Modeling the spatial factors of COVID-19 in New York City

Yefu Chen[1], Junfeng Jiao, Ph.D.[*1], Shunhua Bai[1], Josiah Lindquist[1]

Abstract

The novel coronavirus (COVID-19) is currently being regarded worldwide as a global pandemic, with New York City as one of the epicenters. There is an emerging demand for precise investigation on this disease, specifically how to slow outbreaks and reopen responsibly. This paper explores and models the spatial factors of COVID-19 in New York City through ordinary least squares regression and geographically weighted regression. Results indicate medical density, green space density, mean distance traveled, male percentage, and commuting (walking, carpooling, and public transit) could correlate to higher rates of COVID-19 positive cases. In contrast, areas with high percentages working from home and white only could correlate to lower rates of COVID-19 positive cases. Additionally, there are distinct associations in various zip code areas or clusters. Overall, this study suggests that public sanitation is critical in disease control in areas with high public transportation demand, and the effect of travel reduction is significant in delaying the outbreak. This study advises policymakers to implement unique policies, preventions measures, and reopening strategies based on localized situations considering COVID-19 outbreaks.

Keywords: Geographically weighted regression, COVID-19, spatial factors

[1] Urban Information Lab, The School of Architecture, The University of Texas at Austin, 310 Inner Campus Drive B7500, Austin, Texas, 78712-1009

1 Background

The first reported novel coronavirus (COVID-19) occurred in December 2019. The disease is highly transmissible between human beings and currently regarded worldwide as a global pandemic. As of April 28, 2020, there were 3,136,253 confirmed cases and 221,436 deaths globally. [1]. Nearly 200 countries and territories have reported cases.

The first COVID-19 case in the US was reported in Snohomish County, Washington State, on January 15, 2020 [2,3]. Nationwide, there were 1,035,765 confirmed cases (33.01% of cases globally) and 59,265 deaths (26.76% of deaths globally) as of April 28 [1]. When compared to other cities, the significant and immediate outbreaks in New York City shocked researchers. As of April 28, New York City had confirmed 167,733 cases (16.19% of cases in the US) and 17,904 deaths (30.21% deaths in the US) [4]. New York City's first case emerged in Manhattan on March 1, and the city reached one hundred thousand cases by April 12.

"Shelter-in-place" or "Stay-at-home" orders have been enacted in NYC to slow the outbreaks. On March 18, the mayor of NYC announced the possibility of a "shelter-in-place order" [5], and the governor of New York State officially signed the order later that month. During the "shelter-in-place order", all people of New York State were required to work from home if their work was deemed "non-essential". Also, residents were asked to practice "social distancing" when in public [6]. All commercial, entertainment, and leisure areas were shut down, and bars and restaurants could provide only takeaway or delivery services.

"Shelter-in-place" policies carry the potential to harm the economy, especially small enterprises and hourly laborers. With fears of maintaining a standard of living for America's workers, there have been rising concerns to reopen the economy, terminating "shelter-in-place" orders [7]. Protesters are not the only ones who want to "reopen" the economy — policy makers too, tasked with balancing public health and economic pressures are also deliberating over reopening strategies. Nevertheless, there are two significant questions, how and when to reopen.

To better answer how to reopen and ensure the efficiency of the existing "shelter-in-place" orders, an understanding of the relative impacts of spatial factors on the dispersal of this virus is essential. Until now, however, few studies have investigated the effects of spatial factors on COVID-19. A level of understanding is critical for predicting outbreaks, crafting public health policy, and conducting responsible reopening [8]. Additionally, most studies analyzing COVID-19 have used large-scale data (e.g., county/provenience level). Data at smaller spatial scales (e.g., ZIP code level) has only recently been made available, allowing for neighborhood/zip code-level analyses, which may provide a new but crucial opportunity to explore the spatial factors/associations between socio-demographic, land use, travel behaviors, and COVID-19 spread at granular scales.

Previous studies evaluated the spatial spread of other pandemic diseases, including Lin and Wen's examination and conclusion of the spatial heterogeneity between socio-demographic factors and the spread of dengue [9]. Similarly, Meng et al. found that spatial analysis is essential for understanding pandemic outbreaks and that the effects differ during various spatial diffusion processes of an epidemic [10]. Al-Ahmadi et al. examined the spatial pattern of MERS-COV, a pandemic in the Middle East, and suggested that precise understanding of the spatial patterns of pandemic diseases is essential for prevention, action, and reopening policies [11].

Therefore, this research is a preliminary study exploring and modeling the spatial factors of COVID-19 in New York City. Three primary domains of explanatory variables, land use, travel behaviors, and socio-demographics, are applied to associate with rates of COVID-19 spread by zip code areas. Conclusively, this study purposes to answer the following questions:

- Are there joint or cluster phenomena present in COVID-19 outbreaks?
- What spatial factors could lead to COVID-19 outbreaks?

The general objective is to provide an analytical perspective on the spatial associations between spatial factors and COVID-19 for infection prevention and control.

2 Materials and Methods

2.1 Study area

New York City (NYC) has been regarded as the first epicenter in the US. With a population of 8,398,748 and 27.96 persons per square miles and as a primary international maritime and airport and financial metropolis [12], NYC risks further outbreaks, focusing on the situation critical. There are five boroughs, Brooklyn, Queens, Manhattan, Bronx, and Staten Island, and 248 zip code areas. Figure 1 presents the population's distribution at the zip code level in NYC based on the American Community Survey (ACS) 2018 dataset. In general, there are higher populations in the Bronx and Brooklyn boroughs, many of whom are highly vulnerable.

NYC is located at the southern tip of New York State, neighboring New Jersey. As the center of the regional metropolitan area, the outbreaks of COVID-19 in NYC could quickly spread to other regions in this metropolitan area. Hence, it is essential for policymakers wanting to slow the spread or begin reopening to understand the spatial factors of COVID-19 outbreaks.

