Submitted: 01/25/2022
Accepted: 04/05/2055

Research Article

[1]Department of Geography, McGill University, Montreal, QC, Canada

Keywords

Spatial mobility, COVID-19, demographic, geographic information science (GIS), pandemic

Email Correspondence

emily.chen3@mail.mcgill.ca

Emily Chen[1]

Assessing Early Public Response to COVID-19-Related Restrictions in New York City Using Spatial Analysis of Urban Mobility Data

Abstract

The rapid spread of COVID-19 in the United States initiated shelter-in-place policies that significantly impacted human mobility and daily routines starting in March 2020. Prior literature has examined the differences in lockdown policy efficacy and compliance with government orders[1-6], as well as the effect of mobility changes on case counts[7-12]. However, less attention has been placed on the connection between mobility and socio-demographics after the onset of COVID-19 within a city's borders. This paper focused on how human mobility patterns in New York City during the first three months of the pandemic differed based on socio-demographic factors like age, household income, and method of transportation to work. A secondary analysis determined if the four measurements of mobility used, namely distance traveled from home, home dwell time, non-home dwell time, and percentage time home, yielded significantly different findings. A mobility ratio representing the change in mobility between the first two weeks of February and April 2020 was created using aggregated and anonymized cellphone mobility data from SafeGraph. A Global Moran's Index was calculated for each mobility ratio to test for the presence of spatial autocorrelation, and then two spatial lag models were applied to account for the existence of autocorrelation. That there existed significant differences in mobility patterns based on socio-demographics reinforced the need for physical distancing policies that acknowledge the demographic diversity present not only between but also within cities.

Introduction

Since the United States detected its first case of the 2019 novel coronavirus in January 2020, efforts to contain the virus, such as stay-at-home policies, have greatly restricted human mobility and upended daily routines and momentous occasions alike. This retroactive analysis of the interaction between human mobility patterns during the COVID-19 pandemic, particularly after the implementation of state-level shelter-in-place orders and the socio-demographic differences within a city, contributes to a rapidly growing body of literature examining the effectiveness of lockdown policies. Prior work has investigated the effect of virus mitigation measures[1-6], mobility[7-12], and public gatherings[13-15] on the COVID-19 case positivity growth rate at various geographical scales. This paper focuses specifically on the relationships between average weekly levels of mobility and population demographics within New York City census block groups (CBGs) from February to April 2020. This work aims to provide fine-grained analysis on the socio-demographic effects of lockdown measures for policymakers and inform future strategies for infection mitigation and safe re-opening. To accomplish this goal, this paper raises two research questions:

- ***Research Question 1:*** Which socio-demographic factors have the greatest effect on the change in population mobility in New York City (NYC) before and after the implementation of COVID-19-related lockdown measures in March 2020?
- ***Research Question 2:*** Of the variables measuring the change in population mobility in this research, which one(s) act(s) most robustly as a proxy for physical distancing adherence?

Analyzing population movement to glean human behavior patterns from aggregated smartphone data became increasingly common leading up to the outbreak of COVID-19[16,17]. In the earliest months of the pandemic, several researchers advocated for the analysis of mobile phone surveillance data to predict the spread of COVID-19 and to understand population mobility trends[18]. Academic and industry researchers from wide-ranging disciplines and around the world acted upon these sentiments, producing a staggering number of analyses on spatial mobility trends during the COVID-19 pandemic.

Several studies examined the effects of mobility reduction on case counts outside of the United States[11,19]. In a comprehensive review focused on the geospatial and spatial-statistical analysis of the pandemic, Franch-Pardo et al.[20] evaluated 63 scientific articles on the subject and concluded that interdisciplinary action, proactive planning, and international solidarity were of utmost importance in controlling the spread of COVID-19. One notable paper by Pullano et al.[21] provided a robust overview of the demographic, socioeconomic, and behavioral factors associated with decreased mobility in France prior to and during the early lockdown period in March 2020 based on data from aggregated cellphones.

Several studies have focused on the relationship between mobility and the spread of COVID-19 in the United States[3, 22-24]. Chang et al.[8] sought to understand how the COVID-19 spread in ten of the largest U.S. metropolitan areas by constructing fine-grained dynamic mobility networks derived from geolocation data that mapped the hourly movements of 98 million people from neighborhoods to points of interests between March and May 2020. The authors found that their model simulating the spread of COVID-19 accurately predicted that higher infection rates occurred during the first two months of the pandemic amongst disadvantaged racial and socioeconomic groups because of differences in mobility[8]. Work by Badr et al.[7] investigated the effect of large-scale social distancing adherence on the spread of COVID-19 in 25 U.S. counties with the highest number of confirmed cases as of mid-April 2020. In their analysis, the authors concluded that social distancing had a significant effect on the spread of COVID-19 and that their findings could translate to other U.S. locations, given the geographical diversity of the counties in their sample set.

Within NYC, Lamb et al.[12] conducted an ecological study of residents using data for the number of daily visits to points of interest (POIs). The authors found that the proportion of the population living in households with more than three inhabitants, the proportion of uninsured 18-64-year-olds, the proportion of the population self-identifying as White, and median household income were the four aggregate markers of socioeconomic status that yielded the highest R^2 value across four time periods in April 2020. Their analyses revealed that changes in mobility and SES markers explained 56% of the variability in case positivity through 1 April 2020, but then dropped to a rate of explanation for case positivity variability of just 18% by 30 April 2020.

These findings suggest that after COVID-19 cases peaked on 6 April 2020 in NYC, the SES markers became less predictive due to greater testing capacity, higher SES areas having lower case positivity due to potentially greater engagement with unwarranted testing, and lower SES areas containing a higher number of infections. The authors also found that increased case positivity were independently associated with greater reductions in mobility on 10 April and 20 April but not on 1 April and 30 April. They attributed these mixed findings to the correlation between time and a city-wide decrease in case positivity as testing capacities increased.

