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

by

Emily May-Ying Chen

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of B.A. in Urban Systems Geography

Department of Geography
McGill University
Montréal (Québec) Canada

April 2021

ACKNOWLEDGEMENTS

Writing an honours thesis during a fully virtual academic year would not have been possible without support from several people. I would first like to thank my supervisor, Dr. Grant McKenzie, for his guidance throughout this process. From answering my technical questions on Slack to meeting with me every week, I could not have asked for a more proactive and encouraging thesis advisor. Secondly, I would like to thank Dr. Clio Andris for providing insight and a second opinion as my reader, as well as Dr. Sarah Turner for commenting on two of the chapters. Thirdly, I would like to thank my parents and friends for attending my virtual poster presentation and for supporting me outside of the thesis process. A silver lining of remote work was that we could all attend the presentations, regardless of geographical location! Lastly, I would like to congratulate my fellow Honours students for completing research during a pandemic. Thank you for attending the Zoom work sessions I kept organizing, for spending 3+ hours listening to each other's presentations, and most importantly, for your inspiring resilience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF MAPS

LIST OF TABLES

ABSTRACT

The rapid spread of COVID-19 in the United States initiated shelter-in-place policies that significantly impacted human mobility and daily routines. Prior literature has examined the differences in lockdown policy efficacy and compliance with government orders between cities, as well as the effect of mobility changes on case counts. However, less attention has been placed on the connection between mobility and socio-demographics after the onset of COVID-19 within the same city. This undergraduate thesis 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. Using aggregated and anonymized cellphone mobility data from SafeGraph, I created a mobility ratio representing the change in mobility between the first two weeks of February and April 2020. I calculated a Global Moran's Index for each mobility ratio to test for the presence of spatial autocorrelation, and then I applied two spatial lag models 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.

CHAPTER 1: 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 these lockdown policies. This paper seeks to understand the relationships between average weekly levels of mobility and population demographics within New York City census block groups (CBGs) from February to April 2020, with the intention of providing fine-grained analysis on the socio-demographic effects of lockdown measures for policymakers and informing future strategies for infection mitigation and safe re-opening. Findings from this research reinforce the need for physical distancing policies that acknowledge the existence of demographic diversity between not only geographic regions in the U.S. but also within a single city.

My research is the first to look specifically for the existence of a strong correlation between human mobility levels and socio-demographic characteristics in U.S. CBGs, whereas much of the prior literature examined the effect of mobility and various explanatory variables on the COVID-19 case positivity growth rate (Chen et al., 2020; Lamb et al., 2021; Pullano et al., 2020). A study by Badr et al., which focused exclusively on mobility and COVID-19 case levels, found that decreased mobility, which the authors used as a proxy for increased levels of social distancing, had a positive and significant relationship with reduced case growth in several U.S. counties (2020). Like Badr et al., I employ SafeGraph's aggregated mobility data to measure the effectiveness of social distancing interventions, based on the assumption that fewer trips align with less physical contact and interactions with others (2020). However, my research focuses exclusively on how state-wide and city-wide lockdown measures changed mobility measurement values, such as median distance traveled from home and median non-home dwell time, with respect to socio-demographic characteristics per CBG. Furthermore, my research incorporates a more comprehensive definition of population mobility by testing not one but four different representative variables and their correlations with socio-demographic factors. I also analyze the regression results from these four variables to determine if one or more of these mobility variables most accurately represents physical distancing adherence.

1.1. Research Questions

Given that the overall aim of my research is to investigate how socio-demographic characteristics at the CBG level affect mobility patterns, I propose the following research questions:

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

Research Question 1 relates to my overall research aim by examining the interactions between socio-demographic characteristics and changes in mobility. To answer this question, I employ four models: the Ordinary Least Squares (OLS) linear regression model, the Spatially Lagged X (SLX) model, the Spatial Autoregressive (SAR) model, and the Spatial Error Model (SEM). I hypothesize that age, income, and the number of families with children will yield significant results, with older populations more likely to reduce mobility as a consequence of COVID-19 affecting this group more severely, lower-income residents less likely to be able to reduce mobility due to a higher likelihood of being involved in frontline work, and families with children more likely to reduce their mobility because of school closures.

Research Question 2 provides insight into which mobility measurement(s) best represent(s) public adherence to physical distancing interventions. Results from this analysis may help future researchers choose the most appropriate mobility measurement(s) for their particular research questions. I hypothesize that either median home dwell time or median non-home dwell time will yield the greatest number of significant correlations with socio-demographic factors because of the rigorous methodology that SafeGraph follows to determine a device's home location and dwell status (see Section 3.2 for details). If I included point of interest (POI) data in my analysis, median distance traveled from home might have led to the most informative results, since the model could account for trips' origin and destination, as well as environmental factors like POI density. As for median percentage time spent at home, many of the results from this variable will likely be the same as those obtained from median home dwell time. However, the percentage time at home measurement generalizes dwell time to some value between 0 and 100, whereas median home dwell time is a measurement recorded in minutes with a greater range of possible values. These differing levels of detail may affect the eventual correlation results.

1.2. Overview of Thesis Structure

My review of related literature in Chapter 2 has three sub-sections: population mobility surveillance, patterns of COVID-19 responses and positive case growth, and spatial mobility patterns during COVID-19. I start by examining prior literature on public-health related surveillance methods and the ethical dilemmas surrounding population surveillance. Next, I provide an overview of the various governmental and non-pharmaceutical interventions introduced in response to increasing COVID-19 case rates, both on a worldwide scale and in New York City. Finally, I analyze prior research on spatial mobility patterns during the pandemic, both on a global scale and in the United States.

In Chapter 3, I describe the data and methods used in my research and discuss how I address limitations that arose in prior studies. In Chapter 4, I present my findings, and then in Chapter 5, I summarize the implications of these findings, including how they contribute to a rapidly growing body of literature examining the effects of sociodemographic factors on COVID-19-related mobility. Lastly, in Chapter 6, I acknowledge the strengths and limitations of my work and provide several suggestions for future research.

CHAPTER 2: RELATED LITERATURE

In this section, I will critically examine how prior literature addresses the three tenets leading to my research aims of investigating how demographics at the Census Block Group level in New York City influence human mobility patterns during COVID-19 and to what extent significant results correlate with spatial distance. The organization of this section follows the conceptual framework shown in Figure 2.1.

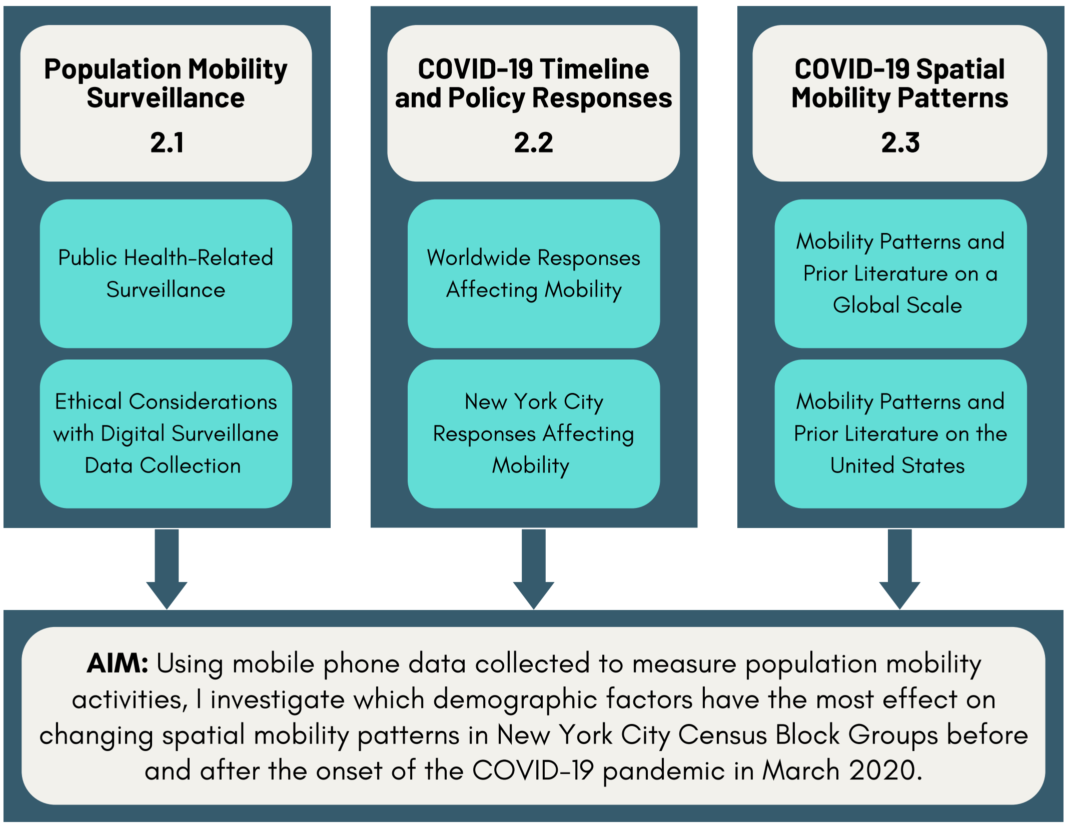


Figure 2.1: Conceptual Framework Guiding This Research

2.1. Tracking Population Mobility through Smartphone Data

Analyzing population movement to glean human behavior patterns from aggregated smartphone data became increasingly common leading up to the outbreak of COVID-19 (Budd et al., 2020; Smith et al., 2015). The increase in academic research interested in utilizing mobility datasets from smartphone tracking for COVID-19 research was a result of several private companies, including the provider of the dataset used for this research, making their previously industry-only data available for academic analysis while this public health crisis continued. In this section on population mobility surveillance, I first explore public health surveillance and its current relation to limiting the spread of COVID-19. The second part of the section discusses the ethical implications of using passive surveillance technologies to collect population-level data.

2.1.1. Public Health Surveillance and COVID-19

Mobile phone surveillance has numerous applications for public health, particularly with regard to mitigating the spread of COVID-19 and understanding population mobility trends. As Buckee et al. argue, “the research and public health response communities can and should use population mobility data collected by private companies, with appropriate legal, organizational, and computational safeguards in place” to “refine interventions” based on “near real-time information about changes in patterns of human movement” (Buckee et al., 2020: 145). Other researchers emphasize the need for regulation and rigorous evaluations of these digital technologies to ensure that they are used for the benefit of public health and not as exploitive tracking mechanisms (Budd et al., 2020). In the next section, I summarize one paper’s suggestions for ethically using aggregated mobility data to combat the spread of COVID-19.

2.1.2. Ethical Considerations of Participatory Surveillance

Researchers must consider representativeness, situational context, and methods of aggregation when working with mobility metrics calculated from GPS-derived aggregated data to analyze the spread of COVID-19 (Kishore et al., 2020). In terms of representativeness, the authors state that data providers must provide information on the fraction of the population represented in the data, the demographic characteristics of the data subjects, and the geographical makeup of the data, including whether or not representative bias exists in favor of urban communities over rural ones. The authors then advise researchers to communicate how the latter group chose the baseline period against which to compare their analyses, which in the context of COVID-19 could be prior to the implementation of physical distancing policies. They also suggest that researchers outline the uncertainty associated with choosing this baseline period. Lastly, the authors note how data aggregation must strike a balance between maintaining an “actionable spatial boundary” on a timescale with epidemiologically relevant information and preventing possible re-identification of individuals from the data (Kishore et al., 2020, p. e623). The authors declare that a pandemic is not justification for ignoring the risks to an individual's privacy associated with using personal data to calculate disease transmission-related metrics like sources of mass infection or mobility habits. They propose statistical thresholds, differential privacy, and appropriate security controls that all stakeholders agree to as appropriate privacy protection measures. This paper's main strength is its actionable guidelines for analyzing aggregated cellphone data. The authors also provide detailed steps for calculating nine metrics that could be used as measurements of physical distancing effectiveness and inputs for models tracking the spread of COVID-19, although the information is evidently written for data providers who have access to disaggregated data. I attribute this approach to the employee relationship that one of the paper's authors has with the data analysis firm Camber Systems, especially since the paper acknowledges that the company may use the metrics described in this paper or ones similar for commercial products in the future. The paper's most noticeable weaknesses are its failures to mention either datasets that successfully adhere to their representativeness and aggregation guidelines or studies that effectively state situational context. Providing examples for both would have enhanced their argument and provided valuable resources for readers.

2.2. Early Spread of COVID-19 and Mitigation Responses

In this subsection, I will first provide an overview of how COVID-19 spread across the world, starting in late December 2019, and the subsequent government mandates enacted to restrict population mobility between and within countries. I will then describe when stay-at-home orders occurred in New York City from March to May 2020. Appendix B.1 provides a visual representation of the COVID-19 timeline both on the world scale and for New York City.

2.2.1. Global State of COVID-19 from December 2019 to May 2020

The government of Wuhan, a city with over 11 million people located in China's Hubei province, first confirmed that their health officials were treating "dozens of cases of pneumonia of unknown cause" on 31 December 2019 (Taylor, 2021). Twelve days later, on 11 January 2020, China reported the first fatality caused by the virus at a time when there were 41 confirmed cases, seven of whom were in severe condition (Qin & Hernández, 2020). The news came right before Chinese passengers were expected to take an estimated three billion trips in the subsequent five weeks to celebrate the Spring Festival, which started on 25 January 2020 and is China's largest holiday (Qin & Hernández, 2020). At that point, the Wuhan Health Commission had not found evidence that the virus spread between humans, so there were no travel restrictions put in place (Qin & Hernández, 2020). However, by 23 January 2020, Chinese authorities canceled all planes and trains leaving Wuhan and suspended bus, subway, and ferry operations within the city as the total number of infections and fatalities reached 570 and 17 respectively in Taiwan, Japan, Thailand, South Korea, and the United States (Taylor, 2021). The United States reported its first confirmed case in Washington state on 21 January 2020 from a man who developed symptoms after returning from Wuhan, and then banned entry to all foreign nationals who had traveled to China within the last 14 days, effective as of 31 January 2020 (Taylor, 2021). The day before, on 30 January 2020, Director-General of the World Health Organization Dr. Tedros Ghebreyesus declared a "public health emergency of international concern over the global outbreak of novel coronavirus" (WHO Director-General, 2020: online), prompting the U.S. State Department to warn Americans against travel to China and U.S. Health and Human Services Secretary Alex M. Azar II to declare a public health emergency for the entire United States on 31 January 2020 (HHS Press Office, 2020; She et al., 2020; Taylor, 2021).

The virus spread across Europe throughout early February 2020, with Italy becoming the site of the first major outbreak on the continent. Italy's number of reported cases jumped from fewer than five to over 150 on 23 February 2020, prompting Italian officials to lock down towns within the Lombardy region, close schools, and cancel sporting and cultural events (Taylor, 2021). In the United States, the number of cases continued to increase rapidly since reporting its first case on 21 January 2020, eventually prompting the Centers for Disease Control and Prevention to advise against gatherings of 50 or more people starting 15 March 2020 and onwards for the subsequent eight weeks (Taylor, 2021). On 16 March 2020, the same day that New York City public schools closed, U.S. President Donald Trump warned citizens against groups of more than 10 people (Taylor, 2021). By 26 March 2020, the United States had the highest number of confirmed COVID-19 cases in the world. Despite reaching this milestone, state-level lockdown policies varied by state throughout March to May 2020 in the absence of an official federal policy, as shown in Figure 2.2.

Figure 2.2: Type and Duration of COVID-19 State and Territorial Stay-At-Home Orders from 1 March to 31 May 2020

From "Timing of State and Territorial COVID-19 Stay-at-Home Orders and Changes in Population Movement — United States, March 1–May 31, 2020" by A. Moreland et al., 2020, *Morbidity and Mortality Weekly Report*, *69(35)*, p. 1200.

2.2.2. State of COVID-19 in New York City from March to May 2020

In spring 2020, New York City (NYC) was the United States epicenter of the COVID-19 outbreak, with approximately 203,000 cases of laboratory-confirmed COVID-19 reported by the

NYC Department of Health and Mental Hygiene between 1 March and 31 May 2020 (Thompson et al., 2020). Governor of New York Andrew Cuomo announced the first case of COVID-19 in New York City on 1 March 2020 after the case was confirmed in a laboratory on 29 February 2020 from a 39-year-old Manhattan woman returning from Iran with mild respiratory symptoms (Thompson et al., 2020; Vasquez et al., 2020). The virus spread rapidly throughout both NYC and New York State (NYS) during the first week of March, prompting Governor Cuomo to declare a state of emergency for NYS on 2020 March 7, when there were 89 cases state-wide and 11 cases in NYC (Vasquez et al., 2020). As the number of positive cases in NYC continued to increase exponentially during the first two weeks of March 2020, NYC Mayor Bill de Blasio shuttered large venues like Barclays Center in Brooklyn, Madison Square Garden in Manhattan, and Radio City Music Hall in Midtown Manhattan on 12 March 2020 (Vasquez et al., 2020). Governor Cuomo also implemented a state-wide ban on gatherings of 500 people or more and announced that venues with capacity below 500 people would have to operate at 50 percent occupancy (Vasquez et al., 2020). Despite the closure of large public arenas by 12 March 2020, a majority of schools in New York City's public school system, the largest school district in the United States with over 1.1 million students, remained open until 16 March 2020, when Mayor de Blasio announced that schools would be closed until at least 20 April 2020, with plans for remote learning to begin on 23 March for kindergarten through 12th grade (Eisenberg & Touré, 2020; NYC Department of Education, n.d.). In addition to closing schools, Governor Cuomo, along with New Jersey Governor Phil Murphy and Connecticut Governor Ned Lamont, lowered the maximum gathering threshold to 50 people, shuttered gyms and casinos, and restricted bars and restaurants to providing just take-out and delivery services across the Tri-State area on 16 March (Vasquez et al., 2020). By 18 March 2020, with 3,437 cases in NYS and more than 1,870 cases in NYC confirmed thus far, Governor Cuomo implemented a statewide mandate that non-essential businesses must have at least 50 percent of their employees working from home (Vasquez et al., 2020). Two days later, on 20 March, Mayor de Blasio mandated that all non-essential businesses in NYC would close at 20:00 on 22 March until further notice and announced fines could be imposed on non-vulnerable individuals who violated rules regarding non-essential gatherings and social distancing, in adherence to the NYS on Pause Program's stipulations that all non-essential workers must stay at home (City of New York, 2020). Twenty-three days after the first laboratory confirmed case of

COVID-19 in NYC, the entire state had been put into strict and unprecedented stay-at-home measures and school closures that Governor Cuomo later extended throughout April and May 2020.

2.3. Spatial Mobility Patterns during the COVID-19 Pandemic

This section will explore prior literature that focuses on spatial mobility trends during the COVID-19 pandemic. I will critically examine research on spatial mobility in countries outside of the United States, followed by research on mobility trends in the U.S. Several of the studies in this latter group also used data from SafeGraph.

2.3.1. Spatial Mobility on a Global Scale during the COVID-19 Pandemic

Research that examined the effects of mobility reduction on case counts outside of the U.S. include an analysis from Kraemer et al. (2020) on the spread of COVID-19 in China and a social network analysis of COVID-19 transmission in India by Saraswathi et al. (2020). In a comprehensive review focused on the geospatial and spatial-statistical analysis of the COVID-19 pandemic, Franch-Pardo et al. (2020) evaluated 63 scientific articles on the subject and concluded that interdisciplinary action, proactive planning, and international solidarity were of utmost importance for controlling the virus.

One particularly notable paper by Pullano et al. (2020) examined how mobility in France changed before and during lockdowns based on aggregated cellphone data from Flux Vision of origin-destination travel flows among 1,436 mainland France geographical areas. The authors segmented their results by trip distance (all trips and long trips, which were defined as more than 100 kilometers of geodesic distance between location centroids), user age (under 18, 18-65, and over 65 years old), residency (residents had French SIM cards while non-residents did not), and time of day (daytime or nighttime and weekend versus weekday, including rush hours). Next, the authors analyzed behavioral responses to announcements of physical distancing interventions and pandemic burden, i.e. COVID-19-related deaths and hospitalization rates. They found that traffic flow significantly decreased from their forecasting model's extrapolated traffic flow, assuming no interference associated with COVID-19 interventions, starting on 14 March 2020. Interestingly, they observed a "pre-lockdown exodus out of Paris" on 16 March 2020, one day before lockdown took effect, which they attributed to relocation caused by fear of the imminent implementation of stricter policies seen prior to that date in Italy, Spain, and Austria (Pullano et al., 2020, p. e642).

Data during the lockdown period also revealed larger mobility reductions in regions more severely affected by the pandemic in terms of number of hospitalizations, suggesting that individuals in these hard-hit regions were more likely to act on concerns about overwhelming the hospital system than those in less affected areas. By testing the effects of several explanatory variables on mobility, this study 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. The authors reported several interesting results, although they acknowledged the limitations of their observational methodology when inferring causal relationships from complex interactions between factors. Another limitation was the study's geographical scale, which at the city level allowed for comparisons in travel between major French cities but mostly ignored mobility differences in sub-regions within cities. Lastly, the authors failed to elaborate on their decision to construct a forecasting model using training data from a period of time (6 January to 9 March 2020) already affected by COVID-19 restrictions (they state that Phase 2 of France's COVID-19 response, which involved social distancing interventions like closing schools, started on 29 February 2020), when they also noted that extrapolated traffic flow after 9 March 2020 assumed no changes due to COVID-19-related interventions.

2.3.2. Spatial Mobility of the U.S. Population during the COVID-19 Pandemic

Aggregated mobility data have been used in prior research as a proxy for quantitatively measuring the effectiveness of social distancing measures (Badr et al., 2020; Buckee et al., 2020). In this section, I will go into detail on four representative publications and summarize the wider literature in Table 2.1.

The first paper, written by Chang et al. (2020), sought to understand how SARS-CoV-2 spread in ten of the largest U.S. metropolitan areas by constructing fine-grained dynamic mobility networks derived from SafeGraph cellphone geolocation data that mapped the hourly movements of 98 million people from neighborhoods to points of interests at the census block group (CBG) level between 1 March and 2 May 2020. The authors found that their metapopulation susceptible-exposed-infectious-removed (SEIR) model simulating the spread of SARS-CoV-2 with the aforementioned mobility networks accurately predicted that higher infection rates occurred during the first two months of the pandemic amongst disadvantaged racial and socioeconomic groups as a result of only differences in mobility. This result not only supports prior literature indicating that

SARS-CoV-2 infections were unevenly distributed across the U.S. population, but also strengthens my hypothesis that income levels will correlate significantly with measures of aggregated mobility at the CBG level. The authors also noted that the model could predict "super-spreader" points of interest (POI) that accounted for a majority of infections. They concluded from this result that varying maximum occupancy levels to increase physical distancing based on POI rather than uniformly reducing mobility across POIs could be a more effective policy measure. This article's exemplary methodological rigor is a result of the authors paying careful attention to every part of the study, including cross-checking mobility trends from SafeGraph data with Google mobility data, building an undirected bipartite graph with 5.4 billion edges between 56,945 CBGs and 552,758 POIs to represent population-level mobility, and providing extensive documentation of the mathematical reasoning behind model initialization, calibration, validation, and data analysis. A limitation to this study, which the authors themselves acknowledged, was that the SafeGraph data underlying their SEIR model did not perfectly represent the population, contain all POIs in the metropolitan areas of interest, or provide context at a geographical scale smaller than the CBG level. Most importantly, the SEIR model did not take into account all real-world factors contributing to SARS-CoV-2 transmission; however, the authors maintained that the predictive accuracy of their model based solely on mobility between POIs robustly supported their broad conclusions on sociodemographic inequities and uneven sources of infection at various POIs.

The second paper by Badr et al. (2020) 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 16 April 2020 using Teralytics' aggregated mobility data from 1 January to 20 April 2020. To evaluate how well decreased mobility affected the rate of new infections, the authors fitted a generalized linear model for each county on a given day by using a lagged mobility ratio (MR) as the predictor for the COVID-19 case growth rate (GR) ratio, and then tested the correlation of the MR and GR at different time lags from both separate models for each county and also from a combined model for all counties. Significant correlations in all 25 counties with Pearson correlation coefficients above 0.7 (out of 1.0) in 20 of the 25 counties led the authors to conclude 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. The authors also discovered that social distancing was evident in early March before any of the first U.S. state-level lockdowns were implemented, which was a phenomenon they partially attributed to county-level restrictions while also noting that all states showed some form of social distancing before these county-level restrictions. By analyzing data at the county level across 11 U.S. states, this study successfully produced results at a geographical scale small enough to account for heterogeneity in the number of confirmed cases and mobility changes within a county, while also providing opportunities to compare entire states and thus generalize findings on a national level. The authors acknowledged that one of their study's most significant limitations was its ignorance of other case mitigation factors like mask-wearing or handwashing that could have significantly contributed to declining COVID-19 case growth rates in March 2020. A follow-up commentary by the first and last authors of this study emphasized the importance of further research on the effect of these other non-pharmaceutical interventions (NPIs), since the strong linear correlation between mobility and case growth rates they had observed in their first paper was absent after April 2020, thus suggesting that this strong correlation after April could be attributed less to mobility having a significant impact on COVID-19 transmission and more towards interventions like mask-wearing and avoiding large gatherings that were adopted in parallel with initial lockdowns (Badr & Gardner, 2020). Further research quantifying the effects of NPIs and their interactions will ultimately determine whether or not restricting mobility alone can affect case growth rates.

For research on patterns in New York City, Lamb et al. (2021) conducted an ecological study of residents in 177 NYC zip code areas using SafeGraph data for the number of daily visits to points of interest (POIs). The authors wanted to determine the extent to which aggregate markers of socioeconomic status (SES) and daily changes in mobility could explain zip code-level COVID-19 case positivity, as well as the extent to which daily changes in mobility independently predicted case positivity. They concluded from ranking univariate analyses and a multivariable prediction model 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 population self-identifying as White, and median household income were the four aggregate markers of SES that yielded the highest R^2 value for all four time periods (1 April, 10 April, 20 April, and 30 April). Their analyses revealed that changes in mobility considered with 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, suggesting that after COVID-19 cases peaked on 6 April 2020 in NYC, these SES markers became less predictive due to several factors, including 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 actual infections. The authors also found that increased case positivity was independently associated with greater reductions in mobility on 10 April and 20 April, but not on 1 April and 30 April, and they attributed these mixed findings to the correlation between time and a city-wide decrease in case positivity as testing capacity increased. The authors acknowledged the limitations of their study, including that its use of zip code areas could not account for the heterogeneity of SES, case positivity, and changing mobility levels within these areas. Furthermore, their use of COVID-19 case positivity as an outcome measurement was highly imprecise, given that the metric was subject to fluctuation based on diagnostic test accessibility. However, this study's most innovative feature was its use of physical check-ins to POIs within a zip code area as its measurement of mobility, which the Center for Disease Control (CDC) also uses to track mobility patterns, rather than a more common metric like average distance displacement.

