

# STA2016 Final Project Proposal

John Lin

Index Terms—COVID-19, NYC Health Dataset



## 1 INTRODUCTION AND MOTIVATION

The COVID-19 pandemic affected communities of all demographics across the globe, with profound health, social and economic consequences. But did it affect all communities equally? Here we will explore New York City as a microcosm of the effects of the COVID-19 pandemic, from infection and death rates, to vaccination rates, based on the demographic makeup of its residents. New York City was chosen since it is a melting pot of different demographic groups stratified across various social, political and economic scales. Understanding the variation in COVID-19 outcomes, such as vaccination coverage and death rates, across different NYC populations can provide key insights in how social determinants impact public health during crises, including a pandemic. Here, we will examine COVID-19 data, including mortality rates and vaccination uptake, in relation to demographic indicators such as political alignment and median family income across NYC neighborhoods. By integrating these datasets, we aim to uncover patterns that may highlight vulnerabilities relating public health to diverse urban populations.

## 2 RELEVANT DATASETS

### 2.1 NYC COVID Vaccinations

The NYC COVID Vaccinations dataset [1], maintained by the NYC Department of Health on GitHub, provides comprehensive vaccination data across New York City's neighborhoods. This dataset uses the Modified Zip Code Tabulation Areas (MODZCTA) for geographic organization and contains detailed immunization records from the Citywide Immunization Registry and Reporting system. The data encompasses vaccination statistics for different age demographics, specifically distinguishing between adult (18+) and pediatric populations.

The dataset also tracks various vaccination metrics, including the percentage of residents who have received their first dose and those who are fully immunized. It provides both relative (percentage-based) and absolute vaccination numbers across neighborhoods, allowing for granular analysis of vaccination patterns at the community level. The geographic granularity of the MODZCTA system enables detailed spatial analysis and visualization of vaccination trends across different New York City neighborhoods.

Key features of this dataset include:

- Temporal vaccination trends
- Neighborhood-level geographic data
- Both cumulative and incremental vaccination counts

- Distinction between partial and full vaccination status

This dataset could be valuable for analyzing vaccination disparities across different communities.

### 2.2 NYC COVID Data

The NYC COVID dataset [2] is also maintained by the NYC Department of Health and available on GitHub, and provides an extensive overview of COVID-19's impact across New York City's neighborhoods. This dataset is also organized using the MODZCTA.

The dataset includes several critical metrics:

- COVID confirmed case count: The total number of confirmed COVID-19 cases in each neighborhood.
- COVID confirmed case rate: The rate of confirmed COVID-19 cases per 100,000 residents, providing a normalized measure of infection spread.
- COVID death count: The total number of confirmed COVID-19 related deaths in each neighborhood.
- COVID death rate: The rate of confirmed COVID-19 related deaths per 100,000 residents, offering a normalized measure of mortality.
- COVID probable case count: The number of probable COVID-19 cases, which includes cases that meet clinical and epidemiological criteria but lack confirmatory laboratory testing.

This dataset also contains detailed neighborhood-level data and allows for geographic granularity using MODZCTA, which could be valuable for analyzing effectiveness of public health measures stratified across the various boroughs.

### 2.3 NYC Political and Administrative Districts

The NYC Political Districts dataset [3], available on the NYC Planning website, provides detailed information on the boundaries of electoral districts within New York City. This dataset includes the geographic delineations for various political districts, such as congressional, state senate, and state assembly districts.

### 2.4 NYC Presidential Election Map (2020, 2024)

The NYC Presidential Election Map 2020 dataset [4], created by Todd W. Schneider and available on GitHub, provides comprehensive election results for New York City during

the 2020 (and most recently, the 2024) presidential election. This dataset includes several key features:

- **Political boundaries:** The dataset delineates the boundaries of electoral divisions within New York City, allowing for precise geographic analysis of voting patterns.
- **Voting results by electoral division:** It provides detailed voting results for each electoral division, indicating the number of votes cast for the Democratic (Dem) and Republican (GOP) candidates. This data is crucial for understanding the political alignment of different neighborhoods.

This dataset is valuable for analyzing the relationship between political alignment and various factors, such as public health metrics or demographic data. It can be used to explore how political preferences may have influenced or been influenced by the COVID-19 pandemic and other socio-economic factors.

### 3 PROPOSED METHODS

#### 3.1 Analyzing the data

Summarizing one of the sheets for the NYC COVID Vaccine dataset we get summary statistics of numerical variables as well as categorical features in TABLE1 and TABLE2.

