Assessing the Impact of Non-Pharmaceutical Interventions on Consumer Mobility Patterns and COVID-19 Transmission in the US

Joseph Zuccarelli
Laura Seaman, Kevin Rader, Kosuke Imai

A thesis presented for the degree of
SM Data Science

Harvard Institute for Applied Computational Science
Draper Scholars Program
Cambridge, MA
May 2023

Assessing the Impact of Non-Pharmaceutical Interventions on Consumer Mobility Patterns and COVID-19 Transmission in the US

Joseph Zuccarelli[1,2]
Laura Seaman[1], Kevin Rader[2], Kosuke Imai[2]

[1] The Charles Stark Draper Laboratory, Inc.
[2] Harvard University

Abstract

The outbreak of the coronavirus disease 2019 (COVID-19) in Wuhan, China, during late December 2019 and the subsequent global pandemic markedly changed consumer mobility patterns worldwide, largely in response to government-ordered non-pharmaceutical interventions (NPIs). In this two-part study, we investigated these changes as they related to the initial spread of COVID-19 within two US states–Massachusetts (MA) and Michigan (MI). Specifically, we used mixed-effects modeling to quantify the relationship between four NPIs (i.e., national emergency, workplace closures, stay-at-home requirements, and gathering restrictions) and point-of-sale (POS) credit card transactions, as well as the relationship between subsequent changes in POS transactions and COVID-19 case growth rates at both the county and individual-level. Both analyses revealed a significant negative association between NPIs and POS transactions, in particular a dose-response relationship. Specifically, gathering restrictions and workplace closures possessed the strongest associations with decreased POS transactions. Both analyses also uncovered a significant positive association between lagged changes in POS transactions compared to pre-pandemic baselines and COVID-19 case growth rates. The optimal consumer mobility lag period varied across states, as the strongest association between changes in POS transactions and COVID-19 case growth rates occurred at a mobility lag of 12-14 days within MA and 8-10 days within MI. Overall, this study supports previous findings that early NPIs effectively reduced human mobility and COVID-19 transmission in the US.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my three committee members–**Laura Seaman**, **Kevin Rader**, and **Kosuke Imai**–for all the time and effort they put forth to make this project possible. Their guidance and encouragement truly inspired me throughout my studies over the past year. I sincerely enjoyed meeting with them each week. I consider them as much more than just advisors–they are mentors for whom I have a great deal of respect and admiration. In addition, I would like to thank the **Harvard Institute of Applied Computational Science** and **Draper Laboratory** for making my graduate school experience both rewarding and enjoyable. This includes my **classmates**, who I am sure will continue to amaze me in all their endeavours beyond graduation. Lastly, I would like to thank my **friends and family** for their unwavering support during all the highs and lows of this journey.

Contents

List of Figures

List of Tables

1 Introduction

In early January 2020, Chinese scientists confirmed that a new pneumonia-like illness identified at a food market in Wuhan, China, was transmittable from human-to-human. Shortly after, three cases of the disease, now known as the Coronavirus Disease 2019 (COVID-19), were reported in Japan and Thailand, causing the Centers for Disease Control and Prevention (CDC) to begin screenings at popular international airports within the United States. On January 20$^{\text{th}}$, 2020, the CDC identified a resident of Washington state as the first person in the United States with a confirmed case of COVID-19 [1]. By the end of the month, the World Health Organization (WHO) was left with no choice but to declare a public health emergency, as the number of cases of COVID-19 worldwide had grown above 9,800, including over 200 deaths. Finally, with the number of cases and related deaths continuing to rise over the entire month of February, Tedros Adhanom Ghebreyesus, Director General of the WHO, officially declared COVID-19 a pandemic on March 11$^{\text{th}}$, 2020. Just two days later, former United States President Donald Trump followed suit, declaring COVID-19 a national emergency [2].

COVID-19 is known to cause severe respiratory system damage as well as other potentially fatal symptoms. Although the death rate of COVID-19 is lower than that of other notable epidemics such as Ebola [3], COVID-19 is highly contagious, allowing it to kill more people than other more deadly diseases [4]. The contagious nature of COVID-19, in conjunction with extensive human mobility, enables the virus to have an extremely high rate of transmission. Therefore, one common approach used to combat the spread of COVID-19 is social distancing recommendations.

Social distancing is defined as a public health practice that aims to prevent infected individuals from coming into contact with healthy individuals to reduce opportunities for disease transmission [5]. This practice hinges on the basic concept that infected particles in the air are less likely to be transmitted with an increased distance between people [6]. Adherence to social distancing recommendations is pertinent to reducing the spread of COVID-19, as COVID-19 is not only highly contagious, but also infects many individuals without showing any common symptoms of the virus. Unfortunately, however, voluntary social distancing guidelines are not sufficient to end the pandemic, as many people may choose not to adhere to these recommendations. Therefore, it is crucial that governments take concrete action to limit close-contact between many people.

The rapid spread of COVID-19 throughout the US required a swift government response. The US federal government chose to defer the responsibility of COVID-19 policy responses primarily to the states [7]. Prior to the development of vaccines, state governments relied heavily on various non-pharmaceutical interventions (NPIs) to control the spread of the disease. Popular types of NPIs included containment measures such as domestic or international travel bans, individual protection measures such as mask wearing requirements, social dis-

tancing measures such as business closures and gathering bans, and health system measures such as testing and contact tracing [8]. Governors across all 50 states chose to implement these different forms of NPIs at various points over time, while learning from each other in the process.

In this two-part study, we measure the impact of four NPIs (i.e., national emergency, workplace closures, stay-at-home requirements, and gathering restrictions) in terms of reducing consumer mobility and controlling the initial outbreak of COVID-19 at both the county and individual-level within urban regions in two states, Massachusetts and Michigan. Unlike previous studies, our mobility measure is point-of-sale (POS) credit card transaction counts, provided by the data broker BDEX. In part one, we investigate the following two research questions: How did early government NPIs impact county-level consumer mobility patterns? How did changes in county-level consumer mobility patterns affect the initial spread of COVID-19? In part two, we investigate the following three research questions: How can we group individuals based on their pre-pandemic consumer behavior? How did early government NPIs impact individual-level consumer mobility patterns? How did changes in individual-level consumer mobility patterns affect the initial spread of COVID-19?

The rest of the paper is organized as follows. First, we provide a thorough review of previous works that explored the effectiveness of government NPIs in terms of reducing human mobility and slowing the spread of COVID-19. Next, we lay out the mathematical details underlying the methods used to carry out our analysis. Then, we divide our study into two parts–a county-level and an individual-level analysis. Each part is structured as follows. First, we describe the data sources used in our analysis and the data aggregation process. Second, we explain our modeling approach. Third, we lay out our modeling results. Fourth and finally, we discuss the implications of our results in a practical context and assess our study's significance.

2 Literature Review

Several researchers conducted studies to estimate the impact of government interventions on reducing human mobility and controlling the COVID-19 outbreak, both on a national and an international scale. These studies typically involved the use of different types of regression analyses on state or county-level data. The outcome of interest in each study tended to be similar, including changes to human mobility patterns, general characteristics of the pandemic, the mortality of the virus, and others. In the following literature review, we focus on studies of the effectiveness of NPIs in terms of human mobility reduction and controlling the COVID-19 within the US, separating them by state and county-level analyses.

2.1 State-Level Studies

Zhang and Warner (2020) used event study regression to study the impact of three inter-

vention policies (i.e., shutdown orders, reopening orders, and mask mandates) on changes in the daily COVID-19 case growth rate across all 50 US states from February - August 2020. Each of these three models was estimated using multilevel mixed effects linear regression nested at the state level. All three models controlled for Medicaid expansion, vulnerable populations (essential workers and the minority population), and the daily test growth rate in each state. Their results highlight the importance of early intervention, as statewide mask mandates and shutdown orders reduced the COVID-19 case growth rate immediately following their implementation. In the long term, masks mandates proved to be more effective than shutdown orders in terms of flattening the curve. On the other hand, reopening orders led to significant increases in new cases 21 days after their declaration. Given these results, Zhang and Warner recommended a dynamic social distancing approach: a shutdown for a short period followed by reopening, combined with universal mask mandates [9]. Although their recommendation appears valid, there were limitations to this study. Specifically, the authors used executive orders as the sole measure of shutdown level, whereas a more comprehensive modeling approach might include data concerning government orders along with measures of human mobility patterns to better understand the disease spread.