The final paper, which was a preprint by Chen et al. (2020), contributed to a growing body of literature examining how to best prevent and control COVID-19 infections by examining and modeling the spatial factors that led to early COVID-19 outbreaks in New York City using land use, travel behaviors, and sociodemographic factors as explanatory variables. The authors categorized land use into three main categories based on points of interest (POI) labels with a high likelihood of congregation: green spaces and/or parks, grocery stores, and medicine-related places. They measured travel activities using the mean distance traveled from home variable from SafeGraph's social distancing metrics. They also used gender, race, poverty, working from home, commuting habits, population, and number of workers as their sociodemographic variables. Using ordinary least squares (OLS) regression for global relations, the authors determined that areas with high medical POI density, green space density, greater median distance traveled from home, percentage of males, and percentage of commuting through walking, carpooling, and public showed higher rates of positive COVID-19 cases. Using geographically weighted regression (GWR) models for local relations, the authors concluded that the effects of working from home varied across postal areas, which led them to suggest that future reopening strategies vary between NYC boroughs. This study's use of GWR modeling provided important insight into local differences in land use, travel behaviors, and sociodemographic factors. However, the study's main limitation was in its scope, as its focus on just New York City prevented comparisons between other U.S. cities. Furthermore, the authors acknowledged that using just SafeGraph and American Community Survey data diminished the reliability of their findings on the effect of public transit on COVID-19 infections. By extending this study's methodology to other U.S. cities and incorporating additional data sources, policymakers might be able to consider reopening strategies not just within but also between cities.

Paper	Study Area Region	Mobility Dataset	Research Question(s)
(Aleta et al., 2020)	Boston, USA	Cuebiq	How does testing, contact tracing, and household quarantine affect the number of second wave COVID-19 cases in the Boston metropolitan area?
(Badr et al., 2020)	25 U.S. counties	Teralytics	What is the effect of large-scale social distancing on the spread of COVID-19 in the USA?
(Bian et al., 2020)	U.S. (county level)	Google Mobility, SafeGraph, de-identified individual cellphone data	How does county-level individualism affect adherence to social distancing?
(Brzezinski et al., 2020)	U.S. (county level)	SafeGraph	What is the difference in cost on the economy between imposing lockdowns and staying open?
(Chang et al., 2020)	10 U.S. counties	SafeGraph	How does mobility during COVID-19 affect case rates and explain racial and socioeconomic inequities/
(Chen et al., 2020)	New York City, USA	SafeGraph	Which spatial factors contribute to the rate of positive COVID-19 cases in NYC?
(Cronin & Evans, 2020)	U.S. (state and county levels)	SafeGraph	What effect do state and local social distancing policies have on foot traffic during COVID-19?
(Dave et al., 2009)	South Dakota, USA	SafeGraph	What were the public health impacts of the Sturgis Motorcycle Rally?
(Dincer & Gillanders, 2020)	U.S. (state level)	SafeGraph	What are the links between corruption and compliance with social distancing during COVID-19 in the U.S.?
(Ding et al., 2020)	U.S. (county level)	SafeGraph	To what degree do social capital characteristics (community engagement and individual commitment to social institutions) account for differences in social distancing adherence?
(Gao et al., 2020)	U.S. (state level)	SafeGraph	What is the association between the rate of human mobility changes and the rate of confirmed COVID-19 cases?
(Holtz et al., 2020)	U.S. (county level)	SafeGraph, Facebook	What is the cost associated with an uncoordinated government response to COVID-19 with regard to stay-at-home orders in the U.S.?

(Janssen & Shapiro, 2020)	Singapore	Lifesight	How does Singapore's transparency of case information affect voluntary activity reductions, particularly in areas with more reported cases?
(Lamb et al., 2021)	New York City, USA	SafeGraph	To what extent can the variability in ZIP-code level case positivity be explained by socioeconomic status and daily change in mobility? To what extent does daily change in mobility independently predict case positivity?
(Li et al., 2020)	Wuhan, China	Self-collected interviews with infected persons, relatives, close contacts, and health care workers	How did human-to-human transmission occur amongst close contacts in Wuhan, China since mid-December 2019?
(Mangrum & Niekamp, 2020)	U.S. (county level)	SafeGraph	How did university spring break travel affect the evolution of confirmed COVID-19 cases and mortality? How did method of travel and destination contribute to the spread of COVID-19?
(Pullano et al., 2020)	France	Flux Vision	How did mobility in France change before and during lockdown by trip distance, user age and residency, and time of day? What spatial heterogeneities exist in the regional data?
(Saraswathi et al., 2020)	Karnataka, India	Public contact tracing data from the Karnataka government	How can social network analysis be used as a tool for outbreak monitoring and control for the COVID-19 outbreak in Karnataka, India?
(Weill et al., 2020)	U.S. (state level)	SafeGraph, Google Mobility, Place IQ	How does income affect responses to social distancing policies?
(Wilder et al., 2020)	Hubei, China Lombardy, Italy New York City, USA	Within-household contact: household distributions from census data Out-of-household contact: age-stratified, country-specific estimated contact matrices	What is the role of transmission due to particular age groups on total COVID-19 infection and deaths? What are the between-population variations in COVID-19 transmission?

Table 2.1: Overview of Prior Literature on COVID-19 and Mobility

CHAPTER 3: DATA AND METHODOLOGY

This section will describe the datasets I used and then go into detail on the data collection, preparation, and annotation processes. I will also discuss the statistical analyses conducted on the cleaned and annotated SafeGraph dataset merged with the American Community Survey and U.S. Census boundary datasets.

3.1. Rationale behind Time Interval Choices

Kishore et al. (2020) argue that researchers who are investigating changes in mobility must clearly establish their rationale for choosing a baseline time period against which to compare their experimental condition(s), as well as acknowledge the uncertainty associated with their decision. In this section, I will explain why I used the first two weeks of February 2020 as my baseline time interval and the first two weeks of April 2020 as my experimental time interval.

To mitigate the influence of outlier data on my analysis, I defined my mobility measurements based on the median of 14 values, which corresponded to 14 consecutive days within a two-week period. Determining which two-week periods to collect data from that represented before and after the onset of COVID-19 required examination of how the virus spread in NYC and the subsequent stay-at-home measures implemented both city- and state-wide that could influence mobility patterns from one day to the next.

Based on the timeline of stay-at-home orders affecting human mobility as outlined in Sections 2.2.1 and 2.2.2, I hypothesized that the two dates in March that affected mobility most significantly in New York City were 16 March 2020, when the NYC school system, gyms, and casinos closed and restaurants and bars started providing only take-out and delivery services, and 22 March 2020, when all non-essential businesses closed and the NYC on Pause Program's stay-at-home orders went into effect. Prior studies examining responses to physical distancing mandates, such as the one by Badr et al. (2020), found that reduced mobility started in early March in the 11 states corresponding to the 25 U.S. counties with the highest number of confirmed cases on 16 April 2020, and thus well before any U.S. state implemented statewide stay-at-home orders. Figure 3.1 shows how the proportion (i.e. mobility ratio) between (1) the sum of total incoming and outgoing trips within a county on a particular day (shown on the x-axis) and (2) the average sum of total incoming and outgoing trips within a county from 8 January to 31 January 2020, which Badr et al. (2020) used as their baseline time period, declined sharply from 1.0 (indicating no change) several days before the first statewide stay-at-home order in California on 19 March. Based on these findings by Badr et al. (2020), I could not assume that the start of a stay-at-home mandate catalyzed the first signs of reduced mobility in NYC, so I chose the first two weeks of February as a baseline window with stable mobility patterns unaffected by COVID-19.

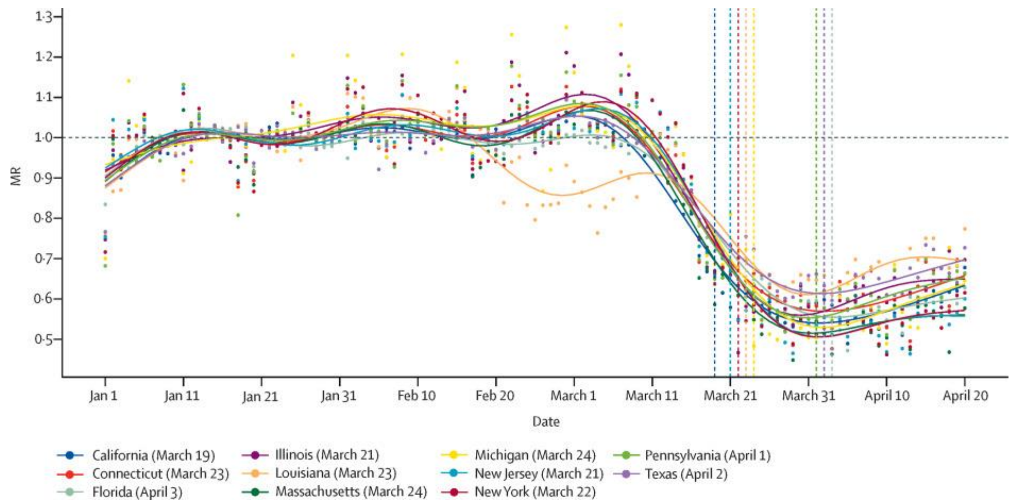


Figure 3.1: Timeseries of Mobility Ratios for U.S. States and Corresponding Dates of Stay-At-Home Mandates[1]

From "Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study" by H. Badr et al., 2020, *The Lancet Infectious Diseases, 3099(20)*, p. 5.

Despite Badr et al.'s findings that populations in eleven states decreased their mobility prior to stay-at-home orders, I hypothesized that mobility would decline further following New York's stay-at-home orders implemented on 22 March 2020. I chose the first two weeks of April as the two-week period that would best represent these orders' effect on the mobility of New York City's population because a week past the start of the stay-at-home orders could account for potential fluctuations in the data as the population adjusted to the new physical distancing measures. Figures Figure 3.2 through Figure 3.9 support this time frame choice, since the histograms for each of the dependent mobility variables differed in shape and peak value between

[1] Notes: Dots represent the raw mobility ratio (MR) data for each day and the vertical dashed lines correspond to state stay-at-home orders. Since some orders occurred on the same day, only eight lines are shown for the 11 states. Plotted lines were smoothed with a generalized additive model.

the first two weeks of February shown on the left and the first two weeks of April shown on the right.

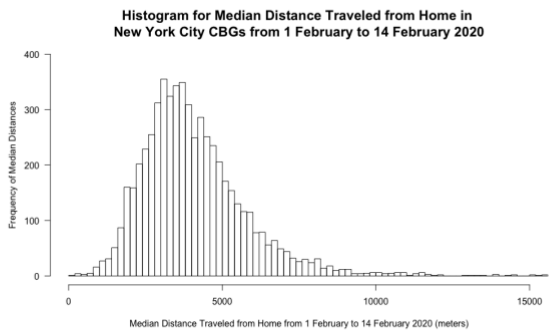


Figure 3.2: Histogram for Median Distance Traveled from Home in NYC, February 2020 (5,000 breaks)

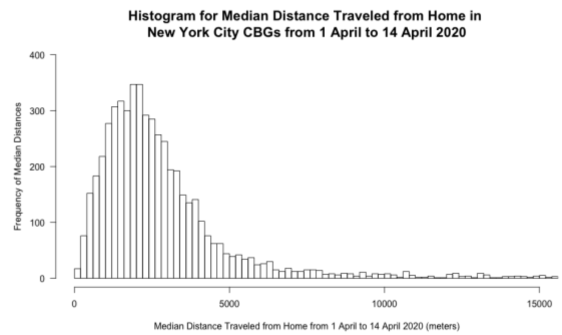


Figure 3.3: Histogram for Median Distance Traveled from Home in NYC, April 2020 (5,000 breaks)

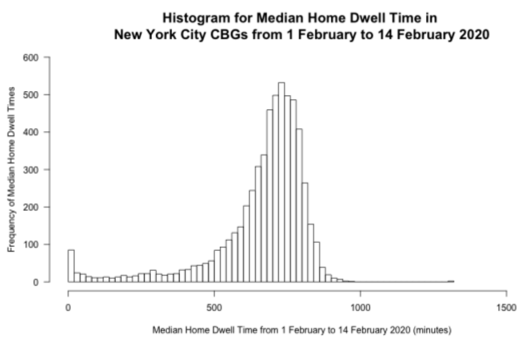


Figure 3.4: Histogram for Median Home Dwell Time in NYC, February 2020 (50 breaks)

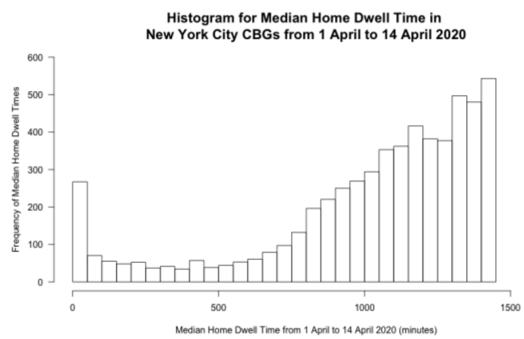


Figure 3.5: Histogram for Median Home Dwell Time in NYC, April 2020 (50 breaks)

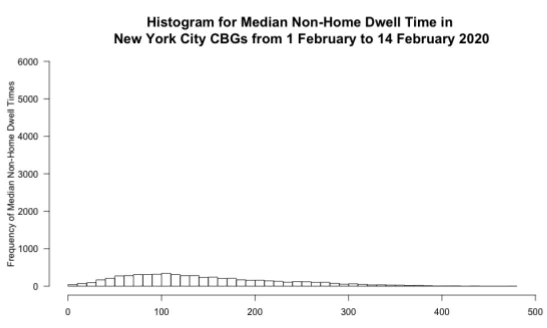


Figure 3.6: Histogram for Median Non-Home Dwell Time in NYC, February 2020

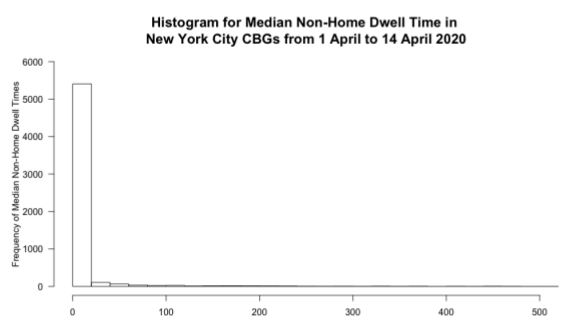


Figure 3.7: Histogram for Median Non-Home Dwell Time in NYC, April 2020

(50 breaks) (50 breaks)

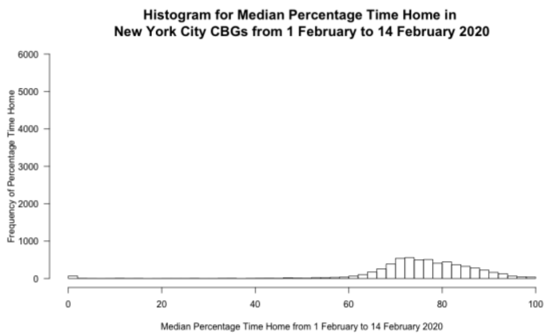


Figure 3.8: Histogram for Median Percentage Time at Home in NYC, February 2020 (50 breaks)

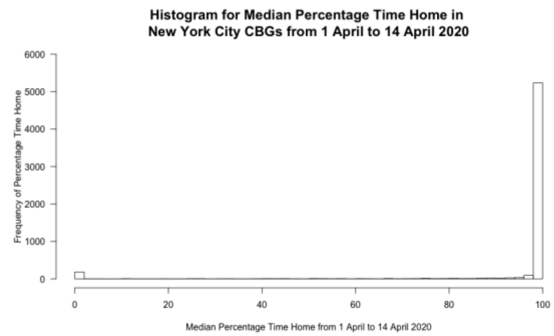


Figure 3.9: Histogram for Median Percentage Time at Home in NYC, April 2020 (50 breaks)

3.2. SafeGraph Social Distancing Metrics Dataset

SafeGraph, which is a private data company that provides location data for around seven million points of interest (POI) and aggregated mobility pattern data for over four million POIs in the United States, created a new dataset called "Social Distancing Metrics" for researchers and industry analysts to track daily physical distancing practices. SafeGraph generates their data using GPS pings from almost 20 million anonymous cellphone devices at the Census Block Group (CBG) level, which is a statistical division of Census Tracts and thus greater in geographical precision and data granularity (SafeGraph, n.d.-b; U.S. Census Bureau, n.d.). For my research purposes, this level of granularity provided data suitable for fine-grained analysis of human behavior. 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 (SafeGraph, n.d.-b). SafeGraph then groups devices into "home" CBGs based on their common nighttime location and provides aggregated data from the devices for each CBG (SafeGraph, n.d.-b).

To maintain individual cellphone users' privacy and ensure "ethical harvesting" of cellphone user data, SafeGraph applies a differential privacy algorithm known as DBSCAN clustering to all device count metrics except the field with the number of devices per CBG, such that there is no personal-identifying information saved from the devices (Goodale-Sussen & Kishore, 2020: online; SafeGraph, n.d.-b). This precaution may cause a discrepancy between the reported number of devices in a CBG and the actual number of devices in that CBG involved in data collection, particularly in sparsely populated CBGs; however, my analysis did not involve variables that included device counts. Table 3.1 shows the dependent variables I used to represent spatial mobility patterns at the CBG level.

Table 3.1: Dependent Variables Used in Regression Equations

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.

3.2.1. SafeGraph Data Collection

The first step of the data collection process was to download the SafeGraph data from their online data catalog, accessible at https://catalog.safegraph.io. Accessing the data for free as an academic researcher required a SafeGraph account, which I obtained by joining the SafeGraph COVID-19 Data Consortium, now called the Placekey community (SafeGraph, n.d.-a). SafeGraph approved my request for data on 15 April 2020, and my signed Non-Commercial Data License Agreement can be found in Appendix E. The data were in comma-separated values (CSV) format, with each file corresponding to a specific date (e.g. 1 February 2020) and each row in a file corresponding to one CBG. I downloaded data for the entire months of February and April 2020.

3.2.2. SafeGraph Data Preparation

I used PostgreSQL 12 to load and merge all of the SafeGraph CSV files onto a remote server in pgAdmin V4. I then queried for all of the rows in the database whose origin census block group started with the State of New York's Federal Information Processing Standards (FIPS) code

(36) and exported the selected rows into a new CSV file. Next, I narrowed the dataset from New York State CBGs to just the CBGs that intersected with a shapefile of NYC CBGs to create a new CSV file with data for just NYC. Lastly, I wrote a series of scripts with Python 3.9 that created one dataset with NYC data from 1 February to 14 February 2020 and another dataset with NYC data from 1 April to 14 April 2020.

3.3. American Community Survey

The American Community Survey (ACS) is an ongoing survey that releases new data every year on population and housing at the national, state, county, Census Tract, and even Census Block Group (CBG) level (United States Census Bureau, 2017). ACS users range from federal, state, and local agencies to educators and journalists, and the data are useful because they are more recent than the Census, whose data are collected every 10 years (United States Census Bureau, 2017). The ACS has three different types of data releases that come out each year: 1-year estimates, 1-year supplemental estimates, and 5-year estimates.

3.3.1. ACS Data Collection

I used demographic data from the ACS rather than the 2010 Census because incorporating the most recent possible demographic data provided a more accurate representation of the areas that I was analyzing in the context of a time-sensitive event like the COVID-19 pandemic. More specifically, I chose to use the ACS 5-year estimates, and while the U.S. Census Bureau released the most recent version, which covered 2015-2019, on 10 December 2020, I opted for the 5-year estimates from 2016 that covered 2012-2016 because SafeGraph had already organized the latter estimates into a set of CSV files, thus eliminating the extensive wrangling process required when working with Open Census Data. As was the case with collecting the SafeGraph mobility data, I accessed the ACS's 2016 5-year estimates by CBG using SafeGraph's online data catalog.

3.3.2. ACS Data Preparation and Annotation

Since the 5-year estimates data were reported at the CBG level, I merged the SafeGraph mobility patterns dataset with the 5-year estimates using the CBG ID code. I then renamed the columns with the full variable name and merged the CSVs for February, March, and April 2020 into one CSV file to facilitate calculations between monthly variables.

3.4. United States Census Bureau Boundaries

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, I used ArcMap v.10.7.1 to reproject both the NYC 2010 census block shapefile and the NY 2020 CBG shapefile to the WGS 1984 UTM Zone 18N coordinate system, which is the one recommended for representing data at a scale smaller than 1:10,000 (New York Standards Work Group, n.d.). Next, I dissolved the NYC census block shapefile into census block groups, and then intersected the result with the NY 2020 CBG shapefile using the GEOID column, which held the census block group codes. Lastly, I linked the shapefile with the CSV file containing SafeGraph and ACS data.

3.5. Statistical Analyses

I ran my analyses using the 3.6.2 version of the R programming language in version 1.2.5033 of RStudio. To compare the differences in distance traveled from home before and after the onset of COVID-19 in New York City, I divided the median distance traveled from home in the first two weeks of February 2020 for each CBG by the median mobility value in the first two weeks of April 2020 for the equivalent CBG to create a mobility ratio (MR). I then repeated the process to compare the differences in median home dwell time, median non-home dwell time, and median percentage time home. This approach loosely follows Badr et al. (2020)'s methodology, as they also created a MR variable to quantify mobility changes from baseline.

To avoid dividing by zero when computing the four mobility ratios, I changed all instances of "0" to "0.1" for the four variables representing changes in mobility between February and April 2020. For median distance traveled from home, 0.1 corresponded to one-tenth of a meter or ten centimeters. For median home dwell time and median non-home dwell time, 0.1 corresponded to one-tenth of a minute or six seconds. Lastly, for median percentage time spent at home, 0.1 corresponded to one-tenth of a percentage.

Next, I prepared the inputs needed for the regression models. First, I converted a shapefile containing NYC block groups to a neighbours list based on queen contiguity, which meant that a block group sharing a single boundary point with another counted as neighbors. Of the 6,863 block groups, only one had zero links to other block groups, which was the block group for Ellis Island. I also created a vector of power traces of the spatial weights matrix to use as input for the spatial lag model rather than the neighbours list (LeSage & Pace, 2009). Using the `trW` function, I set the type for powering the matrix to `moments,` which uses Smirnov and Anselin (2009)'s looping space saving algorithm to transform the matrix values.

My approach to computing the global and local regression models stayed the same for all four mobility ratios. I first fitted an ordinary least squares (OLS) linear regression model to the data in R to determine the global relations between mobility and sociodemographic factors. Table 3.2 shows the explanatory variables used in the regression equation.

Table 3.2: Explanatory Variables Used in Regression Equations

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 worked out to about $18,000/year (New York State Department of Labor, n.d.).
`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 noninstitutionalized population with no health insurance coverage

While each of the non-spatial OLS models yielded reasonable results, my low R^2 values led me to check for spatial autocorrelation using the Global Moran's Index correlation test for regression residuals. Spatial autocorrelation determines "how related the values of a variable are based on the locations where they were measured." (UCLA Institute for Digital Research & Education Statistical Consulting, n.d.: online). The R command from the `spdep` library, `lm.morantest,` required two inputs. The first was the OLS regression equation that provided the residuals for the linear correlation test. The second input was the large `listw` object holding the spatial relationship matrix I calculated based on queen contiguity (Anselin, 2007). I also set the zero policy to `TRUE` so that I could include the island block groups for Ellis Island in my analysis.

Next, I ran Lagrange Multiplier diagnostics for spatial dependence in linear models using the `lm.LMtests` command from the `spdep` library to determine how much my model performance would improve if I used the simple LM test for error dependence (LMerr), the simple LM test for a missing spatially lagged dependent variable (LMlag), the robust version RLMerr, which tests for error dependence in the possible presence of a missing lagged dependent variable and attempts to filter out possible false positives, or the other robust version, RLMlag, which has the same idea but tests the other way around (DataCamp, n.d.). The Lagrange Multiplier diagnostics tests also included the portmanteau test (SARMA) for completeness, even though this test is rarely the most suitable model (BurkeyAcademy, 2018). To determine which model to use, I compared the p-values of LMerr and LMlag. If they were both significant, I chose the model that corresponded to the robust model (RLMerr or RLMlag) with the lower p-value (Anselin, 2003, 2005). The figure in Appendix C.1 illustrates this decision process.

Next, I ran two spatial regression models. Both of these models determined whether the mobility patterns in surrounding CBGs affected the mobility pattern in one CBG (Medina & Solymosi, 2019). However, the first, called Spatially Lagged X (SLX), tested local spatial relations, which meant that surrounding block groups were those immediately adjacent to a block group. The second model, which was the spatial autoregressive (SAR) Spatial Lag model, tested global spatial relations, which meant that surrounding block groups were all of the observations in the data. I used the `spdep` command `lmSLX` to run my SLX models and the `impacts` command to observe the direct, indirect, and total effects. The `spdep` command `lagsarlm` for the SAR models was more complex, as I used the vector of power traces rather than the `listw` object and specified the approximate log-determinant method as `Chebyshev.` To summarize the impacts from the SAR models, I set the number of simulations to 5,000 to compute distributions for the impact measures.

The third and final spatial regression model was the Spatial Error Model (SEM), which I ran using the `spdep` command `errorsarlm.` This model also used the `Chebyshev` method, although I used the `listw` object rather than the vector of power traces as input. After running the model, I conducted a spatial Hausman test to determine if differences existed between the OLS and SEM coefficients. A significant result suggested that neither OLS nor SEM were "yielding regression parameter estimates matching the underlying parameters in the [data generating process]" (Pace & LeSage, 2008: 283). If I obtained a significant result from the spatial Hausman test, I ignored the OLS and SEM results.

