

Vax Populi: COVID-19 Outcomes and Political Alignment in New York City

John Lin

Index Terms—COVID-19, NYC Health Data, Vaccination, Politics, Election Data

1 INTRODUCTION

The COVID-19 pandemic has affected communities across the globe with wide-reaching impacts on public health and socioeconomic consequences. However, the question arises: did the pandemic affect everyone equally? This project aims to explore the disparities in COVID-19 outcomes across different communities, with a particular focus on New York City (NYC).

New York City was chosen for this study due to its unique characteristics - it serves as a microcosm of various demographic and cultural differences, making it an ideal location to study the intersection of public health and socio-political factors. NYC is well-connected through public transit and other transport options, which facilitates the spread of infectious diseases. Additionally, the city is composed of distinct communities, each with its own demographic and cultural profile.

The recent 2024 US election has highlighted heightened polarity on various issues, including healthcare administration and policy. This polarization raises an important question: does political affiliation stratify health outcomes, particularly in the context of COVID-19? The anti-vaccination movement, for instance, has often been framed as a political issue. This project seeks to examine whether political affiliation correlates with COVID-19 outcomes, such as case rates, vaccination rates, and death rates, in NYC neighborhoods.

By analyzing COVID-19 outcomes in relation to election voting ratios per district for the 2020 and 2024 elections, we hope to uncover patterns that highlight issues in public health within diverse urban populations. This analysis aims to provide insights into how political and social factors may influence public health outcomes, and to identify potential areas for targeted public health interventions.

2 METHODS

2.1 Datasets

2.1.1 NYC COVID Data

The NYC COVID Data dataset [1] is publicly available on GitHub and is maintained by the NYC Department of Health. This dataset provides comprehensive information on the impact of COVID-19 across New York City's neighborhoods. The data is organized using Modified Zip Code Tabulation Areas (MODZCTA), which allows for detailed spatial analysis.

Key features of the dataset include:

- **Case Counts:** The total number of confirmed COVID-19 cases in each neighborhood.
- **Case Rates:** The rate of confirmed COVID-19 cases per 100,000 residents, providing a normalized measure of infection spread.
- **Death Counts:** The total number of confirmed COVID-19 related deaths in each neighborhood.
- **Death Rates:** The rate of confirmed COVID-19 related deaths per 100,000 residents, offering a normalized measure of mortality.

To facilitate analysis, we perform some post-processing on the dataset: normalizing the case rates and death rates to a floating-point number between 0 and 1. This allows cases to be represented as percentages of the total population in that neighborhood.

2.1.2 NYC COVID Vaccination Data

The NYC COVID Vaccination Data dataset [2] is publicly available on GitHub and is maintained by the NYC Department of Health. This dataset provides detailed information on COVID-19 vaccination rates across New York City's neighborhoods. Similar to the COVID case data, this dataset uses Modified Zip Code Tabulation Areas (MODZCTA) for spatial organization.

Key features of the dataset:

- **Vaccination Rates:** The dataset includes cumulative vaccination rates as percentages of the total population in each neighborhood.
- **Stratified Data:** The data is stratified into different categories:
 - Individuals vaccinated with only the first dose.
 - Individuals who are fully vaccinated (two or more doses).
 - Individuals who have received additional doses (more than two doses).

To facilitate analysis, we focused on post-processing the following features:

- **Full Vaccination Rates:** We primarily focus on the percentage of individuals who are fully vaccinated (two or more doses), as this aligns with popular policy definitions of a complete vaccination program.
- **Normalization:** Similar to the COVID case data, we normalize the vaccination rates to a floating-point number between 0 and 1, representing the percentage of the total population in each neighborhood.

2.1.3 NYC Electoral Districts

The NYC Electoral Districts dataset [3] provides detailed information on the boundaries of electoral districts within New York City. This dataset is essential for understanding the political landscape of NYC and is available from the NYC Planning website.

Key features of the dataset include:

- **Electoral Boundaries:** The dataset includes the geographic delineations of various political districts, such as congressional, state senate, and state assembly districts.
- **Shapefile Format:** The data is provided in shapefile format, which includes polygon geometry for precise mapping and spatial analysis.
- **District Numbers:** Each electoral division is identified by its district number, facilitating detailed and organized analysis.

2.1.4 NYC Election Results 2020 and 2024

The NYC Election Results dataset, created by Todd W. Schneider and available on GitHub [4], provides detailed election results for New York City during the 2016, 2020, and 2024 presidential elections. This dataset is essential for analyzing the political landscape of NYC and its potential impact on COVID-19 outcomes.

Key features of this dataset include:

- **Election Results:** The dataset includes the number of votes cast for the Democratic (Dem) and Republican (GOP) candidates in each electoral district.
- **Temporal Data:** The dataset covers three election years: 2016, 2020, and 2024. This allows for comparative analysis of voting patterns over time.
- **Geographic Mapping:** The election results are matched with electoral district codes, facilitating geographic analysis.

We also perform some post-processing on this dataset by:

- **Filtering by Year:** We primarily focus on the 2020 election results, as this year coincides with the peak of the COVID-19 pandemic. However, we also consider the 2024 election results to examine changes in voting patterns and correlations over time.
- **Data Integration:** We integrate the election results with the COVID-19 and vaccination datasets based on electoral district codes.
- **Normalization:** We normalize the voting data to represent the proportion of votes for each candidate as a floating-point number between 0 and 1. Since we are mostly focused on Dem vs. GOP voting ratios, we sum the votes for both parties, and take the fraction of each as Democrat-two-party and Republican-two-party fractions.

2.2 Data transformation and cleaning

The data transformation and cleaning process is one of the most computationally intensive parts of this project. It involves several steps to ensure that the datasets are properly aligned and ready for analysis.