Siedner, Harling, et al. (2020) conducted a longitudinal pretest-posttest comparison group study to estimate changes in the average COVID-19 case growth rate before versus after the implementation of social distancing measures across all 50 US states from January - May 2020. The authors applied a broad definition of social distancing measures that included closures of schools, closures of workplaces, cancellations of public events, restrictions on internal movement, and closures of state borders. They fit mixed effects linear regression models, in which the log difference in daily cases was the outcome of interest and state was included as a random effect to allow for within-state correlation of cases over time. Model explanatory variables included time (days relative to intervention), implementation period, and a time-by-implementation-period product term. In fitting these models, the authors adjusted for state-level population density and day of the week, and stratified estimates by the size of the epidemic in the state at the time of intervention implementation. Using this approach, the authors found that the COVID-19 case rate declined by approximately 1% per day beginning 4 days after statewide social distancing measures were implemented. Their model implies that social distancing measures reduced the total number of COVID-19 cases by about 1,600 reported cases at 7 days after implementation, by about 55,000 reported cases at 14 days after implementation, and by about 600,000 reported cases at 21 days after implementation [10]. However, their modeling approach was subject to a few limitations, specifically, the potential bias resulting from the aggregate nature of their study data, the potential confounding caused by other changes that took place during the study period (e.g., increases in testing), and the potential underestimation of social distancing due to spillover effects from neighboring states.

Dreher, Spiera, et al. (2021) measured the impact of social distancing policies (i.e., stay-at-home orders, school closures, non-essential business closures, and bans on social

gatherings) on COVID-19 transmission in US states from March - May 2020 to assess which policies were most effective. In order to measure virus transmissibility, the authors analyzed the average effective reproductive number (R_t) in each state the week following its 500^{th} case and doubling time from 500 to 1,000 cases. They developed linear and logistic regression models to assess the impact of social distancing policies on R_t following 500 total cases, controlling for population density, GDP, and various state-level health metrics. Their results indicated that stay-at-home orders had the largest effect of the four policies included in the study. Using publicly available Google community mobility reports, the authors also found that social distancing adherence drove these effects [11]. Although these findings are noteworthy, there are a few limitations to this study. Specifically, the models included in this study do not account for local variations at the county level as well as periods during which individuals were infectious but asymptomatic.

Li, Rice, et al. (2021) used a regression-based event study framework to assess the effectiveness of seven major public intervention policies (i.e., school closures, workplace closures, public event cancellations, public transport closures, public information campaigns, stay-at-home orders, and international/national travel controls) on human mobility and COVID-19 case growth rates in US states from March - July 2020. The authors applied a fixed effects panel regression model that incorporated state-level policy data, human mobility data, and state-level confirmed case and death data, controlling for testing and demographic differences across states. Their results indicated that policies which created incentives for less movement and prohibited large gatherings (i.e., stay-at-home orders, workplace closures, public event cancellations, and public transport closures) decreased human mobility the most significantly. Of these four policy interventions, only stay-at-home orders and workplace closures led to significant decreases in the COVID-19 confirmed case growth rate, supporting the common assertion that policies which limit human mobility can in turn control disease transmission. Although public information campaigns did not significantly affect mobility, they did produce significant decreases in the confirmed case growth rate, perhaps indicating the importance of implementing policies that may not influence COVID-19 trends directly through mobility [12]. Despite these findings, this study was subject to a few limitations. Most notably, the authors chose to represent all public intervention policies as binary variables, whereas including additional stringency levels defined by whether a policy is recommended or required would perhaps lead to more granular results.

Olney, Smith, et al. (2021) applied an established, semi-mechanistic Bayesian hierarchical model of social distancing interventions on COVID-19 spread in Europe to US states using case fatalities from February - April 2020. The authors studied the effect of six public interventions: self-isolation if ill recommendations, social distancing encouragement, school closures, bans on sporting events, bans on public gatherings, and lockdown orders (note that some more restrictive interventions imply others, e.g., lockdown implies all other interventions, and bans on public events implies bans on sporting events). In order to explore the association between these interventions and fatalities, the authors employed a model that in-

corporated classical Susceptible-Infected-Removed (SIR) concepts in a Bayesian framework in order to estimate the effect of interventions on the time-varying reproduction number for each state. Using this approach, the authors found that school closures and lockdowns were the only two public interventions that had a reliable impact on the time-varying reproduction number. Given these findings, the authors suggested that removing lockdowns, without the implementation of additional, equally effective interventions, would lead to continued, sustained transmission of COVID-19 across the US [13]. However, the authors' modeling approach was based on a few major assumptions, specifically that the effect of each intervention is the same regardless of location and that the implementation of an intervention instantaneously reduces the time-varying reproduction number.

2.2 County-Level Studies

Courtemanche, Garuccio, et al. (2020) evaluated the impact of several social distancing measures (i.e., gathering bans, school closures, entertainment-related business closures, and shelter-in-place orders) on the daily growth rate of confirmed COVID-19 cases across 3,138 US counties between March - May 2020. The authors used an event study regression model with multiple policies to estimate this relationship between social distancing measures and the COVID-19 case growth rate, including control variables related to the availability of COVID-19 tests and fixed effects for geography and time. Using this modeling approach, the authors found evidence that shelter-in-place-orders and entertainment-related business (i.e., gyms, restaurants, bars) closures substantially slowed the spread of COVID-19; however, the same could not be said for bans on social gatherings and school closures. In total, the authors estimated that the adoption of these four government-imposed social distancing measures reduced the daily COVID-19 case rate by 5.4% after one to five days, 6.8% after six to ten days, 8.2% after eleven to fifteen days, and 9.1% after sixteen to twenty days [14]. Although these findings are informative, there are a few limitations to this study. Specifically, available data forced the authors to control for the number of tests performed at the state, opposed to county, level. Additionally, as is typical of observational data analyses, the authors could not rule out all possible threats to causal inference, to include possible confounders such as informal encouragement by government officials to wear masks or improve hygiene, changing business practices, and social norms regarding social distancing.

Ebrahim, Ashworth, et al. (2020) created a large and comprehensive NPI data set for 1,320 US counties across all 50 states from March - July 2020 in order to analyze the relationship between NPI policies and changes in reported COVID-19 cases. This dataset covers the following seven distinct NPI policies: closure of non-essential workplaces, shelter-in-place/stay-at-home orders, enforcement of shelter-in-place/stay-at-home orders, size restrictions on public gatherings, school closures, public transit closures, and publicly available testing. Using this data set, along with a few other publicly available sources of COVID-19 data, the authors conducted time-series correlational analyses to assess the association

between NPI policies, COVID-19 growth rates, and local political alignments. In terms of NPI policies, the authors uncovered significant variation in policy implementation among counties, both within and among states. In terms of COVID-19 case growth rates, the authors identified a statistically significant positive correlation between county workplace closures and lower case growth rates. In terms of political alignments, they found relatively limited differences in NPI policy enforcement by party [15]. Despite these findings, there are a few noteworthy limitations of this study concerning both the data collection process as well as the statistical methods used to analyze the compiled data. Although data collectors were well trained and used standardized collection procedures, the estimated date of policy changes could be highly variable, especially for counties with limited or conflicting information available online. In terms of statistical analysis, the authors were only able to identify correlations in the data set without any implications of causality.