3.6. Methodological Limitations

Prior studies using aggregated and anonymized mobility data have commented on the various limitations involved with this approach, including incomplete population representativeness, differences in individual mobility and SES factors, and perhaps incorrectly inferring causal relationships from complex interactions between factors (Chang et al., 2020; Chen et al., 2020; Lamb et al., 2021; Pullano et al., 2020). My methodology attempted to rectify some of these limitations by correcting for sampling bias, examining data at the CBG level, which is the second most granular geographical scale typically reported, and noting in my discussion when external factors may have influenced observed relationships between mobility and sociodemographic variables.

3.7. Ethics

While I am not personally interacting with the subjects represented in the SafeGraph or ACS datasets, I hold a responsibility as an academic researcher to ensure that my data usage does not overstep the boundaries protecting individual privacy. As Kishore et al. (2020) note, a pandemic is not justification for ignoring the risks to an individual's privacy associated with using personal data to calculate disease transmission-related metrics like sources of mass infection or mobility habits. To ensure that data privacy standards were upheld during my research, I selected data from a provider who applied a differential privacy algorithm to the data, signed a data license agreement, and did not attempt to re-identify subjects during my analysis by reporting findings generalized across NYC CBGs.

CHAPTER 4: RESULTS

This chapter contains all of the results I obtained from calculating descriptive statistics, running four regression models, and visualizing my results. I begin this chapter in Section 4.1 with the descriptive statistics for the sociodemographic factors, mobility variables, and mobility ratios. In Section 4.2, I present the outputs from the four regression models: the Ordinary Least Squares (OLS) linear regression model, the Spatially Lagged X (SLX) model, the Spatial Autoregressive (SAR) model, and the Spatial Error Model (SEM). This section contains the results from regression models with the dependent variable y as the change in median distance traveled from home. In the subsequent sections, I report the results from running the four regression models with the dependent variable y as the change in median home dwell time (Section 4.3), median non-home dwell time (Section 4.4), and median percentage time spent at home (Section 4.5).

4.1. Frequency Distributions and Descriptive Statistics

Both the frequency distribution and descriptive statistics for the change in median distance traveled from home indicate that, overall, most NYC block groups experienced decreased median mobility in the first two weeks of April 2020 compared to the first two weeks of February 2020. Based on how I calculated the ratio of February distances to April distances, a value less than one suggested that people in a block group traveled farther in April than in February, while a ratio value greater than one showed that people in a block group traveled farther in February than in April. Figure 4.1 is the frequency distribution of values for the ratio between median distance traveled from home in NYC during the first two weeks of February and the same data for the first two weeks of April. The histogram shows that most of the ratio values fell between zero and five, but with more values greater than one than less than one. The median values in Table 4.1 support this finding, since M = 3794.50 meters for median distance traveled from home in February versus M = 2448.00 meters for median distance traveled from home in April. Furthermore, in Table 4.2, M = 1.6178 for the change in median distance traveled from home, which indicates that there was a difference in the distance traveled from home between February and April.

The frequency distribution and descriptive statistics for the change in home dwell time indicate that, overall, most NYC block groups experienced increased home dwell time in the first two weeks of April 2020 compared to the first two weeks of February 2020. Based on how I calculated the ratio of February home dwell times to April home dwell times, a value less than one suggested that people in a block group stayed home for longer periods of time in April than in February, while a ratio value greater than one showed that people in a block group stayed home for longer periods of time in February than in April. Figure 4.2 is the frequency distribution of values for the ratio between median home dwell time in NYC during the first two weeks of February and the same data for the first two weeks of April. The histogram shows that almost all of the ratio values fell between zero and one, thus suggesting that median home dwell times across all block groups were mostly greater in April than in February. The median values in Table 4.1 support this finding, since M = 705.50 minutes for median home dwell time in February versus M = 1125.80 minutes for median home dwell time in April. Furthermore, in Table 4.2, M = 0.6170 for the change in median home dwell time, which indicates that there was a difference in the home dwell time between February and April.

The frequency distribution and descriptive statistics for the change in non-home dwell time indicate that, overall, most NYC block groups experienced decreased non-home dwell time in the first two weeks of April 2020 compared to the first two weeks of February 2020. In other words, people spent more time at home in April than in February. Based on how I calculated the ratio of February non-home dwell times to April non-home dwell times, a ratio value greater than one suggested that people in a block group spent time away from home for longer periods of time in February than in April, while a ratio value less than one showed that people in a block group spent time away from home for longer periods of time in April than in February. Figure 4.3 is the frequency distribution of values for the ratio between median non-home dwell time in NYC during the first two weeks of February and the same data for the first two weeks of April. Besides the histogram's large spike at around zero, most of the values were greater than one, thus suggesting that median non-home dwell times across all block groups were mostly greater in February than in April. The median values in Table 4.1 support this finding, since M = 128.00 minutes for median non-home dwell time in February versus M = 0.10 minutes for median non-home dwell time in April. Since I changed all values of "0" to "0.1" during data pre-processing, I interpreted the median non-home dwell time in April of 0.10 minutes to mean that a large majority of block groups experienced essentially no time spent away from home. The median ratio value for the change in median non-home dwell time shown in Table 4.2, M = 1085.0000, further supports that there was a difference in the non-home dwell time between February and April.

Lastly, the frequency distribution and descriptive statistics for the change in percentage time at home indicate that, overall, most NYC block groups experienced an increase in the percentage of time spent at home in the first two weeks of April 2020 compared to the first two weeks of February 2020. Based on how I calculated the ratio of February percentage time at home to April percentage time at home, a ratio value less than one suggested that people in a block group spent more of their time at home in April than in February, while a ratio value greater than one showed that people in a block group spent more of their time at home in February than in April. Figure 4.4 is the frequency distribution of values for the ratio between the median percentage time spent at home in NYC during the first two weeks of February and the same data for the first two weeks of April. Most of the ratio values lie between 0.5 and 1, thus suggesting that the median percentages of time spent at home across all block groups were mostly greater in April than in February. The median values in Table 4.1 support this finding, since M = 76.50 for median non-home dwell time in February versus M = 100.00 for median non-home dwell time in April. The median ratio value for the change in percentage time spent at home shown in Table 4.2, M = 0.7750, further supports that there existed a difference in the percentage time spent at home between February and April.

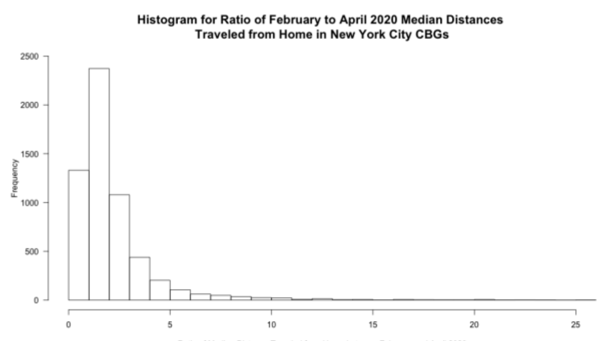


Figure 4.1: Histogram for the Ratio of Median Distance Traveled from Home in NYC between February and April 2020 (50 breaks)

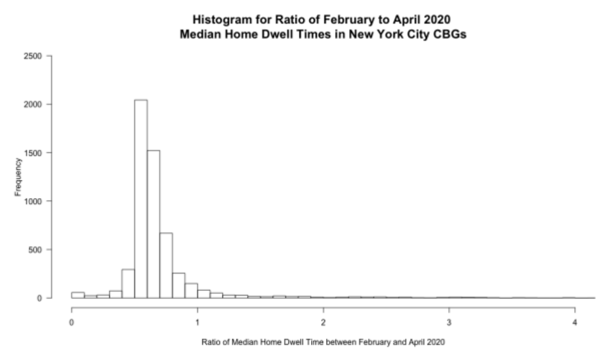


Figure 4.2: Histogram for the Ratio of Median Home Dwell Times in NYC between February and April 2020 (100,000 breaks)

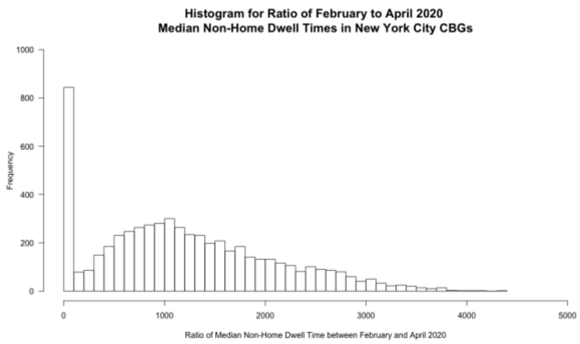


Figure 4.3: Histogram for the Ratio of Median Non-Home Dwell Times in NYC between February and April 2020 (50 breaks)

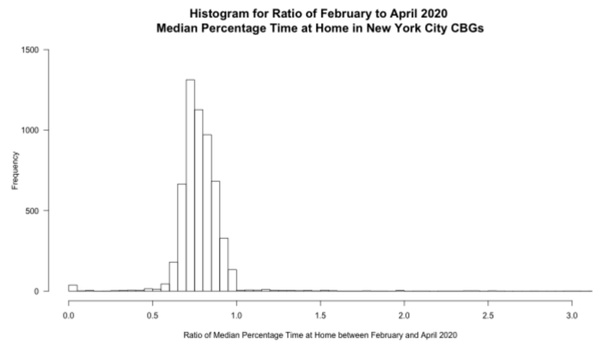


Figure 4.4: Histogram for the Ratio of Median Percentage Time at Home in NYC between February and April 2020 (35,000 breaks)

Table 4.1: Descriptive Statistics for Mobility Variables

Variable	Month (in 2020)	Minimum	1st Quarter	Median	Mean	3rd Quarter	Maximum	NA's
Median distance traveled from home (meters)	Feb	142.50	2944.20	3794.50	4349.20	4908.80	835874.00	1056
	Apr	89.00	1542.00	2448.00	11410.00	3880.00	1166288.00	1062
Median home dwell time (min)	Feb	0.10	622.00	705.50	660.60	762.00	1316.50	1056
	Apr	0.10	874.20	1125.80	1016.60	1307.00	1438.00	1059
Median non-home dwell time (min)	Feb	0.10	81.00	128.00	143.20	193.00	475.50	1056
	Apr	0.10	0.10	0.10	16.31	0.10	1374.00	1059
Median percentage time home	Feb	0.10	71.00	76.50	75.93	83.00	100.00	1056
	Apr	0.10	100.00	100.00	95.15	100.00	100.00	1059

Note: Values of "0" were replaced with "0.1" for median home dwell time, non-home dwell time, and percentage time home. Outlier data were retained.

Table 4.2: Descriptive Statistics for Mobility Ratios Variables between February and April 2020

Ratio Variable	Minimum	1st Quarter	Median	Mean	3rd Quarter	Maximum	NA's
Median distance traveled from home	0.0002	1.0476	1.6178	2.1304	2.4536	41.5222	1075
Median home dwell time	0.0000	0.5590	0.6170	117.1270	0.7360	9265.0000	1073

Median non-home dwell time	0.0000	555.0000	1085.0000	1201.0000	1755.0000	4360.0000	1073
Percentage time home	0.0010	0.7200	0.7750	16.8280	0.8500	1000.0000	1073

Note: Outlier data were retained.

Table 4.3: Descriptive Statistics for Explanatory Variables by Census Block Group

Variable Name	Minimum	1st Quarter	Median	Mean	3rd Quarter	Maximum	NA's
\`age\`	10.30	31.90	36.20	37.40	42.0	87.2	1114
\`race\`	0.00	172.00	481.00	565.90	835.00	6476.00	1035
\`transport\`	0.00	190.00	304.00	344.20	456.00	4107.00	1035
\`female_workers\`	0.00	190.00	268.00	295.40	370.00	3565.00	1035
\`housing_occupancy_rent\`	0.00	0.00	0.00	14.36	20.00	386.00	1035
\`min_wage\`	0.00	54.00	98.00	129.30	170.00	876.00	1035
\`children\`	0.00	121.00	206.00	239.20	319.00	1890.00	1035
\`education\`	0.00	95.00	167.00	185.50	253.00	902.00	1035
\`health_insurance\`	0.00	50.00	110.00	143.30	198.00	1295.00	1035

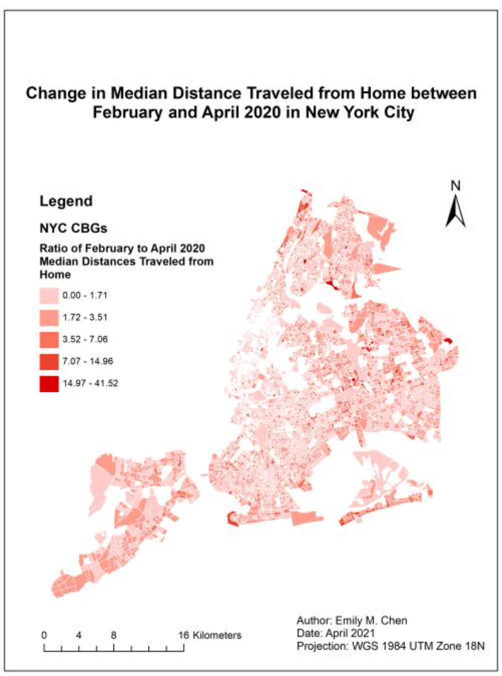
Map 4.1 depicts the change in median distance traveled from home between February and April 2020 at the CBG level in NYC, while Map 4.2 shows the change in median home dwell time during the same time period. Map 4.3 shows the change in median non-home dwell time and Map 4.4 illustrates the change in median percentage time at home. For all four maps, I removed outlier data by excluding the CBGs whose change in median home dwell time were greater than 2.0. Since I had changed "Null" values to -999 for data parsing purposes, I also excluded ratios that were less than 0. I used natural breaks to create the categories.

In Map 4.1, the darker red values correspond to a greater difference in median distance traveled from home between February and April. The smallest category, which includes ratio values from 0.00 to 1.71, shows block groups that had either a greater median distance traveled from home in April, which would make the ratio value less than one, or a slightly larger distance traveled from home in February, which would make the ratio value just above one. However, based on the descriptive statistics for the ratio of median distance traveled from home, I assumed that most of these values belong to the latter category. Block groups in the four largest categories have about a two-fold or greater increase in median travel distance from February to April.

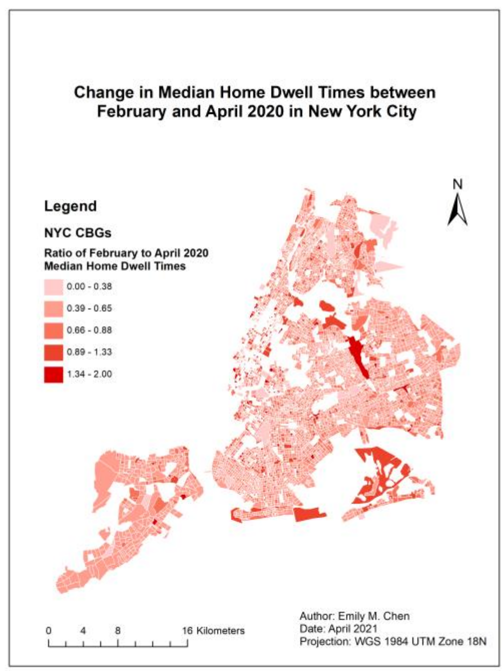
Map 4.2 shows the change in median home dwell time, with the darker red categories corresponding to a greater difference in home dwell time between February and April. The upper ratio value shown is 2.0 because I restricted the outlier values greater than this threshold; however, a majority of the block groups had ratio values less than 1.0, which correspond to more time spent at home in April than in February.

The darker red categories in Map 4.3 correspond to a greater difference in median non-home dwell times between February and April. The first category has ratio values ranging from 0.03 to 495.00. Since I used natural breaks to create the categories, I interpreted this enormous gap to mean that there were few values less than one, which would correspond to more time spent away from home in April, and many more values greater than 495, which would suggest that block group experienced enormous differences in non-home dwell time, with greater time spent away from home in February.

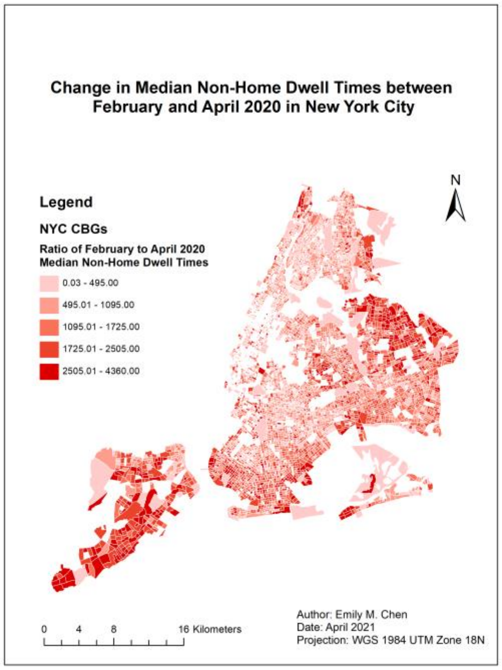
Lastly, Map 4.4 illustrates the change in median percentage time at home. Most values are less than one, thus showing that a majority of block groups experienced a greater percentage of time at home in April than in February.



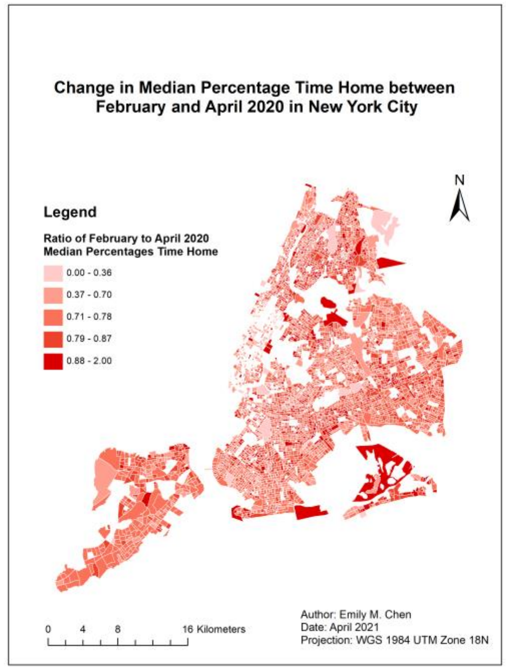
Map 4.1: Change in Median Distance Traveled from Home between February and April 2020 in NYC



Map 4.2: Change in Median Home Dwell Time between February and April 2020 in NYC



Map 4.3: Change in Median Non-Home Dwell Time between February and April 2020 in NYC



Map 4.4: Change in Median Percentage Time at Home between February and April 2020 in NYC

4.2. Effects of Demographic Factors on Distance Traveled from Home in NYC

In this section, I explore the effects of various demographic factors on the change in median distance traveled from home using four regression models. The first model was an ordinary least squares (OLS) regression, also known as a simple linear model, and does not involve a spatial component. Appendix D.1 contains the summary statistics from the OLS regression with change in median distance traveled from home as the dependent variable. Despite including nine explanatory variables in the model, I obtained an R^2 value of 0.02913, which was quite low. This result led me to check if my residuals contained spatial autocorrelation.

To test for spatial autocorrelation, I performed a Global Moran's Index linear correlation for regression residuals test. The null hypothesis is that there is no spatial correlation in the residuals, whereas the alternative hypothesis states that there is spatial correlation in the residuals. Appendix D.2 shows the results of this test. Based on the difference in the observed Moran's I value (0.0661) and the expected value (-0.0005), as well as the significant p-value ($p < 0.001$ for $\alpha = 0.05$), I rejected the null hypothesis and accepted the alternative hypothesis that there existed spatial autocorrelation in the residuals. This result indicated that the OLS model was no longer the appropriate model to use with change in median distance traveled from home as the dependent variable. Instead, I needed to use a spatial regression model.

To determine which spatial model might best fit the data, I ran a Lagrange Multiplier Diagnostic Tests for Spatial Dependence with the dependent variable as median distance traveled from home. Appendix D.3 shows the results of these tests. Since the p-values for both LMerr and LMlag were statistically significant ($p < 0.001$), I compared the p-values for the robust versions, RLMerr and RLMlag. While the p-values for these models were also statistically significant ($p < 0.001$), the p-value for RLMlag was smaller than the one for RLMerr. Therefore, the lag model appeared to be a more appropriate fit for the data.

The first spatial regression model, the Spatial Durbin Model or Spatially Lagged X (SLX) model, was a local spatial regression model. The summary statistics of the model are shown in Appendix D.4. There were six explanatory variables with significant results ($\alpha = 0.05$) both within a block group, which relates to the direct effect, and the neighboring block groups, which relate to the indirect effects (see Appendix D.5). These six variables were median age, number of people who took public transit (excluding taxicabs) to work, number of female workers 16 years and older, number of renter occupied housing units with over 1.5 occupants per room, number of families with children under 18, and number of people 25 years and older whose highest degree earned is a regular high school diploma.

The positive coefficient estimate associated with the original age variable indicated that a block group with a higher median age experienced increased change in distance traveled from home ($p < 0.001$). The neighboring block groups with a greater number of people who take public transit to work also experienced increased changes in distance traveled from home ($p < 0.001$). Additionally, the positive total effect impact measure shown in Appendix D.5 indicates that if the median age in every block group increased, the median change in distance traveled from home would also increase overall ($p < 0.001$). These findings indicate that block groups with an older median age likely experienced a decrease in distance traveled from home between April and February, since a smaller value in the denominator (April) caused the mobility ratio to increase. This result aligned with my expectation that block groups consisting of mostly older people would decrease their mobility more than those primarily made up of younger populations.

The positive coefficient estimate associated with the original transit variable indicated that a block group with a greater number of people who take public transit to work experienced increased change in distance traveled from home ($p < 0.01$). The neighboring block groups with a greater number of people who take public transit to work also experienced increased changes in distance traveled from home ($p < 0.05$). Additionally, the positive total effect impact measure in Appendix D.5 indicates that if the number of public transit users in every block group increased, the median change in distance traveled from home would also increase overall ($p < 0.001$). These findings indicate that block groups with more public transit users likely experienced a decrease in distance traveled from home between April and February.

The negative coefficient estimate associated with the original female workers variable indicated that within a block group with a higher number of female workers over 16, the change in distance traveled from home decreased ($p < 0.001$). The neighboring block groups with a greater number of female workers over 16 also experienced a decrease in the change in distance traveled from home ($p < 0.001$). Additionally, the negative total effect impact measure in Appendix D.5 indicates that if the number of female workers over 16 in every block group increased, the median change in distance traveled from home would decrease overall ($p < 0.001$). These findings indicate that block groups with more working women likely experienced an increase in distance traveled from home between April and February, since a larger value in the denominator (April) caused the mobility ratio to decrease.

The positive coefficient estimate associated with the original housing occupancy variable indicated that a block group with a greater number of renter occupied housing units with over 1.5 occupants per room experienced increased change in distance traveled from home ($p < 0.05$). The neighboring block groups with a greater number of renter occupied housing units with over 1.5 occupants also experienced increased changes in distance traveled from home ($p < 0.05$). Additionally, the positive total effect impact measure in Appendix D.5 indicates that if the number of renter occupied housing units with over 1.5 occupants in every block group increased, the median change in distance traveled from home would also increase overall ($p < 0.01$). These findings indicate that block groups with more crowded rental units likely experienced a decrease in distance traveled from home between April and February.

The positive coefficient estimate associated with the original children variable indicated that a block group with a greater number of families who have children under 18 years old experienced increased change in distance traveled from home ($p < 0.01$). The neighboring block groups with a greater number of families who have children also experienced increased changes in distance traveled from home ($p < 0.001$). Additionally, the positive total effect impact measure in Appendix D.5 indicates that if the number of families who have children in every block group increased, the median change in distance traveled from home would also increase overall ($p < 0.001$). These findings indicate that block groups with a greater number of families with children likely experienced a decrease in distance traveled from home between April and February.

Lastly, the positive coefficient estimate associated with the original education variable indicated that a block group with a greater number of people whose highest degree is a high school diploma experienced increased change in distance traveled from home ($p < 0.01$). The neighboring block groups with a greater number of high school graduates also experienced increased changes in distance traveled from home ($p < 0.01$). Additionally, the positive total effect impact measure in Appendix D.5 indicates that if the number of high school graduates in every block group increased, the median change in distance traveled from home would also increase overall ($p < 0.001$). These findings indicate that block groups with more people who graduated high school likely experienced a decrease in distance traveled from home between April and February.

The second spatial regression model, the Spatial Autoregressive (SAR) Lagged Response model, was a global spatial regression model. The summary statistics of the model are in Appendix D.6, with the rho value indicating that the change in median distance traveled from home in neighboring block groups had a positive effect on the change in median travel distance within a block group ($p < 0.001$). Unlike the SLX model, interpretations of the SAR model must draw upon the impacts reported in Appendix D.7 rather than the coefficient estimates and their p-values because of an infinite feedback loop on the y-value in which an increase in a block group's y-value will affect the neighboring block groups' y-values, which will in turn affect the individual block group's y-value. It is also important to note that since the model simulated the impact measures' p-values ($R = 5{,}000$ iterations), the p-values changed slightly between runs. Therefore, I provided a caveat for variables whose p-values fluctuated around 0.001 or greater and assumed that variables with p-values much lower than 0.001 were always significant (see Appendix D.7). The four variables with consistently significant simulated p-values for the impact measures were median age, number of female workers, number of families with children, and number of people whose highest degree is a high school diploma. The two variables whose simulated p-values fluctuated around $p < 0.001$ were number of White-only residents and number of public transit users. Lastly, the two variables whose simulated p-values fluctuated around $p < 0.05$ were number of renter occupied housing units with over 1.5 occupants per room and number of households who earned less than $25,000 a year.

The positive direct impact value associated with the age variable indicates that if the median age of block group A were to increase, A's change in travel distance from home would also increase. Similarly, the positive indirect impact value shows that an increase in the median age of A's neighboring block groups would increase A's change in travel distance. Since the SAR model is a global spatial model, a second interpretation of the indirect impact is that an increase in A's median age would lead to increased change in travel distance for all block groups in the data. These findings indicate that block groups with an older median age likely experienced shorter travel distances from home in April compared to February.

The negative direct impact value associated with the female workers variable indicates that if the number of female workers in block group A were to increase, A's change in travel distance from home would decrease. Similarly, the negative indirect impact value shows that an increase in the number of female workers in A's neighboring block groups would decrease A's change in travel distance. Additionally, an increase in A's number of female workers would lead to decreased change in travel distance for all block groups in the data. These findings indicate that block groups with more working females likely experienced greater travel distances from home in April compared to February.