2.2.1 Geographic Data Integration

- **MODZCTA Regions:** The NYC COVID data and vaccination data are organized by Modified Zip Code Tabulation Areas (MODZCTA), which consist of 177 regions. (Fig. 1a)
- **Electoral Districts:** The NYC electoral districts dataset contains 4345 distinct regions. These regions are represented as shapefiles with polygon geometry. (Fig. 1b)
- **Spatial Join:** Using geographic information system (GIS) tools, we perform a spatial join to combine the MODZCTA regions with the electoral districts. This results in a polygon map with 177 regions, closely matching the MODZCTA regions. (Fig. 1c)

2.2.2 Data Merging

- **COVID Data Integration:** The COVID data is joined by MODZCTA ID, ensuring that each region's case counts, case rates, death counts, and death rates are accurately represented.
- **Election Data Integration:** The election data is joined by electoral district ID. We filter the election data by year to focus on the 2020 and 2024 election results.
- **Merged Dataset:** The COVID data and election data are merged based on the geographic mapping, resulting in a comprehensive dataset that includes COVID metrics and election results for each region.

2.2.3 Data Cleaning

- **Handling Missing Values:** For certain analyses, rows with missing values (NaN) are dropped to ensure the accuracy of the results.
- **Normalization:** The rates (case rates, death rates, vaccination rates) are normalized to floating-point numbers between 0 and 1. This allows for consistent representation of the data as percentages of the total population in each neighborhood.

This process ensures that the datasets are properly aligned and ready for subsequent analysis. By integrating and normalizing the data, we can accurately explore the relationships between political alignment and COVID-19 outcomes.

2.3 Visualizing Distributions

To understand the distribution of various metrics across NYC neighborhoods, we will create visualizations using histograms and maps.

2.3.1 Histograms

We will generate histograms to show the distributions of the following metrics:

- **Case Rates per District:** The distribution of COVID-19 case rates across different electoral districts.
- **Death Rates per District:** The distribution of COVID-19 death rates across different electoral districts.
- **Vaccination Rates per District:** The distribution of COVID-19 vaccination rates across different electoral districts.

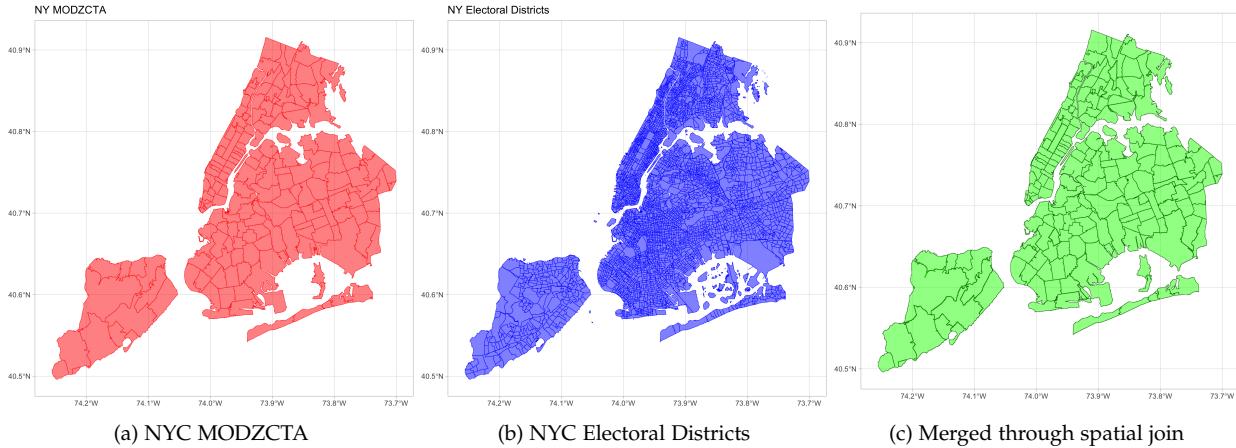


Fig. 1: NYC maps merged through spatial join show similarity to MODZCTA regions

- **Proportion of Democratic Voters (2020/2024):** The distribution of the proportion of votes for Democratic candidates in the 2020 and 2024 elections.
- **Proportion of Republican Voters (2020/2024):** The distribution of the proportion of votes for Republican candidates in the 2020 and 2024 elections.

2.3.2 Mapping Distributions

We will also create maps to visualize the spatial distribution of these metrics:

- **Density Maps:** Mapping the distributions allows us to see the density of the population for each targeted feature at a glance.
- **Overlaying Polygons:** For electoral districts, the overlaid polygons will accumulate results from separate electoral divisions, providing a comprehensive view of the data.

2.4 Correlation Statistics

To explore the relationships between COVID-19 metrics and election results, we will calculate and plot correlation statistics.

2.4.1 Correlation Analysis

We will analyze the correlations between the following pairs of metrics:

- **COVID Case Rate vs. Voting Fraction (2020, 2024):** Examining the relationship between the proportion of votes for Democratic and Republican candidates and the COVID-19 case rates in 2020 and 2024.
- **COVID Death Rate vs. Voting Fraction (2020, 2024):** Investigating the correlation between the voting fractions and COVID-19 death rates in 2020 and 2024.
- **COVID Vaccination Rate vs. Voting Fraction (2020, 2024):** Analyzing how vaccination rates correlate with the voting fractions in 2020 and 2024.
- **COVID Vaccination Rate vs. Case Rates:** Exploring the relationship between vaccination rates and COVID-19 case rates.
- **COVID Vaccination Rate vs. Death Rates:** Assessing the correlation between vaccination rates and COVID-19 death rates.

2.4.2 Statistical Methods

We will use the following statistical methods to calculate and visualize the correlations:

- **Pearson Correlation Coefficient:** To measure the linear correlation between the pairs of metrics.
- **Scatter Plots:** To visualize the relationships between the pairs of metrics.