Jalali, Khoury, et al. (2020) conducted univariate and multivariate statistical analyses to characterize the associations between the differential spread of COVID-19 in the 30 most populous counties in the US and various parameters, including public health interventions, community compliance, and health disparities. The authors grouped these counties into high, mid, and low groups based on their confirmed case growth rates on May 10th, 2020, with ten counties making each spread group. Non-parametric Kruskal-Wallis tests along with multiple regression analyses were employed to compare each group across three categories: public health interventions, community compliance, and health disparities. The authors studied the effectiveness of three policy interventions: restrictions on mass gatherings, stay-at-home orders, and face mask requirements. Their results indicated that the early implementation of public health interventions helped slow the spread of COVID-19. In order to analyze community compliance with these public interventions, the authors used Google community mobility reports to analyze the percent change in mobility across spread groups as well. Their results indicated a low level of compliance with these interventions that likely led to the heightened spread of the virus. The authors also attributed underlying health disparities to the disproportionate distribution of COVID-19, as they used census data to characterize demographic differences, including social determinants of health, to examine associations with the spread of the virus. They found that population density, housing density, African American race, education, and percent uninsured were all associated with amplified case rates of COVID-19 during the early stages of the pandemic [16]. Although these findings are noteworthy, there are a few limitations of this study, most notably the lack of consideration for differences in testing across counties.

Badr, Du, et al. (2020) explored the association between human mobility patterns and COVID-19 transmission in the 25 most affected US counties between January - April 2020 using generalized linear models. Using daily mobility data derived from aggregated and anonymised cell phone data provided by Teralytics, they created a mobility ratio, which quantified the change in mobility patterns as a proxy for social distancing. The authors fitted a generalized linear model for each county, with their lagged mobility ratio as a predictor

of COVID-19 growth rate and tested the correlation between mobility ratio and growth rate at different time lags. Their analysis revealed a strong correlation between lagged mobility patterns and decreased COVID-19 growth rates, with Pearson correlation coefficients above 0.7 for 20 of the 25 counties evaluated. Additionally, the authors found an optimal lag period of 11 days for mobility changes, with a window of 9-12 days [17]. Despite these findings, the authors noted a few limitations of their study, specifically the potential for omitted variable bias (e.g., mask-wearing and hand-washing could have also contributed to a decline in case growth rate and were not accounted for in models), failure to differentiate between low-risk and high-risk trips in creating their mobility metric, as well as the small number of counties included in the study.

Gatala, Tseng, et al. (2020) conducted a very similar study to Badr and colleagues, in which they employed generalized linear models to quantify the associations between phone mobility data and COVID-19 cases in the 25 most affected US counties between January - July 2020. The authors created three mobility metrics using phone mobility data provided by Unacast. The first metric was analogous to the one used by Badr and colleagues, while the other two metrics measured changes in visits to non-essential places and encounter density, which were noted as limitations of the study by Badr and colleagues. Fitting separate generalized linear models for each county using each different mobility metric, the authors identified a strong correlation between decreased mobility and reduced COVID-19 case growth between March 27$^{\text{th}}$ - April 20$^{\text{th}}$, 2020. However, when the authors extended their analysis to later time periods not included in the study by Badr and colleagues (i.e., April 21$^{\text{st}}$ - May 24$^{\text{th}}$ and May 25$^{\text{th}}$ - July 22$^{\text{nd}}$), they uncovered a much weaker correlation between mobility changes and COVID-19 case growth. Perhaps these results suggest that after the initial phase of the COVID-19 pandemic, other individual level factors such as mask-wearing and hand-washing became more important in terms of reducing the disease spread [18]. Although these results are informative, there are a few limitations of this study. Most notably, the authors only evaluated the 25 US counties most affected by COVID-19; therefore, their results may not be generalizable across the country.

McLaren and Wang (2020) investigated the effect of aggregate workplace absence on COVID-19 mortality in 2,830 US counties between February - August 2020 using an instrumental variable approach. The dataset under analysis in this study included COVID-19 deaths reported by the Johns Hopkins University Coronavirus Resource Center, Google user location tracking measures, various demographic and economic controls from the US Census Bureau, and worker surveys conducted by the US Department of Labor. According to the authors, the main challenge in measuring the effectiveness of workplace absence on COVID-19 deaths is the endogeneity problem, specifically the presence of reverse causality. To overcome this challenge, McLaren and Wang employed an instrumental variable strategy based on an index of ability to work from home developed by Dingel and Neiman (2020) [19]. Using this strategy, the authors found no effect of workplace absence on COVID-19 mortality up to mid-May, yet a very strong reductive effect following this period up until August [20]. While

this finding is informative, there are a few study limitations. Specifically, there are a few crucial assumptions associated with the use of the authors' instrumental variable approach, and violations of these assumptions can lead to biased estimates.

Welsch (2022) examined the relationship between mask usage and COVID-19 deaths in 3,079 US counties and county equivalents using a quasi-experimental approach. The main data set considered in this analysis includes COVID-19 county-level deaths as reported by the CDC, information on mask usage from a New York Times survey, 2016 US presidential election results obtained from the MIT Election Data and Science Lab, and various demographic estimates provided by the US Census Bureau. According to Welsch, the main empirical complication when examining the structural relationship between mask usage and COVID-19 deaths is omitted variable bias (endogeneity). In order to address this complication, Welsch incorporated a rich set of county-level controls, including COVID-19 deaths prior to the survey on mask usage, and employed the percentage of individuals who voted for Donald Trump in the 2016 election as an instrumental variable for mask usage. Prior to fitting any models, Welsch presented both empirical evidence and arguments that the assumptions required to carry out an instrumental variable approach were satisfied. The study produced two main findings: 1.) a 1% increase in mask usage would decrease COVID-19 deaths by approximately 10.5% or approximately six people in the average county 2.) Mask usage is more effective at reducing deaths in areas where there are more deaths due to COVID-19 [21]. Despite these findings, Welsch recognizes a few limitations of this study, most notably that the instrumental variable approach may not be able to avoid the issue that mask wearing may be correlated with other COVID-19 prevention methods. If these other preventive measures are negatively correlated with the instrument, the percentage of individuals who voted for Trump in the 2016 election, the effect of mask usage on COVID-19 deaths would be overstated as a result.

3 Methodology

Most statistical learning problems fall into one of two domains: *supervised* or *unsupervised*. In this study, we rely on methods from both domains. The following subsections contain descriptions of the mathematical details underlying these methods.

3.1 Supervised Learning

Supervised learning is defined by the use of labeled datasets (i.e., for each observation of the predictor measurements there is an associated response measurement). Methods within this domain require fitting a model that relates the response to the predictors, with the aim of either accurately predicting the response for future observations (prediction) or better understanding the relationship between the response and the predictors (inference)

[22]. Our analysis involves the use of a few supervised learning methods, which we describe in-depth within the following subsections.

3.1.1 Linear Model

Linear models are central to the practice of statistics; they are the foundation of a broad range of statistical methodologies. A common starting point is the linear regression model, in which y is known as the response variable and $x_1, \ldots, x_p$ are the predictors. This model takes the form

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon,$$

where ϵ is normally distributed (i.e., $\mathcal{N} \sim (0, \sigma_y^2)$) [23]. There are four assumptions associated with linear regression modeling: linearity (i.e., the relationship between the predictors and the means of the response is linear), homoscedasticity (i.e., the variance of the model residuals is constant), independence (i.e., observations are independent of one another), and normality (i.e., the model residuals are normally distributed). Although the standard linear model is very useful, there are also extensions to this framework that enable for more principled modeling when its assumptions are violated.

3.1.2 Poisson Model

The standard linear regression model cannot handle non-normal response variables, such as counts or proportions. The Poisson distribution is commonly used to model count data. A random variable Y is said to follow a Poisson distribution with mean μ if it takes integer values (i.e., $y \in \mathbb{Z}$) with probability

$$f(y; \mu) = \Pr\{Y = y\} = \frac{e^{-u} u^y}{y!}$$

for $\mu > 0$. The mean and variance of a random variable that follows this distribution are defined as follows:

$$\mathbb{E}(Y) = \mathbb{V}(Y) = \mu.$$

Since these two quantities are equal, the usual assumption of homoscedascticity is not appropriate for Poisson data.