The positive direct impact value associated with the children variable indicates that if the number of families with children in block group A were to increase, A's change in travel distance from home would also increase. Similarly, the positive indirect impact value shows that an increase in the number of families with children in A's neighboring block groups would increase A's change in travel distance. Additionally, an increase in A's number of families with children would lead to increased change in travel distance for all block groups in the data. These findings indicate that block groups with more families with children likely decreased their travel distances from home in April compared to February.

The positive direct impact value associated with the education variable indicates that if the number of people with high school diplomas in block group A were to increase, A's change in travel distance from home would also increase. Similarly, the positive indirect impact value shows that an increase in the number of high school graduates in A's neighboring block groups would increase A's change in travel distance. Additionally, an increase in A's number of high school graduates with children would lead to increased change in travel distance for all block groups in the data. These findings indicate that block groups with more high school graduates likely decreased their travel distances from home in April compared to February.

For the variables with p-values fluctuating around 0.001 or greater, the interpretations are very similar. The negative direct and indirect impact values associated with the number of White-only residents indicate that block groups with more White-only residents likely experienced greater travel distances from home in April compared to February. Similarly, the negative direct and indirect impact values associated with the number of households who earned less than $25,000 a year indicate that block groups with more low-income residents likely experienced greater travel distances from home in April. In terms of the number of public transit users and renter occupied housing units, the positive direct and indirect impact values associated with these two variables suggest that block groups with more public transit users and crowded rental units likely experienced shorter travel distances from home in April.

The third and final spatial regression model was the Spatial Error Model (SEM). While the Lagrange tests I ran previously (see Appendix D.3) suggested that the lag model was more appropriate than the error model for a regression with median travel distance as the dependent variable, I included the SEM model in my analysis because the results for LMerr and RMerr were both significant at $\alpha = 0.05$. However, I did not report the summary statistics because the statistically significant spatial Hausman test result ($p < 0.001$ at $\alpha = 0.05$) shown in Appendix D.8 confirmed there were enough differences in the coefficients to establish that neither OLS nor SEM were the right models to use for estimating the coefficients.

4.3. Effects of Demographic Factors on Home Dwell Time in NYC

In this section, I explore the effects of various demographic factors on the change in median home dwell time using four regression models. The first model was an ordinary least squares (OLS) regression. Appendix D.9 contains the summary statistics from the OLS regression with change in median home dwell time as the dependent variable. Despite including nine explanatory variables in the model, I obtained an R^2 value of 0.05419, which was quite low. This result led me to check if my residuals contained spatial autocorrelation.

To test for spatial autocorrelation, I performed a Global Moran's Index linear correlation for regression residuals test. Appendix D.10 shows the results of this test. Based on the difference in the observed Moran's I value (0.1201) and the expected value (-0.0005), as well as the significant p-value ($p < 0.001$ for $\alpha = 0.05$), I rejected the null hypothesis and accepted the alternative hypothesis that there existed spatial autocorrelation in the residuals. This result indicated that the OLS model was no longer the appropriate model to use with change in median home dwell time as the dependent variable. Instead, I needed to use a spatial regression model.

To determine which spatial model might best fit the data, I ran a Lagrange Multiplier Diagnostic Tests for Spatial Dependence with the dependent variable as median home dwell time. Appendix D.11 shows the results of these tests. Since the p-values for both LMerr and LMlag were statistically significant ($p < 0.001$), I compared the p-values for the robust versions, RLMerr and RLMlag. While the p-values for these models were also statistically significant ($p < 0.001$), the p-value for RLMlag was smaller than the one for RLMerr. Therefore, the lag model appeared to be a more appropriate fit for the data.

The first spatial regression model, the Spatial Durbin Model or Spatially Lagged X (SLX) model, was a local spatial regression model. The summary statistics of the model are shown in Appendix D.12. The positive coefficient estimate associated with the original female workers variable indicated that within a block group with a higher number of female workers over 16, the change in median home dwell time increased ($p < 0.01$). The neighboring block groups with a greater number of female workers over 16 also experienced an increase in the change in median home dwell time ($p < 0.001$). Additionally, the positive total effect impact measure in Appendix D.13 indicates that if the number of female workers over 16 in every block group increased, the median change in home dwell time would also increase overall ($p < 0.001$). These findings indicate that block groups with more working women likely experienced shorter home dwell times in April compared to February. The positive coefficient estimate associated with the original female workers variable indicated that within a block group with a higher number of female workers over 16, the change in median home dwell time increased ($p < 0.01$). The neighboring block groups with a greater number of female workers over 16 also experienced an increase in the change in median home dwell time ($p < 0.001$). Additionally, the positive total effect impact measure in Appendix D.13 indicates that if the number of female workers over 16 in every block group increased, the median change in home dwell time would also increase overall ($p < 0.001$). These findings indicate that block groups with more working women likely experienced shorter home dwell times in April compared to February.

The negative coefficient estimate associated with the original children variable indicated that within a block group with a higher number of families with children, the change in median home dwell time decreased ($p < 0.001$). The neighboring block groups with a greater number of families with children also experienced a decrease in the change in median home dwell time ($p < 0.001$). Additionally, the negative total effect impact measure in Error! Reference source not found. indicates that if the number of families with children in every block group increased, the median change in home dwell time would decrease overall ($p < 0.001$). These findings indicate that block groups with more families with children likely experienced longer home dwell times in April compared to February.

Lastly, the negative coefficient estimate associated with the original education variable indicated that within a block group with a higher number of people whose highest degree obtained is a high school diploma, the change in median home dwell time decreased ($p < 0.001$). The neighboring block groups with a greater number of high school graduates also experienced a decrease in the change in median home dwell time ($p < 0.001$). Additionally, the negative total effect impact measure in Appendix D.13 indicates that if the number of high school graduates in every block group increased, the median change in home dwell time would decrease overall ($p < 0.001$). These findings indicate that block groups with more high school graduates likely experienced longer home dwell times in April compared to February.

There were three explanatory variables with significant results ($\alpha = 0.05$) both within a block group, which relates to the direct effect, and the neighboring block groups, which relate to the indirect effects (see Appendix D.13). These three variables were number of female workers 16 years and older, number of families with children under 18, and number of people 25 years and older whose highest degree earned is a regular high school diploma.

The positive coefficient estimate associated with the original female workers variable indicated that within a block group with a higher number of female workers over 16, the change in median home dwell time increased ($p < 0.01$). The neighboring block groups with a greater number of female workers over 16 also experienced an increase in the change in median home dwell time ($p < 0.001$). Additionally, the positive total effect impact measure in Appendix D.13 indicates that if the number of female workers over 16 in every block group increased, the median change in home dwell time would also increase overall ($p < 0.001$). These findings indicate that block groups with more working women likely experienced shorter home dwell times in April compared to February.

The negative coefficient estimate associated with the original children variable indicated that within a block group with a higher number of families with children, the change in median home dwell time decreased ($p < 0.001$). The neighboring block groups with a greater number of families with children also experienced a decrease in the change in median home dwell time ($p < 0.001$). Additionally, the negative total effect impact measure in Appendix D.13 indicates that if the number of families with children in every block group increased, the median change in home dwell time would decrease overall ($p < 0.001$). These findings indicate that block groups with more families with children likely experienced longer home dwell times in April compared to February.

Lastly, the negative coefficient estimate associated with the original education variable indicated that within a block group with a higher number of people whose highest degree obtained is a high school diploma, the change in median home dwell time decreased ($p < 0.001$). The neighboring block groups with a greater number of high school graduates also experienced a decrease in the change in median home dwell time ($p < 0.001$). Additionally, the negative total effect impact measure in Appendix D.13 indicates that if the number of high school graduates in every block group increased, the median change in home dwell time would decrease overall ($p < 0.001$). These findings indicate that block groups with more high school graduates likely experienced longer home dwell times in April compared to February.

The second spatial regression model, the Spatial Autoregressive (SAR) Lagged Response model, was a global spatial regression model. The summary statistics of the model are in Appendix D.14, with the rho value indicating that the change in median home dwell time in neighboring block groups had a positive effect on the change in median home dwell time within a block group ($p < 0.001$). Like Section 4.2, I provided a caveat for variables whose p-values fluctuated around 0.001 or greater and assumed that variables with p-values much lower than 0.001 were always significant (see Appendix D.15). The four variables with consistently significant simulated p-values for the impact measures were median age, number of White-only residents, number of families with children, and number of people whose highest degree is a high school diploma. The two variables whose simulated p-values fluctuated around $p < 0.001$ were number of female workers and number of public transit users.

The negative direct impact values associated with the age, children, and education variables indicate that if the median age, number of families with children, or number of people with a high school diploma in block group A were to increase, A's change in home dwell time would decrease. Similarly, the negative indirect impact values show that an increase in the median age, number of families with children, or number of high school graduates in A's neighboring block groups would decrease A's change in home dwell time. Since the SAR model is a global spatial model, a second interpretation of the indirect impact is that an increase in A's median age, number of families with children, or number of high school graduates would lead to decreased change in home dwell time for all block groups in the data. These findings indicate that block groups with an older median age, greater number of families with children, and high school graduates likely experienced greater home dwell times in April compared to February.

The positive direct impact value associated with the race variable indicates that if the number of White-only residents in block group A were to increase, A's change in home dwell time would also increase. Similarly, the negative indirect impact value shows that an increase in the number of White-only residents in A's neighboring block groups would increase A's change in home dwell time. Additionally, an increase in A's number of White-only residents would lead to increased change in home dwell time for all block groups in the data. These findings indicate that block groups with more White-only residents likely experienced shorter home dwell times in April compared to February.

For the variables with p-values fluctuating around 0.001, the interpretations are very similar. The negative direct and indirect impact values associated with the number of public transit users indicate that block groups with more public transit users likely experienced greater home dwell times in April compared to February. Meanwhile, the positive direct and indirect impact values associated with the number of female workers indicate that block groups with more female workers likely experienced shorter home dwell times in April.

The third and final spatial regression model was the Spatial Error Model (SEM). While the Lagrange tests I ran previously (see Appendix D.11 suggested that the lag model was more appropriate than the error model for a regression with median home dwell time as the dependent variable, I included the SEM model in my analysis because the results for LMerr and RMerr were both significant at $\alpha = 0.05$. However, I did not report the summary statistics because the statistically significant spatial Hausman test result (p < 0.001 at $\alpha = 0.05$) shown in Appendix

D.16 confirmed there were enough differences in the coefficients to establish that neither OLS nor SEM were the right models to use for estimating the coefficients.

4.4. Effects of Demographic Factors on Non-Home Dwell Time in NYC

In this section, I explore the effects of various demographic factors on the change in median non-home dwell time using four regression models. The first model was an ordinary least squares (OLS) regression. Appendix D.17 contains the summary statistics from the OLS regression with change in median non-home dwell time as the dependent variable. Despite including nine explanatory variables in the model, I obtained an R^2 value of 0.1344, which was quite low. This result led me to check if my residuals contained spatial autocorrelation.

To test for spatial autocorrelation, I performed a Global Moran's Index linear correlation for regression residuals test. Appendix D.18 shows the results of this test. Based on the difference in the observed Moran's I value (0.1170) and the expected value (-0.0005), as well as the significant p-value (p < 0.001 for $\alpha = 0.05$), I rejected the null hypothesis and accepted the alternative hypothesis that there existed spatial autocorrelation in the residuals. This result indicated that the OLS model was no longer the appropriate model to use with change in median non-home dwell time as the dependent variable. Instead, I needed to use a spatial regression model.

To determine which spatial model might best fit the data, I ran a Lagrange Multiplier Diagnostic Tests for Spatial Dependence with the dependent variable as median non-home dwell time. Appendix D.19 shows the results of these tests. Since the p-values for both LMerr and LMlag were statistically significant (p < 0.001), I compared the p-values for the robust versions, RLMerr and RLMlag. While the p-values for these models were also statistically significant (p < 0.001), the p-value for RLMlag was smaller than the one for RLMerr. Therefore, the lag model appeared to be a more appropriate fit for the data.

The first spatial regression model, the Spatial Durbin Model or Spatially Lagged X (SLX) model, was a local spatial regression model. The summary statistics of the model are shown in Appendix D.20. There were six explanatory variables with significant results ($\alpha = 0.05$) both within a block group, which relates to the direct effect, and the neighboring block groups, which relate to the indirect effects (see Appendix D.21). These six variables were median age, number of White-only residents, number of public transit users, number of households earning less than

$25,000 a year, number of families with children under 18, and number of people 25 years and older whose highest degree earned is a regular high school diploma.

The positive coefficient estimates associated with the original age, race, children, and education variables indicated that within a block group with a higher median age, number of White-only residents, number of families with children, and number of people with a high school diploma, the change in median non-home dwell time increased ($p < 0.001$ for all). The neighboring block groups with a higher median age ($p < 0.001$), number of White-only residents ($p < 0.05$), number of families with children ($p < 0.05$), and number of high school graduates ($p < 0.001$) also experienced an increase in the change in median non-home dwell time. Additionally, the positive total effect impact measures in Appendix D.21 indicated that if the median age, number of White-only residents, number of families with children, and number of high school graduates in every block group increased, the median change in non-home dwell time would also increase overall ($p < 0.001$ for all). These findings indicate that block groups with a higher median age, greater number of White-only residents, greater number of families with children, and greater number of high school graduates likely experienced shorter non-home dwell times in April compared to February.

The negative coefficient estimate associated with the original transport and minimum wage variables indicated that within a block group with a higher number of public transit users and households earning less than $25,000 a year, the change in median non-home dwell time decreased ($p < 0.001$ for both). The neighboring block groups with a greater number of public transit users ($p < 0.01$) and low-income households ($p < 0.001$) also experienced a decrease in the change in median non-home dwell time. Additionally, the negative total effect impact measure in Appendix D.21 indicated that if the number of public transit users and low-income households in every block group increased, the median change in non-home dwell time would decrease overall ($p < 0.001$ for both). These findings indicate that block groups with more public transit users and low-income households likely experienced greater non-home dwell times in April compared to February.

The second spatial regression model, the Spatial Autoregressive (SAR) Lagged Response model, was a global spatial regression model. The summary statistics of the model are in Appendix D.22, with the rho value indicating that the change in median non-home dwell time in neighboring block groups had a positive effect on the change in median non-home dwell time within a block group ($p < 0.001$). Like the previous sections in this chapter, I provided a caveat for variables whose p-values fluctuated around 0.001 or greater and assumed that variables with p-values much lower than 0.001 were always significant (see Appendix D.23). The six variables with consistently significant simulated p-values for the impact measures were median age, number of White-only residents, number of public transit users, number of households earning less than $25,000 a year, number of families with children, and number of people whose highest degree is a high school diploma. The variable whose simulated p-value fluctuated around $p < 0.001$ was the number of female workers. Lastly, the two variables whose simulated p-values fluctuated around $p < 0.05$ were number of renter occupied housing units with over 1.5 occupants per room and number of people from the civilian noninstitutionalized population without health insurance coverage.

The positive direct impact values associated with the age, race, children, and education variables indicate that if the median age, number of White-only residents, number of families with children, and number of people whose highest obtained degree was a high school diploma in block group A were to increase, A's change in non-home dwell times would also increase. Similarly, the positive indirect impact value shows that an increase in the number of White-only residents in A's neighboring block groups would increase A's change in travel distance. Additionally, an increase in A's median age, number of White-only residents, number of families with children, and number of high school graduates would lead to increased change in median non-home dwell times for all block groups in the data. These findings indicate that block groups with older residents, greater number of White-only residents, greater number of families with children, and greater number of high school graduates likely experienced smaller median non-home dwell times in April compared to February.

The negative direct impact values associated with the transport and minimum wage variables indicate that if the number of public transit users and low-income households in block group A were to increase, A's change in non-home dwell time would decrease. Similarly, the negative indirect impact values show that an increase in the number of public transit users and low-income households in A's neighboring block groups would decrease A's change in non-home dwell time. Since the SAR model is a global spatial model, a second interpretation of the indirect impact is that an increase in A's number of public transit users and low-income households would lead to decreased change in non-home dwell time for all block groups in the data. These findings indicate that block groups with a greater number of public transit users and low-income households likely experienced larger non-home dwell times in April.

For the female worker variable with p-value fluctuating around 0.001, the interpretation is very similar to prior explanations. The positive direct and indirect impact values associated with the number of female workers indicate that block groups with more female workers likely experienced shorter non-home dwell times in April.

Lastly, for the variables with p-values fluctuating around 0.05, the negative direct and indirect impact values associated with the number of renter occupied housing units indicate that block groups with more crowded rental units likely experienced greater non-home dwell times in April than in February. Meanwhile, the positive direct and indirect impact values associated with the number of uninsured people indicate that block groups with more uninsured people likely experienced shorter non-home dwell times in April.

The third and final spatial regression model was the Spatial Error Model (SEM). While the Lagrange tests I ran previously (see Appendix D.19) suggested that the lag model was more appropriate than the error model for a regression with median non-home dwell time as the dependent variable, I included the SEM model in my analysis because the results for LMerr and RMerr were both significant at $\alpha = 0.05$. However, I did not report the summary statistics because the statistically significant spatial Hausman test result ($p < 0.001$ at $\alpha = 0.05$) shown in Appendix D.24 confirmed there were enough differences in the coefficients to establish that neither OLS nor SEM were the right models to use for estimating the coefficients.

4.5. Effects of Demographic Factors on Percentage Time at Home in NYC

In this section, I explore the effects of various demographic factors on the change in median percentage time spent at home using four regression models. The first model was an ordinary least squares (OLS) regression. Appendix D.25 contains the summary statistics from the OLS regression with change in median percentage time spent at home as the dependent variable. Despite including nine explanatory variables in the model, I obtained an R^2 value of 0.05837, which was quite low. This result led me to check if my residuals contained spatial autocorrelation.

To test for spatial autocorrelation, I performed a Global Moran's Index linear correlation for regression residuals test. Appendix D.26 shows the results of this test. Based on the difference in the observed Moran's I value (0.1133) and the expected value (-0.0005), as well as the significant p-value ($p < 0.001$ for $\alpha = 0.05$), I rejected the null hypothesis and accepted the alternative hypothesis that there existed spatial autocorrelation in the residuals. This result indicated that the OLS model was no longer the appropriate model to use with change in median percentage time spent at home as the dependent variable. Instead, I needed to use a spatial regression model.

To determine which spatial model might best fit the data, I ran a Lagrange Multiplier Diagnostic Tests for Spatial Dependence with the dependent variable as median percentage time spent at home. Appendix D.27 shows the results of these tests. Since the p-values for both LMerr and LMlag were statistically significant ($p < 0.001$), I compared the p-values for the robust versions, RLMerr and RLMlag. While the p-values for these models were also statistically significant ($p < 0.001$), the p-value for RLMlag was smaller than the one for RLMerr. Therefore, the lag model appeared to be a more appropriate fit for the data.

The first spatial regression model, the Spatial Durbin Model or Spatially Lagged X (SLX) model, was a local spatial regression model. The summary statistics of the model are shown in Appendix D.28. There were three explanatory variables with significant results ($\alpha = 0.05$) both within a block group, which relates to the direct effect, and the neighboring block groups, which relate to the indirect effects (see Appendix D.29). These three variables were number of female workers, number of families with children under 18, and number of people 25 years and older whose highest degree earned is a regular high school diploma.

The positive coefficient estimate associated with the original female worker variable indicated that within a block group with a greater number of female workers, the change in median percentage time spent at home increased ($p < 0.01$). The neighboring block groups with a greater number of female workers also experienced an increase in the change in median percentage time spent at home ($p < 0.001$). Additionally, the positive total effect impact measures in Appendix D.29 indicated that if the number of female workers in every block group increased, the median change in percentage time at home would also increase overall ($p < 0.001$). These findings indicate that block groups with a higher number of female workers likely experienced smaller percentage times at home in April compared to February.

The negative coefficient estimate associated with the original children and education variables indicated that within a block group with a higher number of families with children and number of high school graduates, the change in median percentage time at home decreased ($p < 0.001$ for both). The neighboring block groups with a greater number of families with children and number of high school graduates also experienced a decrease in the change in the median percentage time at home ($p < 0.001$ for both). Additionally, the negative total effect impact measure in Appendix D.29 indicated that if the number of families with children and high school graduates in every block group increased, the median change in percentage time at home would decrease overall ($p < 0.001$ for both). These findings indicate that block groups with more families with children and high school graduates likely experienced higher percentages of time at home in April compared to February.

The second spatial regression model, the Spatial Autoregressive (SAR) Lagged Response model, was a global spatial regression model. The summary statistics of the model are in Appendix D.30, with the rho value indicating that the change in median percentage time at home in neighboring block groups had a positive effect on the change in median percentage time at home within a block group ($p < 0.001$). Like the previous sections in this chapter, I provided a caveat for variables whose p-values fluctuated around 0.001 or greater and assumed that variables with p-values much lower than 0.001 were always significant (see Appendix D.31). The five variables with consistently significant simulated p-values for the impact measures were median age, number of White-only residents, number of public transit users, number of families with children, and number of people whose highest degree is a high school diploma. The two variables whose simulated p-values fluctuated around $p < 0.05$ were number of female workers and number of renter occupied housing units with over 1.5 occupants per room.

The positive direct impact values associated with the race variable indicate that if the number of White-only residents in block group A were to increase, A's change in percentage time at home would also increase. Similarly, the positive indirect impact value shows that an increase in the number of White-only residents in A's neighboring block groups would increase A's change in percentage time at home. Additionally, an increase in A's number of White-only residents would lead to increased change in median percentage time at home for all block groups in the data. These findings indicate that block groups with a greater number of White-only residents likely experienced smaller median percentages of time spent at home in April compared to February.

The negative direct impact values associated with the age, transport, children, and education variables indicate that if the median age, number of public transit users, number of families with children, and number of people whose highest degree obtained is a high school diploma in block group A were to increase, A's change in percentage time at home would decrease. Similarly, the negative indirect impact values show that an increase in the median age, number of public transit users, number of families with children, and number of high school graduates in A's neighboring block groups would decrease A's change in percentage time at home. Since the SAR model is a global spatial model, a second interpretation of the indirect impact is that an increase in A's median age, number of public transit users, number of families with children, and number of high school graduates would lead to decreased change in percentage time at home for all block groups in the data. These findings indicate that block groups with a greater median age, number of public transit users, number of families with children, and number of high school graduates likely experienced larger percentage times spent at home in April.

For the variables with p-values fluctuating around 0.05, the positive direct and indirect impact values associated with the number of female workers and number of rental units with more than 1.5 occupants per room indicate that block groups with a greater number of female workers and crowded rental units likely experienced smaller percentage times spent at home in April.

The third and final spatial regression model was the Spatial Error Model (SEM). While the Lagrange tests I ran previously (see Appendix D.27) suggested that the lag model was more appropriate than the error model for a regression with median percentage time at home as the dependent variable, I included the SEM model in my analysis because the results for LMerr and RMerr were both significant at $\alpha = 0.05$. However, I did not report the summary statistics because the statistically significant spatial Hausman test result (p < 0.001 at $\alpha = 0.05$) shown in Appendix D.32 confirmed there were enough differences in the coefficients to establish that neither OLS nor SEM were the right models to use for estimating the coefficients.

CHAPTER 5: DISCUSSION

In this chapter, I address the implications of my results and how they relate to my main research questions. In Section 5.1, I explore how my findings help answer my research aim of understanding which sociodemographic factors had the most effect on population changes in mobility. In Section 5.2, I discuss my approach to determining which of the four mobility variables most accurately represented physical distancing adherence.

5.1. Sociodemographic Factors and Their Effects on Changes in Mobility

My main research aim sought to understand which sociodemographic factors had the most effect on population change in mobility in New York City before and after the implementation of COVID-19-related lockdown measures in March 2020. To answer this first question, I chose nine noncollinear explanatory variables and ran four regression models 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, I rejected the null hypothesis and accepted the alternative hypothesis that there existed spatial autocorrelation in the residuals from all OLS models. Similarly, the significant p-value ($p < 0.001$ for $\alpha = 0.05$) obtained for all of 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, I used only the Spatially Lagged X (SLX) and Spatial Autoregressive (SAR) models to interpret my results. Table 5.1 summarizes the findings from these two models.

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 or not the independent variables cause changes in the dependent variable, the interpretations of which explanatory variables have an effect on mobility are still valid.

Table 5.1: Summary of Results from the Spatially Lagged X and Spatial Autoregressive 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.

5.1.1. Effect of Age on Mobility

For the SLX model, the median age of a block group correlated positively with the change in median distance traveled from home ($p < 0.001$) and the change in median non-home dwell time ($p < 0.001$). This result indicated that as the median age of a block group increased, it experienced both comparatively shorter distances traveled from home and also comparatively shorter amounts of time spent away from home in April. Thus, older residents were not only staying home for longer but also traveling shorter distances in April. Given that age alone posed the most significant risk factor for dying from COVID-19, with older populations affected much more severely by the disease compared to younger populations (Mueller et al., 2020; Santesmasses et al., 2020; Williamson et al., 2020), it makes sense that older people would curtail activities and time spent away from their homes to a greater extent if they could. Additionally, the scale of the SLX direct, indirect, and total impacts for age with median home and non-home dwell times as dependent variables were much higher than for any of the other explanatory variables. Whereas the impact scores for the rest of the explanatory variables were all between -3.0 and 2.0 for home dwell time and non-home dwell time, the direct, indirect, and total impact scores for age with home and non-home dwell times were [-6.65, -3.06, -9.71] and [14.52, 11.11, 25.63], respectively. These results indicate that increasing the median age within a block group had a greater effect on mobility defined as home and non-home dwell time than the other explanatory variables within that block group (direct impact), in the block group's immediate neighbors (indirect impact), and in all block groups in the data (total impact) (Golgher & Voss, 2016).

For the SAR model, the median age of a block group correlated positively with the change in median distance traveled from home ($p < 0.001$) and the change in median non-home dwell time ($p < 0.001$) as well, thus supporting the conclusions from the SLX model. The SAR model also yielded significant results with change in median home dwell time or change in median percentage time at home as the dependent variable. That the median age of a block group correlated negatively with both of these variables ($p < 0.001$ for both) indicated that as the median age increased in these block groups, the change in home dwell time and percentage time at home decreased. A decrease in these variables meant that residents of these block groups spent longer amounts of time at home in April. These findings support the earlier conclusions that older populations traveled shorter distances from home and stayed away from home less.

