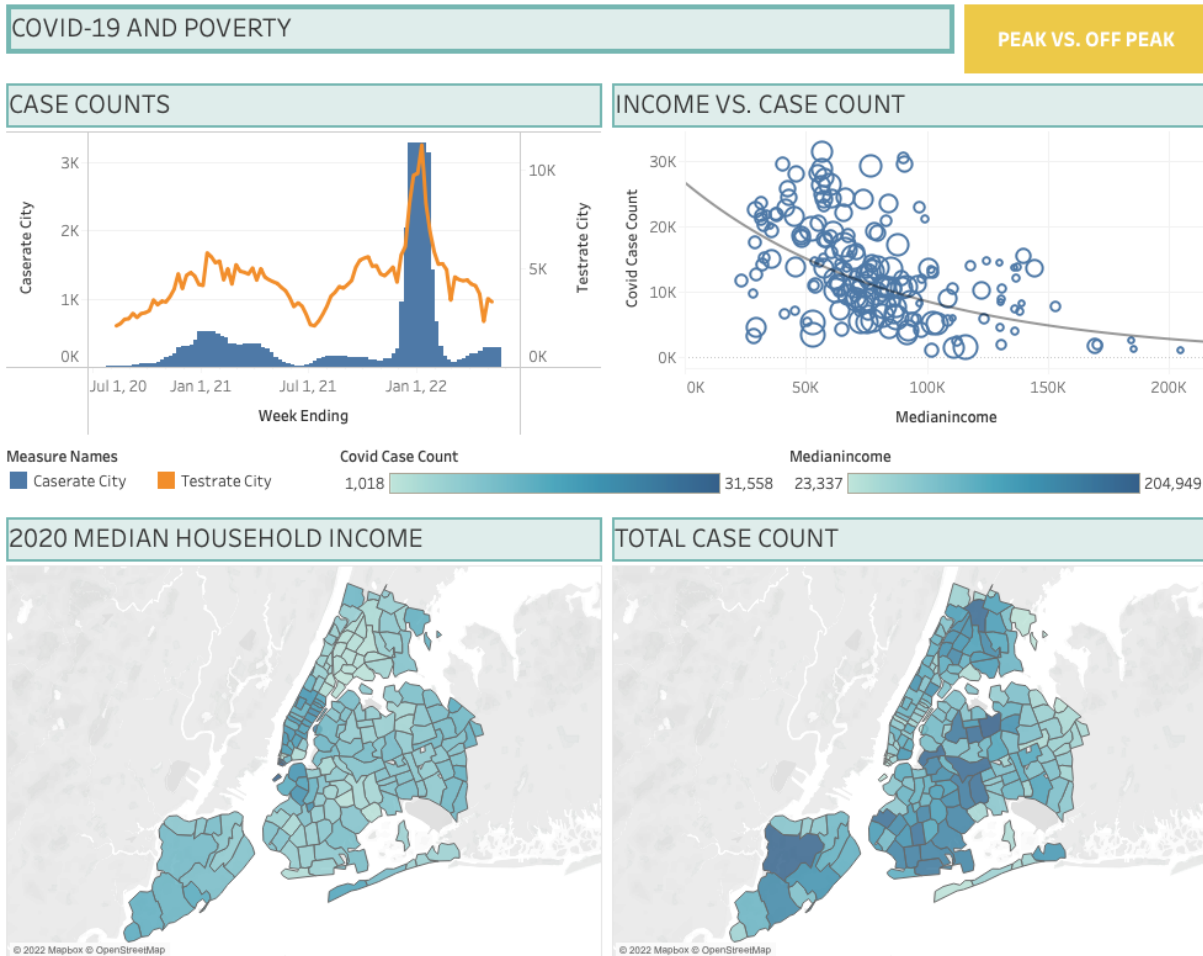


Figure 1: Project Dashboard 1.

ABSTRACT

The purpose of this project is to determine the correlation between the COVID-19 pandemic and economic inequalities in New York City. Although the phrase "COVID-19 does not discriminate" is repeated on various occasions, the disease affects economically disadvantaged people disproportionately, reinforcing and exacerbating the economic inequalities in society through lack of access to resources. The COVID-19 pandemic increases concerns centering around lack of access to safe housing, food, and medical resources, which targets the less privileged group in society. This project studies the relationship between the number of positive cases of coronavirus and average rent in NYC from early 2020 to Spring 2022. Initial results show that the cases in poorer neighborhoods grow faster than in wealthier neighborhoods.

*e-mail: yachun.zhang21@myhunter.cuny.edu

Index Terms: COVID-19 pandemic, modified ZIP Code Tabulation Areas (MODZCTA), poverty threshold, visualization.