One useful property of the Poisson distribution is that the sum of independent Poisson random variables is also Poisson. That is, if $Y_1, \ldots, Y_n$ are independent with $Y_i \sim \text{Pois}(\mu_i)$ for $i = 1, \ldots, n$, then it follows that $\sum_i^n Y_i \sim \text{Pois}(\sum_i^n \mu_i)$. This is useful because it allows for the analysis of individual or clustered data with equivalent results. If it is assumed that the individual-level data is Poisson, then by this property so is the summed data and Poisson regression can still be applied.

Suppose Y_i represents a sample of n count responses $y_1, y_2, \ldots, y_n$ such that $Y_i \sim \text{Pois}(\mu_i)$. In order to model Y_i in terms of a vector of predictors $\mathbf{x}_i$, a simple linear model of the form

$\mu_i = \mathbf{x}_i^T \beta$ is not a valid approach, as the linear predictor on the right hand side of the equality can assume any real value, while the Poisson mean on the left hand side of the equality must be non-negative. A straightforward solution to this issue is to model the logarithm of the Poisson mean using a linear model. The result is the following generalized linear model with a log link function [24]:

$$\log(\mu_i) = \mathbf{x}_i^T \beta.$$

The resulting log-likelihood function is defined as follows:

$$\log L(\beta; y) = \sum_{i=1}^{n} \left(y_i \mathbf{x}_i^T \beta - \exp(\mathbf{x}_i^T \beta) - \log(y_i!) \right).$$

Differentiating with respect to β gives the maximum likelihood estimation as the solution to

$$\sum_{i=1}^{n} \left(y_i - \exp(\mathbf{x}_i^T \hat{\beta}) \right) \mathbf{x}_j = 0 \qquad \forall j,$$

which can be simplified such that

$$\mathbf{X}^T \mathbf{y} = \mathbf{X}^T \hat{\mu},$$

where $\mathbf{X}$ is the model matrix (i.e., a matrix with one row for each observation and one column for each predictor, including the intercept), $\mathbf{y}$ is the response vector, and $\hat{\mu}$ is a vector of fitted values (calculated from the maximum likelihood estimates of $\hat{\beta}$ by exponentiating the linear predictor $\mathbf{x}_i^T \hat{\beta}$. Note that there is no explicit formula for $\hat{\beta}$ in this setting, thus requiring the use of numerical methods to find a solution [25].

Although the preceding paragraph described how Poisson regression is used to model count responses, this form of regression is also appropriate for modeling rate data, where the rate is a count of events divided by a measure of each unit's exposure. For instance, epidemiologists are often interested in modeling the count of a disease within an area divided by the population of that area. In Poisson regression models, this is handled by the inclusion of an offset term, which represents the size, measurement time, or population size of each observational unit. This term is included in Poisson models as a covariate whose regression coefficient is fixed to be one. Consider the Poisson regression model defined previously in matrix form:

$$\log(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p,$$

where μ is the mean of the response variable and $x_1, \ldots, x_p$ are predictors. Suppose that A represents the size, measurement time, or population size of each observational unit. Including an offset term results in the following model:

$$\log(\mu) = \log(A) + \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p,$$

which, by the properties of logarithms, can be re-written as follows [26]:

$$\log(\frac{\mu}{A}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

3.1.3 Negative Binomial Model

As mentioned in the previous section, the Poisson model requires that the mean and variance of the response variable are equal (i.e., $\mathbb{E}(Y) = \mathbb{V}(Y)$). Oftentimes this restriction is not satisfied when using real-world data, resulting in a phenomenon known as over-dispersion. Suppose that $\mathbb{E}(Y) = \Phi\mathbb{V}(Y)$, where Φ represents a term known as the dispersion parameter. When $\Phi > 1$, the data is considered over-dispersed, which can lead to underestimated stand error estimates and inflated test statistics when using a Poisson model [27].

Another type of model that works under such a condition and is commonly used as a distribution for count data is the negative binomial model, which is formulated as follows:

$$Y_i \sim \text{NB}(\mu_i, \mu_i[1 + \alpha\mu_i]),$$

where α represents a scaling parameter that controls for over-dispersion. As this formulation suggests, the negative binomial model is simply a generalization of the Poisson model that relaxes the restrictive assumption that the mean and variance of the response variable must be equal. The negative binomial probability distribution function can be expressed in a variety of ways, with a common parameterization appearing as follows:

$$f(y; \mu, \alpha) = \Pr\{Y = y\} = \binom{y + \frac{1}{\alpha} - 1}{\frac{1}{\alpha} - 1} \left(\frac{1}{1 + \alpha\mu}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\mu}{1 + \alpha\mu}\right)^{y}.$$

The resulting log-likelihood function is defined as follows:

$$\log L(\beta; y, \alpha) = \sum_{i=1}^{n} \left[y_i \log \left(\frac{\alpha \exp(x_i'\beta)}{1 + \alpha \exp(x_i'\beta)} \right) - \left(\frac{1}{\alpha}\right) \log(1 + \alpha \exp(x_i'\beta)) \right]$$

$$+ \log \Gamma(y_i + \frac{1}{\alpha}) - \log \Gamma(y_i + 1) - \log \Gamma(\frac{1}{\alpha}).$$

Parameter estimation can be achieved through likelihood maximization by employing a non-linear optimization method [28].

3.1.4 Mixed Effects Models

The standard linear regression model is also unreliable when there exists a correlation structure in the error term, ϵ. This is often the case when data has a grouped, nested, or hierarchical structure. Mixed effects models, also commonly known as multilevel models, are an extension of the standard regression model, in which data are structured in groups and regression coefficients can vary by group. These models contain both fixed effects and random effects, hence the name "mixed effects models." A fixed effect is an unknown constant that can be estimated using data. In contrast, a random effect is a random variable. Instead of using data to estimate a random effect, one can estimate the parameters that describe the

distribution of a random effect [29].

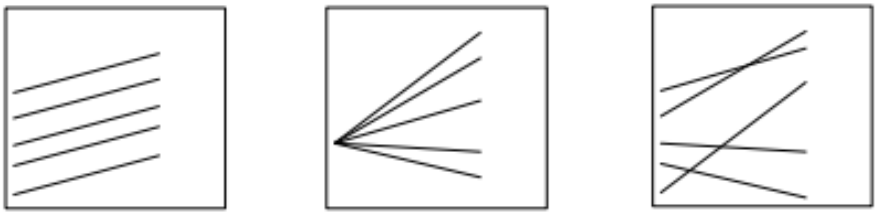
The linear mixed effects model takes the following form:

$$y_{ij} = \beta_1 x_{1ij} + \cdots + \beta_n x_{nij} + \gamma_{i1} z_{1ij} + \cdots + \gamma_{in} z_{nij} + \epsilon_{ij},$$

where y_{ij} represents the response variable for observation i from group j, β_1 through β_n are the fixed effect regression coefficients, x_{1ij} through x_{nij} are the fixed effect predictor variables for observation i from group j (usually the first is reserved for the intercept/constant; $x_{1ij} = 1$), γ_{i1} through γ_{in} are the random effect regression coefficients, which are assumed to be multivariate normally distributed, z_{1ij} through z_{nij} are the random effect predictor variables (usually the first is reserved for the intercept/constant; $z_{1ij} = 1$), and ϵ_{ij} represents the error term for observation i from group j, in which each group's error is assumed to be multivariate normally distributed [30].

There are three types of mixed effects models: the varying-intercept model, the varying-slope model, and the varying-intercept, varying-slope model. The varying-intercept model allows the intercept to vary by group j with constant slope. The varying-slope model allows the slope to vary by group j with constant intercept term. The varying-intercept, varying-slope model allows both the intercept and slope terms to vary by group j [29]. Figure 1 visually depicts these three types of models.

The maximum likelihood estimation of mixed effects models is a bit more complex than that of models with only fixed effects. The expression for the likelihood of a mixed effects model is an integral over the random effects space. For a linear mixed effects model, this integral can be evaluated exactly. A common approach is to fit these models under the restricted maximum likelihood (REML) criterion opposed to the standard log-likelihood criterion for the optimization of parameter estimates, as REML generally produces less biased estimates [31].