In addition to examining the impact values and simulated p-values for R = 5,000 iterations, I also noted the direct, indirect, and total impact scores. Similarly to the SLX model, whereas the impact scores for the rest of the explanatory variables were all between -2.5 and 1.5 for home dwell time and non-home dwell time, the direct, indirect, and total impact scores for age with home and non-home dwell times were [-6.39, -2.81, -9.20] and [15.08, 6.50, 21.58], respectively. These results indicate that increasing the median age within a block group had a greater effect on mobility defined as home and non-home dwell time than the other explanatory variables within that block group (direct impact) and in all block groups in the data (indirect and total impacts) (BurkeyAcademy, 2018; Golgher & Voss, 2016).

5.1.2. Effect of Race on Mobility

For the SLX model, the estimated number of White-only residents in a block group correlated positively with only the change in median non-home dwell time ($p < 0.001$). This result indicated that as the White-only population within a block group increased, those residents experienced comparatively shorter amounts of time spent away from home in April. Thus, block groups with more White-only residents were staying home for longer periods of time in April compared to February. My findings align logically with the fact that 75% of all NYC frontline workers, who held jobs that required them to leave their homes, are people of color (Stringer, 2020). Thus, I inferred that a greater proportion of non-White residents could not switch to remote work, which in turn led to longer times spent away from home compared to White-only residents.

Another possible explanation for longer times spent at home regardless of race was the unprecedented 14.8% national unemployment rate in April 2020 (Falk et al., 2021). However, a report from the Congressional Research Service found that national unemployment rates for White workers (14.2%) were lower than for Black workers (16.7%), and lower for non-Hispanics (13.6%) than for Hispanic workers (18.9%) (Falk et al., 2021). Furthermore, in NYC areas, Asian, Black, and Hispanic/Latinx adults experienced a 48%, 67%, and 68% loss of income since 13 March 2020 respectively, compared with 45% of White adults (Nischan, 2020). Notably, these numbers do not include income loss statistics for noncitizens, which is a group who experience income loss at a higher rate than citizens (Nischan, 2020). Thus, the observed difference in non-home dwell times between White-only and non-White residents cannot be explained entirely by unemployment rates.

For the SAR model, the estimated number of White-only residents in a block group correlated positively with the change in median home dwell time ($p < 0.001$), the change in median non-home dwell time ($p < 0.001$), and the change in percentage time spent at home ($p < 0.001$). These results were particularly interesting, since a positive correlation for both home and non-home dwell times meant that an increase in the number of White-only residents correlated with both less time spent at home and also less time spent away from home in April. While seemingly contradictory, one possible explanation is that the differences arose based on income level, where higher income White-only residents spent less time away from home and lower income White-only residents spent less time at home. Further research on the interaction between race and income would clarify these results. The SAR model also yielded significant results with change in median distance traveled from home as the dependent variable. That the number of White-only residents correlated negatively with travel distance ($p < 0.05$) suggested that block groups with more White-only residents experienced increased travel distance from home in April. These findings prompt the need for future research that examines the origin and destination of trips made from these block groups, as one explanation could be that higher income White residents sheltered in place at a second home outside of the city during early lockdown measures (Frank, 2020; Gordon, 2020; Tully & Stowe, 2020). However, more detailed analyses are necessary to confirm this hypothesis.

5.1.3. Effect of Transport Method to Work on Mobility

For the SLX model, the estimated number of people who use public transit to travel to work in a block group correlated positively with the change in median distance traveled from home ($p < 0.001$) and negatively with the change in median non-home dwell time ($p < 0.001$). This result showed that an increase in the number of public transit users within a block group led to shorter distances traveled from home and longer non-home dwell times in April. While seemingly contradictory at first, one possible explanation for these phenomena could be that the overall decrease in public transit ridership, particularly on buses and subways, after the March 2020 lockdowns (Penney, 2021) contributed to a decrease in distances traveled from home for certain NYC populations, whereas essential workers, 55% of whom used the subway, bus, or rail to travel to work prior to the pandemic, continued to use public transit to get to work during the pandemic (Stringer, 2020). Analysis of trips taken from block group could reveal another possible scenario, which is if block groups with more public transit users correlated with a greater number of people spending time outside in nearby green spaces, which would contribute both to greater time away from home and also decreased travel distance from home.

For the SAR model, the estimated number of public transit users in a block group correlated positively with the change in median distance traveled from home ($p < 0.05$) and negatively with the change in median home dwell time ($p < 0.05$), the change in non-home dwell time ($p < 0.001$), and the change in percentage time spent at home ($p < 0.001$). These results were particularly interesting, since a negative correlation for both home and non-home dwell times meant that an increase in the number of public transit users correlated with both more time spent at home and also more time spent away from home in April. While seemingly contradictory, one possible explanation is that the differences arose based on other related explanatory factors, similarly to the SAR model results associated with White-only residents as a factor.

5.1.4. Effect of Female Worker Share on Mobility

For the SLX model, the estimated number of female workers correlated negatively with the change in median distance traveled from home ($p < 0.001$) and positively with the change in median home dwell time ($p < 0.001$) and change in percentage time at home ($p < 0.001$). This result showed that an increase in the number of female workers within a block group led to longer distances traveled from home, shorter home dwell times, and smaller percentage times at home in April. At first, I found this result surprising, given reports that women were more likely than men to quit their jobs and assume childcare responsibilities when schools closed (Bateman & Ross, 2020; Gogoi, 2020). However, 60% of all frontline workers in NYC are women, with the highest percentages in healthcare (74% are women) and childcare, homeless, food, and family services (81% are women) (Stringer, 2020). Thus, one possible explanation could be the decomposition of NYC frontline workers by sex, as it suggests that these women continued to leave home for work, which contributed to less time spent at home and greater distances traveled from home.

For the SAR model, the estimated number of female workers in a block group correlated negatively with the change in median distance traveled from home ($p < 0.01$) and positively with the change in median home dwell time ($p < 0.05$), the change in non-home dwell time ($p < 0.05$), and the change in percentage time spent at home ($p < 0.05$). These results were particularly interesting, since a positive correlation for both home and non-home dwell times meant that an increase in the number of public transit users correlated with both more time spent at home and also more time spent away from home in April. While seemingly contradictory, one possible explanation is that the differences arose based on other related explanatory factors, similarly to the SAR model results associated with the number of White-only residents and public transit users as factors.

5.1.5. Effect of Housing Occupancy on Mobility

For the SLX model, the estimated number of renter occupied housing units with more than 1.5 residents per room correlated negatively with the change in median distance traveled from home ($p < 0.01$). This result indicated that block groups with a greater proportion of crowded rental homes correlated with shorter travel distances in April, which was surprising since over half of all NYC frontline workers are renters (Stringer, 2020). However, since travel distance was the only mobility variable to yield a significant result for the SLX model, perhaps a more appropriate measurement would have been the total number of rental units or the number of renters rather than rental units with 1.5 or more people per room.

For the SAR model, the estimated number of renter occupied housing units with more than 1.5 residents per room in a block group correlated negatively with the change in median non-home dwell time ($p < 0.05$) and positively with the change in median distance traveled from home ($p < 0.05$) and the change in percentage time spent at home ($p < 0.05$). The results for non-home dwell time and percentage time home, which suggested that people in block groups with a higher number of crowded rental units spent less time at home in April, appeared contradictory to the result for distance traveled for home, which implied that people traveled shorter distances in these block groups. However, these differences may relate to other related factors, such as limited access to green space in crowded neighborhoods contributing to lower travel distances, and a high percentage (59%) of frontline workers being renters leading to more time spent away from home (Stringer, 2020).

5.1.6. Effect of Annual Income on Mobility

For the SLX model, the estimated number of households making less than $25,000 a year correlated negatively with the change in median non-home dwell time ($p < 0.001$). This result indicated that block groups with a greater proportion of households whose annual income amount corresponded to a full-time minimum wage job experienced greater time spent away from home in April. This result could come in part from the statistic that 8% of all frontline workers are at or below the poverty line, which is defined as $26,200 for a family of four (Stringer, 2020). However, 8% is a small proportion of frontline workers compared with the 24% of all frontline workers at or below twice the poverty line, which is defined as $52,400 for a family of four (Stringer, 2020). There are particularly high percentages of grocery, convenience, and drug store workers (35%), childcare, homeless, food, and family services workers (34%), and building cleaning services (39%) that fit into this latter category of twice the poverty line (Stringer, 2020). Thus, a more appropriate statistic to assess mobility trends might have been the number of households making less than \$55,000 a year, since this income bracket would include 32% of all frontline workers compared to just 8%.

For the SAR model, the estimated number of households making less than \$25,000 a year in a block group correlated negatively with the change in median distance traveled from home ($p < 0.05$) and median non-home dwell time ($p < 0.001$). These results indicated that an increase in the number of low-income households within a block group led to an increase in distance traveled from home and non-home dwell time in April. If only 8% of NYC frontline workers are at or below the federal poverty line (Stringer, 2020), this conclusion is somewhat surprising. Furthermore, a survey conducted in October 2020 found that 80% of NYC adults earning less than \$35,000 a year experienced a loss of income since 13 March 2020 compared with the 54% NYC area average (Nischan, 2020), suggesting that low-income households were the most affected by the staggering unemployment rate and thus did not have reason to leave their homes. Additional research exploring where residents of block groups with a higher number of low-income households may help clarify the correlation with increased non-home dwell time and travel distance from home.

5.1.7. Effect of Families with Children on Mobility

For the SLX model, the estimated number of families with children under age 18 correlated positively with the change in median distance traveled from home ($p < 0.001$) and the change in median non-home dwell time ($p < 0.001$). These results indicated that block groups with a greater number of families with children experienced shorter travel distances from home and less time spent away from home in April. Similarly, the number of families with children correlated negatively with the change in median home dwell time ($p < 0.001$) and median percentage time at home ($p < 0.001$), meaning that block groups with more families with children experienced greater amounts of time at home. These conclusions make 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 with regard to childcare left parents, particularly working mothers, with home school responsibilities that prompted some mothers to leave their jobs entirely (Bateman & Ross, 2020). 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 (Heggeness & Fields, 2020). While these results are based on national data, there is little reason to believe that this phenomenon did not extend 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 (Heggeness & Fields, 2020). Given that NY state was one of the first states to implement stay-at-home measures (see Figure 2.2), it seems likely that NYC working mothers fit into the category of being more likely to leave their jobs.

For the SAR model, the results were the same: the estimated number of families with children in a block group correlated positively with the change in median distance traveled from home ($p < 0.001$) and the change in median non-home dwell time ($p < 0.001$). These results indicated that an increase in the number of families with children within a block group led to an increase in distance traveled from home and non-home dwell time in April for all block groups. The SAR model also found that the estimated number of families with children in a block group correlated negatively with the change in median home dwell time ($p < 0.001$) and median percentage time at home ($p < 0.001$), meaning that block groups with more families with children experienced greater amounts of time at home. Thus, the SAR model results confirm the conclusions made from the SLX model results.

5.1.8. Effect of Educational Attainment on Mobility

For the SLX model, the estimated number of people with just a high school diploma correlated positively with the change in median distance traveled from home ($p < 0.001$) and the change in median non-home dwell time ($p < 0.001$). These results indicated that block groups with a greater number of high school graduates experienced shorter travel distances from home and less time spent away from home in April. Similarly, the number of people with a high school diploma correlated negatively with the change in median home dwell time ($p < 0.001$) and median percentage time at home ($p < 0.001$), meaning that block groups with more high school graduates experienced greater amounts of time at home. 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 (Nischan, 2020). 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 (Nischan, 2020). Without a job to go to, this demographic traveled shorter distances and stayed at home for longer periods of time.

For the SAR model, the results were the same: the estimated number of people with just a high school diploma in a block group correlated positively with the change in median distance traveled from home ($p < 0.001$) and the change in median non-home dwell time ($p < 0.001$). These results indicated that an increase in the number of high school graduates within a block group led to an increase in distance traveled from home and non-home dwell time in April for all block groups. The SAR model also found that the estimated number of people with just a high school diploma in a block group correlated negatively with the change in median home dwell time ($p < 0.001$) and median percentage time at home ($p < 0.001$), meaning that block groups with more high school graduates experienced greater amounts of time at home. That the SAR and SLX models yielded the same results for all mobility variables seemed to strengthen the findings that educational attainment significant correlated with mobility.

5.1.9. Effect of Health Insurance Status on Mobility

For the SLX mode, there were no significant correlations between the estimated number of people from the civilian noninstitutionalized population with no health insurance coverage and mobility. However, for the SAR model, the estimated number of non-insured people in a block group correlated positively with the change in median non-home dwell time ($p < 0.05$). This finding suggested that as the number of non-insured people in a block group increased, the time not spent at home in April decreased across all block groups. While 11% of all NYC frontline workers are uninsured, with higher percentages for grocery, convenience, and drug store workers (12.1%), trucking, warehouse, and postal service workers (14.8%), and building cleaning services (29.1%) (Stringer, 2020), the finding that uninsured workers spent more time at home was likely more related to the share of low-income workers who lost jobs due to COVID-19. Nationally, low- and middle-income workers are more likely to be uninsured (Institute of Medicine (US) Committee on the Consequences of Uninsurance, 2001), and 80% of NYC workers who earned less than $35,000 a year reported losing income due to COVID-19, which could have been from either reduced work or unemployment (Nischan, 2020).

5.2. Differences in Population Effects

Having obtained several significant results from the SLX and SAR regression models, I wanted to determine which of the variables measuring population mobility (median distance traveled from home, median home dwell time, median non-home dwell time, and median percentage time home) most accurately served as a proxy for physical distancing adherence. I used a three-step process to examine possible answers to this question. First, I examined the results from the four regression models in the context of the nine explanatory variables for each of the mobility measures and chose the mobility measurement that yielded the greatest number of logical results. Next, I evaluated how SafeGraph collected this data to see if their methodology helped create a more accurate measurement. Finally, I compared my findings with my hypothesis in Chapter 1.

The mobility measurement that yielded the greatest number of results from the SLX and SAR models was median non-home dwell time. Not only did every explanatory variable have a significant correlation with non-home dwell time for the SAR model, but also the SLX model's R^2 value (0.162) with this dependent factor was the highest. SafeGraph defined this measurement as the aggregated median dwell time of devices at places outside of their Geohash-7 home for the entire 24-hour period (SafeGraph, n.d.-b). The specificity of 153 meters by 153 meters to which devices were tracked outside of the home certainly made the measurement more accurate, as a smaller radius meant SafeGraph got as close as it could to determining when a device was outside the home without actually knowing the device's true home.

This finding that median non-home dwell time was the most accurate proxy for physical distancing adherence aligned with my original hypothesis that either home dwell time or non-home dwell time would be the most accurate factors. The other dependent variables, while also highly accurate, had limitations or dependencies. Median distance traveled from home could have provided greater insight into where people were going during the pandemic stay-at-home orders, but I had not included place visits in my model, so this variable seemed less likely to be the most accurate measurement. For percentage time spent at home, since the measurement was based on home dwell time calculations, it made sense that the results for home dwell time and percentage time at home aligned almost exactly, with the exception of housing occupancy and percent time at home for the SAR model. Therefore, if median home dwell time had been the most accurate measurement, there was a high chance that median percent time home would also be highly accurate. Since home dwell time did not yield the greatest number of significant results, non-home dwell time surpassed both home dwell time and percent time home.

CHAPTER 6: CONCLUSION

This chapter will address limitations of the research, suggest pathways for future research, and summarize the research findings.

6.1. Limitations of this Research

As was the case with prior literature using aggregated cellphone mobility data, the representativeness of SafeGraph's data made it challenging to draw definitive conclusions from regression models. Despite its exceptional size and granularity, the data came from fewer than 500,000 devices and accounted for only one-ninth of the NYC population. On the one hand, this share of the population may seem small, but on the other hand, mobility data from 500,000 devices is a staggeringly large sample size compared to early mobility research that relied on recruiting participants to self-report data. Furthermore, a vast majority of block groups contained data. Therefore, the limitation was worth noting in the sense that any conclusions drawn from my findings must acknowledge that they illustrate general population mobility trends from aggregated data. I could also strengthen my findings by running the four regression models using different mobility datasets and comparing the results. Additionally, summarizing differences in mobility datasets and providing a comprehensive evaluation of the strengths and limitations of each could help researchers choose the most appropriate datasets for their research questions (Dodge, 2021).

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 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.

Lastly, aggregated mobility data collected from everyday human behavior patterns are inherently messy. The motivations, desires, and beliefs of every human differ, so the findings that apply to one person might be entirely misaligned with the behavior of another person who has a similar sociodemographic profile. Thus, drawing conclusions other than sweeping generalizations can be difficult when using data collected from human interactions taking place in the real-world as opposed to a strictly controlled laboratory setting. One possible way to add an individualistic component and validate the broad population findings would be to conduct interviews with residents from the study area.

6.2. Pathways for Future Research

In addition to the research opportunities already discussed, there are five additional possibilities I will present. The first is to extend my methodology to data from other cities. A between-city comparison might provide greater insight into how stay-at-home policies affected regions differently based on sociodemographic patterns, public transit infrastructure, or population density. In addition to comparing cities, there are other explanatory factors that could be added to the regression models, such as the number of households who own second homes or citizenship status. Instead of simply sociodemographic variables, one could also use points of interest (POI) data to expand on the findings in this paper. For example, to better understand 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% of them had lost income during the pandemic (Nischan, 2020). Therefore, if younger workers were more likely to experience unemployment and 37% of NYC frontline workers are over 50 years old (Stringer, 2020), 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, there are several types of datasets that could be used to cross-reference these 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 be a fascinating next research topic. This last potential avenue of research 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 (Badr & Gardner, 2020). Their findings suggest that other non-pharmaceutical interventions like mask-wearing or handwashing played a significant role in mitigating the spread of the virus early on in the pandemic, so future research should add these factors to their models when looking at the relationship between mobility and case positivity.

6.3. Concluding Thoughts

My intention for this research was to provide fine-grained analysis for policymakers on the varying effects of lockdown measures and inform future strategies for infection mitigation and safe re-opening. My findings that there exist significant differences in mobility based on sociodemographic factors, particularly age, education level, and whether or not families have children, reinforce the need for physical distancing policies that acknowledge the demographic diversity present not only between but also within cities. Future research can both confirm these findings and also examine 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 sociodemographic effects on different measurements of mobility, this paper emphasizes that stay-at-home policies introduce unevenly distributed effects to different groups and that there are several ways to measure mobility patterns within a city.

REFERENCE LIST

Aleta, A., Martín-Corral, D., Pastore y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N. E., Halloran, M. E., Longini, I. M., Merler, S., Pentland, A., Vespignani, A., Moro, E., & Moreno, Y. (2020). Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nature Human Behaviour*, *4*(9), 964–971.

Anselin, L. (2003). *An Introduction to Spatial Regression Analysis in R* (pp. 1–25). http://labs.bio.unc.edu/Buckley/documents/AnselinIntroSpatRegres.pdf

Anselin, L. (2005). Exploring Spatial Data with GeoDaTM: A Workbook. In *Center for Spatially Integrated Social Science* (pp. 1–226).

Anselin, L. (2007). *Spatial Regression Analysis in R: A Workbook* (pp. 1–134). Center for Spatially Integrated Social Science. https://dces.wisc.edu/wp-content/uploads/sites/128/2013/08/W14_Anselin2007.pdf

Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., & Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*, *3099*(20), 1–8.

Badr, H. S., & Gardner, L. M. (2020). Limitations of using mobile phone data to model COVID-19 transmission in the USA. *The Lancet Infectious Diseases*, *3099*(20), 30861.

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

Bian, B., Li, J., Xu, T., & Foutz, N. (2020). Individualism During Crises. *SSRN Electronic Journal*, 1–57.

Brzezinski, A., Kecht, V., & Van Dijcke, D. (2020). The Cost of Staying Open: Voluntary Social Distancing and Lockdowns in the US. *SSRN Electronic Journal*, 1–34.

Buckee, C. O., Balsari, S., Chan, J., Crosas, M., Dominici, F., Gasser, U., Grad, Y. H., Grenfell, B., Halloran, M. E., Kraemer, M. U. G., Lipsitch, M., Metcalf, C. J. E., Meyers, L. A., Perkins, T. A., Santillana, M., Scarpino, S. V., Viboud, C., Wesolowski, A., & Schroeder, A. (2020). Aggregated mobility data could help fight COVID-19. *Science*, *368*(6487), 145–146. https://science.sciencemag.org/content/368/6487/146/tab-pdf

Budd, J., Miller, B. S., Manning, E. M., Lampos, V., Zhuang, M., Edelstein, M., Rees, G., Emery, V. C., Stevens, M. M., Keegan, N., Short, M. J., Pillay, D., Manley, E., Cox, I. J., Heymann, D., Johnson, A. M., & McKendry, R. A. (2020). Digital technologies in the public-health response to COVID-19. *Nature Medicine*, *26*(8), 1183–1192.

BurkeyAcademy. (2018). *Spatial Regession in R 1: The Four Simplest Models*. YouTube. https://www.youtube.com/watch?v=b3HtV2Mhmvk

Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2020). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, *589*(7840), 82–87.

Chen, Y., Jiao, J., Bai, S., & Lindquist, J. (2020). Modeling the Spatial Factors of COVID-19 in New York City. *SSRN Electronic Journal*.

City of New York. (2020). *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

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

DataCamp. (n.d.). *lm.LMtests: Lagrange Multiplier diagnostics for spatial dependence in linear models*. RDocumentation. Retrieved April 15, 2021, from https://www.rdocumentation.org/packages/spdep/versions/1.1-5/topics/lm.LMtests

Dave, D., Friedson, A., McNichols, D., & Sabia, J. (2009). The Contagion Externality of a Superspreading Event: The Sturgis Motorcycle Rally and COVID-19. *National Bureau of Economic Research*, *October*, 1–39.

Dincer, O. C., & Gillanders, R. (2020). Shelter in Place? Depends on the Place: Corruption and Social Distancing in American States. *Social Science & Medicine*, *November*, 113569.

Ding, W., Levine, R. E., Lin, C., & Xie, W. (2020). Social Distancing and Social Capital: Why U.S. Counties Respond Differently to COVID-19. *SSRN Electronic Journal*. http://www.nber.org/papers/w27393

Dodge, S. (2021). Using movement as a marker to trace human activity and contact patterns. *McGill Department of Geography GeoSpectives Seminar Series*. https://www.mcgill.ca/geography/geospectives

Eisenberg, A., & Touré, M. (2020). *De Blasio: NYC schools will close as of Monday, may not reopen this year*. Politico. https://www.politico.com/states/new-york/albany/story/2020/03/15/de-blasio-nyc-schools-will-close-as-of-monday-may-not-reopen-this-year-1267141

Falk, G., Carter, J. A., Nicchitta, I. A., Nyhof, E. C., & Romero, P. D. (2021). *Unemployment Rates During the COVID-19 Pandemic: In Brief*. https://fas.org/sgp/crs/misc/R46554.pdf

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

Frank, R. (2020). *Coronavirus: Wealthy New Yorkers flee Manhattan for suburbs and beyond*. CNBC. https://www.cnbc.com/2020/04/30/wealthy-new-yorkers-flee-manhattan-for-suburbs-and-beyond.html

Gao, S., Rao, J., Kang, Y., Liang, Y., Kruse, J., Dopfer, D., Sethi, A. K., Mandujano Reyes, J. F., Yandell, B. S., & Patz, J. A. (2020). 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*(9), e2020485.

Gogoi, P. (2020). *Stuck-At-Home Moms: The Pandemic's Devastating Toll On Women*. NPR. https://www.npr.org/2020/10/28/928253674/stuck-at-home-moms-the-pandemics-devastating-toll-on-women

Golgher, A. B., & Voss, P. R. (2016). How to Interpret the Coefficients of Spatial Models: Spillovers, Direct and Indirect Effects. *Spatial Demography*, *4*(3), 175–205.

Goodale-Sussen, G., & Kishore, A. (2020). *Why and when to use SafeGraph data in your analysis*. ArcGIS Blog. https://www.esri.com/arcgis-blog/products/bus-analyst/data-management/why-and-when-to-use-safegraph-data-in-your-analysis/

Gordon, A. L. (2020). *They're the Last Rich People Left on the Upper East Side*. Bloomberg. https://www.bloomberg.com/news/articles/2020-03-27/they-re-the-last-rich-people-left-on-the-upper-east-side

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

HHS Press Office. (2020). *Secretary Azar Declares Public Health Emergency for United States for 2019 Novel Coronavirus*. United States Health and Human Services. https://www.hhs.gov/about/news/2020/01/31/secretary-azar-declares-public-health-emergency-us-2019-novel-coronavirus.html

Holtz, D., Zhao, M., Benzell, S. G., Cao, C. Y., Rahimian, M. A., Yang, J., Allen, J., Collis, A., Moehring, A., Sowrirajan, T., Ghosh, D., Zhang, Y., Dhillon, P. S., Nicolaides, C., Eckles, D., & Aral, S. (2020). Interdependence and the cost of uncoordinated responses to COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(33), 19837–19843.

Institute of Medicine (US) Committee on the Consequences of Uninsurance. (2001). Who Goes Without Health Insurance? Who Is Most Likely to Be Uninsured? In *Coverage Matters: Insurance and Health Care*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK223657/

Janssen, A., & Shapiro, M. H. (2020). Does Precise Case Information Limit Precautionary Behavior? Evidence from COVID-19 in Singapore. *SSRN Electronic Journal*, 1–25.

Kishore, N., Kiang, M. V., Engø-Monsen, K., Vembar, N., Schroeder, A., Balsari, S., & Buckee, C. O. (2020). Measuring mobility to monitor travel and physical distancing interventions: a common framework for mobile phone data analysis. *The Lancet Digital Health*, *2*(11), e622–e628.

Kraemer, M. U. G., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., du Plessis, L., Faria, N. R., Li, R., Hanage, W. P., Brownstein, J. S., Layan, M., Vespignani, A., Tian, H., Dye, C., Pybus, O. G., & Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497.

Lamb, M. R., Kandula, S., & Shaman, J. (2021). Differential COVID-19 case positivity in New York City neighborhoods: Socioeconomic factors and mobility. *Influenza and Other Respiratory Viruses*, *15*(2), 209–217.

LeSage, J., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman & Hall/CRC.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., … Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*, *382*(13), 1199–1207.