Methods

Data

This paper's area of interest is NYC because it was the epicenter of the COVID-19 outbreak in the United States, with approximately 203,000 laboratory-confirmed cases reported between 1 March and 31 May 2020[25]. On 16 March 2020, the NYC school system, gyms, and casinos closed, and restaurants and bars were restricted to take-out and delivery services[26]. On 22 March 2020, all non-essential businesses closed, and the NYC on Pause Program's stay-at-home orders went into effect[27]. Building off these key dates, February 2020 was identified as the "before" time period and April 2020 as the "after" time period for analysis. Mobility patterns were retrieved from the "Social Distancing Metrics" dataset provided by the place-based data collection platform SafeGraph[28]. SafeGraph collects data using GPS pings from 20 million anonymous cellphone devices across the US. To calculate a mobile device's home, SafeGraph determines the device's common nighttime location to a Geohash-7 granularity of about 153 meters by 153 meters, and then groups devices into "home" CBGs. It also provides aggregated data, every 24-hours, for each CBG[28]. Table 1 describes the mobility variables and how they are reported by SafeGraph. To compare the differences in distance traveled from home before and after the onset of COVID-19 in NYC, the median distance traveled from home in the first two weeks of February 2020 for each CBG was divided by the median mobility value in the first two weeks of April 2020 for the equivalent CBG to create a mobility ratio (MR). The process was repeated to compare the differences in median home dwell time, median non-home dwell time, and median percentage time home. The latest available socioeconomic and demographic data was accessed from the 2016 5-year estimates in the American Community Survey (ACS)[29]. ACS data at the CBG level were the highest resolution available for the selection of socioeconomic and demographic variables. These data were cleaned to remove null and erroneous values.

Analysis

Global and local regression models were computed the same way for all four mobility ratios. Table 2 describes the explanatory variables used in all regression models. First, an ordinary least squares (OLS) linear regression model was fitted to the data to determine the global relations between mobility and socio-demographic factors. Next, the Global Moran's Index correlation test for regression residuals was used to check for spatial autocorrelation.

Table 1: Dependent Variables Used in Regression Models

Variable Name	Metadata
distance_traveled_from_home	Reported as an Integer. The value represents the median distance (in meters) of the median distance per device in a CBG traveled from the device's calculated "home" (i.e. Geohash-7 common nighttime location) within a 24-hour period. SafeGraph excluded distances equal to 0.
median_home_dwell_time	Reported as an Integer. The value represents the median time (in minutes) of the sum of all total time per device in a CBG spent at the device's Geohash-7 common nighttime location within a 24-hour period. Included in the total time are time ranges that may or may not have stopped or started within the 24-hour period.
median_non_home_dwell_time	Reported as an Integer. The value represents the median time (in minutes) of the sum of all total time per device in a CBG spent outside of the device's Geohash-7 common nighttime location within a 24-hour period.
median_percentage_time_home	Reported as an Integer. The value represents the ratio between median percentage of time spent at "home" for all devices in a CBG and the median total time observed within a 24-hour period.

Table 2: Explanatory Variables Used in Regression Models

Variable Name	Metadata
age	Estimated median age of the population.
race	Estimated number of people who identify as only White.
transport	Estimated number of workers 16 years and older who use public transportation (excluding taxicabs) to travel to work.
female_workers	Estimated number of female workers 16 years and older
housing_occupancy_rent	Estimated number of renter occupied housing units with over 1.5 occupants per room.
min_wage	Estimated number of households that earned less than $25,000 a year in 2016 (accounting for inflation) *Note: The base minimum wage in New York City from 12/31/15 to 12/31/16 was $9.00/hour, which is equivalent to $18,000/year [17].*
children	Estimated number of families with children under the age of 18.
education	Estimated number of people 25 years and older with a regular high school diploma.
health_insurance	Estimated number of people from the civilian non-institutionalized population with no health insurance coverage.

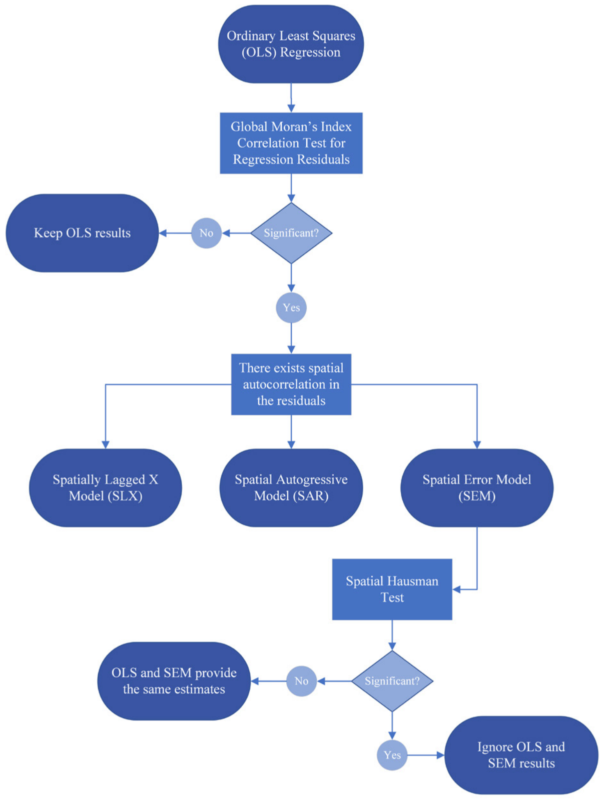

Figure 1: Spatial Regression Decision Process Flowchart

Two spatial regression models were run that determined whether the mobility patterns in surrounding CBGs affected the mobility pattern in one CBG[33]. The Spatially Lagged X (SLX) model tested local spatial relations, which meant that surrounding CBGs were those immediately adjacent to a CBG. The spatial autoregressive (SAR) Spatial Lag model tested global spatial relations, which meant that surrounding CBGs were all the observations in the data. To summarize the impacts from the SAR models, the number of simulations was set to 5,000 to compute distributions for the impact measures.