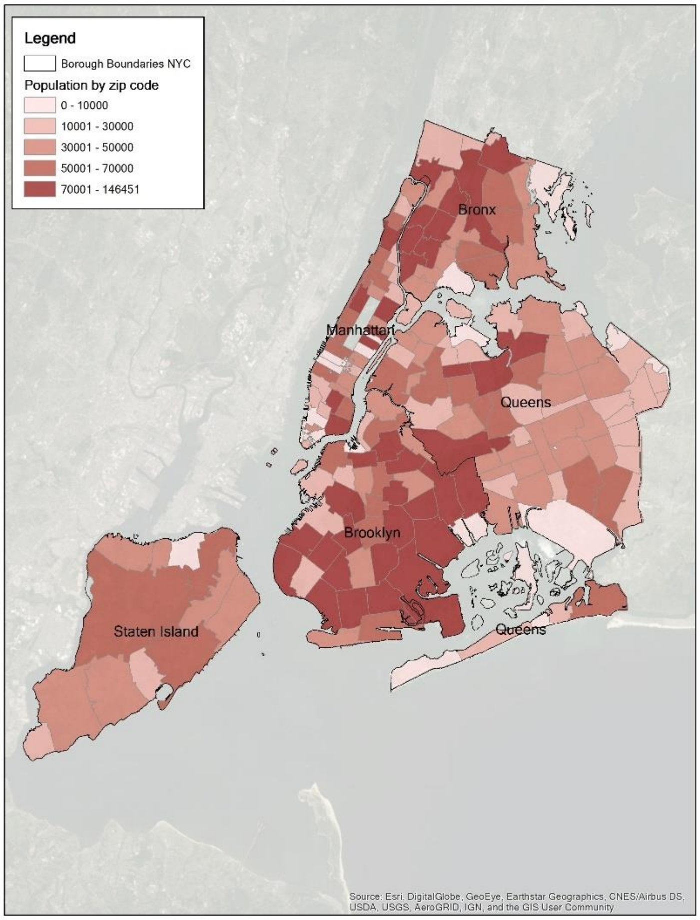


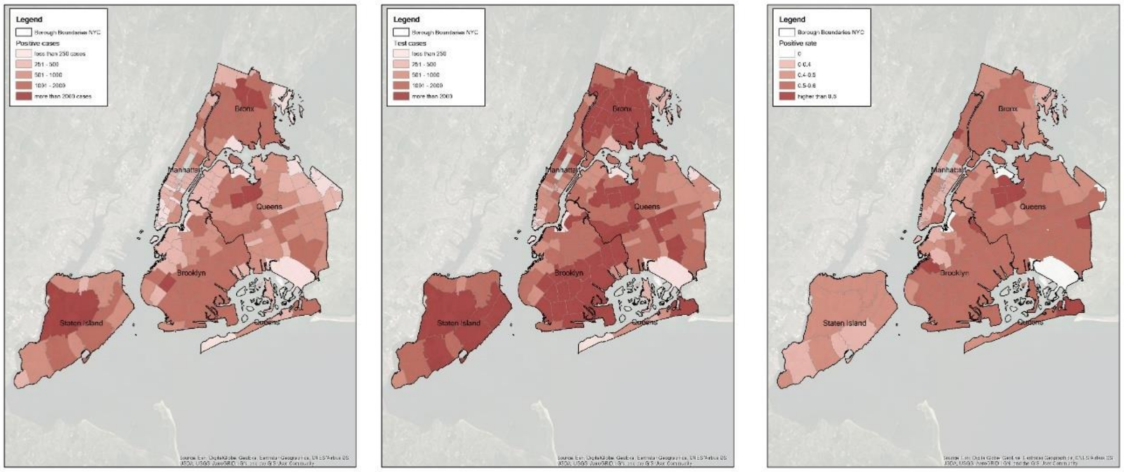
Figure 1. population by zip code areas of NYC

2.2 Data and measurements

2.2.1 Response variables of COVID-19

On April 2, the New York City Department of Health began releasing COVID-19 coronavirus testing data, the daily number of tests, and the number of positive cases by zip code [13]. Figures 2 (a), (b), and (c) present the number of COVID-19 tests, the number of positive cases, and the rate of positive cases by zip code in NYC. Relatively, there are more tests in the Bronx, Staten Island, and Brooklyn boroughs, while the rate of positive cases is highest in Brooklyn, Queens, and the Bronx.

Figure 2 (b) indicates that there are relatively more COVID-19 cases occurring in the Bronx and Brooklyn compared to other boroughs. Hence, there should be more significant concerns for these two boroughs [14]; however, the rates of COVID-19 cases are not consistently high across all zip code areas of the Bronx and Brooklyn. There is value for researchers in focusing on small "clusters" with higher positive rates in Brooklyn, Queens, and the Bronx. Figure 2 (c) shows the zip code level distribution of rates of COVID-19 cases in NYC. The outcome of interest in this study is the rate of COVID-19 cases (the number of COVID-19 positive cases divided by the number of tests to avoid potential bias).



(a) (b) (c)

Figure 2. number of tests and positive cases by zip code areas in NYC

2.2.2 Explanatory variables

The explanatory variables have been compiled from three datasets. Table 1 indicates the related variables and definitions. In general, there are three distinct domains of explanatory variables in this study, land use, travel behaviors, and socio-demographics.

For land-use factors, we first gathered the green spaces and parks from NYC Open Data Portal [15], given the likelihood that they are often places of the congregation [16], and as a result, potential centers of disease spread. While some believe parks and abundant green spaces do not provide reasonable space for social distancing, there are counterarguments that they are conducive to social distancing [17]. The researchers assume that green spaces serve both residents and those from surrounding areas.

For other land use (grocery and medicine), we introduce points of interest (POIs) instead of traditional areas of zoning because these two types of POIs are now severely limited for customers. For instance, only up to 50 customers are allowed inside Whole Foods stores at a time (Fortney 2020). We gathered these POIs from SafeGraph, a company providing information and locations for POIs and activities in the US [18].

For current travel activities, we also utilized a measure, mean distance traveled, from social distancing metrics provided by SafeGraph, which is based on GPS information of anonymous mobile devices [19]. The locations of homes have been defined as places where devices stay overnight for six weeks. The travel activities variable in this study is defined as the median distance traveled from the home by the devices at the block group level initially from April 1 to April 14. Then, this variable is integrated at the zip code level.

Socio-demographic variables are introduced based on an ACS 2018 dataset at the block group level then integrated into zip code areas. In general, socio-demographics in this study cover gender, race, poverty, and commuting behaviors. It should be noted that the commuting behaviors are not current but based on ACS 5-year estimate. In other words, it represents the previous commuting choices of residents.

The initial dataset for this study includes POIs of commercial areas, POIs of nursing homes, POIs of restaurants, hospital beds, median age, median household income, population density, and US citizens' rate in the dataset. Then we apply correlation analysis and stepwise analysis. The final dataset is summarized in Table 1.

Table 1. Definitions of variables

Sources	**Variable name**	**Definitions**
New York City	Green spaces	Green spaces and parks in NYC
Safegraph group	POIs of grocery	POIs include • Bakeries and Tortilla • Beer, Wine, and Liquor Stores • Drinking Places (Alcoholic Beverages) • Grocery and Related Product Merchant • Grocery Stores • Other Miscellaneous Store Retailers • Specialty Food Stores
	POIs of medical	POIs include: • Drugs and Druggists' Sundries • General Medical and Surgical Hospitals • Health and Personal Care Stores • Home Health Care Services • Medical and Diagnostic Laboratories • Nursing Care Facilities (Skilled Nursing Facilities) • Offices of Dentists • Offices of Other Health Practitioners • Offices of Physicians • Other Ambulatory Health Care Services • Personal Care Services
	Distance traveled	Median distance traveled from the home by the devices (excluding any distances of 0) from April 1 to April 14, 2020
2018 American Community Survey of Census Bureau	Gender	Number of males at the tract-level originally, and integrated to zip-code level

	Race	Number of white only people at the tract-level originally, and integrated to zip-code level
	Poverty	Number of people living in below poverty line in NYC
	Working from home	Number of workers who used to work from home at the tract-level originally, and integrated to zip-code level
	Commuting through carpooling, public transit, and walking	Number of workers who used to work through carpooling, public transit, and walking at the tract-level originally, and integrated to zip-code level
	Population	Population at the tract-level originally, and integrated to zip-code level
	Workers	Number of total workers at the tract-level originally, and integrated to zip-code level

2.3 Methods

In this study, we use spatial factors, including land use, travel, and socio-demographic, as the explanatory variables for predicting COVID-19 cases. We first apply the Ordinary Least Squares (OLS) for the global relations and Geographically Weighted Regression (GWR) models for local situations. Table 2 presents the statistical description of the variables in this study.