Summarizing the COVID-19 data categorized by MOD-ZCTA we get Table 3, with geographic locations given by MODZCTA as well as latitude and longitude, as well as various COVID-19 statistics such as case rate and death rate (Table 4).

Looking at the NYC electoral maps for 2020 and 2024 provided by Schneider, we can see the delineations in electoral zones in Figure 1 and in Figure 2. Something interesting we can see at a glance is the trending of certain electoral districts to become more conservative in 2024 compared to previously in 2020, especially in the Queens borough. It will be interesting to compare COVID vaccination, infection and death rates to political alignment and perhaps other confounding variables such as socioeconomic status and race.

#### 3.2 Further Analyses

Depending on the initial findings, we may conduct additional analyses, such as investigating how demographic factors (e.g., age, income, race) interact with political alignment and COVID-19 metrics.

By employing these methods, we aim to uncover meaningful insights into how political alignment may influence or be influenced by COVID-19 vaccination rates, infection rates, and death rates across New York City.

### REFERENCES

- [1] "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>
- [2] "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>
- [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://toddschneider.com/maps/nyc-presidential-election-results/#10.23/40.7053/-73.975>

TABLE 1  
Quantifiable features in the COVID Vaccine dataset

Statistic	Min	Q1	Median	Mean	Q3	Max
MODZCTA	10001	10301	11109	10810	11361	11697
POP_DENOMINATOR	2685	21576	34454	37473	53118	83122
COUNT_PARTIALLY_CUMULATIVE	176	1902	3351	3873	5209	15216
COUNT_FULLY_CUMULATIVE	3229	20097	30674	33752	47501	88982
COUNT_1PLUS_CUMULATIVE	3405	22606	33699	37625	51827	101463
COUNT_ADDITIONAL_CUMULATIVE	1837	10355	16772	17883	23260	50372
COUNT_BIVALENT_ADDITIONAL_CUMULATIVE	637	3445	5647	7028	9321	31476
PERC_PARTIALLY	4.54	7.50	9.14	11.11	11.40	79.84
PERC_FULLY	69.97	82.89	89.55	91.91	96.70	168.61
PERC_1PLUS	78.67	91.28	98.29	103.03	107.46	248.45
PERC_ADDITIONAL	27.96	39.05	47.13	49.63	59.49	83.28
PERC_BIVALENT_ADDITIONAL	6.33	11.54	15.85	20.25	26.69	48.34

TABLE 2  
Summary of Categorical Variables in the COVID Vaccine Dataset

Variable	Length	Class	Mode
DATE	177	character	character
NEIGHBORHOOD_NAME	177	character	character
BOROUGH	177	character	character
AGE_GROUP	177	character	character
Label	177	character	character

TABLE 3  
Summary of Location Data

Statistic	MODIFIED_ZCTA	lat	lon
Min	10001	40.51	-74.24
Q1	10301	40.67	-73.98
Median	11109	40.73	-73.92
Mean	10810	40.73	-73.92
Q3	11361	40.78	-73.85
Max	11697	40.90	-73.71

TABLE 4  
COVID-19 Statistics

Statistic	Confirmed Cases	Probable Cases	Total Cases	Confirmed Case Rate	Case Rate	Death Rate
Min	1,108	228	1,441	20,992	27,441	11.42
Q1	9,316	2,345	11,482	31,236	39,126	393.13
Median	14,842	3,567	18,085	33,599	41,699	507.31
Mean	16,150	3,721	19,871	34,400	42,685	522.32
Q3	22,643	5,067	27,606	36,815	44,864	645.65
Max	37,419	10,264	46,933	52,114	65,505	1,597.26

Note: Population denominator ranges from 2,972 to 110,370 (mean: 47,076)