Lastly, a Spatial Error Model (SEM) and a spatial Hausman test were used to detect predictor variables in a regression model and were run to determine if differences existed between the OLS and SEM coefficients. A significant result suggested that neither OLS nor SEM yielded regression parameter estimates that matched the linear model parameters with independent identically distributed disturbances[34]. Thus, if a significant result was obtained from the spatial Hausman test, the OLS and SEM results were not considered. Figure 1 illustrates the analysis flow explained in this section. All analyses were performed using the 3.6.2 version of the R programming language[35] in version 1.2.5033 of RStudio[36].

Mapping

To create a map for each mobility ratio, R was used to remove outlier data by excluding the CBGs whose change in median home dwell time were greater than 2. Since "Null" values were changed to -999 for data parsing purposes, ratios that were less than 0 were also excluded. The categories were delineated by natural bins.

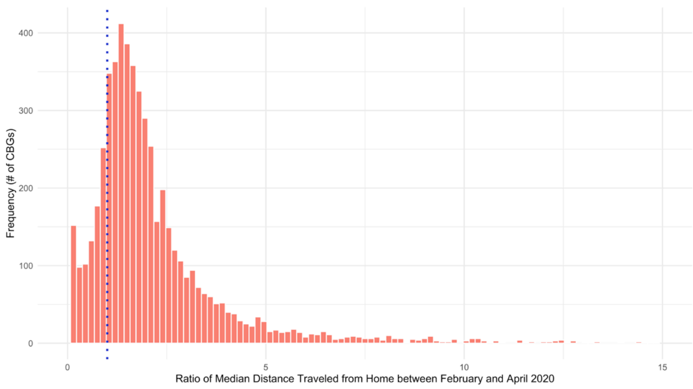

Figure 2a: Histogram for the Mobility Ratio of Median Distance Traveled from Home in NYC between February and April 2020 (bin size = 0.15). CBGs to the right of the dotted blue line at x = 1 indicate those with residents who traveled greater distances from home in February compared to April.

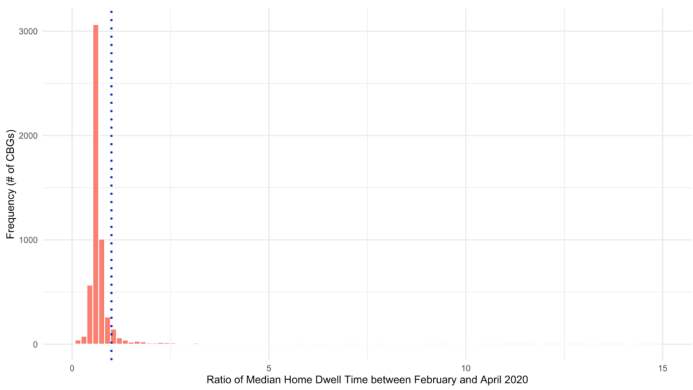

Figure 2b: Histogram for the Mobility Ratio of Median Home Dwell Times in NYC between February and April 2020 (bin size = 0.15). CBGs to the right of the dotted blue line at x = 1 indicate those with residents who stayed at home for longer in February compared to April.

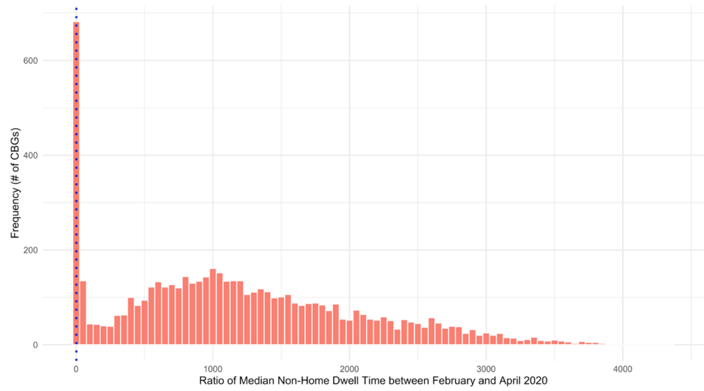

Figure 2c: Histogram for the Mobility Ratio of Median Non-Home Dwell Times in NYC between February and April 2020 (bin size = 50). CBGs to the right of the dotted blue line at x = 1 indicate those with residents who spent more time away from home in February compared to April.

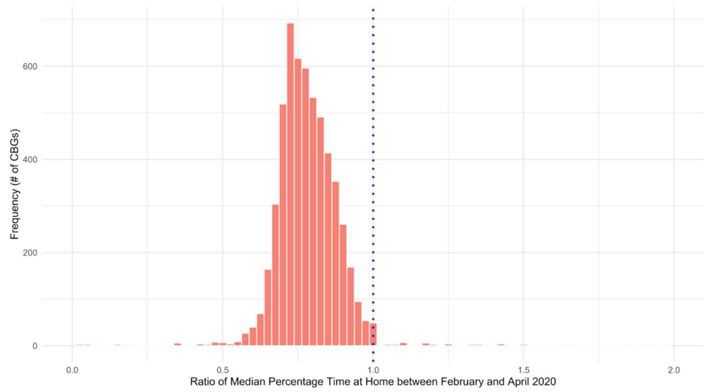

Figure 2d: Histogram for the Mobility Ratio of Median Percentage Time at Home in NYC between February and April 2020 (bin size = 0.025). CBGs to the right of the dotted blue line at x = 1 indicate those with residents who stayed at home for a higher percentage of time in February compared to April.