1 INTRODUCTION

The project will focus on the relationship between the cases of coronavirus and the economic factors in New York City. Since the beginning of 2020, the COVID-19 pandemic has brought both positive and negative changes to society, influencing over 70 million people. In the city of New York, these changes hit hard on people who are less privileged and completely altered their way of living. This project will look into the average rent in New York City before and after the COVID-19 pandemic and compare it to case rates in each neighborhood, analyzing the connections between the spread of virus and the economic advantages.

2 PREREQUISITES

The project works with publicly accessible data sources, details will be provided in the bibliography section. The project also requires the 2022.1 version of Tableau Desktop. Visualization of case rates

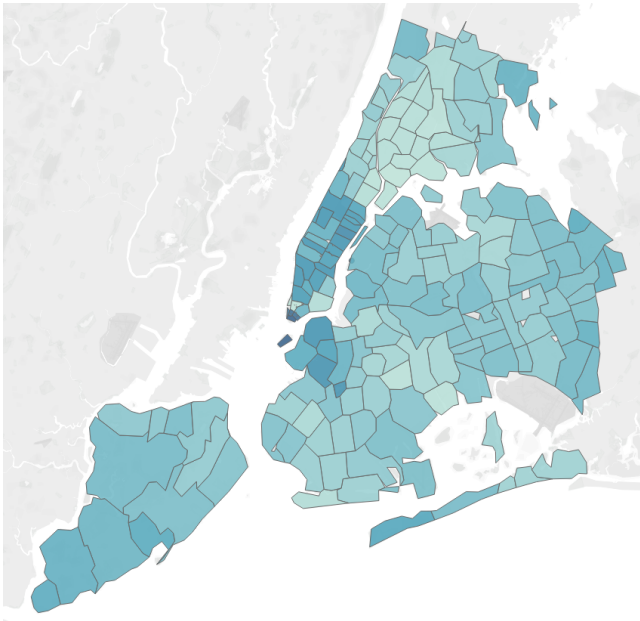


Figure 2: A visualization of the current NYC median household income.

trend and average rent trend will be created by Tableau to analyze the correlation of poverty and spread of COVID-19.

3 METHODOLOGY

The methodology in this project includes two parts: pre-processing and methods. Pre-processing cleans out any useless data to enhance performance, and the visualizations are made following steps in methods.

Live animations are available here :

https://public.tableau.com/views/PovertyIncomeandCOVID-19inNewYorkCity/Dashboard1?:language=en-US&display_count=n&origin=viz_share_link

3.1 Pre-processing

A preliminary processing of data is done at first in order to prepare the data sets for primary processing and further analysis. Methods for data pre-processing include data cleaning, data integration, data reduction, and data transformation. During the data selection period, the focus was on the completeness and inclusiveness on topics for each data sets, i.e., the complete list of daily case rates by zip-code is chosen over list of monthly case rates by boroughs. Both Excel and Python MySQL were used to reorganize data sets. The data file containing case rates by MODZCTA contains columns such as "CASERATE_10001". Such column names are altered by eliminating "CASERATE_" to keep only the clean zip-code. The original data set is up to date, which is then cut short to fit the narrative of the research time frame. Similar actions are performed on the median income data set to clean up column names and unwanted areas.

3.2 Methods

The visualizations are created with Tableau. After importing the first data source of case rate by MODZCTA, a new data set of median income data is connected with a common factor of zip-code.

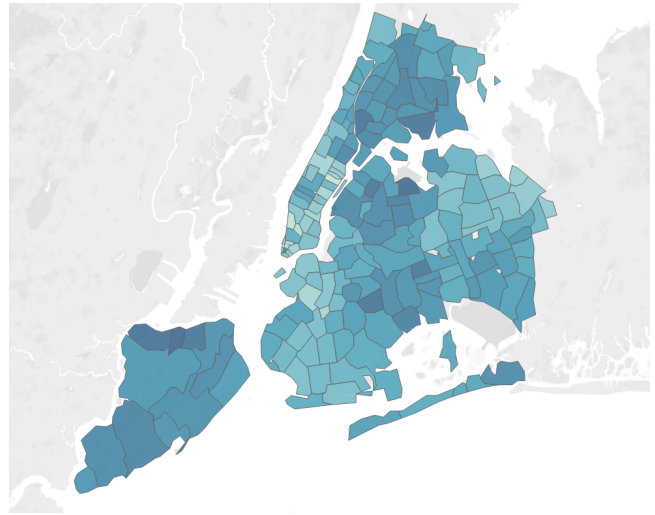


Figure 3: A visualization of the NYC case rate map in January 2022.