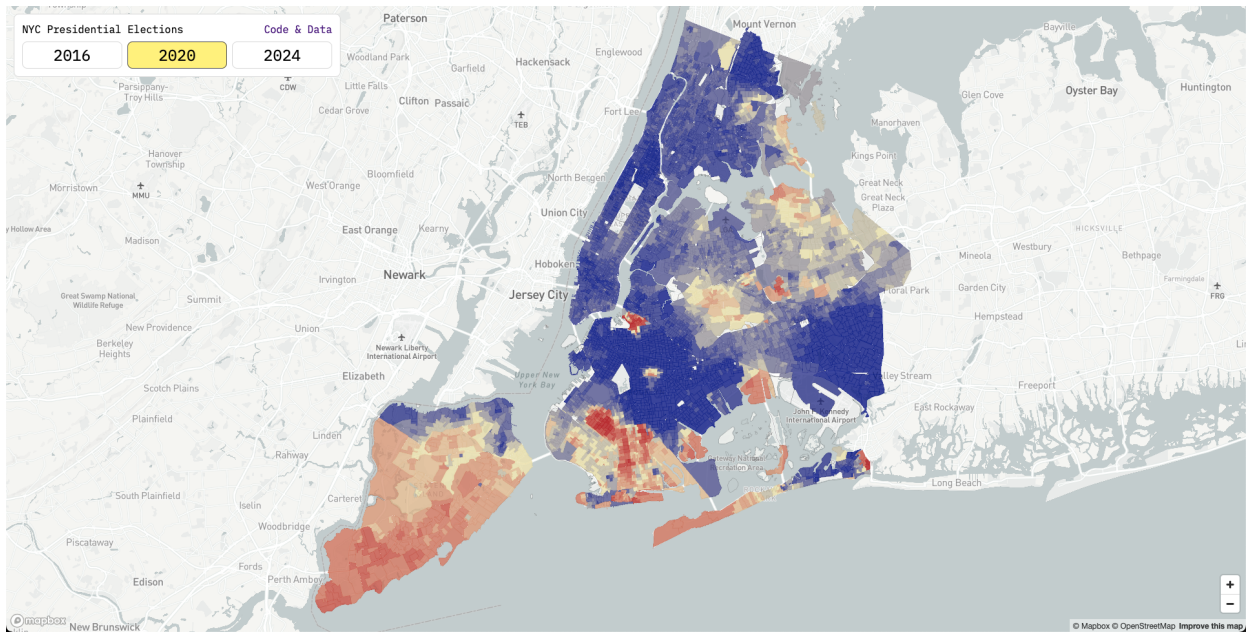


Fig. 1. NYC electoral map 2020

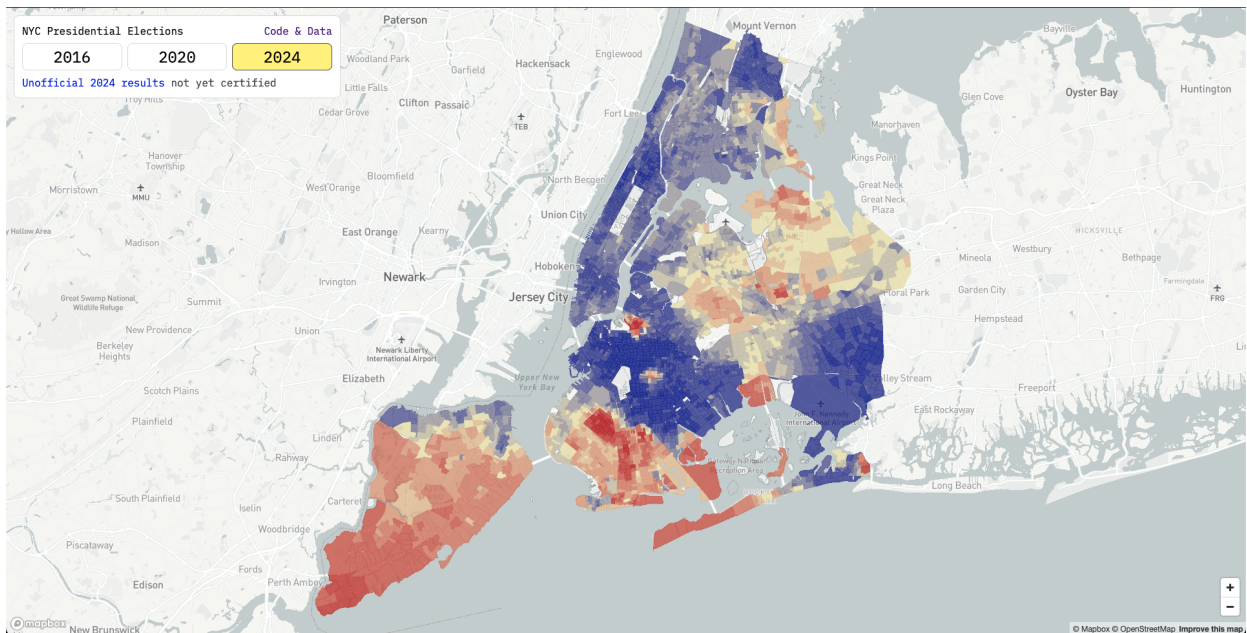


Fig. 2. NYC electoral map 2024