Table 2. Statistical description of outcome and explanatory variables

Variables	**Statistics**	**Units**	**Mean**	**St. Dev.**	**Min**	**Max**
Outcome	Rate of positive cases	%	0.36	0.237	0	0.71
Explanatory variables	POIs of grocery density	1000*POIs per people	0.001	0.008	0	0.107
	POIs of medicine density	1000*POIs per people	0.007	0.051	0	0.611
	Green spaces density	Square mile per people	0.001	0.008	0	0.118
	Distance traveled	1000 miles	0.01	0.013	0.001	0.07
	Percentage of male	%	0.335	0.22	0	0.58
	Percentage of white only	%	0.257	0.283	0	0.974
	Percentage of poverty	%	0.129	0.121	0	0.495
	Percentage of commuting through walking,	%	0.405	0.290	0	0.789

	carpooling, and public transit					
	Percentage of working from home	%	0.027	0.025	0	0.12

First, we applied OLS regression to explain the global relations between the rate of COVID-19 and our chosen spatial factors. For the pandemic, factors that could influence outbreaks were grouped into three domains: (1) land use, (2) travel active index, (3) socio-demographic factors.

Then we calculate Moran's index before GWR to test the spatial autocorrelations [8]. Moran's index is a metric representing spatial autocorrelation. This metric presents whether similar values appear in adjacent regions; if this metric is not significant (p-value >0.1), the variable is considered randomly distributed [20].

If the spatial correlation is statistically significant, we define the rate of COVID-19 positive cases as clustered. We applied GWR multivariate models to examine the spatial correlation between factors and COVID-19 cases by the postal code areas. Unlike OLS, GWR can observe changes in relations and create maps for investigating and interpreting spatial non-stationarity through changes in estimated coefficients depending on local situations [21].

Through GWR, the spatial influences among postal areas could be assessed. The GWR model can be given by

$$Rate_j = \beta_{0j} + \beta_{ij} * Factor_{ij} + \varepsilon_j$$

where j indicated the parameters that are estimated at each postal area.

Based on ArcGIS 10.5, we apply the adaptive kernel with AICc estimated bandwidth and adaptive kernels settings for GWR models because postal areas are spatially inhomogeneous in the study area.

3 Results

3.1 OLS model

Table 3 presents the results from the OLS regression. In general, areas with higher POIs of medicine, green spaces density, median distance traveled, male percentage, percentage of commuting through walking, carpooling, and public transit are significantly associated with a higher rate of COVID-19 positive cases. However, areas with a higher percentage of white only, and residents working from home were associated with a lower rate of positive COVID-19 cases.

For every one-unit increase of medical POIs density, there was a 0.463 increase in the average rate of COVID-19. Also, a one-unit green spaces density increase is significantly associated with a 3.780 increase in the average rate of positive cases.

The current travel behavior measurement is also significant. For each unit increase in median distance traveled, we saw a 1.060 increase in the average rate of positive cases. It indicates that the intensity of travel behavior could be crucially associated with the COVID-19. This argument is consistent with previous studies [22,23].

The role of socio-demographics is most significant. First, the male percentage is significantly positive. A one-percent increase in the male population is significantly associated with a 0.898 increase in the average rate of positive cases, which indicates higher rates of COVID-19 cases in areas associated with higher male percent.

Similar to gender, a one percent increase of residents commuting through walking, carpooling, and public transit is significantly associated with a 0.350 increase in the average rate of positive cases, indicating the disinfection of public transportation and pedestrian facilities is necessary.

Moreover, areas with higher percentages of white only see lower positive rates of COVID-19. It could indicate the need for concerns on the minority or vulnerable groups residentially gathering. The rate at which a population works from home seems to be an influential factor in lower positive rates of COVID-19, which may be evidence that working from home could dramatically slow the outbreaks of COVID-19.

Table 3. OLS regression results

	Outcome: Rate of COVID-19 positive cases
POIs of grocery density	1.136
	(1.086)
POIs of medicine density	0.463***
	(0.165)
Green spaces density	3.780***
	(0.652)
Median distance traveled	1.060**
	(0.517)
Percentage of male	0.898***
	(0.081)
Percentage of white only	-0.118***
	(0.033)
Percentage of poverty	-0.045
	(0.085)
Percentage of commuting through walking, carpooling, and public transit	0.350***
	(0.061)
Percentage of working from home	-2.364***
	(0.399)
Constant	0.002
	(0.012)
Observations	248
Adjusted R square	0.605
Akaike Inf. Crit. (AIC)	-423.366

Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

3.2 GWR models

Before applying GWR models, we investigated positive spatial relations among the distributions of positive cases rate (Moran's I = 0.637, $p < 0.01$). The results of GWR are listed in Table 4. In general, GWR is regarded as more suitable than OLS (AIC = -0.560.12 < -423.36). Also, the adjusted R-square indicates that around 92% of positive cases rates are fit.

Table 4. GWR results

Statistic	Mean	St. Dev.	Min	Pctl (25)	Pctl (75)	Max
LocalR2	0.894	0.016	0.828	0.893	0.901	0.918
Intercept	0.003	0.007	-0.017	0.001	0.007	0.038
POIs of grocery density	1.006	0.817	-1.803	0.96	1.474	1.964
POIs of medicine density	0.458	0.058	0.403	0.419	0.481	0.708
Green spaces density	5.971	2.915	1.968	3.794	6.974	18.514
Median distance traveled	0.915	0.528	-1.599	0.593	1.233	2.24
Percentage of male	0.846	0.105	0.598	0.767	0.921	1.096
Percentage of white only	-0.134	0.043	-0.187	-0.167	-0.104	-0.004
Percentage of poverty	-0.076	0.055	-0.119	-0.103	-0.065	0.327
Percentage of commuting through walking, carpooling, and public transit	0.369	0.042	0.199	0.331	0.407	0.418
Percentage of working from home	-2.109	0.385	-2.983	-2.397	-1.842	-0.002
Adjusted R square	0.928					
AIC	-560.12					

In detail, Figures 3, 4, and 5 present local coefficients and p-values among all the explanatory variables and outcome variables for each postal area generated through the GWR models. For land-use factors, the association between grocery POIs density and rate of positive cases is only significant and positive in middle and upper Manhattan, most of the Bronx, and along the boundaries between Brooklyn and the Bronx (Figure 3 (a)).