2.5 Areal Statistics

To account for spatial dependencies in the data, we will perform areal statistics analysis. This involves calculating spatial autocorrelation metrics and analyzing the spatial relationships between neighboring regions.

2.5.1 Moran's I

Moran's I is a measure of spatial autocorrelation, which indicates the degree to which a variable is similar to itself in nearby locations. We will calculate Moran's I using different spatial weights:

- **6 Nearest Neighbors (6NN):** Considering the six closest neighboring regions for each area.
- **Queen Contiguity:** Considering regions that share a common boundary or vertex.
- **Inverse Distance Weighting (IDW):** Considering the influence of neighboring regions based on the inverse of their distance.

2.5.2 Correlograms

Correlograms will be generated to visualize the autocorrelative effects across neighboring regions. These plots will help us understand how the spatial relationships between regions influence the distribution of COVID-19 metrics and political alignment.

2.5.3 Other areal analysis

We will also perform the following additional analyses:

- **Getis-Ord Gi* Statistic:** The Getis-Ord Gi* statistic identifies clusters of high or low values, detecting hotspots and cold spots within the dataset.

- **Spatial Autoregressive (SAR) Model:** The SAR model accounts for spatial dependencies by incorporating spatial lag-errors, analyzing the influence of neighboring regions on COVID-19 case rates, vaccination rates, and political alignment.
- **Conditional Autoregressive (CAR) Model:** The CAR model considers the conditional distribution of each region given its neighbors, helping to understand the influence of the spatial structure on COVID-19 case rates, vaccination rates, and political alignment.

2.6 Code

All code and analyses is available in a public repository at <https://github.com/john-s-lin/vax-populi>.

3 RESULTS

3.1 Distributions and Maps of COVID Case Rates, Death Rates, and Vaccinations

3.1.1 Case Rates

The analysis of COVID-19 case rates across New York City reveals significant insights into the spread of the virus in different neighborhoods. For one, the average case rate per district is approximately 0.4 (Fig. 2a), indicating that, on average, 2 out of 5 people in NYC have had COVID-19 at some point contracted COVID-19. Additionally, the case counts did not exceed 50,000 cases per 100,000 people, resulting in a maximum case rate of 0.5.

The map (Fig. 2b) also shows several hotspots:

- **Staten Island:** Significant hotspots are observed in Staten Island.
- **Midtown Manhattan:** A hotspot is identified in Midtown Manhattan, an area known for its high density due to tourist attractions.
- **Rockaway Beach:** Another hotspot is located in Rockaway Beach at the southern border of Queens.

3.1.2 Death Rates

The analysis of COVID-19 death rates across New York City provides insights into the mortality impact of the virus in different neighborhoods.

- **Average Death Rate:** The average death rate per district is approximately 0.5%, indicating that the mortality rate due to COVID-19 was relatively low across NYC (Fig. 3a).
- **Spatial Distribution:** The death rate is relatively evenly spread across all districts, with a few exceptions (Fig. 3b):
 - **Southern Brooklyn:** Isolated districts in southern Brooklyn show slightly higher death rates.
 - **Queens:** Similar isolated districts in Queens also exhibit higher death rates.

3.1.3 Vaccination Rates

The analysis of COVID-19 vaccination rates across New York City provides insights into the uptake of vaccines in different neighborhoods.

- **High Vaccination Rates:** The percentage of fully vaccinated individuals in each district is very high, with many districts having vaccination rates greater than 2 doses per person, indicating booster vaccinations (Fig. 4a).
- **Average Vaccination Rate:** The average vaccination rate across districts is approximately 0.9. This high average suggests that a significant portion of the population received full vaccination, including booster doses.
- **Spatial Distribution:** The vaccination rates are relatively high across most districts, with some variation (Fig. 4b):
 - **Higher Rates:** Many districts have vaccination rates above 1, indicating that a substantial number of individuals received more than the initial two doses.
 - **Variation:** While the overall vaccination rates are high, there are variations across different neighborhoods, reflecting differences in vaccine uptake.

3.2 Distributions of Election Results

3.2.1 2020

For the most part, NYC was predominantly left-leaning in 2020, with many districts showing a higher Democrat-Republican voting ratio. However, Staten Island and southern Brooklyn had more communities voting significantly to the right (Fig. 5a).

3.2.2 2024

In 2024, NYC shifted more towards the right, with more districts in northern Queens becoming conservative. The ratio of Democrat to Republican voters decreased, with Staten Island, southern, and central Brooklyn becoming more conservative, while Manhattan, Queens, and the Bronx still had a higher Democrat voting ratio but showed a shift towards conservatism (Fig. 5b).

3.3 Correlations

There were several notable correlations we observed:

- **Political Alignment vs. Case Rate:** In 2020, Democrat voters and case rates were negatively correlated (-0.37, p-value = 1.13E-7), while Republican voters showed a positive correlation. In 2024, the negative correlation for Democrat voters slightly decreased to -0.33 but remained statistically significant.
- **Political Alignment vs. Death Rate:** In 2020, Democrat voting fraction and death rates were negatively correlated (p-value = 0.04). In 2024, the correlation became stronger at -0.2395 (p-value = 1.6E-3).
- **Political Alignment vs. Vaccinations:** There was no significant correlation between vaccination rates