(a) Varying-intercept (b) Varying-slope (c) Varying-intercept, Varying-slope

Figure 1: Mixed-Effect Models [29]

3.1.5 Negative Binomial Mixed Effects Models

Negative binomial mixed effects models combine the statistical properties introduced in the previous two subsections to enable for the modeling of over-dispersed count or rate data when there exists correlation in the error term. These models follow a very similar format as the standard linear mixed effects model introduced in the previous section, with the exception that the response variable is assumed to follow a negative binomial distribution, requiring the inclusion of a log link function [32]:

$$Y_i \sim \text{NB}(\mu_i, \mu_i[1 + \alpha\mu_i])$$

$$\log(\mu_{ij}) = \beta_1 x_{1ij} + \cdots + \beta_n x_{nij} + \gamma_{i1} z_{1ij} + \cdots + \gamma_{in} z_{nij} + \epsilon_{ij}.$$

Note that the random effects regression coefficients as well as each group's error is assumed to follow a multivariate normal distribution, as was the case with the standard linear mixed effects model. Also note that the inclusion of an offset term in a negative binomial mixed effects model follows the same logic as the model with only fixed effects; it is included as a fixed effect whose regression coefficient is fixed to be one [33].

In terms of maximum likelihood estimation, recall from the previous section that the expression for the likelihood of a mixed effects model is an integral over the random effects space. Unlike linear mixed effects models, however, for generalized linear mixed effects models, the integral must be approximated, as the likelihood function does not have a closed form solution. There are a few different approximation methods for maximum likelihood estimation with respect to generalized linear mixed models–Laplace Approximation, Adaptive Gaussian Quadrature, and Penalized Quasi likelihood are the most common techniques. Laplace Approximation is known to produce the most reliable parameter estimates in a computationally efficient manner [34, 35].

3.1.6 Covariance Structure

Recall from the previous section describing mixed effects models that the random effects regression coefficients as well as each group's error is assumed to follow a multivariate normal distribution. Typically, random effects are fit with unstructured variance-covariance matrices, which is a reasonable approach when there does not exist dependence among within-group observations. However, observations of a time series tend to exhibit dependence upon one another, which is typically referred to as autocorrelation or serial correlation. In order to model the covariance pattern of observations from the same group, random effects can be fit with a structured variance-covariance matrix. An autoregressive (order 1) covariance structure specifies homogeneous variances and correlations that decline exponentially with time.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

In other words, this means that the variability of an observation will be constant regardless of the time of an observation, and temporally close observations will be highly correlated, yet as the time between observations increases they will be less correlated [36, 37, 38].

3.2 Unsupervised Learning

Unsupervised learning is defined by the use of unlabeled datasets (i.e., one observes a vector of predictor measurements, but no associated response). Given the lack of a response variable to "supervise" analysis, it is not possible to fit regression models within this setting. Instead, the goal of unsupervised learning is often to develop an understanding of the relationships between variables or observations, a task that can be achieved using a method known as *clustering*, for which the purpose is to determine whether the individuals within a dataset fall into relatively distinct groups [22].

Cluster analysis is characterized by four steps: feature selection, algorithm selection, cluster validation, and results interpretation [39]. There are two main types of clustering algorithms–hierarchical and partitional. Hierarchical clustering involves iteratively merging (or dividing) the data, typically one observation at a time, then deciding on partitions afterward. Partitional clustering requires specifying the number of clusters in advance, then invoking an algorithm to partition the data [40]. In the following subsections, we describe common proximity measures used to perform clustering, methods for choosing the optimal number of clusters, two hierarchical clustering algorithms (i.e., agglomerative and divisive clustering), and one partitional clustering algorithm (i.e., K-means clustering).

3.2.1 Proximity Measures

Clustering requires knowledge on how "close" individuals in the dataset are to one another. Most clustering investigations begin with an $n \times n$ one-mode matrix containing elements that represent a quantitative measure of closeness, more commonly referred to within this context as *dissimilarity*, *distance* or *similarity*, with a general term being *proximity*. Two individuals are considered close when their dissimilarity is small or their similarity is large. When all observed features are continuous, proximities between individuals are typically quantified by dissimilarity measures [41].

Dissimilarity measures are broadly divided into two categories: distance-measures and correlation-type measures. Two common distance measures are Euclidean distance and Manhattan distance. Let x_{ik} and x_{jk} represent, respectively, the kth feature value of the p-dimensional observations for individuals i and j. Euclidean distance is defined as follows [41]:

$$d_{ij}(x_{ik}, x_{jk}) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}.$$

Similarly, Manhattan distance is defined as follows [41]:

$$d_{ij}(x_{ik}, x_{jk}) = \sum_{k=1}^{p} |x_{ik} - x_{jk}|.$$

3.2.2 Determining Optimal Number of Clusters

In general, there are two types of methods for determining the optimal number of clusters: direct and testing methods. Direct methods (e.g., elbow and silhouette methods) involve optimizing a particular criterion. Testing methods (e.g., gap statistic method) evaluate evidence against a null hypothesis [42]. In the following paragraphs, we describe the technical details behind each specific method.

The elbow method is defined as follows. Let $W(C_k)$ represent the within-cluster sum of squared distances between all pairs of observations within cluster k for a given cluster. Let $T_K = \sum_{k=1}^{K} W(C_k)$ represent the total within-cluster variation for the clustering with K total clusters. For a particular clustering method, let K vary over a range of values (e.g., $K \in \{1, \ldots, 10\}$). For each K, compute T_K and plot it against K. The value of K where the "elbow" or "knee" of the curve occurs is considered the optimal number of clusters, as clustering with an additional cluster no longer results in a significant reduction in the total within-cluster variation [43].

The silhouette method is defined as follows. Let $s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$ be the silhouette for observation i, where a_i represents the average dissimilarity between observation i and the other observations in the same cluster, and b_i represents the average dissimilarity between observation i and the other observations in the next closest cluster. Note that observations for which $s_i \approx 1$ are well clustered, while observations for which $s_i < 0$ are not well-clustered. Let $S_K = \frac{1}{n} \sum_{i=1}^{n} s_i$ represent the average silhouette for the clustering with K total clusters. For a particular clustering method, let K vary over a range of values (e.g., $K \in \{1, \ldots, 10\}$). For each K, compute S_K and plot it against K. The value of K that maximizes S_K is considered the optimal number of clusters [44].

The gap statistic method is defined as follows. Cluster the data at a varying number of total clusters (e.g., $K \in \{1, \ldots, 10\}$). Let T_K represent the total within-cluster sum of squared distances for the clustering with K total clusters. Generate B reference datasets of size n, with the simulated values of variable j uniformly generated over the range of the observed x_j. For each reference dataset $b = 1, \ldots, B$, perform the clustering for each K and compute the total within-cluster sum of squared distances $T_K^{(b)}$. Compute the gap statistic:

$$\text{Gap}(K) = \left(\frac{1}{B} \sum_{b=1}^{B} \log(T_K^{(b)}) \right) - \log(T_K),$$

for each value of K, which in essence detects whether the data clustered into K groups is significantly better than if they were generated at random. The value of K that maximizes $\text{Gap}(K)$ is considered the optimal number of clusters [45].

3.2.3 Agglomerative Clustering

Agglomerative clustering is a bottom-up approach that is particularly effective when it comes to identifying small clusters. The basic algorithm is defined as follows. Initially, each observation starts as its own cluster (i.e., leaves). At each step of the algorithm, the two clusters that are most similar are merged to form a larger cluster (i.e., nodes). This process of merging clusters is repeated until all observations are members of one large cluster (i.e., root). Using the approaches described in Section 3.2.2, one can then decide on the optimal number of clusters to extract [46].

Agglomerative clustering analysis is typically displayed visually using a dendrogram, which is a diagram of the resulting tree. The amount of clustering structure found in a given dataset is reflected by the agglomerative coefficient (AC). Specifically, the AC measures the dissimilarity of an object to the first cluster it joins, divided by the dissimilarity of the final merger in the cluster analysis, averaged across all samples. Values closer to one reflect a more balanced clustering structure, while values closer to zero reflect less well-formed clusters [40, 47, 48].