Shapefiles from the United States Census Bureau have cartographic boundary levels at the 2020 CBG level for each state. However, the NYC Department of City Planning provides shapefiles for the NYC boundary at only the 2010 census block level, which is at an even higher resolution than the CBG level. To obtain a shapefile with NYC CBGs, ArcMap v.10.7.1[37] was used to reproject both the NYC 2010 census block shapefile and the NY 2020 CBG shapefile to the WGS 1984 UTM Zone 18N coordinate system. The NYC census block shapefile was dissolved into census block groups and then intersected with the NY 2020 CBG shapefile. Lastly, R was used to link the shapefile with the CSV file containing SafeGraph and ACS data.

Results

Descriptive Statistics and Spatial Visualization

The frequency distributions for all four mobility ratios indicate that, overall, most NYC CBGs experienced decreased mobility and more time spent at home in the first two weeks of April 2020 compared to the first two weeks of February 2020. Based on the mobility ratio calculations, a ratio value less than 1 for a CBG suggested that people in that CBG traveled farther if they left home, stayed at home for longer, spent less time outside

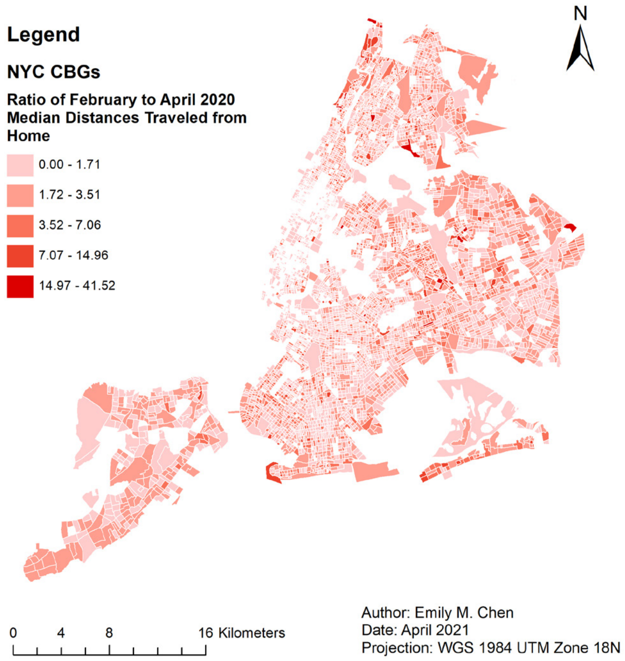

Figure 3a: Change in Median Distance Traveled from Home between February and April 2020 in New York City

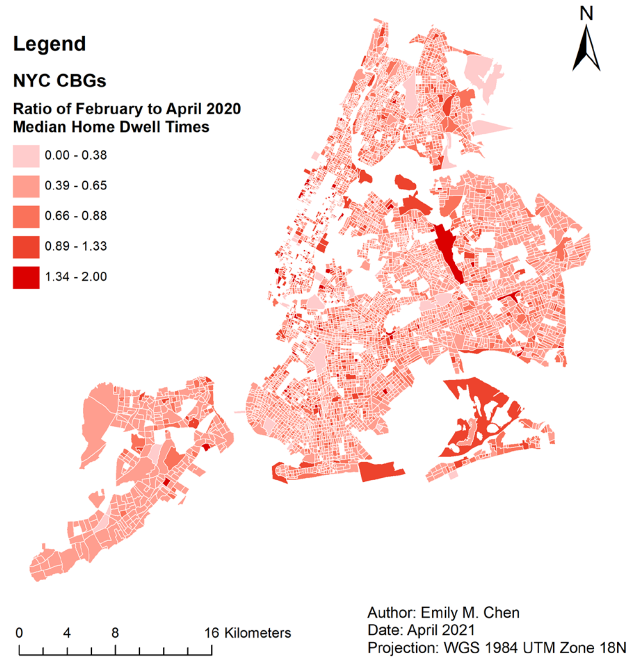

Figure 3b: Change in Median Home Dwell Time between February and April 2020 in New York City

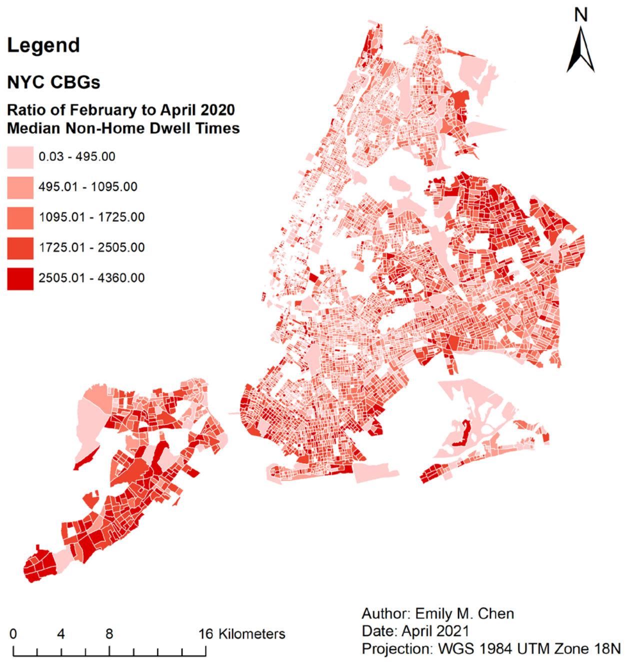

Figure 3c: Change in Median Non-Home Dwell Time between February and April 2020 in New York City