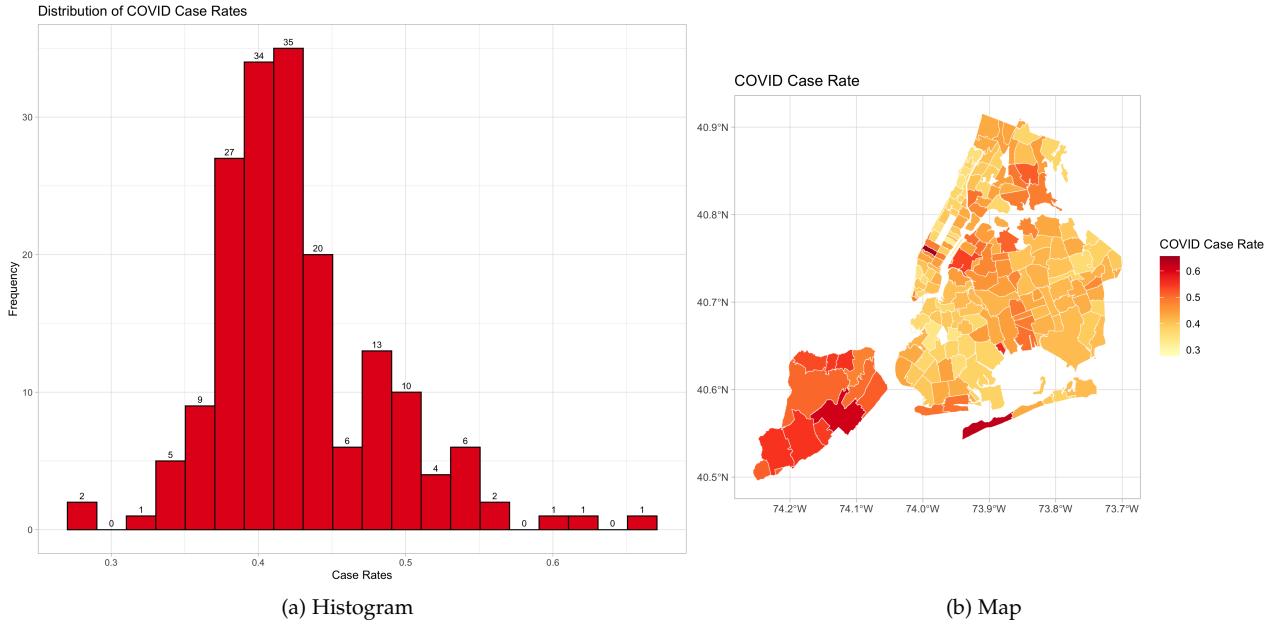


Fig. 2: NYC COVID Case Rates

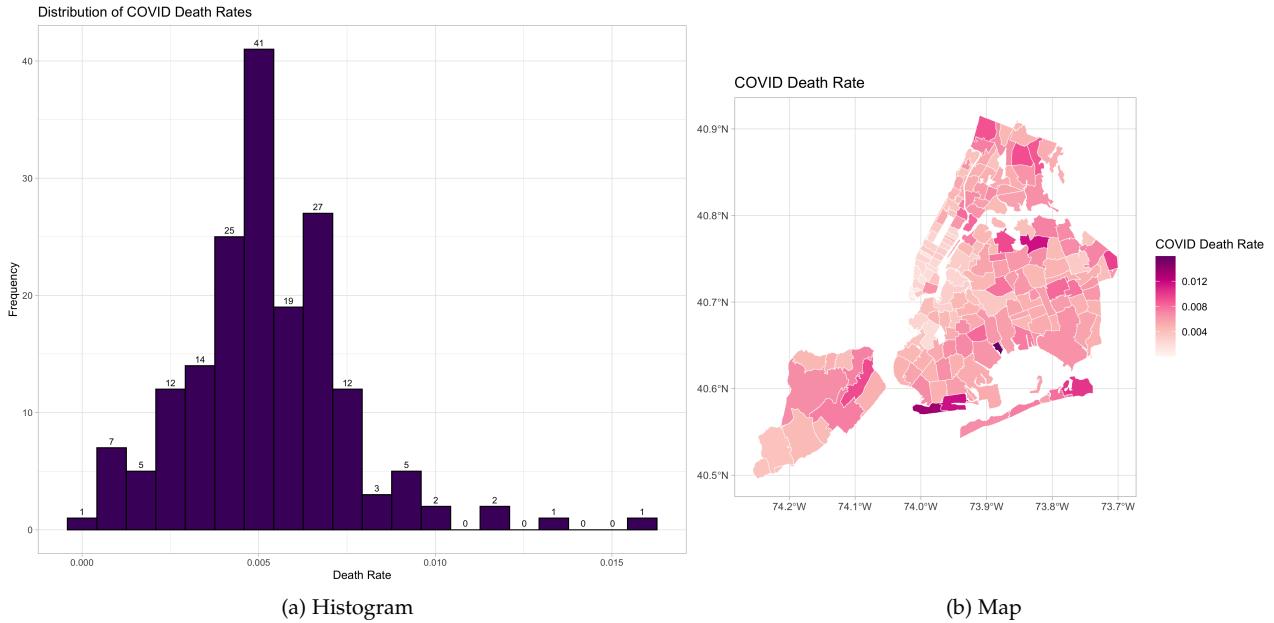


Fig. 3: NYC COVID Death Rates

and election voting patterns in both 2020 and 2024, indicating equal vaccination behavior across political alignments.

- **Case Rates vs. Fully Vaccinated:** A positive correlation (0.33 , $p\text{-value} = 6.2\text{E-}06$) was observed, suggesting that higher case rates may lead to increased vaccination rates or that vaccinated individuals might engage in riskier behaviors.
- **Death Rates vs Fully Vaccinated:** There was a negative correlation observed (-0.195 , $p\text{-value} = 0.0093$), which is in line with other studies that show that vaccination prevents severe symptoms and provide

better outcomes [5].

Correlations are summarized in Table 1 and visualized in Fig. 6.

3.4 Areal Analysis

3.4.1 Moran's I and Correlogram Lag Structures

To understand the spatial dependencies in the data, we performed Moran's I analysis and examined correlogram lag structures.