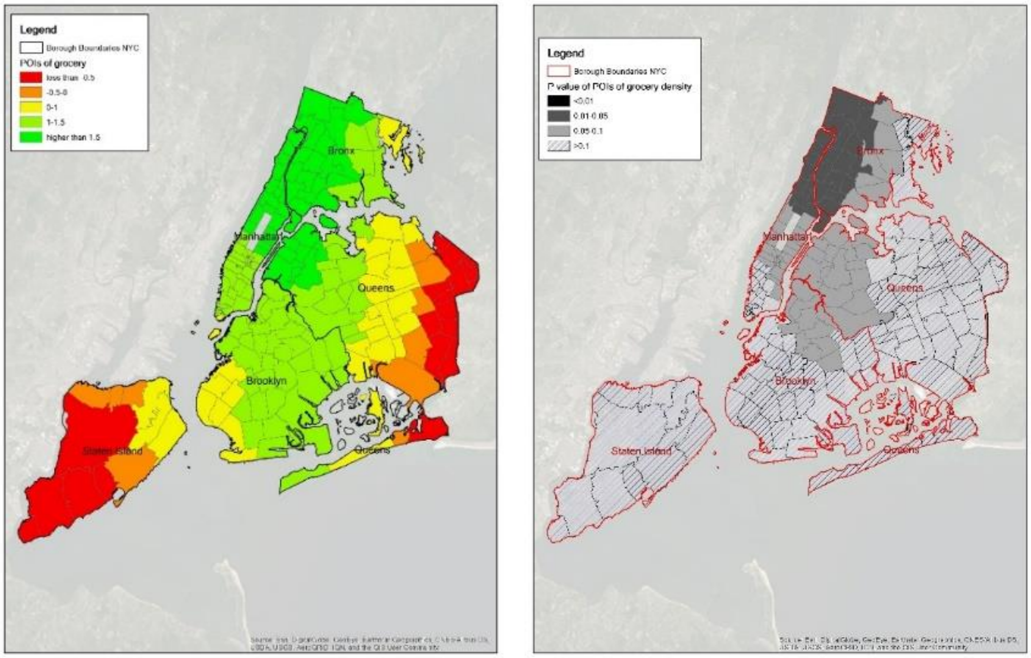
Additionally, both medical POIs density and green spaces density are globally (from OLS) and locally positively correlated with COVID-19, as shown in Figure 3 (b) and (c). Brooklyn, Queens, and Staten Island have more significant coefficients for medical POIs density and rates of positive cases than Manhattan and surrounding areas. In other words, in Brooklyn, Queens, and Staten Island, the positive correlations surpass those in other regions. In contrast, Manhattan, Staten Island, and Brooklyn have more significant coefficients for green spaces' density and higher rates of positive cases than others. Regarding the association between green space density and COVID-19, with a high population density in Manhattan, there could be a high COVID-19 risk for Manhattan residents outside their homes and visiting recreational areas.

Figure 4 demonstrates the coefficient and p-value between median distance traveled and the rate of positive cases. It indicates that areas with higher distance traveled, or more activity, are associated with higher positive rates. This effect is significant in most of Manhattan, west of the Bronx, and Queens. A one-unit increase in the median distance traveled in the Bronx and Queens yields higher increases in the rate of positive cases.

Furthermore, most socio-demographic factors, excluding poverty percent, are significant in this study. Figure 5 presents coefficients and p-values for socio-demographic factors and rates of positive cases. Postal areas with higher male percentages have significantly higher rates of COVID-19 cases.

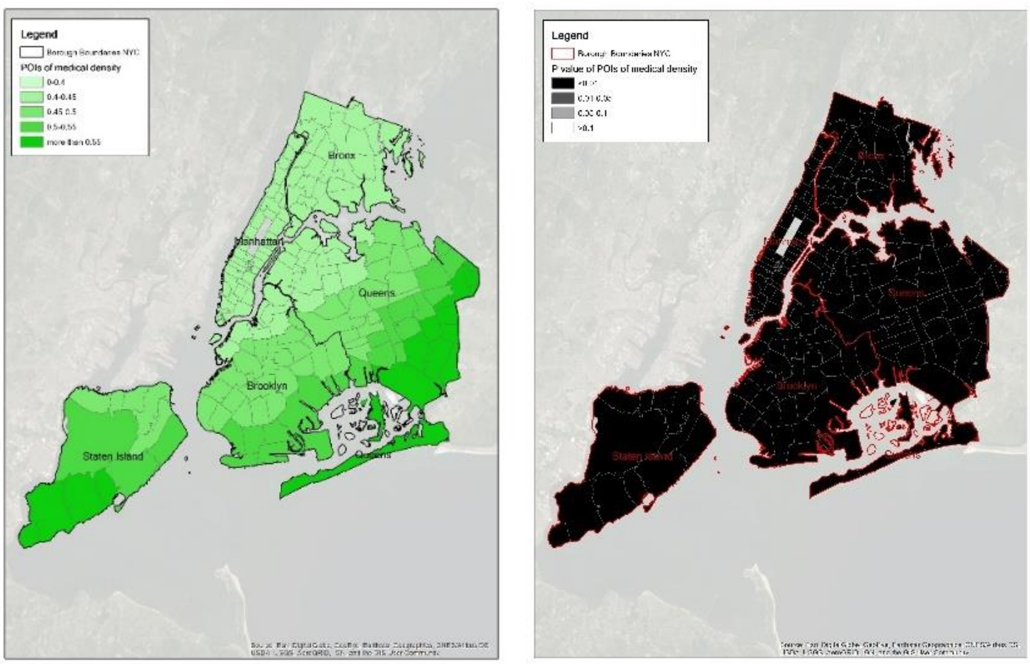
In terms of the spatial variation of coefficients, a one-unit increase of male percentage is associated with higher rates of positive cases in some parts of the Bronx and Brooklyn, and all parts of Queens, as shown in Figure 5 (a). In contrast, a one-unit increase of white only percentage is associated with lower rates of positive cases in some parts of the Bronx, Brooklyn, and all parts of Queens, as presented in Figure 5 (b). Although the poverty rate is not significant across the whole of NYC, Figure 5 (c) indicates that the poverty percentage is significantly negative in north Queens. It indicates that there is no evidence of a higher poverty rate leading to a higher rate of positive COVID-19 cases.

Figure 5 (d) demonstrates that postal areas with a higher percentage of commuting through walking, carpooling, and public transit have higher rates of positive cases, excluding Staten Island. Also, a one-unit increase in commuting through walking, carpooling, and public transit percentage is associated with more rates of positive cases in Manhattan. The percentage of residents working from home could also play a significant and negative role in associations with rates of positive cases in most of NYC, excluding Staten Island. Working from home percentages in Brooklyn and Queen are associated with lower rates of positive cases.