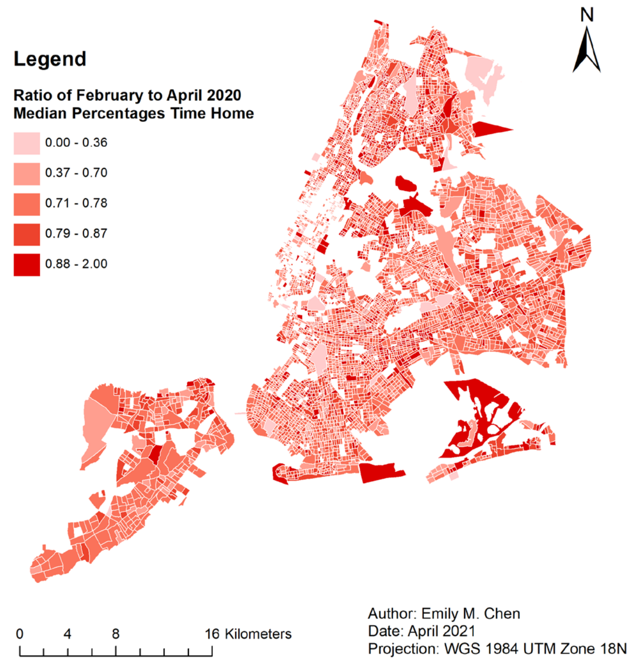

Figure 3d: Change in Median Percentage Time at Home between February and April 2020 in NYC 2020 in New York City

of home, or spent a greater percentage of their time at home in April than in February, and vice versa for a ratio value greater than 1. For distance traveled from home (Figure 2a), the histogram shows that most of the ratio values fall between 0 and 5, but with more values greater than 1 (M_{ratio}=1.6178). Thus, distances traveled from home dropped in April 2020. For home dwell time (Figure 2b), almost all ratio values fall between 0 and 1 (M_{ratio}=0.6170), so median home dwell times across all CBGs were mostly greater in April than in February. Conversely, for non-home dwell time (Figure 2c), most of the values are much greater than 1 (M_{ratio}=1085.0000), so median non-home dwell times across all CBGs were mostly greater in February than in April. Since all values of "0" were changed to "0.1" during data pre-processing to avoid dividing by 0, a median non-home dwell time value of 0.10 minutes in April was interpreted to mean that a large majority of CBGs experienced essentially no time spent away from home. A majority of April non-home dwell times close to 0 led to larger ratio values for median non-home dwell time compared to the other dependent variables. Lastly, for percentage time at home (Figure 2d), most of the ratio values lie between 0.5 and 1 (M_{ratio}=0.7750), therefore the median percentages of time spent at home across all CBGs were mostly greater in April than in February.

The maps of mobility ratio values for each CBG illustrate spatial variability across NYC. Category cut-off values were determined by natural breaks. For distance traveled from home (Figure 3a), darker red corresponds to a greater difference in median distance traveled from home between February and April. CBGs in the four largest categories had about a two-fold or greater increase in median travel distance from February to April. For home dwell time (Figure 3b), most CBGs had ratio values less than 1.0, which meant more time spent at home in April than in February. For non-home dwell time (Figure 3c), the enormous range in values for the smallest category suggests most CBGs had populations that spent almost no time away from home in April. Lastly, for percentage time at home (Figure 3d), most values were less than 1.0, thus showing that most CBGs experienced a greater percentage of time at home in April than in February.

RQ1: Effects of Socio-Demographic Factors on Mobility

Our first research question sought to understand which socio-demographic factors had the most effect on the change in population mobility in New York City before and after the implementation of COVID-19-related lockdown measures in March 2020. Nine noncollinear explanatory variables were chosen and four regression models were run with the four different measurements of change in mobility from February to April 2020: change in median distance traveled from home, change in median home dwell time, change in median non-home dwell time, and change in median percentage of time spent at home.

Based on the difference between the observed and expected Moran's I value as well as the significant p-value ($p<0.001$ for $\alpha=0.05$) for each of the Global Moran's Index linear correlation for regression residuals tests, we accepted the alternative hypothesis that there existed spatial autocorrelation in the residuals from all OLS model. Similarly, the significant p-value ($p<0.001$ for $\alpha=0.05$) obtained for all the spatial Hausman tests confirmed there were enough differences in the Standard Error Model (SEM) regression coefficients such that neither OLS nor SEM were appropriate models. Thus, only the Spatially Lagged X (SLX) and Spatial Autoregressive (SAR) models were used to interpret the results. Figure 4 summarizes the findings from these two models.

For the SLX model, a positive coefficient estimate associated with an explanatory variable meant that as the value for that variable within a CBG increased, so did the mobility ratio in that CBG (direct effect) and in neighboring CBGs (indirect effect). Median age had a positive direct and indirect (both $p<0.001$) value associated with distance traveled from home and non-home dwell time. Number of white-only residents had positive direct ($p<0.001$) and indirect ($p<0.05$) values for non-home dwell time. Number of families with children and high school graduates both had

Summary of Results from Spatially Lagged X and Spatial Autoregressive Regression Models

		Travel distance	Home dwell	Non-home dwell	Percent home
Age	SLX	***		***	
	SAR	***	***	***	***
Race	SLX			***	
	SAR	*	***	***	***
Transport	SLX	***		***	
	SAR	*	*	***	***
Female workers	SLX	***	***		***
	SAR	***	*	*	*
Housing	SLX	**			
	SAR	*		*	*
Income	SLX			***	
	SAR	*		***	
Children	SLX	***	***	***	***
	SAR	***	***	***	***
Education	SLX	***	***	***	***
	SAR	***	***	***	***
Health insurance	SLX				
	SAR			*	

Significance codes: p < 0.001 '***', p < 0.01 '**', p < 0.05 '*'

Notes:

- For Spatially Lagged X (SLX) models, green represents a positive coefficient estimate and red a negative coefficient.
- For Spatial Autoregressive (SAR) models, green represents a positive total estimate value and red a negative total estimate. P-values reported at R = 5,000 simulations, with '***' denoting p-values much less than 0.001 and '*' denoting p-values around 0.001 or greater.