Moran's I is a measure of spatial autocorrelation, indicating the degree to which a variable is similar to itself

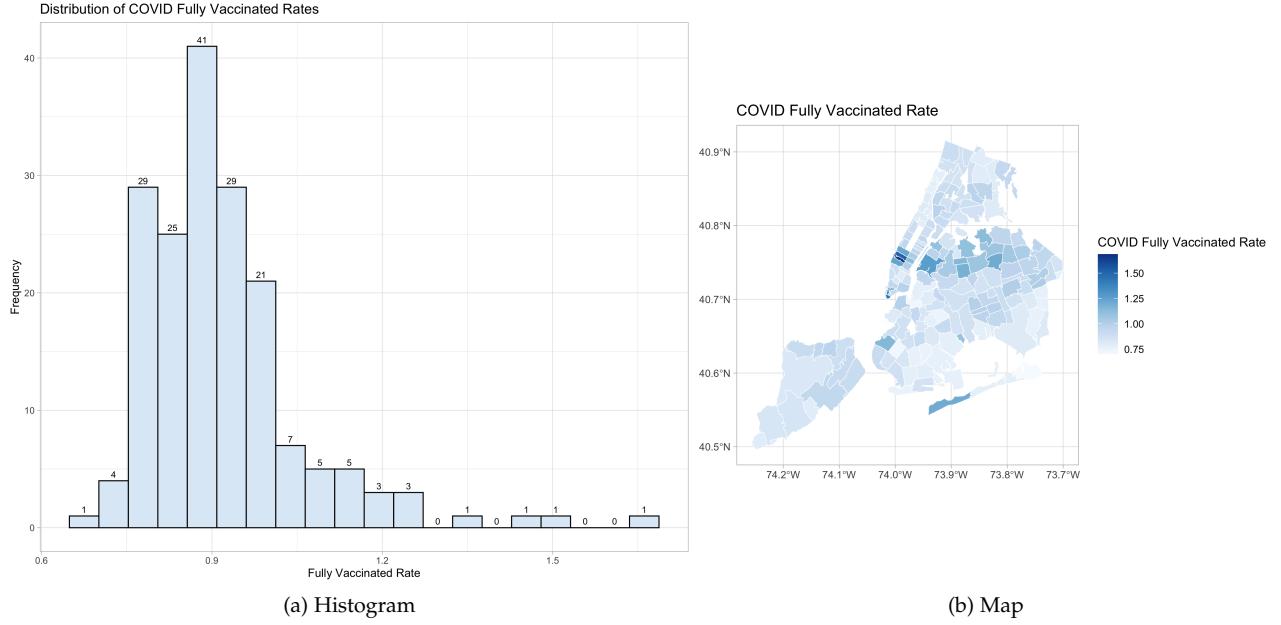


Fig. 4: NYC COVID Full Vaccination Rates

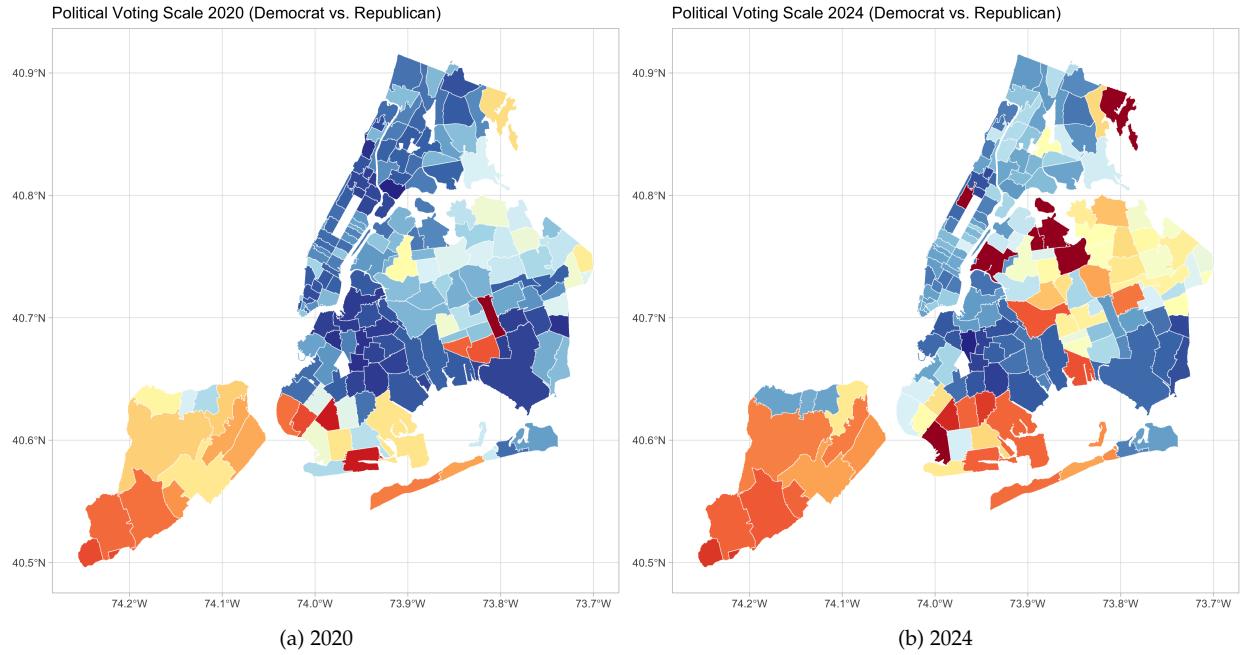


Fig. 5: NYC Election Results

in nearby locations. We calculated Moran's I for various metrics using different spatial weights: Queen contiguity, six nearest neighbors (6-NN), and inverse distance weighting (IDW). Results of Moran's I can be seen in table 2.