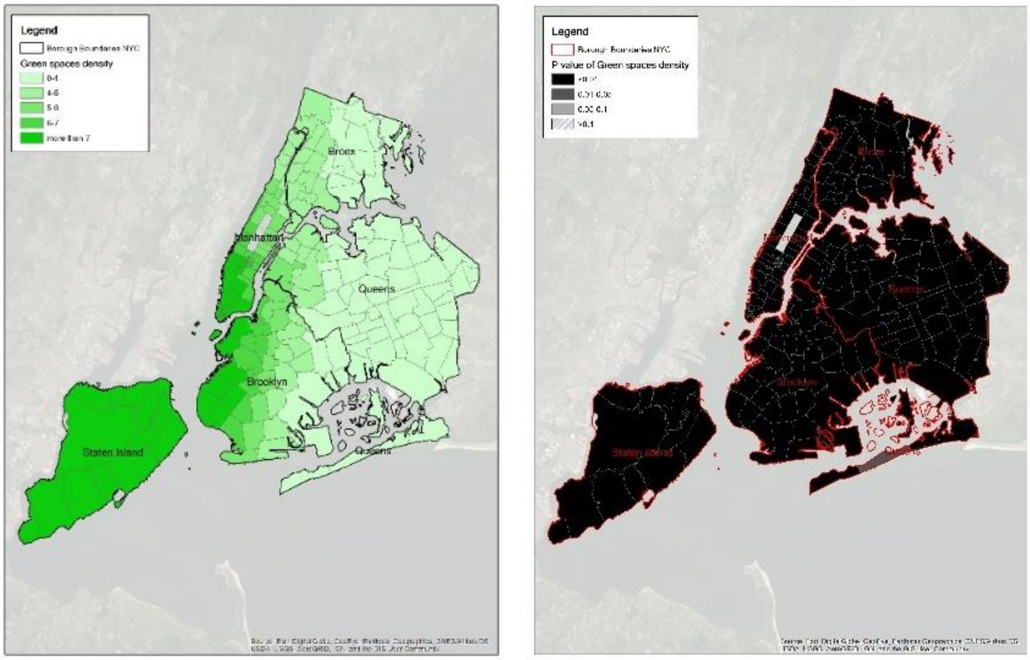
Legend
Borough Boundaries NYC
POIs of grocery
-0.5-0
0-1
1-1.5
higher than 1.5
Bronx
Manhattan
Queens
Brooklyn
Staten Island
Legend
Borough Boundaries NYC
P value of POIs of grocery density
<0.01
0.01-0.05
0.05-0.1
>0.1

(a)



Legend
Borough Boundaries NYC
POIs of medical density
0.4-0.45
0.45-0.5
more than 0.55
Bronx
Manhattan
Queens
Brooklyn
Staten Island
Legend
P value of POIs of medical density
>0.1

(b)



(c)

Figure 3. Coefficients and p values between land use factors and rates of COVID-19 positive cases by zip code areas

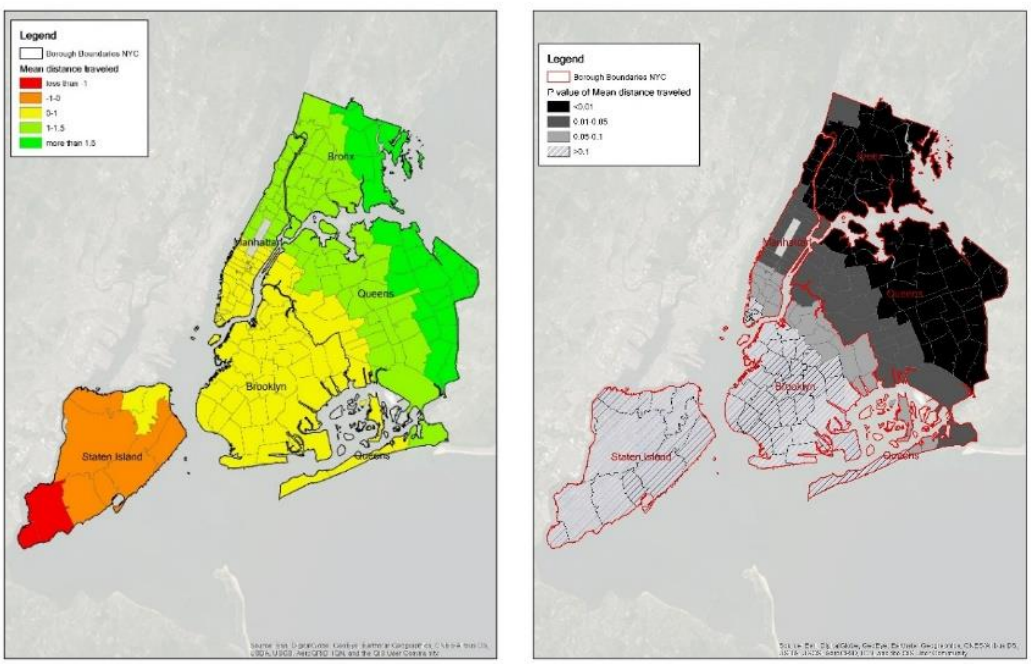
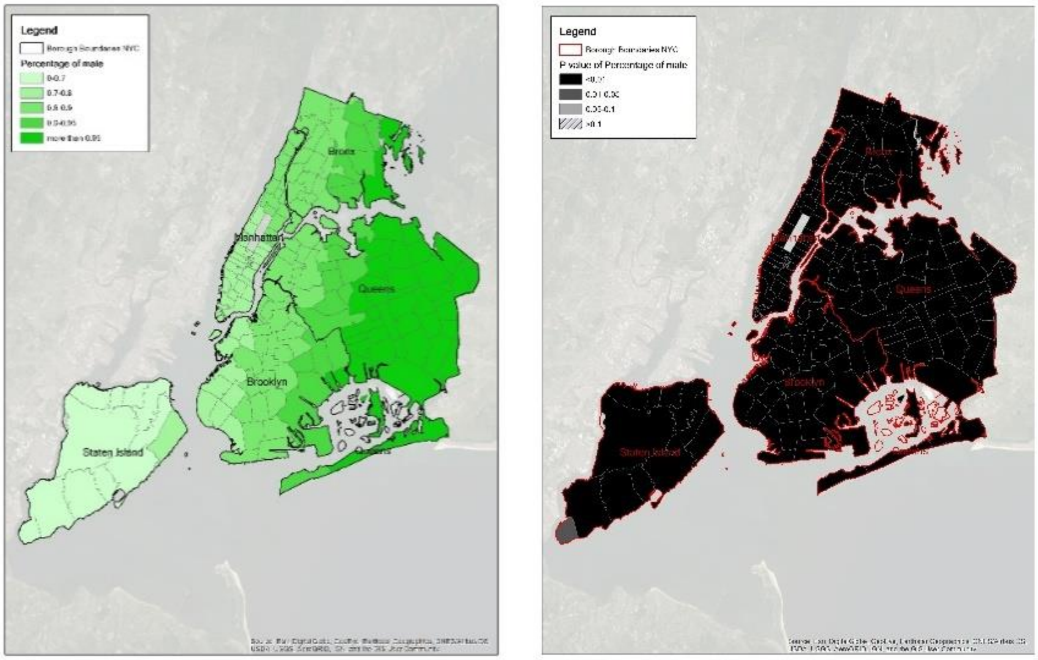
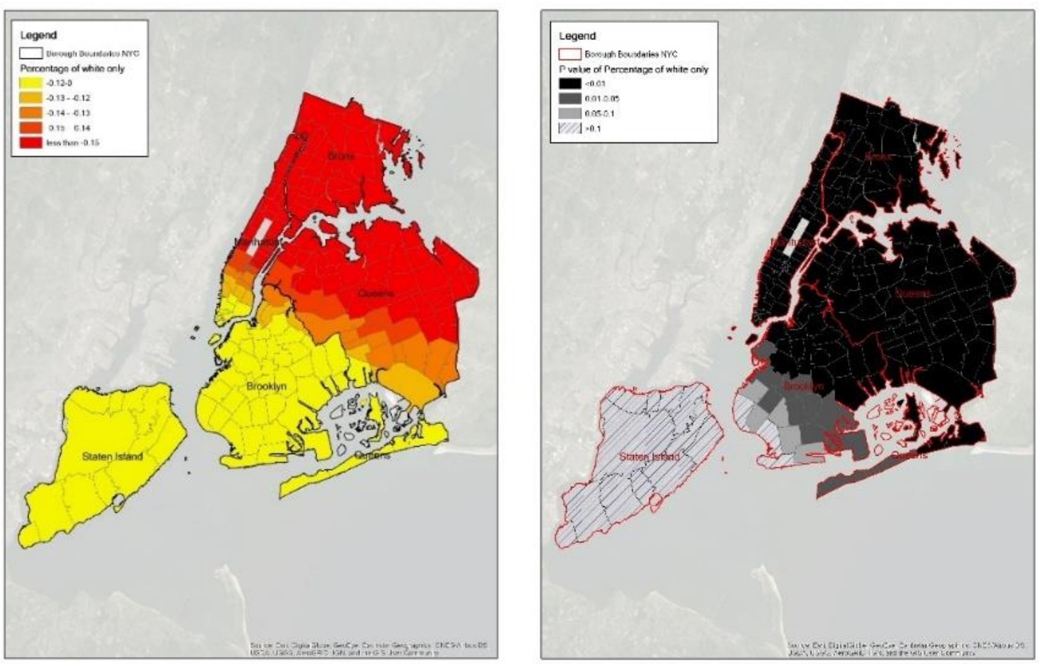