Mangrum, D., & Niekamp, P. (2020). College Student Contribution to Local COVID-19 Spread: Evidence from University Spring Break Timing. *SSRN Electronic Journal*, 1–38.

Medina, J., & Solymosi, R. (2019). *Spatial Regression Models*. Crime Mapping in R.

https://maczokni.github.io/crimemapping_textbook_bookdown/spatial-regression-models.html

Mueller, A. L., McNamara, M. S., & Sinclair, D. A. (2020). Why does COVID-19 disproportionately affect older people? *Aging*, *12*(10), 9959–9981.

New York Standards Work Group. (n.d.). *Datum and Coordinate System Standard*. Gis.Ny.Gov. Retrieved April 14, 2021, from https://gis.ny.gov/coordinationprogram/workgroups/wg_1/related/standards/datum.htm

New York State Department of Labor. (n.d.). *Minimum Wage Lookup*. Retrieved April 15, 2021, from https://webapps.labor.ny.gov/dolweb/minimum-wage-lookup/

Nischan, U. (2020). *Uneven Impact: What Job Loss During COVID-19 Means for New Yorkers Now and into the Future*. https://www1.nyc.gov/assets/dca/downloads/pdf/partners/Uneven_Impact.pdf

NYC Department of Education. (n.d.). *DOE Data at a Glance*. Retrieved March 20, 2021, from https://www.schools.nyc.gov/about-us/reports/doe-data-at-a-glance

Pace, R. K., & LeSage, J. P. (2008). A spatial Hausman test. *Economics Letters*, *101*(3), 282–284.

Penney, V. (2021). *How Coronavirus Has Changed New York City Transit, in One Chart*. The New York Times.

Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S., & Colizza, V. (2020). 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*(12), e638–e649.

Qin, A., & Hernández, J. C. (2020). *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

SafeGraph. (n.d.-a). *SafeGraph COVID-19 Data Consortium*. Retrieved March 21, 2021, from https://www.safegraph.com/covid-19-data-consortium

SafeGraph. (n.d.-b). *Social Distancing Metrics*. SafeGraph Documentation; SafeGraph. Retrieved March 21, 2021, from https://docs.safegraph.com/docs/social-distancing-metrics

Santesmasses, D., Castro, J. P., Zenin, A. A., Shindyapina, A. V., Gerashchenko, M. V., Zhang, B., Kerepesi, C., Yim, S. H., Fedichev, P. O., & Gladyshev, V. N. (2020). COVID-19 is an emergent disease of aging. *Aging Cell*, *19*(10).

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

She, J., Jiang, J., Ye, L., Hu, L., Bai, C., & Song, Y. (2020). 2019 novel coronavirus of pneumonia in Wuhan, China: emerging attack and management strategies. *Clinical and Translational Medicine*, *9*(19), 1–7.

Smirnov, O. A., & Anselin, L. E. (2009). An parallel method of computing the Log-Jacobian of the variable transformation for models with spatial interaction on a lattice. *Computational Statistics & Data Analysis*, *53*(8), 2980–2988.

Smith, C. M., Le Comber, S. C., Fry, H., Bull, M., Leach, S., & Hayward, A. C. (2015). Spatial methods for infectious disease outbreak investigations: systematic literature review. *Eurosurveillance*, *20*(39), 1–21.

Stringer, S. M. (2020). *New York City's Frontline Workers*. https://comptroller.nyc.gov/wp-content/uploads/documents/Frontline_Workers_032020.pdf

Taylor, D. B. (2021, March 17). *A Timeline of the Coronavirus Pandemic*. The New York Times. https://www.nytimes.com/article/coronavirus-timeline.html

Thompson, C. N., Baumgartner, J., Pichardo, C., Toro, B., Li, L., Arciuolo, R., Chan, P. Y., Chen, J., Culp, G., Davidson, A., Devinney, K., Dorsinville, A., Eddy, M., English, M., Fireteanu, A. M., Graf, L., Geevarughese, A., Greene, S. K., Guerra, K., … Fine, A. (2020). COVID-19 Outbreak — New York City, February 29 – June 1, 2020. *Morbidity and Mortality Weekly Report*, *69*(46), 1725–1729.

Tully, T., & Stowe, S. (2020). *The Wealthy Flee Coronavirus. Vacation Towns Respond: Stay Away*. The New York Times. https://www.nytimes.com/2020/03/25/nyregion/coronavirus-leaving-nyc-vacation-homes.html

U.S. Census Bureau. (n.d.). *United States Census Bureau Geography Program Glossary*. United States Census Bureau. Retrieved December 14, 2020, from https://www.census.gov/programs-surveys/geography/about/glossary.html#par_textimage_4

UCLA Institute for Digital Research & Education Statistical Consulting. (n.d.). *How Can I Calculate Moran's I in R?* Retrieved March 17, 2021, from https://stats.idre.ucla.edu/r/faq/how-can-i-calculate-morans-i-in-r/

United States Census Bureau. (2017). *American Community Survey Information Guide*.

https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf

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

Weill, J. A., Stigler, M., Deschenes, O., & Springborn, M. R. (2020). 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*(33), 19658–19660.

WHO Director-General. (2020). *WHO Director-General's statement on IHR Emergency Committee on Novel Coronavirus (2019-nCoV)*. https://www.who.int/director-general/speeches/detail/who-director-general-s-statement-on-ihr-emergency-committee-on-novel-coronavirus-(2019-ncov)

Wilder, B., Charpignon, M., Killian, J. A., Ou, H. C., Mate, A., Jabbari, S., Perrault, A., Desai, A. N., Tambe, M., & Majumder, M. S. (2020). Modeling between-population variation in COVID-19 dynamics in Hubei, Lombardy, and New York City. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(41), 25904–25910.

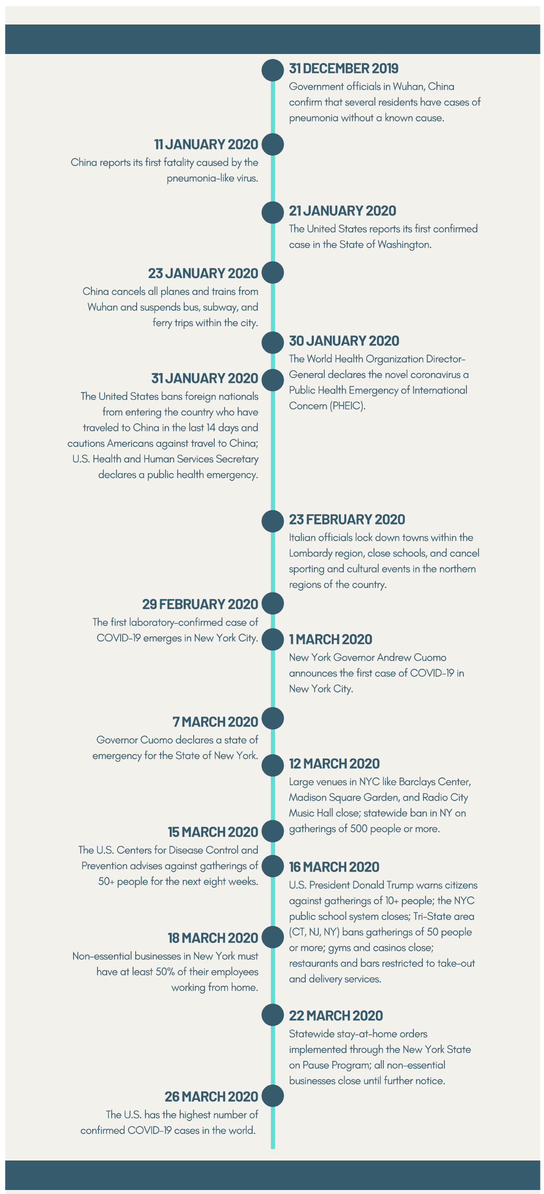
Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., Curtis, H. J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H. I., MacKenna, B., Tomlinson, L., Douglas, I. J., Rentsch, C. T., Mathur, R., Wong, A. Y. S., Grieve, R., … Goldacre, B. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature*, *584*(7821), 430–436.

APPENDIX A: GITHUB REPOSITORY

A repository with all aspects of the data collection, cleaning, and analysis processes exists at https://github.com/emilyemchen/honours-research. The repository's `README.md` file provides a broad overview of the file structure and contents.

APPENDIX B: LITERATURE REVIEW-RELATED FIGURES

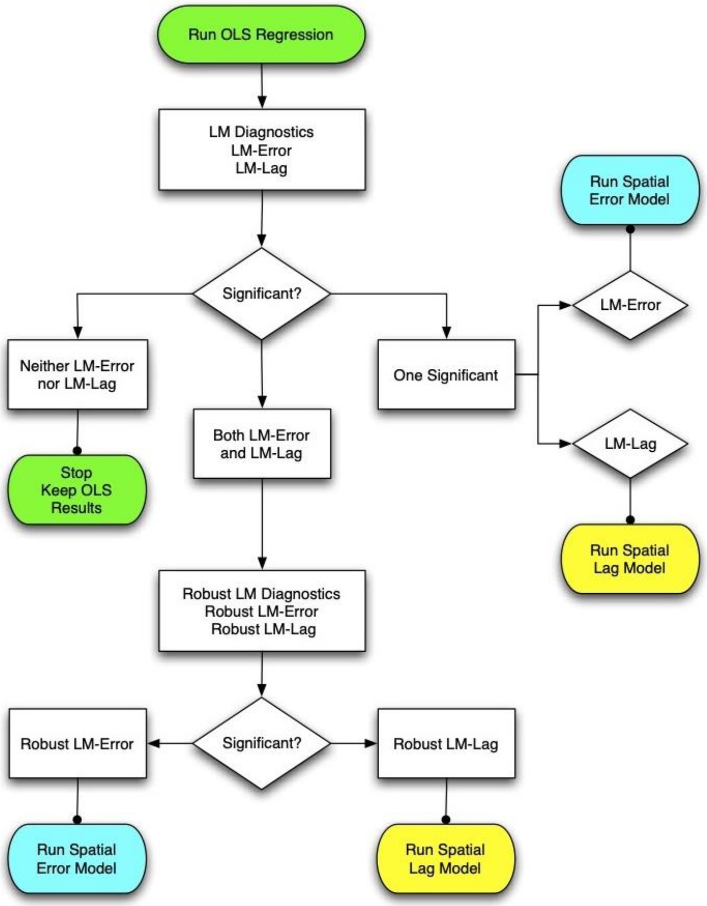
Appendix B.1: Timeline of the COVID-19 pandemic from December 2019 to March 2020



Source: Information retrieved from Taylor, 2021 and Qin & Hernández, 2020

APPENDIX C: METHODOLOGY-RELATED FIGURES AND TABLES

Appendix C.1: Spatial Regression Decision Process Flowchart



Source: From "Exploring Spatial Data with GeoDa™: A Workbook" by L. Anselin, 2005, *Center for Spatially Integrated Social Science*, p. 199.

Appendix C.2: Mathematical Equations for Regression Models

Model Name	Equation	Variable Definitions
Ordinary Least Squares Model	$y = X\beta + \varepsilon$	X = independent/explanatory variables β = slope coefficient ε = random error term (residuals)
Spatial Durbin Model (Spatially Lagged X)	$y = X\beta + \varepsilon + \theta WX$	X = matrix of observations for explanatory variables β = parameter vector ε = vector of error terms

		θ = vector of response parameters WX = weights matrix W of exogenous spatial lags for explanatory variables X
Spatial Autoregressive Lagged Response Model (SAR)	$y = X\beta + \varepsilon + \rho W y$	X = matrix of observations for explanatory variables β = parameter vector ε = vector of error terms ρ = spatial autocorrelation parameter Wy = spatially lagged dependent variable y for matrix W
Spatial Error Model (SEM)	$y = X\beta + \varepsilon + \lambda W u$	X = matrix of observations for explanatory variables β = parameter vector ε = vector of spatially autocorrelated error terms λ = autoregressive coefficient (indicates existence of stochastic shock to neighbors) W = spatial weights matrix u = vector of independent identically distributed (i.i.d.) errors

APPENDIX D: RESULTS-RELATED FIGURES AND TABLES

Appendix D.1: Results of Ordinary Least Squares Regression with Median Distance Traveled from Home as the Dependent Variable

```
Coefficients:
                          Estimate  Std. Error t value Pr(>|t|)
(Intercept)             1.33044625  0.19252860   6.910 5.36e-12 ***
age                     0.02131064  0.00443277   4.808 1.57e-06 ***
race                   -0.00028701  0.00007339  -3.911 9.30e-05 ***
transport               0.00085691  0.00025090   3.415 0.000642 ***
female_workers         -0.00234845  0.00033416  -7.028 2.34e-12 ***
housing_occupancy_rent  0.00323193  0.00142172   2.273 0.023048 *
min_wage               -0.00077529  0.00030858  -2.512 0.012016 *
children                0.00118689  0.00024657   4.814 1.52e-06 ***
education               0.00155179  0.00030695   5.056 4.43e-07 ***
health_insurance        0.00028636  0.00029970   0.955 0.339367
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.263 on 5727 degrees of freedom
  (1126 observations deleted due to missingness)
Multiple R-squared:  0.02913,   Adjusted R-squared:  0.0276
F-statistic: 19.09 on 9 and 5727 DF,  p-value: < 2.2e-16
```

Appendix D.2: Results of Global Moran I test for Regression Residuals with Median Distance Traveled from Home as the Dependent Variable

Moran I Statistic Standard Deviate	p-value	Observed Moran I	Expectation	Variance
8.8898	< 2.2e-16	0.0661	-0.0005	0.0001

Appendix D.3: Results of Lagrange Multiplier Diagnostic Tests for Spatial Dependence with Median Distance Traveled from Home as the Dependent Variable

Test	Test Statistic	p-value	df
LMerr	77.65	< 2.2e-16	1
LMlag	106.30	< 2.2e-16	1
RLMerr	75.01	< 2.2e-16	1
RLMlag	103.66	< 2.2e-16	1
SARMA	181.31	< 2.2e-16	2

Appendix D.4: Results of Spatial Durbin Model (Spatially Lagged X Model) with Median Distance Traveled from Home as the Dependent Variable

```
Coefficients:
                            Estimate  Std. Error t value     Pr(>|t|)
(Intercept)               0.17204039  0.38363225   0.448     0.653845
age                       0.02087746  0.00458903   4.549 0.0000054905 ***
race                      0.00003684  0.00008827   0.417     0.676419
transport                 0.00077025  0.00027603   2.790     0.005280 **
female_workers           -0.00195471  0.00035060  -5.575 0.0000000258 ***
housing_occupancy_rent    0.00282303  0.00142641   1.979     0.047851 *
min_wage                 -0.00045253  0.00033014  -1.371     0.170520
children                  0.00077537  0.00025079   3.092     0.002000 **
education                 0.00085205  0.00032980   2.583     0.009805 **
health_insurance         -0.00018900  0.00030869  -0.612     0.540383
lag.age                   0.03179430  0.00891418   3.567     0.000364 ***
lag.race                 -0.00058029  0.00013820  -4.199 0.0000272224 ***
lag.transport             0.00103696  0.00048760   2.127     0.033490 *
lag.female_workers       -0.00291046  0.00068541  -4.246 0.0000220801 ***
lag.housing_occupancy_rent  0.00751830  0.00315441   2.383     0.017185 *
lag.min_wage             -0.00139381  0.00060059  -2.321     0.020337 *
lag.children              0.00182390  0.00053098   3.435     0.000597 ***
lag.education             0.00157441  0.00060850   2.587     0.009695 **
lag.health_insurance      0.00105318  0.00060728   1.734     0.082927 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.243 on 5718 degrees of freedom
Multiple R-squared:  0.04784,   Adjusted R-squared:  0.04484
F-statistic: 15.96 on 18 and 5718 DF,  p-value: < 2.2e-16
```

Appendix D.5: Impact Measures for Spatially Lagged X Model with Median Distance Traveled from Home as the Dependent Variable

```
Impact measures (SLX, estimable, n-k):
                             Direct      Indirect          Total
age                     0.02087745518  0.0317942952  0.0526717504
race                    0.00003684238 -0.0005802885 -0.0005434462
transport               0.00077024770  0.0010369630  0.0018072107
female_workers         -0.00195471343 -0.0029104561 -0.0048651695
housing_occupancy_rent  0.00282302778  0.0075183018  0.0103413296
min_wage               -0.00045252591 -0.0013938079 -0.0018463338
children                0.00077537169  0.0018239040  0.0025992757
education               0.00085204580  0.0015744134  0.0024264592
health_insurance       -0.00018899954  0.0010531755  0.0008641760
========================================================
Standard errors:
                            Direct      Indirect         Total
age                    0.00458903186 0.0089141839 0.0089543066
race                   0.00008827144 0.0001381984 0.0001196394
transport              0.00027602679 0.0004875993 0.0004655346
female_workers         0.00035059686 0.0006854141 0.0006845234
housing_occupancy_rent 0.00142640994 0.0031544133 0.0032874082
min_wage               0.00033014045 0.0006005945 0.0005782219
children               0.00025079033 0.0005309843 0.0005431107
education              0.00032980408 0.0006084964 0.0005949798
health_insurance       0.00030868605 0.0006072788 0.0006114668
```

```
Z-values:
                           Direct  Indirect     Total
age                     4.5494248  3.566708  5.882281
race                    0.4173759 -4.198953 -4.542367
transport               2.7904817  2.126670  3.882012
female_workers         -5.5753878 -4.246274 -7.107382
housing_occupancy_rent  1.9791139  2.383423  3.145739
min_wage               -1.3707073 -2.320714 -3.193123
children                3.0917129  3.434949  4.785904
education               2.5834908  2.587383  4.078221
health_insurance       -0.6122711  1.734254  1.413284

p-values:
                       Direct         Indirect    Total
age                    0.000005379276 0.00036149  4.0465e-09
race                   0.6764035      0.000026815 5.5626e-06
transport              0.0052630      0.03344748  0.0001036
female_workers         0.000000024698 0.000021736 1.1826e-12
housing_occupancy_rent 0.0478032      0.01715246  0.0016567
min_wage               0.1704662      0.02030230  0.0014074
children               0.0019901      0.00059267  1.7022e-06
education              0.0097806      0.00967080  4.5382e-05
health_insurance       0.5403584      0.08287308  0.1575723
```

Appendix D.6: Results of Spatial Autoregressive Lagged Response Model with Median Distance Traveled from Home as the Dependent Variable

```
Regions with no neighbours included:
 3374
Coefficients: (numerical Hessian approximate standard errors)
                        Estimate  Std. Error z value         Pr(>|z|)
(Intercept)           0.96509907  0.19418119  4.9701 0.00000066919941
age                   0.02001987  0.00438409  4.5665 0.00000495975950
race                 -0.00022088  0.00007288 -3.0307        0.0024399
transport             0.00082252  0.00024806  3.3159        0.0009136
female_workers       -0.00215358  0.00033096 -6.5070 0.00000000007668
housing_occupancy_rent  0.00287263  0.00140594  2.0432        0.0410318
min_wage             -0.00065053  0.00030532 -2.1306        0.0331205
children              0.00105116  0.00024416  4.3052 0.00001668418093
education             0.00135396  0.00030415  4.4516 0.00000852171454
health_insurance      0.00013699  0.00029669  0.4617        0.6442790

Rho: 0.19253, LR test value: 88.022, p-value: < 2.22e-16
Approximate (numerical Hessian) standard error: 0.020298
    z-value: 9.4848, p-value: < 2.22e-16
Wald statistic: 89.961, p-value: < 2.22e-16

Log likelihood: -12777.99 for lag model
ML residual variance (sigma squared): 5.0066, (sigma: 2.2375)
Number of observations: 5737
Number of parameters estimated: 12
AIC: 25580, (AIC for lm: 25666)
```

Appendix D.7: Impact Measures for Spatial Autoregressive Lagged Response Model with Median Distance Traveled from Home as the Dependent Variable

```
Impact measures (lag, trace):
                              Direct        Indirect          Total
age                     0.0201778511   0.00461536543   0.0247932165
race                   -0.0002226193  -0.00005092066  -0.0002735400
transport               0.0008290078   0.00018962247   0.0010186303
female_workers         -0.0021705691  -0.00049648346  -0.0026670525
housing_occupancy_rent  0.0028952944   0.00066225296   0.0035575474
min_wage               -0.0006556616  -0.00014997227  -0.0008056339
children                0.0010594541   0.00024233342   0.0013017875
education               0.0013646388   0.00031213963   0.0016767785
health_insurance        0.0001380683   0.00003158094   0.0001696492
```

Simulation results (mixed Hessian approximation variance matrix):
Direct:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
age	0.0202375	0.00442075	0.000062519	0.000064558
race	-0.0002251	0.00007404	0.000001047	0.000001047
transport	0.0008269	0.00025374	0.000003588	0.000003588
female_workers	-0.0021701	0.00033646	0.000004758	0.000004758
housing_occupancy_rent	0.0028966	0.00143511	0.000020295	0.000020295
min_wage	-0.0006537	0.00030740	0.000004347	0.000004166
children	0.0010674	0.00024621	0.000003482	0.000003482
education	0.0013593	0.00030486	0.000004311	0.000004311
health_insurance	0.0001348	0.00029970	0.000004238	0.000004238

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
age	0.01150871	0.0173190	0.0201998	0.0231434	0.02913623
race	-0.00036897	-0.0002761	-0.0002247	-0.0001764	-0.00007681
transport	0.00034724	0.0006571	0.0008209	0.0009961	0.00133408
female_workers	-0.00281885	-0.0024024	-0.0021723	-0.0019393	-0.00151349
housing_occupancy_rent	0.00005612	0.0019358	0.0028884	0.0038664	0.00567672
min_wage	-0.00125525	-0.0008574	-0.0006508	-0.0004449	-0.00005029
children	0.00058616	0.0009019	0.0010612	0.0012343	0.00155074
education	0.00077646	0.0011508	0.0013560	0.0015664	0.00196149
health_insurance	-0.00044218	-0.0000646	0.0001361	0.0003420	0.00072345

Indirect:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
age	0.00462336	0.00117186	0.0000165726	0.0000162261
race	-0.00005128	0.00001783	0.0000002521	0.0000002521
transport	0.00018900	0.00006332	0.0000008956	0.0000008956
female_workers	-0.00049526	0.00009649	0.0000013646	0.0000013646
housing_occupancy_rent	0.00066166	0.00034083	0.0000048201	0.0000048201
min_wage	-0.00014907	0.00007267	0.0000010277	0.0000009727
children	0.00024350	0.00006306	0.0000008917	0.0000008637
education	0.00031009	0.00007857	0.0000011111	0.0000011111
health_insurance	0.00003020	0.00006888	0.0000009741	0.0000009741

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
age	0.00248037	0.00383170	0.00456941	0.00535254	0.00709867
race	-0.00008785	-0.00006251	-0.00005093	-0.00003937	-0.00001727
transport	0.00007612	0.00014528	0.00018588	0.00022889	0.00032306
female_workers	-0.00069505	-0.00055701	-0.00049138	-0.00042766	-0.00031992
housing_occupancy_rent	0.00001173	0.00043543	0.00065114	0.00087810	0.00136187
min_wage	-0.00030137	-0.00019488	-0.00014708	-0.00009980	-0.00001171
children	0.00013072	0.00019938	0.00024025	0.00028396	0.00037422
education	0.00017065	0.00025608	0.00030515	0.00035905	0.00047877
health_insurance	-0.00010236	-0.00001449	0.00003014	0.00007646	0.00016532

```
Total:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                            Mean         SD    Naive SE Time-series SE
age                    0.0248609 0.00546474 0.000077283    0.000079272
race                  -0.0002764 0.00009078 0.000001284    0.000001284
transport              0.0010159 0.00031302 0.000004427    0.000004427
female_workers        -0.0026654 0.00041477 0.000005866    0.000005866
housing_occupancy_rent  0.0035582 0.00176553 0.000024968    0.000024968
min_wage              -0.0008028 0.00037766 0.000005341    0.000005105
children               0.0013109 0.00030270 0.000004281    0.000004190
education              0.0016694 0.00037479 0.000005300    0.000005300
health_insurance       0.0001650 0.00036806 0.000005205    0.000005205

2. Quantiles for each variable:

                            2.5%         25%        50%        75%       97.5%
age                    0.0141515  0.02120145  0.0248414  0.0284351  0.03582909
race                  -0.0004549 -0.00033876 -0.0002762 -0.0002161 -0.00009556
transport              0.0004234  0.00080802  0.0010082  0.0012254  0.00164318
female_workers        -0.0034592 -0.00294693 -0.0026670 -0.0023845 -0.00185147
housing_occupancy_rent  0.0000674  0.00238415  0.0035493  0.0047511  0.00701348
min_wage              -0.0015500 -0.00105197 -0.0008000 -0.0005467 -0.00006124
children               0.0007251  0.00110572  0.0013079  0.0015160  0.00191298
education              0.0009526  0.00141548  0.0016637  0.0019242  0.00241930
health_insurance      -0.0005419 -0.00007797  0.0001663  0.0004195  0.00088479

Simulated standard errors
                            Direct       Indirect        Total
age                    0.00442075402 0.00117185743 0.0054647412
race                   0.00007404103 0.00001782574 0.0000907829
transport              0.00025374311 0.00006332499 0.0003130184
female_workers         0.00033646493 0.00009649420 0.0004147702
housing_occupancy_rent 0.00143510568 0.00034083195 0.0017655307
min_wage               0.00030740018 0.00007266801 0.0003776633
children               0.00024620824 0.00006305507 0.0003026992
education              0.00030485656 0.00007856640 0.0003747894
health_insurance       0.00029969671 0.00006888275 0.0003680572

Simulated z-values:
                           Direct   Indirect       Total
age                     4.5778388  3.9453288  4.5493211
race                   -3.0400723 -2.8764728 -3.0442441
transport               3.2589961  2.9845863  3.2456458
female_workers         -6.4497730 -5.1325275 -6.4261639
housing_occupancy_rent  2.0183766  1.9413092  2.0153962
min_wage               -2.1266566 -2.0513288 -2.1257048
children                4.3352380  3.8617037  4.3306069
education               4.4588459  3.9468505  4.4542297
health_insurance        0.4497284  0.4384403  0.4482539

Simulated p-values:
                       Direct           Indirect      Total
age                    0.00000469804400 0.00007969054 0.00000538192624
race                   0.0023652        0.0040215     0.0023327
transport              0.0011181        0.0028396     0.0011718
female_workers         0.00000000011202 0.00000028588 0.00000000013086
housing_occupancy_rent 0.0435521        0.0522208     0.0438632
min_wage               0.0334486        0.0402349     0.0335278
children               0.00001456024705 0.0001126     0.00001486989518
education              0.00000824021360 0.00007918597 0.00000841949436
health_insurance       0.6529063        0.6610671     0.6539700
```