Correlograms were generated to visualize the autocorrelative effects across neighboring regions using the 6-NN spatial weight, which we observed to be overall the best. Lag plots can be viewed in figure 7.

- **Case Rates:** Neighbor effects were observed up to lag 2, indicating that case rates are autocorrelated within two neighboring regions.

- **Death Rates:** Due to the low death rates, autocorrelation extended up to lag 6, showing spatial dependencies up to six neighboring regions.
- **Vaccination Rates:** Similar to case rates, vaccination rates showed autocorrelation within 1-2 lag distances.
- **Voting Fraction 2020:** Autocorrelation was observed up to lag 3, indicating spatial dependencies within three neighboring regions.
- **Voting Fraction 2024:** Similar to 2020, autocorrelation extended up to lag 3.

X	Y	R^2	p-value
Democrat Fraction 2020	Case Rate	-0.3869	1.1336e-07
Democrat Fraction 2024	Case Rate	-0.3304	9.5669e-06
Democrat Fraction 2020	Death Rate	-0.1545	0.0406
Democrat Fraction 2024	Death Rate	-0.2395	0.0016
Democrat Fraction 2020	Percent Fully Vaccinated	-0.0225	0.7665
Democrat Fraction 2024	Percent Fully Vaccinated	0.0177	0.8180
Percent Fully Vaccinated	Case Rate	0.3324	6.1842e-06
Percent Fully Vaccinated	Death Rate	-0.195	0.0093

TABLE 1: Correlation between Political Alignment and COVID-19 Metrics

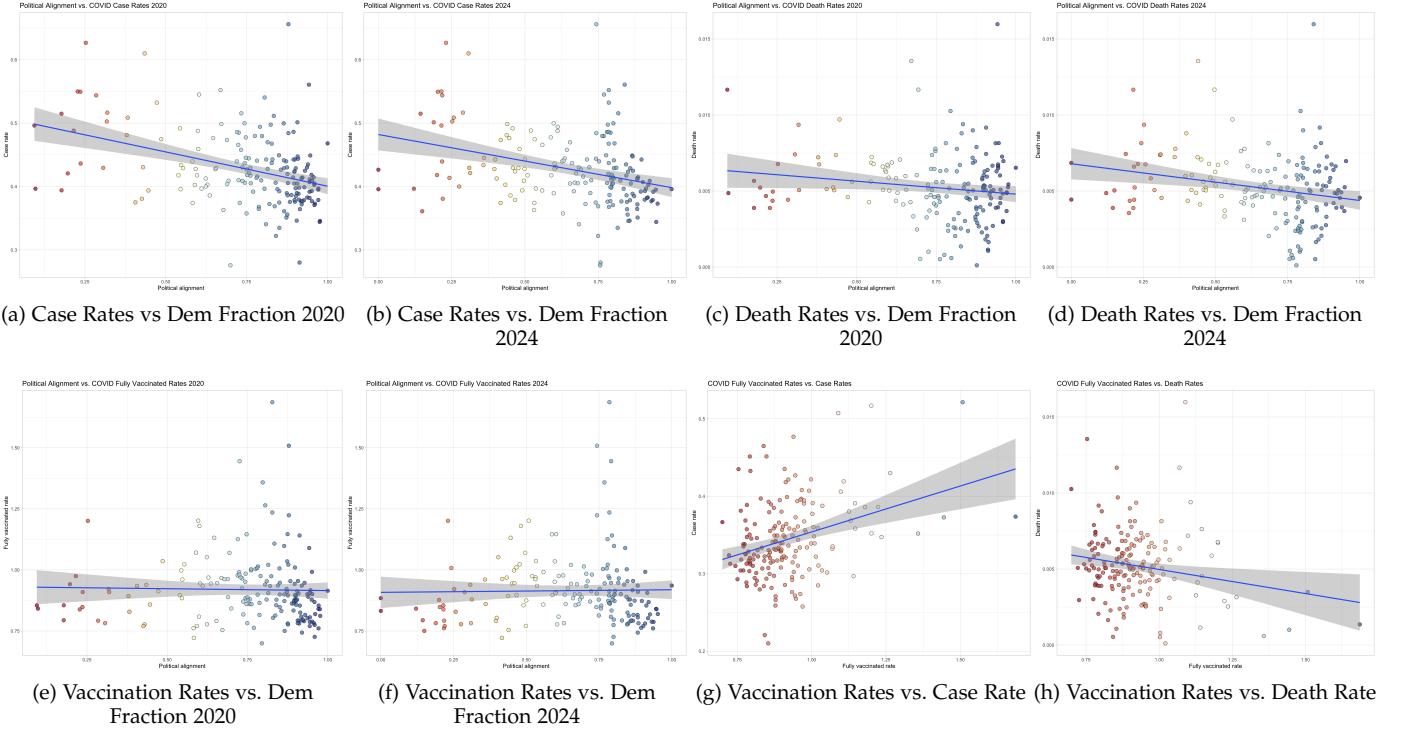


Fig. 6: Correlation plots

Method	Metric	Statistic	P-value
Queen	Case Rate	8.5820	4.6627e-18
Queen	Death Rate	8.5888	4.3948e-18
Queen	Percent Fully Vaccinated	7.9847	7.0413e-16
Queen	Dem Fraction 2020	9.6518	2.416e-22
Queen	Dem Fraction 2024	8.6571	2.4186e-18
IDW	Case Rate	10.7632	2.5686e-27
IDW	Death Rate	11.2114	1.7932e-29
IDW	Percent Fully Vaccinated	6.1612	3.6087e-10
IDW	Dem Fraction 2020	14.9204	1.2136e-50
IDW	Dem Fraction 2024	12.1339	3.4901e-34
6-NN	Case Rate	10.7859	2.0074e-27
6-NN	Death Rate	11.1769	2.644e-29
6-NN	Percent Fully Vaccinated	9.9011	2.0588e-23
6-NN	Dem Fraction 2020	13.0456	3.3644e-39
6-NN	Dem Fraction 2024	10.5155	3.665e-26

TABLE 2: Moran's I Results for Different Metrics and Methods

3.4.2 Getis-Ord G-star

The Getis-Ord Gi* statistic is used to identify clusters of high or low values in spatial data, helping to detect hotspots and cold spots within the dataset. This analysis provides insights into the spatial distribution of COVID-19 metrics and political alignment.