There are several different agglomeration methods. The complete (or maximum) linkage method specifies that the distance between two clusters is the largest dissimilarity between any observation in the first cluster and any observation in the second cluster. The single linkage method specifies that the distance between two clusters is the smallest dissimilarity between any observation in the first cluster and any observation in the second cluster. The average linkage method specifies that the distance between two clusters is the average of the dissimilarities between observations in the first cluster and observations in the second cluster [49, 50]. Ward's method, unlike the others mentioned thus far, requires using information about the variance of observations within clusters, rather than the distance between clusters. Specifically, at each step, Ward's method involves joining the two clusters whose merged cluster possess the smallest within-cluster sum of squared distances [51].

3.2.4 Divisive Clustering

Divisive clustering is a top-down approach that is particularly effective when it comes to identifying large clusters. The basic algorithm is defined as follows. Initially, all observations are contained by one large cluster (i.e., root). At each step of the algorithm, the cluster with the largest diameter is selected. The diameter of a cluster is defined as the largest dissimilarity between any two of its observations. In order to divide the selected cluster, the algorithm identifies the observation with the largest average disimilarity to the other observations within the selected cluster. This observation initiates the splinter group.

In the steps to follow, the algorithm reassigns observations that are more similar to the splinter group than the old group. The result is a division of the selected cluster into two new clusters (i.e., nodes). This process is repeated until each observation represents its own cluster (i.e. leaves). Using the approaches described in Section 3.2.2, one can then decide on the optimal number of clusters to extract [52].

Divisive clustering analysis also tends to be displayed visually using a dendrogram. The amount of clustering structure found in a given dataset is represented by the divisive coefficient (DC). Specifically, the DC measures the diameter of the last cluster to which each observation belongs (before being split off as a single observation), divided by the diameter of the whole dataset, averaged across all observations. Values closer to one reflect a stronger clustering structure, while values closer to zero reflect less well-formed clusters [40, 47, 48].

3.2.5 K-Means Clustering

K-means clustering is the most widely used partitional clustering method. Its goal is to determine the assignment of observations to K clusters such that the total within-cluster variation is minimized. When this goal is achieved, observations within the same cluster are as similar as possible (i.e., high intra-class similarity) and observations from different clusters are as dissimilar as possible (i.e., low inter-class similarity). Each cluster is represented by its center (i.e., centroid), which corresponds to the mean of observations assigned to the cluster [40, 53, 54].

The basic algorithm behind K-means clustering is defined as follows. First, K observations are chosen as the initial centroids. One can determine the most appropriate value of K using the methods described in Section 3.2.2. Next, the remaining observations are assigned to the closest centroid based on a particular proximity measure. Then, once the K clusters are formed, the centroid of each cluster is recomputed. The algorithm iteratively repeats this process until the centroids do not change or any other alternative relaxed convergence criterion is met. Note that K-means clustering is a greedy algorithm, which means that it is guaranteed to be locally optimal, but not globally optimal [55].

4 County-Level Study

4.1 Data

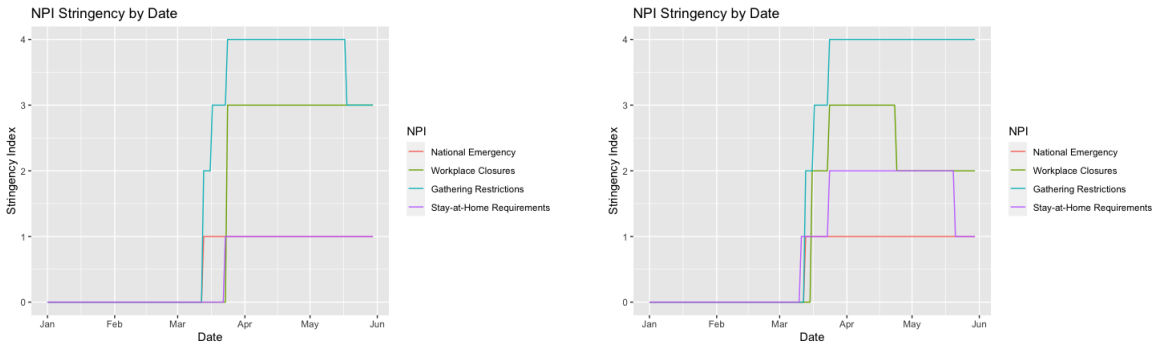
The population under observation in this study is the counties surrounding Ann Arbor, Michigan (MI), and Boston, Massachusetts (MA), during the time period from January 1st, 2020, through May 30th, 2020. The final data set used in our statistical analysis is the aggregate of data from multiple unique sources. In the following subsections, we describe each data source in detail.

4.1.1 Non-Pharmaceutical Intervention Data

The Oxford COVID-19 Government Response Tracker (OxCGRT) collected systematic information on the various policy measures implemented by government leaders to control the spread of COVID-19. For each US state, OxCGRT used publicly available sources such as news articles and government press releases to record data on indicators of government response to the outbreak of COVID-19 dating back to the beginning of 2020 [56]. In our analysis, we focus on four of the NPIs recorded by OxCGRT: national emergency, workplace closures, stay-at-home requirements, and gathering restrictions. Note that each of these NPIs is represented as a categorical variable with multiple stringency levels. Table 1 provides an in-depth description of each NPI variable. Figure 2 displays how the intensity of each NPI variable changed over time within each state. Both state governments enacted NPIs around the same time period (mid-March), yet the stringency level of each NPI varied across states. The MI government was more active in terms of altering the stringency of each NPI.

Policy	Description	Stringency Index
National Emergency	Record national emergency status	0 - not effective 1 - effective
Workplace Closures	Record closing of workplaces	0 - no measures 1 - recommend closing 2 - require closing (some sectors) 3 - require closing (all but essential)
Stay-at-Home Reqs.	Record orders to "shelter in place"	0 - no measures 1 - recommend not leaving home 2 - require not leaving home (with exceptions) 3 - require not leaving home (minimal exceptions)
Gathering Restrictions	Record limits on gatherings	0 - no measures 1 - restrict gatherings $>$ 1000 people 2 - restrict gatherings 101-1000 people 3 - restrict gatherings 11-100 people 4 - restrict gatherings $\leq$ 10 people

Table 1: Non-Pharmaceutical Intervention Data Features



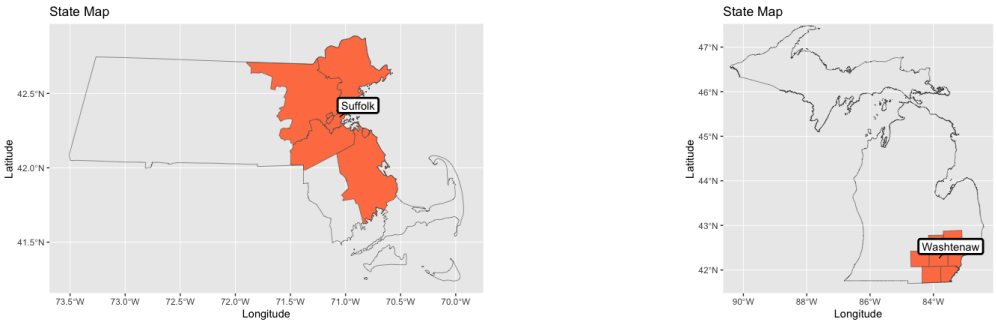
(a) Massachusetts (b) Michigan

Figure 2: Non-Pharmaceutical Intervention Stringency by Date

4.1.2 Point-of-Sale Transaction Data

BDEX, a popular data exchange platform, was established within the past decade in an effort to empower business-to-consumer companies to utilize the power of data to understand consumer behavior and intentions [57]. This platform privately made available point-of-sale (POS) credit card transactions carried out by two random samples of 2,000 individuals from Ann Arbor, MI, and Boston, MA, during the time period from January 1st, 2020, through May 30th, 2020. Each observation within this collection of data represents an individual POS transaction, including a unique individual identifier of the transactor along with the date, time, type of business (i.e., grocery, convenience, restaurant, health, hotel) and location of the transaction. Grocery encompasses businesses that retail food and household supplies (e.g., markets). Convenience includes smaller retail stores that sell everyday items (e.g., mini markets, quick stops). Restaurant encompasses businesses that prepare and sell food and drinks (e.g., fast food chains, pizzerias). Health includes businesses that promote physical and/or mental well-being (e.g., counseling centers, rehabilitation facilities, dental clinics, fitness centers, physicians offices). Hotel encompasses businesses that offer lodging to travelers (e.g., motels, inns, bed and breakfasts).