Appendix D.8: Results of Spatial Hausman Test with Median Distance Traveled from Home as the Dependent Variable

Hausman Test Statistic	p-value	df
86.183	3.053e-14	10

Appendix D.9: Results of Ordinary Least Squares Regression with Home Dwell Time as the Dependent Variable

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            440.29069   68.38638   6.438 1.31e-10 ***
age                     -6.29183    1.57389  -3.998 6.48e-05 ***
race                     0.28752    0.02607  11.027  < 2e-16 ***
transport               -0.30830    0.08914  -3.459 0.000547 ***
female_workers           0.45031    0.11873   3.793 0.000151 ***
housing_occupancy_rent   0.43147    0.50473   0.855 0.392664
min_wage                 0.13975    0.10963   1.275 0.202457
children                -0.60517    0.08759  -6.909 5.39e-12 ***
education               -0.86088    0.10907  -7.893 3.52e-15 ***
health_insurance         0.02648    0.10650   0.249 0.803614
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 804.3 on 5729 degrees of freedom
  (1124 observations deleted due to missingness)
Multiple R-squared:  0.05419,   Adjusted R-squared:  0.0527
F-statistic: 36.47 on 9 and 5729 DF,  p-value: < 2.2e-16
```

Appendix D.10: Results of Moran's I Calculations with Home Dwell Time as the Dependent Variable

Moran I Statistic Standard Deviate	p-value	Observed Moran I	Expectation	Variance
16.125	< 2.2e-16	0.1201	-0.0005	0.0001

Appendix D.11: Results of Lagrange Multiplier Diagnostic Tests for Spatial Dependence with Home Dwell Time as the Dependent Variable

Test	Test Statistic	p-value	df
LMerr	257.14	< 2.2e-16	1
LMlag	312.46	< 2.2e-16	1
RLMerr	29.56	5.4e-08	1
RLMlag	84.88	< 2.2e-16	1
SARMA	342.02	< 2.2e-16	2

Appendix D.12: Results of Spatial Durbin Model (Spatially Lagged X Model) with Home Dwell Time as the Dependent Variable

```
Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)               645.97058  135.88816   4.754 2.05e-06 ***
age                        -6.64974    1.62531  -4.091 4.35e-05 ***
race                        0.23594    0.03128   7.544 5.29e-14 ***
transport                  -0.44832    0.09779  -4.585 4.64e-06 ***
female_workers              0.34667    0.12424   2.790 0.005285 ** 
housing_occupancy_rent      0.18863    0.50504   0.374 0.708786    
min_wage                   -0.06225    0.11699  -0.532 0.594663    
children                   -0.44636    0.08884  -5.024 5.20e-07 ***
education                  -0.39129    0.11689  -3.347 0.000821 ***
health_insurance            0.16210    0.10939   1.482 0.138445    
lag.age                    -3.06295    3.15448  -0.971 0.331597    
lag.race                    0.03654    0.04898   0.746 0.455649    
lag.transport              -0.05593    0.17279  -0.324 0.746177    
lag.female_workers          0.95963    0.24286   3.951 7.86e-05 ***
lag.housing_occupancy_rent  1.21562    1.11707   1.088 0.276546    
lag.min_wage                0.47670    0.21292   2.239 0.025200 *  
lag.children               -0.66508    0.18814  -3.535 0.000411 ***
lag.education              -1.41516    0.21568  -6.561 5.80e-11 ***
lag.health_insurance       -0.18472    0.21525  -0.858 0.390826    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 795 on 5720 degrees of freedom
Multiple R-squared:  0.07735,	Adjusted R-squared:  0.07445 
F-statistic: 26.64 on 18 and 5720 DF,  p-value: < 2.2e-16
```

Appendix D.13: Impact Measures for Spatially Lagged X Model with Home Dwell Time as the Dependent Variable

```
Impact measures (SLX, estimable, n-k):
                            Direct    Indirect       Total
age                    -6.64973636 -3.06294771 -9.71268406
race                    0.23594485  0.03653967  0.27248452
transport              -0.44831626 -0.05593345 -0.50424971
female_workers          0.34666716  0.95962680  1.30629395
housing_occupancy_rent  0.18863441  1.21561912  1.40425353
min_wage               -0.06225124  0.47670369  0.41445245
children               -0.44636247 -0.66508024 -1.11144271
education              -0.39129081 -1.41516457 -1.80645537
health_insurance        0.16209825 -0.18472349 -0.02262524
========================================================
Standard errors:
                           Direct   Indirect      Total
age                    1.62530564 3.15447647 3.16785700
race                   0.03127788 0.04897548 0.04240065
transport              0.09778594 0.17279420 0.16499061
female_workers         0.12424413 0.24285860 0.24250493
housing_occupancy_rent 0.50503829 1.11707498 1.16423930
min_wage               0.11698718 0.21291739 0.20495483
children               0.08883824 0.18813884 0.19246757
education              0.11689257 0.21568373 0.21089585
health_insurance       0.10939175 0.21524881 0.21673044
```

```
Z-values:
                           Direct   Indirect      Total
age                    -4.0913759 -0.9709845 -3.0660109
race                    7.5435057  0.7460810  6.4264235
transport              -4.5846699 -0.3236998 -3.0562329
female_workers          2.7902094  3.9513808  5.3866697
housing_occupancy_rent  0.3735052  1.0882162  1.2061554
min_wage               -0.5321202  2.2389138  2.0221648
children               -5.0244407 -3.5350502 -5.7747012
education              -3.3474395 -6.5612951 -8.5656278
health_insurance        1.4818143 -0.8581859 -0.1043935

p-values:
                       Direct     Indirect          Total            
age                    4.2882e-05 0.3315560         0.0021694        
race                   4.5741e-14 0.4556185         0.00000000013064 
transport              4.5470e-06 0.7461653         0.0022414        
female_workers         0.00526740 0.000077701573975 0.00000007177517 
housing_occupancy_rent 0.70877250 0.2764997         0.2277576        
min_wage               0.59464277 0.0251615         0.0431593        
children               5.0490e-07 0.0004077         0.00000000770897 
education              0.00081562 0.000000000053342 < 2.22e-16       
health_insurance       0.13838972 0.3907898         0.9168571        
```

Appendix D.14: Results of Spatial Autoregressive Lagged Response Model with Home Dwell Time as the Dependent Variable

```
Regions with no neighbours included:
 3374
Coefficients: (numerical Hessian approximate standard errors)
                        Estimate Std. Error z value       Pr(>|z|)
(Intercept)           394.044230  66.270576  5.9460 0.000000002748
age                    -6.242689   1.523866 -4.0966 0.000041923793
race                    0.242147   0.025391  9.5366      < 2.2e-16
transport              -0.316812   0.086303 -3.6709      0.0002417
female_workers          0.357819   0.115093  3.1089      0.0018776
housing_occupancy_rent  0.321551   0.488728  0.6579      0.5105794
min_wage                0.071857   0.106225  0.6765      0.4987491
children               -0.509782   0.084994 -5.9979 0.000000001999
education              -0.629270   0.106515 -5.9078 0.000000003467
health_insurance        0.083436   0.103167  0.8088      0.4186577

Rho: 0.32132, LR test value: 261.17, p-value: < 2.22e-16
Approximate (numerical Hessian) standard error: 0.019259
    z-value: 16.684, p-value: < 2.22e-16
Wald statistic: 278.36, p-value: < 2.22e-16

Log likelihood: -46401.71 for lag model
ML residual variance (sigma squared): 606480, (sigma: 778.77)
Number of observations: 5739
Number of parameters estimated: 12
AIC: 92827, (AIC for lm: 93087)
```

Appendix D.15: Impact Measures for Spatial Autoregressive Lagged Response Model with Home Dwell Time as the Dependent Variable

```
Impact measures (lag, trace):
                            Direct    Indirect      Total
age                    -6.39013419 -2.80816927 -9.1983035
race                    0.24786635  0.10892583  0.3567922
transport              -0.32429524 -0.14251280 -0.4668080
female_workers          0.36627014  0.16095884  0.5272290
housing_occupancy_rent  0.32914608  0.14464452  0.4737906
min_wage                0.07355408  0.03232363  0.1058777
children               -0.52182282 -0.22931706 -0.7511399
education              -0.64413277 -0.28306664 -0.9271994
health_insurance        0.08540715  0.03753251  0.1229397
```

```
Simulation results (mixed Hessian approximation variance matrix):
Direct:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                           Mean      SD  Naive SE Time-series SE
age                    -6.39025 1.57071 0.0222132      0.0222132
race                    0.24766 0.02618 0.0003703      0.0003595
transport              -0.32531 0.08974 0.0012692      0.0012692
female_workers          0.36944 0.12005 0.0016977      0.0016483
housing_occupancy_rent  0.32346 0.50162 0.0070939      0.0070939
min_wage                0.07267 0.10842 0.0015333      0.0015333
children               -0.52065 0.08629 0.0012204      0.0012204
education              -0.64384 0.10710 0.0015146      0.0014441
health_insurance        0.08377 0.10657 0.0015071      0.0015071

2. Quantiles for each variable:

                          2.5%        25%      50%     75%   97.5%
age                    -9.4432 -7.4551386 -6.38983 -5.3379 -3.2815
race                    0.1950  0.2301554  0.24805  0.2654  0.2984
transport              -0.5014 -0.3843406 -0.32486 -0.2641 -0.1511
female_workers          0.1360  0.2895769  0.36916  0.4500  0.6100
housing_occupancy_rent -0.6585 -0.0200138  0.32662  0.6710  1.2821
min_wage               -0.1324 -0.0005433  0.07138  0.1465  0.2891
children               -0.6892 -0.5789422 -0.51982 -0.4613 -0.3524
education              -0.8538 -0.7144464 -0.64472 -0.5722 -0.4355
health_insurance       -0.1268  0.0126351  0.08419  0.1557  0.2894

Indirect:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                           Mean      SD  Naive SE Time-series SE
age                    -2.81271 0.73259 0.0103605      0.0103605
race                    0.10892 0.01400 0.0001980      0.0001980
transport              -0.14328 0.04165 0.0005890      0.0005890
female_workers          0.16244 0.05421 0.0007667      0.0007457
housing_occupancy_rent  0.14243 0.22161 0.0031340      0.0029710
min_wage                0.03181 0.04790 0.0006774      0.0006774
children               -0.22901 0.04164 0.0005889      0.0005889
education              -0.28301 0.05064 0.0007161      0.0006670
health_insurance        0.03694 0.04724 0.0006681      0.0006774

2. Quantiles for each variable:

                           2.5%        25%      50%      75%    97.5%
age                    -4.32451 -3.2989002 -2.78641 -2.31253 -1.40549
race                    0.08284  0.0991294  0.10844  0.11822  0.13768
transport              -0.22939 -0.1699632 -0.14227 -0.11462 -0.06564
female_workers          0.05928  0.1255713  0.16135  0.19688  0.27419
housing_occupancy_rent -0.28487 -0.0086362  0.14148  0.29101  0.56731
min_wage               -0.06000 -0.0002358  0.03176  0.06407  0.12955
children               -0.31357 -0.2561181 -0.22684 -0.20052 -0.15181
education              -0.38750 -0.3153295 -0.28198 -0.24770 -0.18759
health_insurance       -0.05746  0.0054847  0.03656  0.06830  0.12918
```

```
Total:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                         Mean      SD  Naive SE Time-series SE
age                   -9.2030 2.27541 0.0321791      0.0321791
race                   0.3566 0.03795 0.0005366      0.0005366
transport             -0.4686 0.13012 0.0018402      0.0018402
female_workers         0.5319 0.17298 0.0024463      0.0023748
housing_occupancy_rent  0.4659 0.72244 0.0102169      0.0102169
min_wage               0.1045 0.15616 0.0022085      0.0022085
children              -0.7497 0.12471 0.0017636      0.0017636
education             -0.9268 0.15373 0.0021741      0.0020271
health_insurance       0.1207 0.15363 0.0021727      0.0022030

2. Quantiles for each variable:

                          2.5%         25%     50%     75%   97.5%
age                   -13.6242 -10.7586419 -9.2091 -7.6724 -4.7159
race                    0.2821   0.3309376  0.3572  0.3822  0.4310
transport              -0.7274  -0.5557953 -0.4689 -0.3802 -0.2178
female_workers          0.1960   0.4163135  0.5323  0.6467  0.8831
housing_occupancy_rent  -0.9386  -0.0285074  0.4700  0.9643  1.8618
min_wage               -0.1918  -0.0007718  0.1034  0.2114  0.4186
children               -0.9936  -0.8340338 -0.7480 -0.6652 -0.5066
education              -1.2265  -1.0262486 -0.9262 -0.8233 -0.6311
health_insurance       -0.1832   0.0180230  0.1211  0.2245  0.4166
 Simulated standard errors
                           Direct   Indirect      Total
 age                    1.57070747 0.73259494 2.27540684
 race                   0.02618218 0.01399993 0.03794623
 transport              0.08974471 0.04165084 0.13012041
 female_workers         0.12004752 0.05421264 0.17298089
 housing_occupancy_rent 0.50161652 0.22160901 0.72244268
 min_wage               0.10842266 0.04790204 0.15616435
 children               0.08629320 0.04164341 0.12470651
 education              0.10710118 0.05063561 0.15373267
 health_insurance       0.10657082 0.04724146 0.15363284

 Simulated z-values:
                           Direct   Indirect      Total
 age                    -4.0683925 -3.8393861 -4.0445379
 race                    9.4590463  7.7800674  9.3969515
 transport              -3.6248305 -3.4399280 -3.6011667
 female_workers          3.0774556  2.9963654  3.0748008
 housing_occupancy_rent  0.6448359  0.6427151  0.6448841
 min_wage                0.6702043  0.6641542  0.6690367
 children               -6.0334937 -5.4992168 -6.0113588
 education              -6.0115277 -5.5890750 -6.0289589
 health_insurance        0.7860277  0.7819549  0.7856934

 Simulated p-values:
                        Direct          Indirect   Total
 age                    0.0000473385981 0.00012334 0.0000524264189
 race                   < 2.22e-16      7.3275e-15 < 2.22e-16
 transport              0.00028915      0.00058187 0.00031679
 female_workers         0.00208776      0.00273219 0.00210643
 housing_occupancy_rent 0.51903355      0.52040899 0.51900230
 min_wage               0.50272754      0.50659164 0.50347209
 children               0.0000000016045 3.8148e-08 0.0000000018397
 education              0.0000000018378 2.2828e-08 0.0000000016502
 health_insurance       0.43185126      0.43424106 0.43204712
```

Appendix D.16: Results of Spatial Hausman Test with Home Dwell Time as the Dependent Variable

Hausman Test Statistic	p-value	df
97.136	< 2.2e-16	10

Appendix D.17: Results of Ordinary Least Squares Regression with Non-Home Dwell Time as the Dependent Variable

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            490.90319   69.12322   7.102 1.38e-12 ***
age                     17.13034    1.59085  10.768  < 2e-16 ***
race                     0.14984    0.02635   5.686 1.37e-08 ***
transport               -0.89200    0.09010  -9.901  < 2e-16 ***
female_workers           0.57243    0.12001   4.770 1.89e-06 ***
housing_occupancy_rent  -1.49497    0.51017  -2.930  0.00340 **
min_wage                -1.71649    0.11081 -15.490  < 2e-16 ***
children                 0.45702    0.08853   5.162 2.52e-07 ***
education                1.16949    0.11025  10.608  < 2e-16 ***
health_insurance         0.32927    0.10764   3.059  0.00223 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 813 on 5729 degrees of freedom
  (1124 observations deleted due to missingness)
Multiple R-squared:  0.1344,    Adjusted R-squared:  0.1331
F-statistic: 98.86 on 9 and 5729 DF,  p-value: < 2.2e-16
```

Appendix D.18: Results of Moran's I Calculations with Non-Home Dwell Time as the Dependent Variable

Moran I Statistic Standard Deviate	p-value	Observed Moran I	Expectation	Variance
15.705	< 2.2e-16	0.1170	-0.0005	0.0001

Appendix D.19: Results of Lagrange Multiplier Diagnostic Tests for Spatial Dependence with Non-Home Dwell Time as the Dependent Variable

Test	Test Statistic	p-value	df
LMerr	243.86	< 2.2e-16	1
LMlag	366.98	< 2.2e-16	1
RLMerr	38.95	4.3e-10	1
RLMlag	162.07	< 2.2e-16	1
SARMA	405.93	< 2.2e-16	2

Appendix D.20: Results of Spatial Durbin Model (Spatially Lagged X Model) with Non-Home Dwell Time as the Dependent Variable

```
Coefficients:
                          Estimate Std. Error t value       Pr(>|t|)    
(Intercept)              110.48119  136.86850   0.807       0.419581    
age                       14.51725    1.63703   8.868        < 2e-16 ***
race                       0.10556    0.03150   3.351       0.000811 ***
transport                 -0.52613    0.09849  -5.342 0.000000095564 ***
female_workers             0.39943    0.12514   3.192       0.001421 ** 
housing_occupancy_rent    -1.04930    0.50868  -2.063       0.039178 *  
min_wage                  -1.27086    0.11783 -10.785        < 2e-16 ***
children                   0.35261    0.08948   3.941 0.000082202827 ***
education                  0.72048    0.11774   6.119 0.000000001001 ***
health_insurance           0.18415    0.11018   1.671       0.094706 .  
lag.age                   11.11101    3.17723   3.497       0.000474 ***
lag.race                   0.10517    0.04933   2.132       0.033051 *  
lag.transport             -0.46289    0.17404  -2.660       0.007843 ** 
lag.female_workers        -0.12404    0.24461  -0.507       0.612123    
lag.housing_occupancy_rent -1.43531   1.12513  -1.276       0.202122    
lag.min_wage              -1.31363    0.21445  -6.125 0.000000000964 ***
lag.children               0.42129    0.18950   2.223       0.026241 *  
lag.education              1.40334    0.21724   6.460 0.000000000113 ***
lag.health_insurance       0.30175    0.21680   1.392       0.164030    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 800.8 on 5720 degrees of freedom
Multiple R-squared:  0.1616,	Adjusted R-squared:  0.1589 
F-statistic: 61.24 on 18 and 5720 DF,  p-value: < 2.2e-16
```

Appendix D.21: Impact Measures for Spatially Lagged X Model with Non-Home Dwell Time as the Dependent Variable

```
Impact measures (SLX, estimable, n-k):
                           Direct   Indirect      Total
age                    14.5172521 11.1110118 25.6282639
race                    0.1055618  0.1051680  0.2107299
transport              -0.5261253 -0.4628938 -0.9890191
female_workers          0.3994317 -0.1240355  0.2753962
housing_occupancy_rent -1.0492998 -1.4353074 -2.4846073
min_wage               -1.2708607 -1.3136325 -2.5844931
children                0.3526130  0.4212898  0.7739028
education               0.7204845  1.4033417  2.1238262
health_insurance        0.1841514  0.3017497  0.4859011
========================================================
Standard errors:
                           Direct  Indirect      Total
age                    1.63703101 3.1772337 3.19071075
race                   0.03150352 0.0493288 0.04270654
transport              0.09849139 0.1740408 0.16618089
female_workers         0.12514046 0.2446106 0.24425442
housing_occupancy_rent 0.50868177 1.1251339 1.17263842
min_wage               0.11783116 0.2144534 0.20643343
children               0.08947914 0.1894961 0.19385608
education              0.11773586 0.2172397 0.21241731
health_insurance       0.11018093 0.2168017 0.21829399
```

```
Z-values:
                           Direct   Indirect      Total
age                      8.868037  3.4970710   8.032149
race                     3.350794  2.1319801   4.934370
transport               -5.341841 -2.6596860  -5.951461
female_workers           3.191867 -0.5070734   1.127497
housing_occupancy_rent  -2.062782 -1.2756771  -2.118818
min_wage               -10.785438 -6.1254905 -12.519741
children                 3.940728  2.2232108   3.992151
education                6.119499  6.4598760   9.998367
health_insurance         1.671354  1.3918237   2.225902

p-values:
                       Direct          Indirect        Total     
age                    < 2.22e-16      0.0004704       8.8818e-16
race                   0.0008058       0.0330085       8.0410e-07
transport              0.0000000920076 0.0078214       2.6576e-09
female_workers         0.0014136       0.6121034       0.259532  
housing_occupancy_rent 0.0391333       0.2020697       0.034106  
min_wage               < 2.22e-16      0.00000000090405 < 2.22e-16
children               0.0000812346081 0.0262016       6.5477e-05
education              0.0000000009387 0.00000000010479 < 2.22e-16
health_insurance       0.0946517       0.1639758       0.026021  
```

Appendix D.22: Results of Spatial Autoregressive Lagged Response Model with Non-Home Dwell Time as the Dependent Variable

```
Regions with no neighbours included:
 3374
Coefficients: (numerical Hessian approximate standard errors)
                        Estimate Std. Error  z value  Pr(>|z|)
(Intercept)           210.844039  68.626013   3.0724 0.0021237
age                    14.740204   1.542523   9.5559 < 2.2e-16
race                    0.108733   0.025562   4.2537 2.103e-05
transport              -0.660785   0.088002  -7.5087 5.973e-14
female_workers          0.446700   0.116137   3.8463 0.0001199
housing_occupancy_rent  -1.194454   0.493050  -2.4226 0.0154106
min_wage               -1.415268   0.108385 -13.0578 < 2.2e-16
children                0.397485   0.085575   4.6449 3.403e-06
education               0.896063   0.107609   8.3271 < 2.2e-16
health_insurance        0.251789   0.104062   2.4196 0.0155377

Rho: 0.31691, LR test value: 291.81, p-value: < 2.22e-16
Approximate (numerical Hessian) standard error: 0.017965
    z-value: 17.64, p-value: < 2.22e-16
Wald statistic: 311.18, p-value: < 2.22e-16

Log likelihood: -46447.89 for lag model
ML residual variance (sigma squared): 616630, (sigma: 785.26)
Number of observations: 5739
Number of parameters estimated: 12
AIC: 92920, (AIC for lm: 93210)
```

Appendix D.23: Impact Measures for Spatial Autoregressive Lagged Response Model with Non-Home Dwell Time as the Dependent Variable

```
Impact measures (lag, trace):
                           Direct    Indirect      Total
age                    15.0779188  6.50080933 21.5787281
race                    0.1112239  0.04795394  0.1591779
transport              -0.6759244 -0.29142321 -0.9673476
female_workers          0.4569342  0.19700612  0.6539403
housing_occupancy_rent -1.2218201 -0.52678486 -1.7486049
min_wage               -1.4476937 -0.62416973 -2.0718634
children                0.4065915  0.17530098  0.5818925
education               0.9165933  0.39518704  1.3117803
health_insurance        0.2575578  0.11104546  0.3686033
```

```
Simulation results (mixed Hessian approximation variance matrix):
Direct:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                          Mean      SD  Naive SE Time-series SE
age                    15.0739 1.58053 0.0223521      0.0223521
race                    0.1115 0.02594 0.0003669      0.0003822
transport              -0.6764 0.09005 0.0012735      0.0012735
female_workers          0.4581 0.11918 0.0016855      0.0016180
housing_occupancy_rent -1.2055 0.50590 0.0071545      0.0071545
min_wage               -1.4480 0.11029 0.0015597      0.0015597
children                0.4063 0.08689 0.0012289      0.0012289
education               0.9194 0.11100 0.0015697      0.0015697
health_insurance        0.2557 0.10660 0.0015075      0.0015075

2. Quantiles for each variable:

                           2.5%      25%     50%     75%   97.5%
age                    11.97651 14.03299 15.0611 16.1082 18.1656
race                    0.05939  0.09467  0.1118  0.1289  0.1613
transport              -0.85313 -0.73707 -0.6770 -0.6161 -0.4973
female_workers          0.22429  0.37835  0.4586  0.5395  0.6901
housing_occupancy_rent -2.19978 -1.54787 -1.2019 -0.8796 -0.2255
min_wage               -1.66176 -1.52423 -1.4473 -1.3744 -1.2302
children                0.23763  0.34752  0.4055  0.4645  0.5756
education               0.70142  0.84228  0.9194  0.9965  1.1354
health_insurance        0.03684  0.18830  0.2560  0.3271  0.4613

Indirect:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                           Mean      SD  Naive SE Time-series SE
age                     6.51740 0.83412 0.0117962      0.0117962
race                    0.04819 0.01165 0.0001648      0.0001706
transport              -0.29215 0.04254 0.0006016      0.0006016
female_workers          0.19791 0.05317 0.0007520      0.0007212
housing_occupancy_rent -0.52098 0.22255 0.0031474      0.0031474
min_wage               -0.62587 0.06436 0.0009102      0.0010105
children                0.17565 0.03974 0.0005620      0.0005620
education               0.39717 0.05373 0.0007599      0.0007599
health_insurance        0.11050 0.04673 0.0006608      0.0006608

2. Quantiles for each variable:

                           2.5%      25%      50%      75%    97.5%
age                     4.97119  5.94029  6.49632  7.06323  8.20576
race                    0.02542  0.04050  0.04804  0.05603  0.07153
transport              -0.37815 -0.31927 -0.29086 -0.26380 -0.21134
female_workers          0.09546  0.16148  0.19686  0.23405  0.30395
housing_occupancy_rent -0.96751 -0.66935 -0.51647 -0.37729 -0.09686
min_wage               -0.75688 -0.66753 -0.62249 -0.58175 -0.50328
children                0.10081  0.14881  0.17419  0.20213  0.25866
education               0.29827  0.35993  0.39477  0.43253  0.50635
health_insurance        0.01595  0.08009  0.11028  0.14175  0.20245
```

```
Total:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                          Mean      SD  Naive SE Time-series SE
age                    21.5913 2.29415 0.0324442       0.032444
race                    0.1597 0.03713 0.0005251       0.000547
transport              -0.9685 0.12789 0.0018086       0.001809
female_workers          0.6560 0.17062 0.0024129       0.002313
housing_occupancy_rent -1.7265 0.72533 0.0102577       0.010258
min_wage               -2.0739 0.15961 0.0022572       0.002494
children                0.5820 0.12483 0.0017653       0.001765
education               1.3166 0.15781 0.0022317       0.002232
health_insurance        0.3662 0.15264 0.0021587       0.002159

2. Quantiles for each variable:

                           2.5%     25%     50%     75%   97.5%
age                    17.06925 20.0320 21.5633 23.1375 26.0486
race                    0.08544  0.1357  0.1600  0.1849  0.2322
transport              -1.21810 -1.0551 -0.9678 -0.8827 -0.7135
female_workers          0.32015  0.5411  0.6578  0.7733  0.9921
housing_occupancy_rent -3.12948 -2.2166 -1.7211 -1.2553 -0.3241
min_wage               -2.38149 -2.1808 -2.0749 -1.9649 -1.7554
children                0.33885  0.4993  0.5802  0.6643  0.8261
education               1.01294  1.2079  1.3160  1.4243  1.6262
health_insurance        0.05332  0.2693  0.3668  0.4694  0.6630

Simulated standard errors
                           Direct   Indirect      Total
age                    1.58053380 0.83411679 2.29414950
race                   0.02594224 0.01165314 0.03712738
transport              0.09004764 0.04254304 0.12788858
female_workers         0.11918205 0.05317247 0.17061710
housing_occupancy_rent 0.50589912 0.22255475 0.72532583
min_wage               0.11028894 0.06435758 0.15960661
children               0.08689346 0.03974221 0.12482721
education              0.11099579 0.05373403 0.15780834
health_insurance       0.10659835 0.04672611 0.15264138

Simulated z-values:
                           Direct  Indirect      Total
age                      9.537227  7.813531   9.411465
race                     4.299880  4.135327   4.302433
transport               -7.511484 -6.867107  -7.573304
female_workers           3.843715  3.722019   3.844930
housing_occupancy_rent  -2.382903 -2.340916  -2.380297
min_wage               -13.129191 -9.724959 -12.993694
children                 4.675883  4.419712   4.662067
education                8.283251  7.391457   8.342897
health_insurance         2.398315  2.364923   2.398826

Simulated p-values:
                       Direct     Indirect   Total
age                    < 2.22e-16 5.5511e-15 < 2.22e-16
race                   1.7089e-05 3.5445e-05 1.6893e-05
transport              5.8398e-14 6.5516e-12 3.6415e-14
female_workers         0.00012119 0.00019764 0.00012059
housing_occupancy_rent 0.01717671 0.01923650 0.01729871
min_wage               < 2.22e-16 < 2.22e-16 < 2.22e-16
children               2.9269e-06 9.8832e-06 3.1305e-06
education              2.2204e-16 1.4522e-13 < 2.22e-16
health_insurance       0.01647070 0.01803382 0.01644775
```