Figure 4: Summary of Results from the Spatially Lagged X and Spatial Autoregressive Models

positive direct (both $p<0.01$) and indirect ($p<0.001$, $p<0.01$ respectively) values for distance traveled from home, negative direct and indirect (all $p<0.001$) values for home dwell time, positive direct (both $p<0.001$) and indirect ($p<0.05$, $p<0.001$ respectively) values for non-home dwell time, and negative direct and indirect (all $p<0.001$) values for percent time at home. Number of public transit users had positive direct ($p<0.01$) and indirect ($p<0.05$) values for distance traveled from home and negative direct ($p<0.001$) and indirect ($p<0.01$) values for non-home dwell time. Number of female workers had negative direct and indirect (both $p<0.001$) values for distance traveled from home and positive direct ($p<0.01$) and indirect ($p<0.001$) values for both home dwell time and percent time at home. Lastly, number of occupied renter units had positive direct and indirect (both $p<0.05$) values for distance traveled from home, while household income had negative direct and indirect (both $p<0.001$) values for non-home dwell time.

Interpretation of the SAR model relies on the impact measures' p-values and the direction of the direct impact value. Median age had positive direct impact values and consistently significant simulated p-values ($p<0.001$ for every run) for distance traveled from home and non-home dwell time, and negative values for home dwell time and percent time at home. Number of white-only residents had positive values for home dwell time, non-home dwell time, and percent time at home. Number of families with children and number of high school graduates had positive values for distance traveled from home and non-home dwell time and negative values for home dwell time and percent time at home. Number of public transit users had negative values for non-home dwell time and percent time at home. Lastly, number of female workers and household income both had negative values for distance traveled from home and non-home dwell time.

A caveat for the strength of the findings is that the SLX multiple R^2 values, while larger than the OLS multiple R^2 values for each dependent variable, were still quite low despite including nine explanatory variables (R^2=0.048 for distance traveled from home, R^2=0.077 for home dwell time, R^2=0.162 for non-home dwell time, and R^2=0.081 for percentage time spent at home). These low R^2 values indicate that the proportions of the variance in the dependent variables predictable from the explanatory variables were quite low. Solutions for increasing the R^2 value in future research include using other data sources and adding more explanatory variables. Importantly, since the R^2 value is not an indicator of whether the independent variables cause changes in the dependent variable, the interpretations of which explanatory variables affect mobility remain valid.

RQ2: Mobility Variables as a Proxy for Physical Distancing Adherence

This second research question asked which of the variables measuring population mobility served most robustly as a proxy for physical distancing adherence. To propose possible answers, the results from the four regression models were examined in the context of the nine explanatory variables for each of the mobility measures and found that median non-home dwell time yielded the greatest number of significant correlations with the explanatory variables from the SLX and SAR models. Furthermore, the SLX model's R^2 value (0.162) with this dependent factor was the highest of all mobility measures.

Discussion

Implications of Findings

Several of the results from the SLX and SAR models have interesting implications. Since non-home dwell time was the most accurate proxy for adherence to physical distancing, interpretations for some of the socio-demographic effects on this mobility measurement are presented here. For example, the SLX and SAR models found that median age of a CBG correlated positively with the change in median non-home dwell time (both $p<0.001$). This result indicated that the older the median age of a CBG, the less likely its residents were to spend time away from home in April. That CBGs with older populations saw less time spent away from homes suggests that older people were particularly careful about staying at home due to a combination of retirement, fewer reasons to leave the home, and knowledge that the elderly were affected more severely by the disease compared to younger populations[38-40]. Additionally, the scale of the SLX and SAR direct, indirect, and total impacts for age with non-home dwell times as the dependent variable were much higher than for any of the other explanatory variables (SLX: 14.52, 11.11, and 25.63 for age (respectively) compared to impact measures > -2.58 and < 2.12 for all other variables; SAR: 15.08, 6.50, 21.58 for age (respectively) compared to impact measures > -2.07 and < 1.31 for all other variables). These results indicate that a higher median age within a CBG had a greater effect on mobility defined as home and non-home dwell time than the other explanatory variables within that CBG (direct impact), in the CBG's immediate neighbors (indirect impact), and in all CBGs in the data (total impact)[41].

Another strong finding from the SLX and SAR models concerned the estimated number of families with children under age 18, which correlated positively with the change in median non-home dwell time (both $p<0.001$). These results indicated that CBGs with a greater number of families with children experienced less time spent away from home in April. This finding makes sense given that once schools closed, many parents stayed home to take care of young children while juggling full-time jobs. School closures and uncertainty about childcare left parents, particularly working mothers, with home school responsibilities that prompted some mothers to leave their jobs entirely[42]. Research by the U.S. Census Bureau and Federal Reserve found that of the adults not working, women ages 25-44 were almost three times as likely as men (32.1% compared to 12.1%) to not be working due to childcare demands[43]. While these results are based on national data, this phenomenon likely extended to NYC families as well. Furthermore, the U.S. Census study also found that working mothers in states with early stay-at-home orders and school closures were 68.8% more likely to leave their jobs than working mothers in states with later closures[44]. Given that NY state was one of the first states to implement stay-at-home measures, it seems likely that NYC working mothers fit into the category of being more likely to leave their jobs.

Lastly, both the SLX and SAR models found that the estimated number of people with only a high school diploma correlated positively with the change in median non-home dwell time (both $p<0.001$), showing that CBGs with a greater number of high school graduates spent more time at home in April compared to in February. One explanation for this finding was the 15% seasonally adjusted unemployment rate in April and that those more likely to face unemployment due to COVID-19 in NYC were workers with lower educational attainment (i.e., without a bachelor's degree)[44]. As confirmation, 61% NYC adults without a bachelor's degree experienced a loss in income since 13 March 2020 compared with 45% of adults with more than a bachelor's degree[44]. Without a job to go to, this demographic traveled shorter distances and stayed at home for longer periods of time.