Based on the maps we observe the following:

- Case Rates:** The analysis shows clustering of hotspots in Midtown Manhattan and Staten Island, indicating areas with significantly higher case rates.
- Death Rates:** Clustering of higher death rates is observed in southern Queens, highlighting regions with elevated mortality.
- Vaccination Rates:** Hotspots for vaccination rates are also found in Midtown Manhattan, suggesting areas with higher vaccine uptake.
- Political Alignment (2020 and 2024):** The analysis reveals clustering of Republican voters in southern Staten Island, Brooklyn, and Queens in 2020, with these clusters migrating north in 2024.

3.5 SAR and CAR Models

To account for spatial dependencies, we applied both the Spatial Autoregressive (SAR) model and the Conditional Autoregressive (CAR) model to analyze the relationship between COVID-19 case rates, vaccination rates, and political alignment.

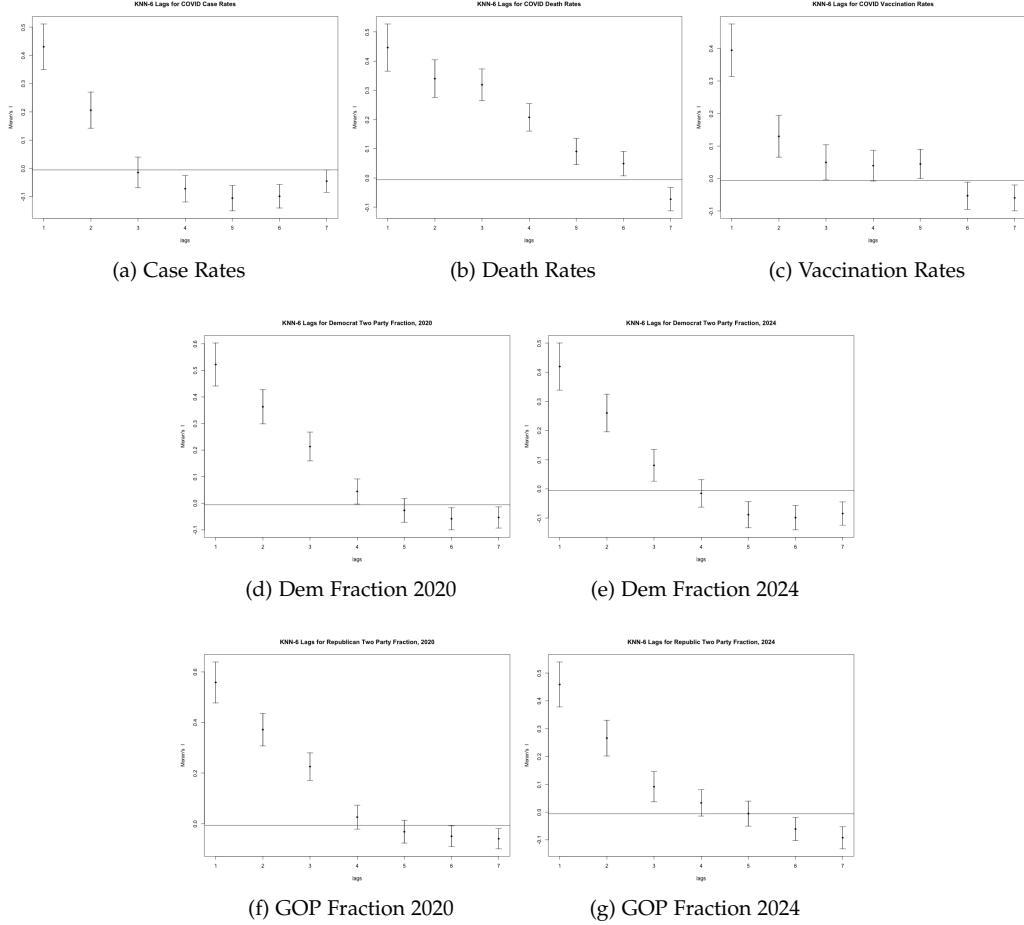


Fig. 7: Lag plots

The SAR model incorporates spatial lag-errors to capture the influence of neighboring regions on the variable of interest. However, in the analysis, the CAR model considers the conditional distribution of each region given its neighbors, providing a better fit for irregular spatial layouts (Table 3).

Model	Variable	Estimate	P-value
SAR	Intercept	-3.46721	1.724e-11
SAR	Democrat Fraction	-0.11151	0.7111
SAR	Percent Fully Vaccinated	4.06339	< 2.2e-16
SAR	AIC	368.21	-
CAR	Intercept	15.875	< 2.2e-16
CAR	Democrat Fraction	9.1847	< 2.2e-16
CAR	Percent Fully Vaccinated	12.666	< 2.2e-16
CAR	AIC	-1681.7	-

TABLE 3: SAR and CAR Model Results

The SAR model shows that case rates and vaccination rates capture most of the spatial behavior, but not voting behavior, with an AIC of 368.21. The CAR model, with an AIC of -1681.7, indicates that vaccination rates and voting fractions better capture the autoregressive behavior in COVID case rates, making it more suitable for the irregular spatial layouts of NYC in this particular context.