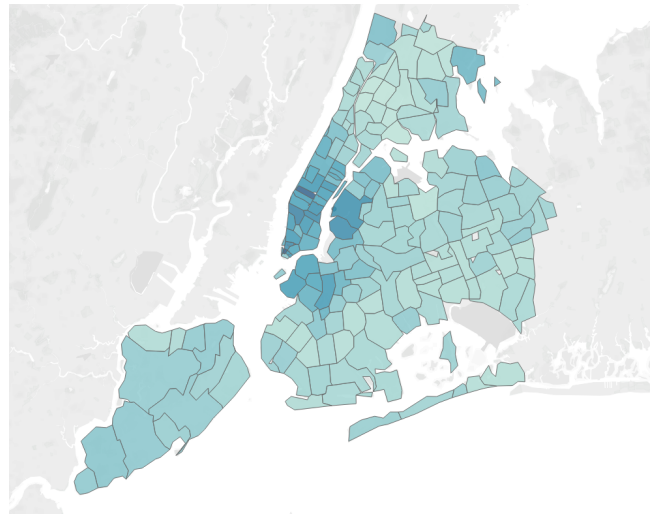


Figure 4: A visualization of the NYC case rate in April 2022.

4 INITIAL RESULTS

Figure 2 is a choropleth map showing the 2020 median household income in NYC. Wealth was more concentrated in most parts of Manhattan and downtown Brooklyn. Figure 4 is the case rate in NYC on April 2nd, 2022 when positive cases are slowing down while Figure 3 shows the case rate on the week ending on January 1st, 2022 when the Omicron variant peaked in NYC. In both Figures 3 and 4, the density of color represents the case rates per 100,000 people, the darkest being the most severe. Figure 3 confirmed my hypothesis of the high case rates in underprivileged neighborhoods such as Melrose, New York (zip-code 10451) where its 2020 median income is around \$32,589, almost half of the New York City median of \$67,046. In April 2022, the growth of case rates starts to slow down. One thing I notice about Figure 4 is that this map looks similar to the average income choropleth map of NYC. Neighborhoods with higher COVID-19 case rates are also neighborhoods with higher median incomes when the pandemic is off-peak. One explanation is that the rising number of people returning to the office and tourists visiting the city leads to the increase in case rates in downtown and midtown Manhattan.

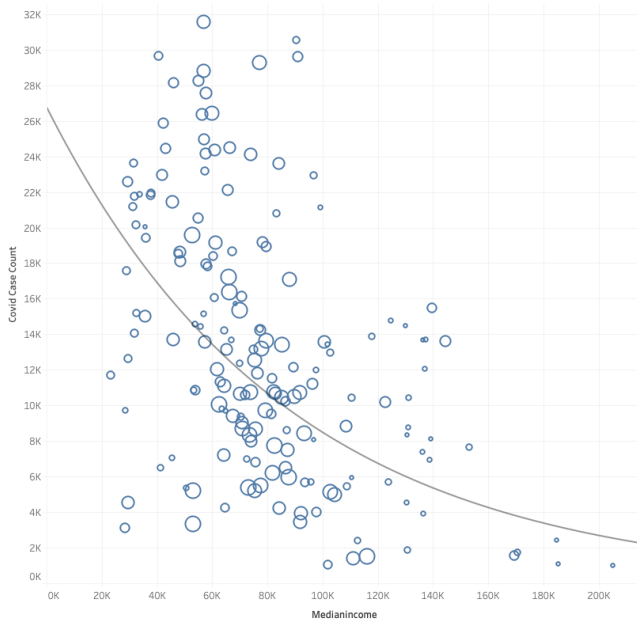


Figure 5: A scatter plot of median income and case counts.

5 FURTHER ANALYSIS

To further analyze the relationship between economic inequalities and the COVID-19 pandemic, figure 5 shows a scatter plot of the median income and current total case count in NYC with the best fit line. As shown in the graph, income and case count has a weak negative correlation where R-Squared equals 0.300591. This partially supports the original hypothesis that income inequality does affect people differently during the pandemic, however, such correlation isn't strong enough to draw a conclusion between the two events. Possible causes include different levels of access to safe housing, nutritious food, and medical support.

6 LIMITATION AND CHALLENGES

One of the major challenges of this project was the cleaning and connecting process of data sets. For both case rate and test rate data sets, I first transposed the entire data sets to get zip-codes in a column so that it can be used to join two data sets. Then I encountered the error where zip-code columns in two data sets were different in data type. This was fixed by changing one of them into string type. Another issue I saw was while creating the scatter plot for income and case count, the plots showed a weak pattern, and the best fit line did not suggest a strong relationship. This led to further analysis of the pattern.

7 FUTURE WORK

If given more time, this project can be polished for a better interactive dashboard which connects each zip code to its corresponding case count line graph.

8 DATASETS

<https://github.com/izhang024/Poverty-Income-and-COVID-19-in-NYC/tree/main>

This project was supervised by Professor Oyewole Oyekoya, as an Advanced Visualization project at Hunter College, Spring 2022.