Appendix D.24: Results of Spatial Hausman Test with Non-Home Dwell Time as the Dependent Variable

Hausman Test Statistic	p-value	df
134.69	< 2.2e-16	10

Appendix D.25: Results of Ordinary Least Squares Regression with Percentage Time Spent at Home as the Dependent Variable

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            61.407858   8.383115   7.325 2.72e-13 ***
age                    -0.890047   0.192935  -4.613 4.05e-06 ***
race                    0.036739   0.003196  11.494  < 2e-16 ***
transport              -0.041412   0.010927  -3.790 0.000152 ***
female_workers          0.056000   0.014555   3.847 0.000121 ***
housing_occupancy_rent  0.122420   0.061872   1.979 0.047908 *
min_wage                0.021296   0.013439   1.585 0.113101
children               -0.081407   0.010737  -7.582 3.94e-14 ***
education              -0.108450   0.013371  -8.111 6.09e-16 ***
health_insurance        0.001344   0.013055   0.103 0.918008
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 98.6 on 5729 degrees of freedom
  (1124 observations deleted due to missingness)
Multiple R-squared:  0.05837,   Adjusted R-squared:  0.05689
F-statistic: 39.46 on 9 and 5729 DF,  p-value: < 2.2e-16
```

Appendix D.26: Results of Moran’s I Calculations with Percentage Time Spent at Home as the Dependent Variable

Moran I Statistic Standard Deviate	p-value	Observed Moran I	Expectation	Variance
15.205	< 2.2e-16	0.1133	-0.0005	0.0001

Appendix D.27: Results of Lagrange Multiplier Diagnostic Tests for Spatial Dependence with Percentage Time Spent at Home as the Dependent Variable

Test	Test Statistic	p-value	df
LMerr	228.54	< 2.2e-16	1
LMlag	284.76	< 2.2e-16	1
RLMerr	32.59	1.1e-08	1
RLMlag	88.82	< 2.2e-16	1
SARMA	317.35	< 2.2e-16	2

Appendix D.28: Results of Spatial Durbin Model (Spatially Lagged X Model) with Percentage Time Spent at Home as the Dependent Variable

```
Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              99.4304668 16.6614571   5.968 2.55e-09 ***
age                      -0.9150041  0.1992812  -4.592 4.49e-06 ***
race                      0.0301627  0.0038350   7.865 4.38e-15 ***
transport                -0.0563862  0.0119897  -4.703 2.63e-06 ***
female_workers            0.0415895  0.0152338   2.730 0.006351 **
housing_occupancy_rent    0.0973353  0.0619235   1.572 0.116037
min_wage                  0.0008224  0.0143440   0.057 0.954280
children                 -0.0610821  0.0108926  -5.608 2.15e-08 ***
education                -0.0544048  0.0143324  -3.796 0.000149 ***
health_insurance          0.0177214  0.0134127   1.321 0.186474
lag.age                  -0.6912135  0.3867752  -1.787 0.073971 .
lag.race                  0.0049162  0.0060050   0.819 0.412999
lag.transport            -0.0143470  0.0211866  -0.677 0.498322
lag.female_workers        0.1289507  0.0297773   4.331 1.51e-05 ***
lag.housing_occupancy_rent  0.0887697  0.1369663   0.648 0.516938
lag.min_wage              0.0443206  0.0261061   1.698 0.089617 .
lag.children             -0.0948064  0.0230680  -4.110 4.01e-05 ***
lag.education            -0.1596432  0.0264453  -6.037 1.67e-09 ***
lag.health_insurance     -0.0200953  0.0263920  -0.761 0.446440
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 97.48 on 5720 degrees of freedom
Multiple R-squared:  0.08103,   Adjusted R-squared:  0.07814
F-statistic: 28.02 on 18 and 5720 DF,  p-value: < 2.2e-16
```

Appendix D.29: Impact Measures for Spatially Lagged X Model with Percentage Time Spent at Home as the Dependent Variable

```
Impact measures (SLX, estimable, n-k):
                              Direct      Indirect        Total
age                    -0.9150041322 -0.691213451 -1.606217583
race                    0.0301627497  0.004916182  0.035078932
transport              -0.0563862007 -0.014347011 -0.070733211
female_workers          0.0415894858  0.128950704  0.170540190
housing_occupancy_rent  0.0973353363  0.088769710  0.186105046
min_wage                0.0008224192  0.044320635  0.045143054
children               -0.0610820825 -0.094806354 -0.155888436
education              -0.0544047792 -0.159643218 -0.214047997
health_insurance        0.0177213829 -0.020095274 -0.002373891
========================================================
Standard errors:
                           Direct    Indirect       Total
age                    0.199281228 0.386775220 0.388415828
race                   0.003835029 0.006004959 0.005198809
transport              0.011989685 0.021186563 0.020229752
female_workers         0.015233765 0.029777267 0.029733903
housing_occupancy_rent 0.061923523 0.136966285 0.142749174
min_wage               0.014343978 0.026106129 0.025129828
children               0.010892594 0.023067993 0.023598745
education              0.014332378 0.026445314 0.025858265
health_insurance       0.013412690 0.026391988 0.026573653
```

```
Z-values:
                           Direct    Indirect       Total
age                    -4.5915219 -1.7871193 -4.13530415
race                    7.8650648  0.8186870  6.74749413
transport              -4.7028927 -0.6771750 -3.49649418
female_workers          2.7300859  4.3305084  5.73554672
housing_occupancy_rent  1.5718637  0.6481136  1.30372065
min_wage                0.0573355  1.6977099  1.79639329
children               -5.6076712 -4.1098658 -6.60579343
education              -3.7959352 -6.0367299 -8.27773993
health_insurance        1.3212400 -0.7614157 -0.08933251

p-values:
                       Direct     Indirect   Total
age                    4.4003e-06 0.073918   3.5448e-05
race                   3.7748e-15 0.412965   1.5042e-11
transport              2.5650e-06 0.498295   0.00047141
female_workers         0.00633178 1.4877e-05 9.7198e-09
housing_occupancy_rent 0.11598217 0.516911   0.19232885
min_wage               0.95427794 0.089563   0.07243199
children               2.0507e-08 3.9589e-05 3.9539e-11
education              0.00014709 1.5727e-09 2.2204e-16
health_insurance       0.18642136 0.446409   0.92881766
```

Appendix D.30: Results of Spatial Autoregressive Lagged Response Model with Percentage Time Spent at Home as the Dependent Variable

```
Regions with no neighbours included:
 3374
Coefficients: (numerical Hessian approximate standard errors)
                        Estimate Std. Error z value        Pr(>|z|)
(Intercept)           54.6052306  8.1536611  6.6970 0.00000000002127
age                   -0.8646352  0.1874001 -4.6138 0.00000395286854
race                   0.0314528  0.0031224 10.0733        < 2.2e-16
transport             -0.0421589  0.0106128 -3.9724 0.00007113787987
female_workers         0.0448765  0.0141544  3.1705         0.001522
housing_occupancy_rent  0.1107175  0.0600992  1.8422         0.065439
min_wage               0.0140925  0.0130611  1.0790         0.280602
children              -0.0696336  0.0104547 -6.6605 0.00000000002729
education             -0.0820529  0.0130933 -6.2668 0.00000000036859
health_insurance       0.0081590  0.0126871  0.6431         0.520161

Rho: 0.30435, LR test value: 235.71, p-value: < 2.22e-16
Approximate (numerical Hessian) standard error: 0.019227
    z-value: 15.829, p-value: < 2.22e-16
Wald statistic: 250.56, p-value: < 2.22e-16

Log likelihood: -34368.54 for lag model
ML residual variance (sigma squared): 9171.3, (sigma: 95.767)
Number of observations: 5739
Number of parameters estimated: 12
AIC: 68761, (AIC for lm: 68995)
```

Appendix D.31: Impact Measures for Spatial Autoregressive Lagged Response Model with Percentage Time Spent at Home as the Dependent Variable

```
Impact measures (lag, trace):
                             Direct     Indirect       Total
age                    -0.882763999 -0.360149998 -1.24291400
race                    0.032112292  0.013101171  0.04521346
transport              -0.043042895 -0.017560637 -0.06060353
female_workers          0.045817442  0.018692597  0.06451004
housing_occupancy_rent  0.113038889  0.046117598  0.15915649
min_wage                0.014387998  0.005870014  0.02025801
children               -0.071093644 -0.029004780 -0.10009842
education              -0.083773302 -0.034177826 -0.11795113
health_insurance        0.008330091  0.003398510  0.01172860
```

```
Simulation results (mixed Hessian approximation variance matrix):
Direct:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                            Mean       SD   Naive SE Time-series SE
age                    -0.884857 0.193707 0.00273944     0.00268171
race                    0.032083 0.003213 0.00004544     0.00004544
transport              -0.043191 0.010908 0.00015426     0.00015426
female_workers          0.046098 0.014443 0.00020425     0.00020425
housing_occupancy_rent  0.112305 0.061411 0.00086849     0.00088884
min_wage                0.014466 0.013328 0.00018849     0.00018849
children               -0.070990 0.010723 0.00015165     0.00015165
education              -0.083919 0.013150 0.00018596     0.00018596
health_insurance        0.008277 0.012864 0.00018193     0.00017504

2. Quantiles for each variable:

                           2.5%        25%       50%      75%    97.5%
age                    -1.26477 -1.0148898 -0.883838 -0.75216 -0.51211
race                    0.02577  0.0299544  0.032102  0.03429  0.03831
transport              -0.06408 -0.0505321 -0.043137 -0.03593 -0.02154
female_workers          0.01759  0.0363047  0.046386  0.05578  0.07369
housing_occupancy_rent -0.01034  0.0705985  0.113178  0.15463  0.22956
min_wage               -0.01185  0.0054542  0.014482  0.02367  0.04047
children               -0.09177 -0.0782505 -0.070958 -0.06371 -0.05014
education              -0.10994 -0.0927758 -0.084118 -0.07510 -0.05764
health_insurance       -0.01696 -0.0004992  0.008223  0.01700  0.03402
```

```
Indirect:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                            Mean       SD   Naive SE Time-series SE
age                    -0.361857 0.085646 0.00121122     0.00118532
race                    0.013111 0.001669 0.00002361     0.00002361
transport              -0.017671 0.004747 0.00006714     0.00006714
female_workers          0.018832 0.006087 0.00008608     0.00008608
housing_occupancy_rent  0.045910 0.025448 0.00035989     0.00035986
min_wage                0.005899 0.005486 0.00007759     0.00007759
children               -0.029011 0.004954 0.00007005     0.00007005
education              -0.034261 0.005846 0.00008267     0.00008267
health_insurance        0.003394 0.005293 0.00007486     0.00007222

2. Quantiles for each variable:

                            2.5%        25%       50%       75%     97.5%
age                    -0.536243 -0.4173085 -0.357746 -0.303394 -0.202739
race                    0.009956  0.0119576  0.013043  0.014202  0.016524
transport              -0.027106 -0.0207716 -0.017663 -0.014404 -0.008604
female_workers          0.007110  0.0147691  0.018828  0.022792  0.030930
housing_occupancy_rent -0.004091  0.0285549  0.045817  0.062978  0.096976
min_wage               -0.004822  0.0021988  0.005837  0.009523  0.016694
children               -0.039028 -0.0323307 -0.028847 -0.025571 -0.019709
education              -0.046143 -0.0381748 -0.034093 -0.030318 -0.023297
health_insurance       -0.006996 -0.0002074  0.003323  0.006928  0.014024
```

```
Total:

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                          Mean       SD  Naive SE Time-series SE
age                   -1.24671 0.275180 0.0038916      0.0038057
race                   0.04519 0.004589 0.0000649      0.0000649
transport             -0.06086 0.015476 0.0002189      0.0002189
female_workers         0.06493 0.020363 0.0002880      0.0002880
housing_occupancy_rent  0.15821 0.086583 0.0012245      0.0012526
min_wage               0.02036 0.018784 0.0002656      0.0002656
children              -0.10000 0.015214 0.0002152      0.0002152
education             -0.11818 0.018442 0.0002608      0.0002608
health_insurance       0.01167 0.018139 0.0002565      0.0002470

2. Quantiles for each variable:

                           2.5%        25%      50%      75%    97.5%
age                    -1.78675 -1.4313663 -1.24177 -1.05887 -0.72038
race                    0.03620  0.0420554  0.04518  0.04832  0.05431
transport              -0.09054 -0.0713874 -0.06077 -0.05053 -0.03033
female_workers          0.02476  0.0510646  0.06548  0.07878  0.10407
housing_occupancy_rent -0.01434  0.0991228  0.15921  0.21867  0.32511
min_wage               -0.01666  0.0076394  0.02041  0.03329  0.05686
children               -0.12967 -0.1103793 -0.09999 -0.08967 -0.07027
education              -0.15357 -0.1308288 -0.11831 -0.10580 -0.08133
health_insurance       -0.02378 -0.0007041  0.01163  0.02394  0.04801

Simulated standard errors
                            Direct    Indirect       Total
age                    0.193707370 0.085646337 0.275179744
race                   0.003212833 0.001669292 0.004589211
transport              0.010907706 0.004747477 0.015475726
female_workers         0.014442595 0.006086638 0.020362550
housing_occupancy_rent 0.061411396 0.025448264 0.086582526
min_wage               0.013328433 0.005486385 0.018784030
children               0.010722980 0.004953554 0.015213567
education              0.013149645 0.005845757 0.018442305
health_insurance       0.012864043 0.005293060 0.018138643

Simulated z-values:
                           Direct   Indirect      Total
age                    -4.5680077 -4.2250155 -4.5305437
race                    9.9857444  7.8539479  9.8476765
transport              -3.9597032 -3.7221139 -3.9327348
female_workers          3.1918105  3.0939368  3.1886822
housing_occupancy_rent  1.8287280  1.8040572  1.8273302
min_wage                1.0853290  1.0752026  1.0841502
children               -6.6203689 -5.8566561 -6.5731688
education              -6.3818542 -5.8607666 -6.4080781
health_insurance        0.6434443  0.6412227  0.6434509

Simulated p-values:
                       Direct            Indirect   Total
age                    0.00000492381947 2.3892e-05 0.000005883207486
race                   < 2.22e-16       3.9968e-15 < 2.22e-16
transport              0.00007504296913 0.00019756 0.000083984870710
female_workers         0.0014138        0.00197520 0.0014292
housing_occupancy_rent 0.0674404        0.07122234 0.0676501
min_wage               0.2777760        0.28228403 0.2782982
children               0.00000000003583 4.7228e-09 0.000000000049255
education              0.00000000017496 4.6074e-09 0.000000000147365
health_insurance       0.5199359        0.52137799 0.5199316
```

Appendix D.32: Results of Spatial Hausman Test with Percentage Time Spent at Home as the Dependent Variable

Hausman Test Statistic	p-value	df
97.53	< 2.2e-16	10

APPENDIX E: SAFEGRAPH NON-COMMERICAL DATA LICENSE AGREEMENT

START DATE: 4/15/2020
LICENSE PERIOD: Minimum of 1 year or until COVID-19 (Coronavirus) global response has subsided
DESCRIPTION OF DATA TO BE PROVIDED: SafeGraph Patterns data (or other as mutually agreed upon)
USAGE: SafeGraph data is to be used for COVID-19 (Coronavirus) response
PUBLISHING: Company must credit SafeGraph if it publishes anything or creates content using SafeGraph data

Non-Commercial DATA LICENSE AGREEMENT

This DATA LICENSE AGREEMENT ("Agreement") is entered into as of start date set forth above (the
"Effective Date"), by and between SafeGraph, Inc., a Delaware corporation, with its principal place of business at 182 Howard Street, Suite 842, San Francisco CA 94105 ("Licensor") and the company identified below ("Company") (each referred to herein as a "Party" and collectively as the "Parties").

WHEREAS, Licensor has compiled anonymized information and is willing to make available the data set described above (the "Data"); and

WHEREAS, Company wishes to use the Data in connection with Company's products or services in accordance to the terms and conditions herein as well as the Usage described.

NOW, THEREFORE, in consideration of the mutual promises, agreements and conditions stated herein, the Parties agree as follows:

1. Limited License. Subject to the terms and conditions of this Agreement, Licensor hereby grants Company a temporary, limited, royalty free, non-exclusive, non transferable, non-sublicensable, revocable, license to the Data during the License Period solely for the purpose of developing response to helping fight Coronovirus and its first-order and second-order effects and in accordance with the terms and conditions of this Agreement.

Company must mention SafeGraph as originator of the data in any work product. The Data is provided for noncommercial purposes only and Company may not authorize another to use the Data for any commercial, resale, distribution or other purpose. For further clarity, Company shall not: (i) sell, rent, lease, sublicense, distribute, transfer or otherwise provide the Data or any portions or copies thereof to any third party or enable any third party to do any of those acts; (ii) copy, adapt, translate, reverse engineer, or create derivative works therefrom (other than as expressly authorized herein). UNLESS OTHERWISE AGREED BY A SEPARATE WRITING, COMPANY AGREES AND UNDERSTANDS THAT IT IS NOT AUTHORIZED TO DISTRIBUTE OR OTHERWISE USE THE DATA.

2. Further Obligations. Company agrees that it is responsible for any acts or omissions of its agents or permitted subcontractors that access or use any of the Data and Company will ensure that such agents and permitted subcontractors comply with the terms of this Agreement. SafeGraph may use Company logo on corporate website and in marketing materials, and companies will work together on co-marketing initiatives.

3. Ownership. As between the Parties, Licensor shall own and retain all right, title and interest in and to the Data, together with all intellectual property rights therein and thereto. Licensor reserves all rights not expressly granted hereunder. Nothing contained in this Agreement shall be construed as transferring any right, title, or interest in the Data except as expressly set forth herein.

4. Confidentiality. Data shall constitute confidential information belonging to Licensor, and accordingly, Company shall not disclose the Data to any third party, except with Licensor's prior written consent and as permitted under the next sentence. Company may disclose the Data to its employees, consultants or other agents who have a bona fide need to know the Data under the limited license rights herein, provided, that each such employee, consultant or agent is bound by confidentiality obligations at least as protective as those set forth herein. Company shall protect the confidentiality the Data in the same manner that it protects the confidentiality of its own confidential information of like kind (but in no event using less than with reasonable care). Company shall promptly notify Licensor if it becomes aware of any actual or suspected breach of confidentiality of the Data. If Company is compelled by law or legal process to disclose the Data, it shall provide Licensor with prompt prior notice of such compelled disclosure (to the extent legally permitted) and provide reasonable assistance, at Licensor’s expense, if Licensor wishes to contest the proposed disclosure. Company acknowledges and agrees that any disclosure or use or breach of the Data would result in irreparable injury to Licensor for which money damages would be inadequate and in such event Licensor shall have the right, in addition to other remedies available at law and in equity, to seek immediate

injunctive relief. Upon any termination of this Agreement, to the extent that any Data is retained, Company shall continue to maintain the confidentiality of the Data.

5. Term and Termination. The license rights in section 1 is limited in duration to a time period starting from the Effective Date and continuing for period set forth above (the "License Period"), unless terminated herein. Company may terminate this Agreement at any time by notifying Licensor. Licensor may terminate this Agreement immediately if it has reason to believe that Company is not in compliance with the terms of this Agreement. Upon expiration or termination of this Agreement, the license rights stated in section 1 shall terminate and Company shall immediately discontinue all use of the Data and take steps to remove or destroy all copies of the Data from Company (including employees') hardware. Company shall not disclose, retain or use the Data or Test Analytics after the expiration or termination of this Agreement.

6. DISCLAIMERS. TO THE FULLEST EXTENT PERMISSIBLE PURSUANT TO APPLICABLE
LAW, LICENSOR MAKES NO WARRANTIES OR REPRESENTATIONS, EXPRESS, IMPLIED,
ORAL, WRITTEN, OR OTHERWISE, AND LICENSOR EXPRESSLY DISCLAIMS (I) ANY
IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR
NONINFRINGEMENT, (II) ANY WARRANTY REGARDING CORRECTNESS, QUANTITY,
QUALITY, ACCURACY, COMPLETENESS, RELIABILITY, PERFORMANCE, TIMELINESS OR
CONTINUED AVAILABILITY OF THE DATA. UNDER NO CIRCUMSTANCES SHALL
LICENSOR BE LIABLE FOR ANY INDIRECT, PUNITIVE, INCIDENTAL, SPECIAL,
CONSEQUENTIAL OR EXEMPLARY DAMAGES, INCLUDING WITHOUT LIMITATION
DAMAGES, FOR LOSS OF PROFITS, GOODWILL USE, OR OTHER INTANGIBLE LOSSES
THAT RESULT FROM THE USE OF OR INABILITY TO USE THE DATA. TO THE MAXIMUM
EXTENT PERMITTED BY APPLICABLE LAW, LICENSOR ASSUMES NO LIABILITY OR
RESPONSIBILITY FOR (I) ANY PERSONAL INJURY OR PROPERTY DAMAGE, OF ANY NATURE WHATSOEVER, RESULTING FROM COMPANY'S ACCESS TO AND USE OF THE
DATA; (II) ANY ERRORS OR OMISSIONS IN, OR ANY LOSS OR DAMAGE INCURRED AS A

RESULT OF THE USE OF THE DATA. IN NO EVENT SHALL LICENSOR, ITS DIRECTORS, EMPLOYEES, AFFILIATES OR LICENSORS BE LIABLE TO COMPANY FOR ANY CLAIMS, PROCEEDINGS, LIABILITIES, OBLIGATIONS, DAMAGES, LOSSES OR COSTS ARISING UNDER OR RELATING TO THIS AGREEMENT FOR MORE THAN $1,000. THIS LIMITATION OF LIABILITY APPLIES WHETHER THE ALLEGED LIABILITY IS BASED ON CONTRACT, TORT, NEGLIGENCE, STRICT LIABILITY, OR ANY OTHER BASIS, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

7. General. This Agreement shall be governed by the laws of California, except for its conflicts of laws principles. All disputes arising under or relating to this Agreement shall be within the exclusive jurisdiction of the state or federal courts located in San Francisco, California and each Party hereby consents to such exclusive jurisdiction and venue. Neither Party may assign this Agreement to any third party without the prior written consent of the other Party. Nothing in this Agreement is intended to confer any rights or remedies on any person or entity that is not a party to this Agreement. No modification of this Agreement or waiver of the terms and conditions hereof shall be binding upon the Parties unless approved in writing by each of the Parties. Except as otherwise provided herein, the failure of either Party to enforce at any time any provision of this Agreement shall not be constituted to be a present or future waiver of such provision, nor in any way affect the ability of either Party to enforce each and every such provision thereafter. If any provision of this Agreement is held invalid or unenforceable at law, such provision will be deemed stricken from this Agreement and the remainder of this Agreement will continue in effect and be valid and enforceable to the fullest extent permitted by law. This Agreement represents the entire agreement between the Parties and supersedes any and all prior understanding, agreements, or representations by or among the Parties, written or oral, related to the subject matter hereof. This Agreement may be executed in counterparts with the same force and effect as if each of the signatories had executed the same instrument.

IN WITNESS WHEREOF, each of the Parties hereto has caused this Agreement to be executed as of the Effective Date.

SafeGraph, INC.

DocuSigned by:

By: Joshua Levitz ______________________________

81300BE19B68464...

Joshua Levitz

Name: ______________________________

Title: ______________________________ BDR

4/15/2020

Date Signed:
____

McGill University
COMPANY: ______________________

DocuSigned by:
By: Emily Chen
12D35990D37D427...

Emily Chen
Name: __________________________

Undergraduate student
Title: __________________________

4/15/2020
Date Signed: ______________________

Address: _________________________ 805 Sherbrooke St W, Montreal, Quebec H3A 2K6, Canada

State of Incorporation: ______________ Montreal, QC