Given the overwhelmingly large size of the original collection of data, we chose to aggregate observations by the date and county of purchase, creating separate datasets for each sample (MA and MI). Note that we restricted our analysis to the county of the city where each sample was drawn from and bordering counties to avoid introducing a major spatial component to the dataset. Boston, MA, is located within Suffolk County, which is bordered by the following four counties: Essex, Middlesex, Norfolk, and Plymouth. Ann Arbor, MI, is located within Washtenaw County, which is bordered by the following seven counties: Monroe, Wayne, Oakland, Livingston, Jackson, Lenawee, and Monroe. Figure 3 displays these two regions of interest on state maps. Table 2 provides the average daily POS transaction rate within each county before versus during the COVID-19 national emergency along with the percent change between the two time periods.



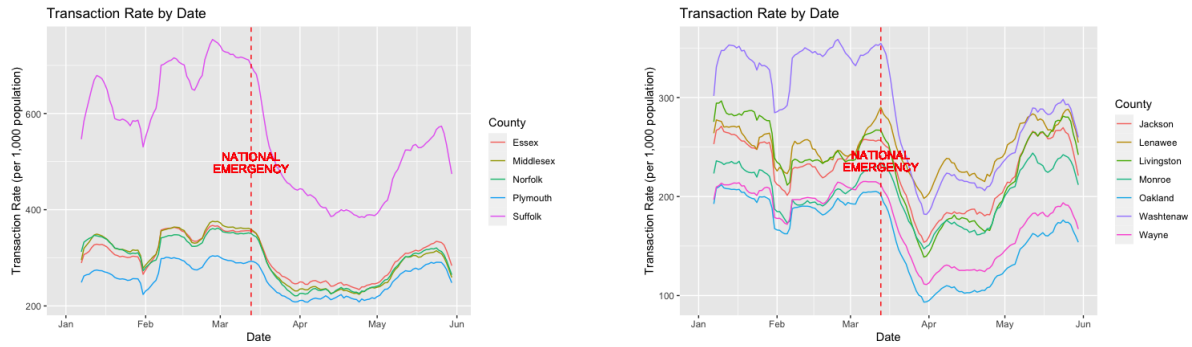
(a) Massachusetts (b) Michigan

Figure 3: State Maps with Counties of Interest

County (MA)	Avg. Daily Transaction Rate			County (MI)	Avg. Daily Transaction Rate		
	Before	*During*	*% Change*		*Before*	*During*	*% Change*
Suffolk	657	462	-29.72	Oakland	194	132	-32.02
Middlesex	339	260	-23.25	Washtenaw	336	244	-27.53
Norfolk	332	261	-21.51	Wayne	203	151	-25.67
Essex	331	274	-17.34	Livingston	257	209	-18.80
Plymouth	276	239	-13.51	Jackson	242	212	-12.57
				Monroe	212	196	-07.82
				Lenawee	255	244	-04.27

Table 2: Average Daily Transaction Rate (per 1,000 population) by County

Figure 4 displays the POS transaction rate per 1,000 population (i.e., $\frac{1000 \times transactions}{population}$) by county over time. Note that both plots display the transaction rate as a 7-day moving average in order to smooth out short-term fluctuations caused by weekly spending patterns. Looking within states, both resident counties (Suffolk, MA, and Washtenaw, MI) exhibited the highest transaction rate. Looking across states, the daily transaction rate was higher among MA counties compared to MI counties. Counties from both states exhibited a sharp decline in the daily transaction rate following the declaration of COVID-19 as a national emergency, followed by a slow and steady increase approximately 1.5-2 months later.

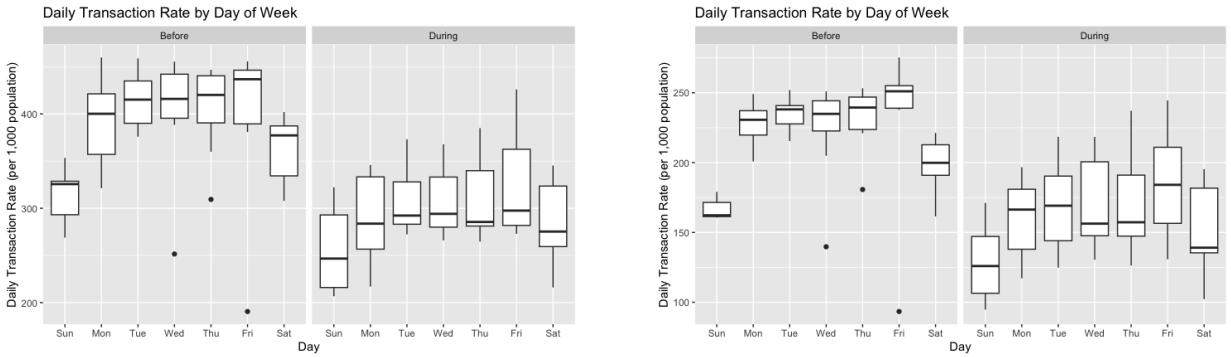


(a) Massachusetts

(b) Michigan

Figure 4: Transaction Rate (per 1,000 population) by Date

Figure 5 displays the distribution of the daily POS transaction rate per 1,000 population for each sample by the day of the week before versus during the declaration of COVID-19 as a national emergency. Both plots suggest that the daily transaction rate within both states exhibited a seasonal pattern, as individuals from both states made more purchases during weekdays (i.e., Monday-Friday) than weekends (i.e., Saturday-Sunday) both before and during the national emergency. However, this seasonal pattern appears to become less pronounced among both states following the national emergency announcement.