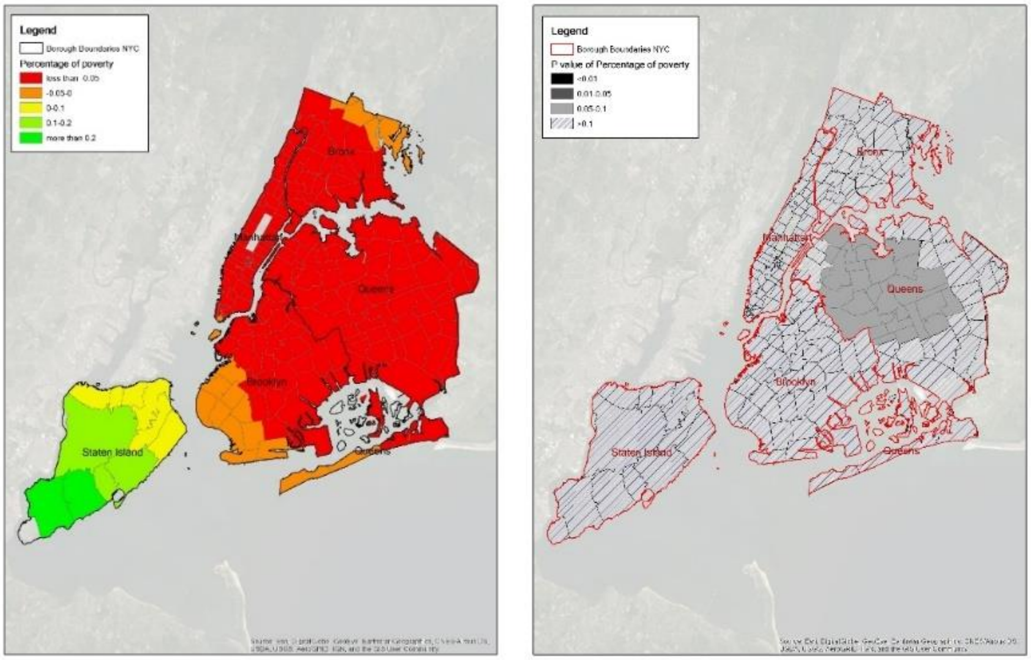
Figure 4. Coefficients and p values between timely travel behavior and rates of COVID-19 positive cases by zip code areas

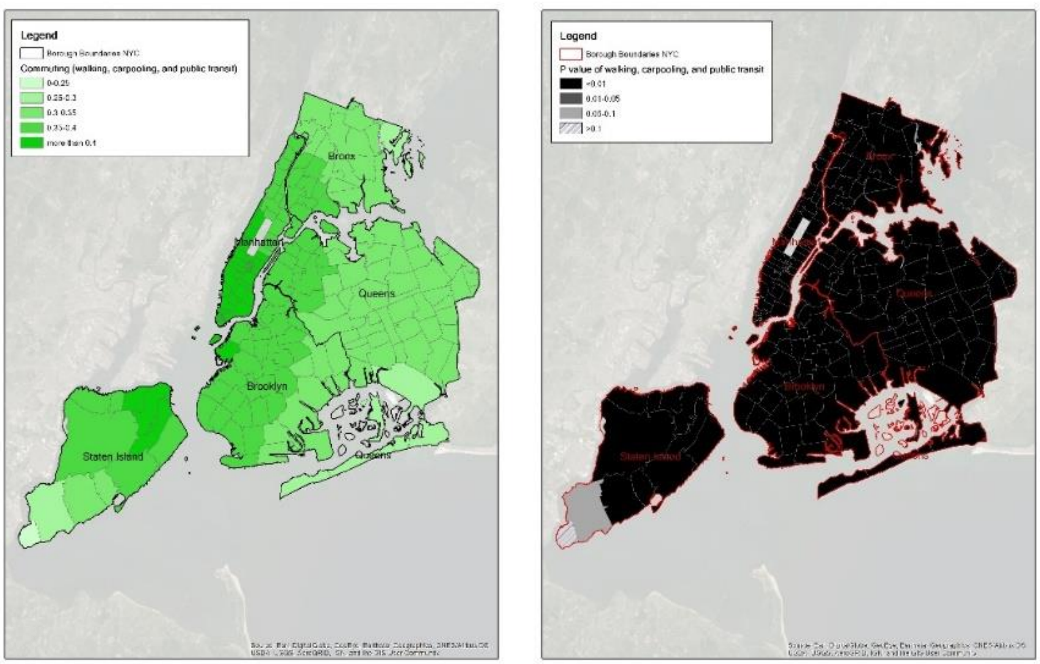

Legend
Percentage of male
0-0.7
0.7-0.8
Bronx
Manhattan
Queens
Brooklyn
Staten Island
Legend
Borough Boundaries NYC
P value of Percentage of male
0.05-0.1
>0.1