Research Limitations

Similar to prior literature using aggregated cellphone mobility data[3, 8, 12, 22-24], the unknown representativeness of SafeGraph's data made it challenging to draw definitive conclusions from regression models. The dataset is certainly one of the largest available, as it came from 500,000 devices in almost every NYC CBG and accounted for one-ninth of the NYC population, which is a staggeringly large sample size compared to early mobility research that relied on participants to self-report data. However, this limitation is still worth noting because any conclusions drawn from these findings must acknowledge that they illustrate general population mobility trends from aggregated data. Running the four regression models using different mobility datasets and comparing the results could also strengthen these findings.

A second limitation to this work was the potential for additional factors, besides stay-at-home restrictions, to influence mobility patterns. For example, warmer weather in April could have contributed to greater time spent away from home for some demographics. To account for this seasonal change, an alternative baseline could have been April 2019, assuming that weather patterns were similar at that time to those observed in April 2020.

Future Directions

There are several ways to build upon the findings in this paper. The first is to extend the methodology to data from other cities, both in the U.S. and internationally. A between-city comparison might provide greater insight into how stay-at-home policies affected regions differently based on socio-demographic patterns, public transit infrastructure, or population density. An international comparison of cities could yield insight into the extent to which government stay-at-home orders reduced population mobility compared with other cross-cultural factors. In addition to comparing cities, other explanatory factors could be added to the regression models, such as the number of households who own second homes or citizenship status. Instead of mobility variables, one could also use points of interest (POI) data. For example, to better determine the large-scale impact of age, it would be useful to understand where younger people were going. National data indicated that younger workers were more likely to face unemployment due to COVID-19, and a survey of NYC metro adults found that 56% had lost income during the pandemic[44]. Therefore, if younger workers were more likely to experience unemployment and 37% of NYC frontline workers are over 50 years old[45], where were the younger age groups going? In addition to POI data, this question could be answered by using age-bracketed data to determine which age group left home the most. Lastly, several types of datasets that could be used to cross-reference our findings and evaluate how other non-pharmaceutical interventions affected mobility. For example, the Delphi Group at Carnegie Mellon University provides a variety of real-time COVID-19 indicators at the U.S. County and state level. Comparing their data on vaccine acceptance or the proportion of mask-wearers with mobility trends at the county level could help illuminate other aspects of disease spread patterns. Exploring other human behavior indicators and non-pharmaceutical interventions has particularly important implications, as researchers found that mobility and infection rates did not positively correlate as strongly after April 2020[46]. Their findings suggest that other non-pharmaceutical interventions like mask-wearing or hand washing played a significant role in mitigating the spread of COVID-19 early in the pandemic, therefore future research should consider these factors in their models when exploring the relationship between mobility and case positivity.

Conclusion

This paper's intent was to provide fine-grained analysis on the varying effects of lockdown measures and to inform future strategies for infection mitigation and safe re-opening. Our findings that there exist significant differences in mobility based on socio-demographic factors, particularly age, education level, and whether families have children, reinforce the need for physical distancing policies that acknowledge the demographic diversity present not only within, but also between cities. Providing resources for populations less able to stay at home (e.g., healthcare workers, service workers) to safely continue working is as important as providing support for populations who end up needing to stay at home (e.g., parents of young children, elderly populations) to minimize the effects of increased childcare demands and isolation. Future research may examine both these findings and the implications of reduced mobility on the spread of COVID-19 compared with other non-pharmaceutical interventions. By providing a detailed analysis of the various socio-demographic effects on different measurements of mobility, this paper emphasizes that there are several ways to measure mobility patterns within a city and that stay-at-home policies introduce unevenly distributed effects to different groups.

Acknowledgments

This research was conducted as part of the author's honours work towards a Bachelor of Arts (Honours) in Urban Systems Geography in the Department of Geography. The author wishes to thank Dr. Grant McKenzie (McGill University) for his guidance throughout the research process as the thesis supervisor, Dr. Clio Andris (Georgia Institute of Technology) for her insight and edits as the thesis reader, and Dr. Sarah Turner (McGill University) for her comments on the work as the professor for GEOG381: Geographic Thought and Practice and GEOG491: Honours Research.

Statement on Open Science

The author is committed to contributing towards the replicability and reproducibility of scientific research. A repository with all aspects of the data collection, cleaning, and analysis processes exists at https://github.com/emilyemchen/covid19mobility. The repository's README.md file provides a broad overview of the file structure and contents.

References

1. Aleta, A. et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nature Human Behaviour* **4**, 964–971 (2020).

2. Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet* **395**, 931–934 (2020).

3. Gao, S. et al. Association of mobile phone location data indications of travel and stay-at-home mandates with COVID-19 infection rates in the US. *JAMA Network Open* **3**, e2020485 (2020).

4. Graff Zivin, J. & Sanders, N. The spread of COVID-19 shows the importance of policy coordination. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 32842–32844 (2020).

5. Holtz, D. et al. Interdependence and the cost of uncoordinated responses to COVID-19. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 19837–19843 (2020).

6. Pei, S., Kandula, S. & Shaman, J. Differential effects of intervention timing on COVID-19 spread in the United States. *Science Advances* **6**, eabd6370 (2020).

7. Badr, H. S. et al. Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases* **3099**, 1–8 (2020).

8. Chang, S. et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2020).

9. Chen, Y., Jiao, J., Bai, S. & Lindquist, J. Modeling the spatial factors of COVID-19 in New York City. *SSRN Electronic Journal*. https://ssrn.com/abstract=3606719 (2020).

10. Cronin, C. J. & Evans, W. N. Private precaution and public restrictions: What drives social distancing and industry foot traffic in the COVID-19 era? *NBER Working Paper Series*, 1–39 (2020).

11. Kraemer, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).

12. Lamb, M. R., Kandula, S. & Shaman, J. Differential COVID-19 case positivity in New York City neighborhoods: Socioeconomic factors and mobility. *Influenza and Other Respiratory Viruses* **15**, 209–217 (2021).