4 CONCLUSION

This study aimed to explore the relationship between political alignment and COVID-19 outcomes, including case rates, death rates, and vaccination rates, across New York City neighborhoods. Despite showing no significant correlation between vaccination rates and political alignment, we observed a negative correlation between Democrat voters and COVID-19 case rates, and a positive correlation for Republican voters. This indicates that areas with higher Republican support experienced higher case rates. Similarly, death rates were negatively correlated with Democrat voters.

The Moran's I analysis and correlogram lag structures revealed significant spatial autocorrelation for COVID-19 metrics and political alignment. The 6-NN method provided the best results, due to the irregular spatial shape of the MODZCTA regions and due to the local geography. Additionally, the CAR model outperformed the SAR model, capturing the autoregressive behavior in COVID-19 case rates more effectively. This indicates that the CAR model is better suited for the irregular spatial layouts of NYC. However, we only explored two relevant features: voting behaviour and vaccination behaviour on COVID-19 case rates, and future work may want to consider other features as well.

Overall, our findings suggest that political alignment does influence COVID-19 outcomes, particularly case rates

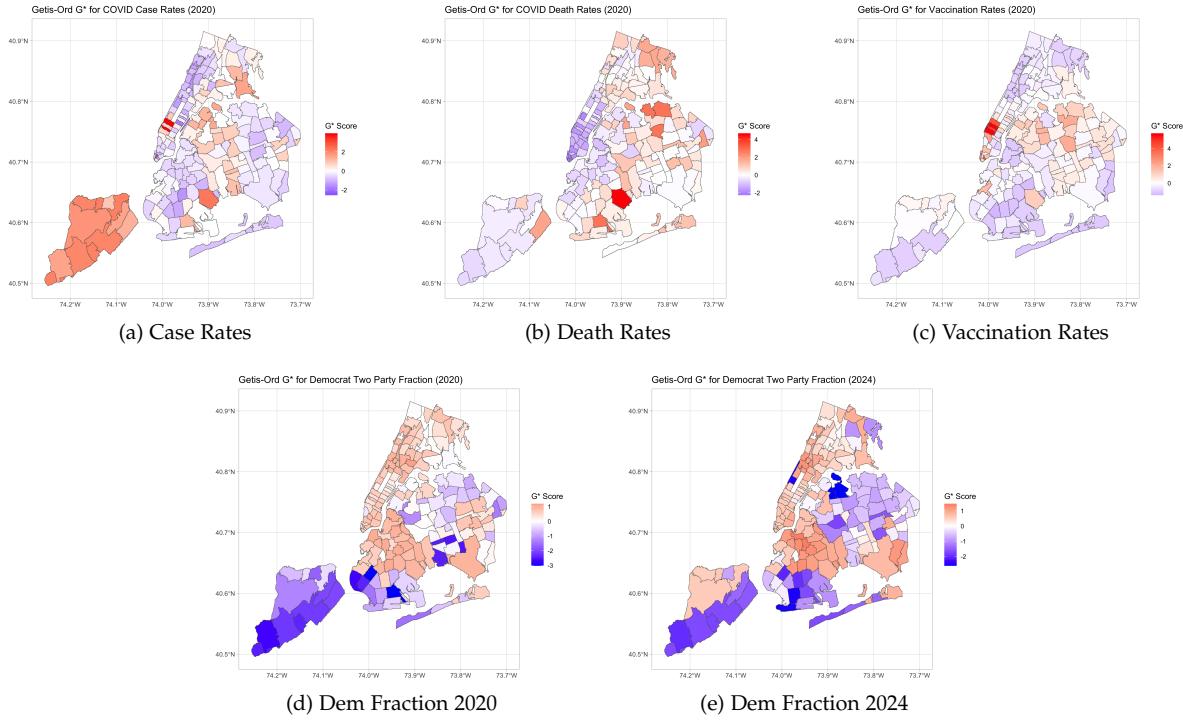


Fig. 8: Getis-Ord G_i^* Mappings

and death rates. However, NYC is a large, well-connected, diverse city, and as a result, the interpretability of COVID-19 case rates cannot be simply chalked-up to so few factors. While this work provides initial insights into the relationship between political alignment and COVID-19 outcomes, it also underscores the complexity of the factors at play. Continued research and more comprehensive data are essential to fully understand and address the public health challenges posed by COVID-19 or similar public health emergencies in a dense urban environment like New York City.

REFERENCES

- [1] "NYC Coronavirus Disease 2019 (COVID-19) Data," NYC Department of Health and Mental Hygiene, 2024, accessed: 2024-11-06. [Online]. Available: <https://github.com/nychealth/coronavirus-data>
- [2] "COVID-19 Vaccination Reporting," NYC Department of Health and Mental Hygiene, 2024, accessed: 2024-11-06. [Online]. Available: <https://github.com/nychealth/covid-vaccine-data>
- [3] "Political and administrative districts," NYC Department of City Planning, 2024, accessed: 2024-11-06. [Online]. Available: <https://www.nyc.gov/site/planning/data-maps/open-data/districts-download-metadata.page>
- [4] T. W. Schneider, "NYC Presidential Elections," 2024, accessed: 2024-11-06. [Online]. Available: <https://toddwschneider.com/maps/nyc-presidential-election-results/#10.23/40.7053/-73.975>
- [5] I. Mohammed, A. Nauman, P. Paul, S. Ganesan, K.-H. Chen, S. M. S. Jalil, S. H. Jaouni, H. Kawas, W. A. Khan, A. L. Vattoth, Y. A. Al-Hashimi, A. Fares, R. Zeghlache, and D. Zakaria, "The efficacy and effectiveness of the COVID-19 vaccines in reducing infection, severity, hospitalization, and mortality: a systematic review," *Hum. Vaccin. Immunother.*, vol. 18, no. 1, p. 2027160, Dec. 2022.