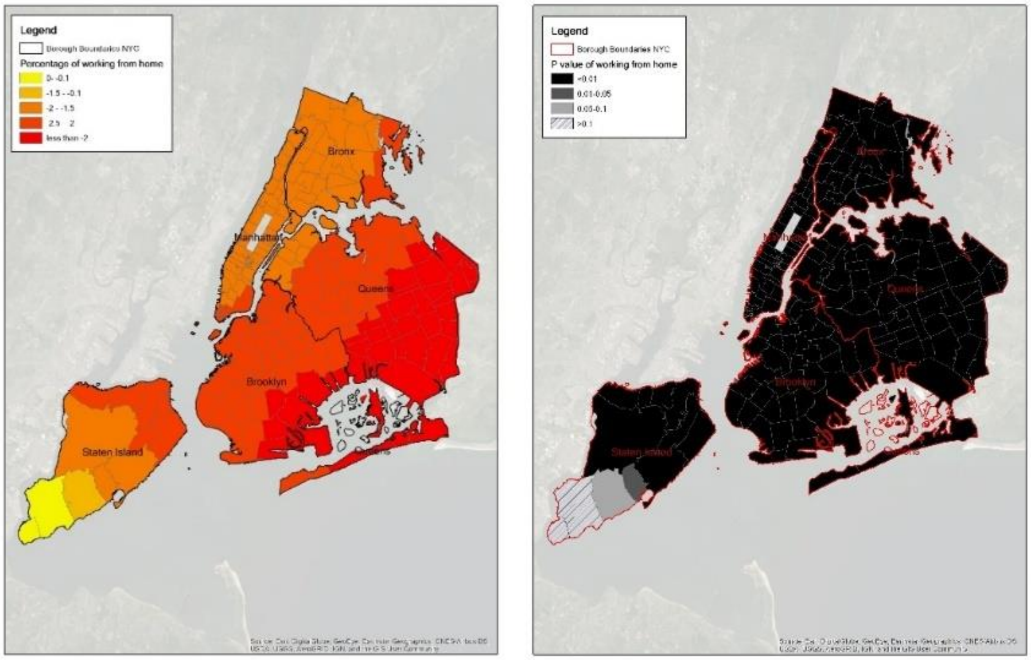
(a)


Legend
Percentage of white only
-0.12-0
-0.13 - -0.12
-0.14 - -0.13
less than -0.15
Bronx
Queens
Brooklyn
Staten Island
Legend
Borough Boundaries NYC
P value of Percentage of white only
<0.01
0.01-0.05
0.05-0.1
>0.1

(b)


Legend
Borough Boundaries NYC
Percentage of poverty
less than -0.05
-0.05-0
0-0.1
0.1-0.2
more than 0.2
Bronx
Manhattan
Queens
Brooklyn
Staten Island
Queens
Legend
Borough Boundaries NYC
P value of Percentage of poverty
<0.01
0.01-0.05
0.05-0.1
>0.1
Bronx
Manhattan
Queens
Brooklyn
Staten Island
Queens
(c)


Legend
Borough Boundaries NYC
Commuting (walking, carpooling, and public transit)
0-0.25
0.25-0.3
0.3-0.35
0.35-0.4
more than 0.4
Bronx
Manhattan
Queens
Brooklyn
Staten Island
Queens
Legend
Borough Boundaries NYC
P value of walking, carpooling, and public transit
<0.01
0.01-0.05
0.05-0.1
>0.1
(d)



(e)

Figure 5. Coefficients and p values between socio-demographics and rates of COVID-19 positive cases by zip code areas

4 Discussion

This study is an explanatory study that models the spatial factors of COVID-19 cases at the zip code level in NYC. Overall, OLS results indicate that areas with high POIs of medical density, green spaces density, median distance traveled, male percentage, percentage of commuting through walking, carpooling, and public transit all have a higher rate of COVID-19 positive cases, relatively. In comparison, areas with a higher percentage of white only and the percentage of residents working from home is associated with a lower rate of positive COVID-19 cases. The results of the GWR are consistent with the results of OLS and examine the models' various local coefficients for each postal area. It implies that the degree of impact of land use, travel behaviors, and socio-demographics on rates of COVID-19 differ for each zip code area in the study area. Critical limitations of this study, such as biases of both positive case rates and median distance traveled, are crucial. Also, information that is currently lacking about COVID-19 patients is essential for studies in the future.

Albeit the limitations, this study is meaningful for policymakers who wish to slow outbreaks and reopen responsibly. First, factors whose positive correlations in the core area, Manhattan, surpass other regions, indicate that policies involving COVID-19 prevention should be based on local conditions in the region. For instance, some associations (green spaces density and commuting through walking, carpooling, and public transit) are generally strong (with high absolute coefficient values) in Manhattan, the heart of NYC. Also, the primary green space of NYC, Central Park, is in Manhattan, and the public transit system is concentrated in Manhattan. These could lead to rising risks of further COVID-19 outbreak and disease spread. Hence, public sanitation is critical in disease control in areas (e.g., Manhattan) with high public transportation demand. Policymakers have already noted this risk. Beginning on May 2, operators of the NYC subway began disinfecting trains each night [24]. Furthermore, there is a need for strict social distancing measures within green spaces in Manhattan.

Some associations (POIs of medical density, median distance traveled, male percent, white only percent, and working from home) is generally strong in peripheral areas (the Bronx, Brooklyn, and Queens). For POIs of medical density, this study suggests that POIs of medical density are significantly and positively correlated with rates of COVID-19 positive cases, and this effect is more potent in peripheral areas. The influence of median distance traveled is also worth noting. It is not surprising to find that the coefficient of median distance traveled is significant in peripheral areas. People who live in these areas may have to spend more time on public transportation, which leads to a higher risk of contracting the virus. Also, areas with a high percentage of white only have shown lower rates of COVID-19, and this effect is also more significant in Upper Manhattan, Queens, and the Bronx, where more minorities reside. This information is critical for policymakers, and the researchers suggest more attention be paid to where the minority live.

We want to emphasize that working from home percentage is a crucial factor, as seen both in the OLS and GWR. Areas with a high percentage of residents working from home have a much lower rate of COVID-19 positive cases. This influence is more significant in peripheral areas than in other areas, especially in Brooklyn and Queens. First, this result indicates that with fewer people needing to leave the house for work, the risk of exposure to COVID-19 is reduced. Results of GWR present that working from home effects should be varied across postal areas. Policymakers may need to think about different strategies for different areas. For instance, it is not wise to treat Brooklyn and Staten Island similarly.

Staten Island is usually regarded as a recreational center in this region, not only for visitors from NYC but also for New Jersey. As a result, areas with high green space density in Staten Island are more likely to be exposed to the virus. Also, green spaces areas of Staten Island could be possible points for COVID-19 transmission. Similarly, commuting through walking, carpooling, and public transit should be worth further considerations in some neighborhoods (e.g., St. George) of Staten Island.