13. Chande, A. et al. Real-time, interactive website for US-county-level COVID-19 event risk assessment. *Nature Human Behaviour* **4**, 1313–1319 (2020).

14. Dave, D., Friedson, A., McNichols, D. & Sabia, J. The contagion externality of a superspreading event: The Sturgis Motorcycle Rally and COVID-19. *Southern Economic Journal* **87**, 1–39 (2020).

15. Mangrum, D. & Niekamp, P. JUE Insight: College student travel contributed to local COVID-19 spread. *Journal of Urban Economics* **127**, 103311 (2022).

16. Budd, J. et al. Digital technologies in the public-health response to COVID-19. *Nature Medicine* **26**, 1183–1192 (2020).

17. Smith, C. M. et al. Spatial methods for infectious disease outbreak investigations: Systematic literature review. *Eurosurveillance* **20**, 1–21 (2015).

18. Buckee, C. O. et al. Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146 (2020).

19. Saraswathi, S., Mukhopadhyay, A., Shah, H. & Ranganath, T. S. Social Network Analysis of COVID-19 Transmission in Karnataka, India. *Epidemiology and Infection* **148**, 1–10 (2020).

20. Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F. & Billa, L. Spatial analysis and GIS in the study of COVID-19. A review. *Science of the Total Environment* **739**, 140033 (2020).

21. Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S. & Colizza, V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: A population-based study. *The Lancet Digital Health* **2**, e638–e649 (2020).

22. Bian, B., Li, J., Xu, T. & Foutz, N. Individualism During Crises. *The Review of Economics and Statistics* **104**, 1–18 (2022).

23. Brzezinski, A., Kecht, V. & Van Dijcke, D. The cost of staying open: Voluntary social distancing and lockdowns in the US. *SSRN Electronic Journal*, 1–34. https://ssrn.com/abstract=3614494 (2020).

24. Weill, J. A., Stigler, M., Deschenes, O. & Springborn, M. R. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 19658–19660 (2020).

25. Thompson, C. N. et al. COVID-19 Outbreak — New York City, February 29 – June 1, 2020. *Morbidity and Mortality Weekly Report* **69**, 1725–1729 (2020).

26. Vasquez, J., Shea, T., Rajamani, M., Price, B. & Intarasuwan, K. Timeline: Tracking the Spread of COVID-19 in Tri-State. https://www.nbcnewyork.com/news/local/timeline-tracking-the-spread-of-covid-19-in-tri-state/2313123/ (2020).

27. City of New York. Mayor de Blasio Issues New Guidance to New Yorkers. https://www1.nyc.gov/office-of-the-mayor/news/173-20/mayor-de-blasio-issues-new-guidance-new-yorkers (2020).

28. SafeGraph. *Social Distancing Metrics* https://docs.safegraph.com/docs/social-distancing-metrics (2021).

29. United States Census Bureau. *American Community Survey Information Guide* tech. rep. (2017), 1–15. https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf.

30. Taylor, D. B. A Timeline of the Coronavirus Pandemic. *The New York Times*. https://www.nytimes.com/article/coronavirus-timeline.html (2021).

31. Qin, A. & Hernández, J. C. China Reports First Death From New Virus. *The New York Times*. https://www.nytimes.com/2020/01/10/world/asia/china-virus-wuhan-death.html (2020).

32. New York State Department of Labor. *Minimum Wage Lookup* https://webapps.labor.ny.gov/dolweb/minimum-wage-lookup/ (2021).

33. Medina, J. & Solymosi, R. *Crime Mapping in R* chap. Spatial Regression Models. https://maczokni.github.io/crimemapping_textbook_bookdown/spatial-regression-models.html (2019)

34. Pace, R. K. & LeSage, J. P. A spatial Hausman test. *Economics Letters* **101**, 282–284 (2008).

35. R Core Team. *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2019). https://www.r-project.org/.

36. RStudio Team. *RStudio: Integrated Development Environment for R* (RStudio, Inc., Boston, MA, 2020). http://www.rstudio.com/.

37. Environmental Systems Research Institute. *ArcMap: Version 10.7.1* (Redlands, CA, 2019).

38. Mueller, A. L., McNamara, M. S. & Sinclair, D. A. Why does COVID-19 disproportionately affect older people? *Aging* **12**, 9959–9981 (2020).

39. Santesmasses, D. et al. COVID-19 is an emergent disease of aging. *Aging Cell* **19** (2020).

40. Williamson, E. J. et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436 (2020).

41. Golgher, A. B. & Voss, P. R. How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects. *Spatial Demography* **4**, 175–205 (2016).

42. Bateman, N. & Ross, M. *Why has COVID-19 been especially harmful for working women?* tech. rep. (Brookings, 2020). https://www.brookings.edu/essay/why-has-covid-19-been-especially-harmful-for-working-women/.

43. Heggeness, M. L. & Fields, J. M. Working Moms Bear Brunt of Home Schooling While Working During COVID-19. https://www.census.gov/library/stories/2020/08/parents-juggle-work-and-child-care-during-pandemic.html (2020).

44. Nischan, U. *Uneven Impact: What Job Loss During COVID-19 Means for New Yorkers Now and into the Future* tech. rep. (New York City Department of Consumer and Worker Protection, New York City, 2020), 1–19.https://www1.nyc.gov/assets/dca/downloads/pdf/partners/Uneven_Impact.pdf.

45. Stringer, S. M. *New York City's Frontline Workers* tech. rep. (Office of the New York City Comptroller, New York City, 2020), 1–15. https://comptroller.nyc.gov/wp-content/uploads/documents/Frontline_Workers_032020.pdf.

46. Badr, H. S. & Gardner, L. M. Limitations of using mobile phone data to model COVID-19 transmission in the USA. *The Lancet Infectious Diseases* **3099**, 30861 (2020).