5 Conclusion

This study explores and models the spatial factors of COVID-19 by postal areas in NYC through OLS and GWR. Results indicate that high of levels of medical density, green space density, mean distance traveled, male percentage, commuting (walking, carpooling, and public transit) may correlate to higher rates of COVID-19 positive cases; while areas with higher a working from home percentage and white only percentage often show lower rates. Also, it suggests that future study considers taking a local approach (e.g., GWR), allowing an efficient investigation of local variations across zip codes. Additionally, it may be more productive for policymakers to implement strategies depending on local situations instead of globally.

Moreover, this study suggests that policymakers may need to pay attention to policy involving public transit to keep it from becoming a hotbed for this disease. We found that travel reduction generally delayed outbreaks of COVID-19. Specifically, Manhattan ought to be most cautious of public transit and social distancing in green spaces given the concentration. For peripheral areas, medical POIs, distance traveled, and working from home measurements should be essential for policymakers right now and in reopening stages. Results on Staten Island differ largely from the rest of NYC. Green spaces density and public transit on this island should be more thoroughly considered.

Some detailed limitations may be worth further research. First, this study is based on COVID-19 data from NYC. It may not include the number of suspected cases or cases who chose to not be treated, which should lead to statistical bias. Second, this study mainly uses ACS data and SafeGraph data. It would be more reliable if current transportation ridership data and travel behaviors data were utilized. Third, this study chooses NYC as a case study but ignores the effects of surrounding cities (e.g., Newark). Also, it could be more effective to explore a greater number of cities, relating COVID-19 cases by zip code areas or community-level around the US or even worldwide.

Acknowledgement

This project was supported by the Good System Grand Challenge and the Cooperative Mobility for Competitive Megaregions (CM2) center both at The University of Texas at Austin.

References

1. WorldMeter: COVID-19 Coronavirus Pandemix. In: WorldMeter. (Accessed May 2, 2020) Available at: https://www.worldometers.info/coronavirus/
2. Washington State Department of Health: 2019 Novel Coronavirus Outbreak (COVID-19). (Accessed January 30, 2020) Available at: https://www.doh.wa.gov/emergencies/coronavirus
3. Centers for Disease Control and Prevention: First Travel-related case of 2019 novel coronvirus in the US. In: CDC newsroom. (Accessed January 21, 2020) Available at: https://www.cdc.gov/media/releases/2020/p0121-novel-coronavirus-travel-case.html
4. Dong, E., Du, H., Gardner, L.: An interactive web-based dashboard to track COVID-19 in real time. The Lancet (2020)
5. Kopecki, D., Lovelace, B.: NYC mayor pushes for 'shelter-in-place' order as coronavirus cases surge to almost 1,900. (Accessed March 18, 2020) Available at: https://www.cnbc.com/2020/03/18/nyc-mayor-pushes-for-shelter-in-place-order-as-coronavirus-cases-surge-to-almost-1900.html
6. Plltt, A.: New York issues a 'pause' on nonessential services and gatherings. Here's what that means. (Accessed April 7, 2020) Available at: https://ny.curbed.com/2020/3/20/21187022/coronavirus-new-york-shutdown-shelter-in-place
7. BBC: Coronavirus: US faced with protests amid pressure to reopen. (Accessed April 29, 2020) Available at: https://www.bbc.com/news/world-us-canada-52348288#
8. Kang, D., Choi, H., Kim, J.-H., Choi, J.: Spatial epidemic dynamics of the COVID-19 outbreak in China. International Journal of Infectious Diseases, 96-102 (2020)
9. Lin, C.-H., Wen, T.-H.: Using Geographically Weighted Regression (GWR) to Explore Spatial Varying Relationships of Immature Mosquitoes and Human Densities with the Incidence of Dengue. International Journal of Environmental Research and Public Health, 2798-2815 (2011)
10. Meng, B., Wang, J., Liu, J., Wu, J., Zhong, E.: Understanding the spatial diffusion process ofsevere acute respiratory syndrome in Beijing. Journak if the royal institute of public health, 1080-1087 (2005)
11. Al-Ahmadi, K., Alahmadi, S., Al-Zahrani, A.: Spatiotemporal Clustering of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Incidence in Saudi Arabia, 2012–2019. International Journal of Environment Research and Public Health (2019)
12. World Population Review: New York City Population. In: World Population Review. (Accessed April 20, 2020) Available at: http://worldpopulationreview.com/us-cities/new-york-city/
13. NYC Department of Health and Mental Hygiene: Coronavirus Data. In: GitHub. (Accessed April 24, 2020) Available at: https://github.com/nychealth/coronavirus-data
14. Morgan, R.: The Bronx, long a symbol of American poverty, is now New York City's coronavirus capital. (Accessed April 21, 2020) Available at: https://www.washingtonpost.com/national/bronx-new-york-yankee-stadium-coronavirus/2020/04/21/3b38b460-8182-11ea-a3ee-13e1ae0a3571_story.html

15. NYC OpenData: Green Spaces. In: NYC OpenData. (Accessed April 24, 2020) Available at: https://data.cityofnewyork.us/Recreation/Green-Spaces/mwfu-376i
16. Moon, J.: People are still gathering, breaking coronavirus isolation rules, some say. (Accessed April 1, 2020) Available at: https://www.thetelegraph.com/news/article/Residents-say-COVID-19-rules-not-being-followed-15172735.php
17. Carmichael, S.: Parks and green spaces can be great calmers only if there is social distancing. In: The Print. (Accessed April 6, 2020) Available at: https://theprint.in/opinion/parks-and-green-spaces-can-be-great-calmers-only-if-there-is-social-distancing/395945/
18. SafeGraph: SafeGraph POIs dataset. (Accessed May 2, 2020) Available at: https://www.safegraph.com/?utm_campaign=branded_exact&utm_term=safegraph&rm_ci=8481213813&rm_gid=88800951800&utm_medium=cpc&utm_source=google&utm_content=404850132212&utm_term=safegraph&rm_kw=safegraph&rm_mt=e&rm_n=g&rm_cr=404850132212device=c&creative=40
19. SafeGraph: Social Distancing Metrics. (Accessed April 17, 2020) Available at: https://docs.safegraph.com/docs/social-distancing-metrics
20. Moran, P.: Notes on Continuous Stochastic Phenomena. Biometrika, 17-23 (1950)
21. Fotheringham, A. S., Brunsdon, C., Charlton, M.: Quantitative Geography: Perspectives on Spatial Data Analysis. Sage:, Newbury Park (2000)
22. Vaidya, R., Herten-Crabb, A., Spencer, J., Moon, S., Lillywhite, L.: Travel restrictions and infectious disease outbreaks. (2020)
23. Merinuk, M.: Coronavirus outbreak leads to travel deals and fewer crowds — with a dose of risk. (Accessed March 6, 2020) Available at: https://www.today.com/money/travel-deals-flights-hotels-rise-coronavirus-spreads-t175484
24. Dwyer, C.: 'A Monumental Undertaking': NYC Subway To Disinfect Every Train Each Night. (Accessed May 2, 2020) Available at: https://www.npr.org/sections/coronavirus-live-updates/2020/05/02/849589416/a-monumental-undertaking-nyc-subway-to-disinfect-every-train-each-night