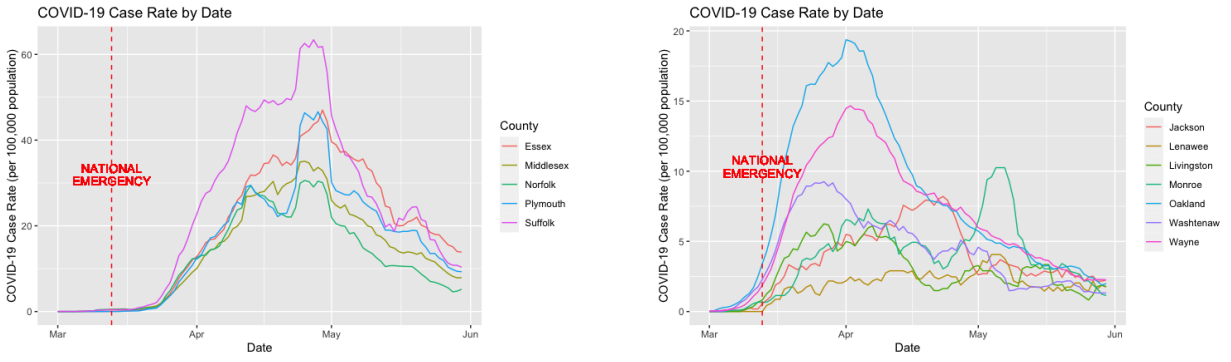


(a) Massachusetts (b) Michigan

Figure 5: Daily Transaction Rate (per 1,000 population) by Day of Week

4.1.3 COVID-19 Case & Testing Data

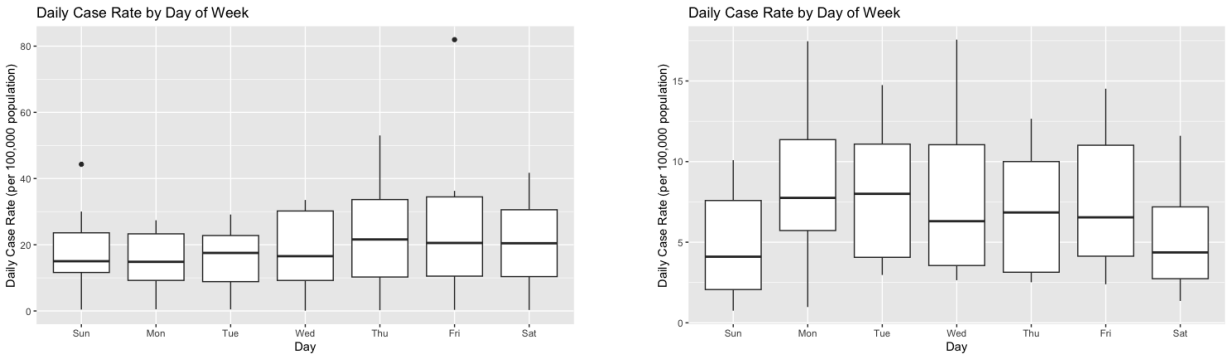
Both the MA and MI state governments kept well-organized COVID-19 databases dating all the way back to the beginning of the pandemic in early 2020 [58, 59]. These two databases contain daily and cumulative confirmed case counts at the county-level. Figure 6 displays the COVID-19 case rate (per 100,000 population) within each state by county over time. Note that both plots display case rates as 7-day moving averages in order to smooth out short-term fluctuations caused by weekly patterns. Looking within states, most counties experienced one prominent spike in case rate around the same time frame, with the exception of Monroe County, MI, which endured two distinct spikes. Looking across states, MI counties were most affected by the outbreak of COVID-19 during late March through early April, while MA counties experienced a bit more pronounced, yet delayed COVID-19 outbreak during the initial phase of the pandemic.



(a) Massachusetts (b) Michigan

Figure 6: COVID-19 Case Rate (per 100,000 population) by Date

Figure 7 displays the distribution of the daily COVID-19 case rate (per 100,000 population) for each state by day of the week during the period in which COVID-19 was considered a national emergency. In MA, the daily COVID-19 case rate did not exhibit a major seasonal pattern, as the distribution is fairly similar across all seven days of the week. However, in MI, the daily COVID-19 case rate distributions appear slightly lower on weekends compared to weekdays, perhaps suggesting the presence of a minor seasonal pattern.



(a) Massachusetts (b) Michigan

Figure 7: Daily COVID-19 Case Rate (per 100,000 population) by Day of Week

Table 3 displays the average daily COVID-19 case rate (per 100,000 population) within each county during the period in which COVID-19 was declared a national emergency (March 13^{th} - May 30^{th}, 2020). Note that the average daily case rate in every MA county was larger than that of every MI county, again suggesting that MA was more affected by the initial outbreak of COVID-19.

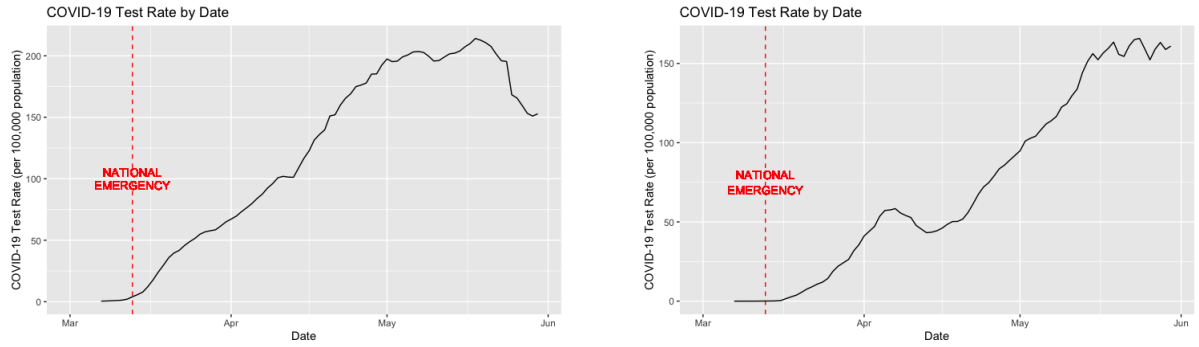
County (MA)	**Avg. Case Rate**
Suffolk	28.11
Essex	22.62
Plymouth	18.99
Middlesex	16.55
Norfolk	14.31

County (MI)	**Avg. Case Rate**
Oakland	8.86
Wayne	7.39
Washtenaw	4.57
Monroe	4.57
Jackson	4.24
Livingston	3.27
Lenawee	2.21

Table 3: Average Daily COVID-19 Case Rate (per 100,000 population) by County

Each state database also contains the daily number of total COVID-19 tests administered, yet at the state-level. Figure 8 displays the COVID-19 testing rate (per 100,000 population) within each state over time. Note that both plots display the testing rate as

a 7-day moving average in order to smooth out short-term fluctuations caused by weekly patterns. Within both states, COVID-19 testing steadily increased during the first three months of the pandemic, with the exception of a short period of decline during April in MI.

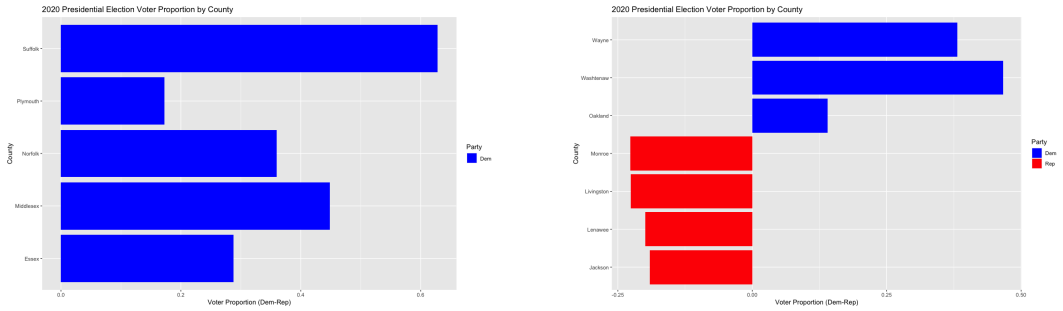


(a) Massachusetts (b) Michigan

Figure 8: COVID-19 Test Rate (per 100,000 population) by Date

4.1.4 Political and Demographic Data

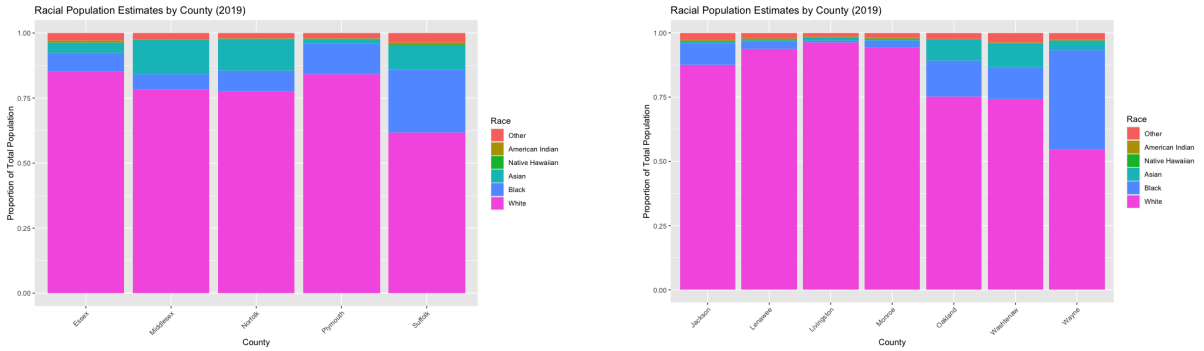
Our study explores how the effectiveness of the four NPIs described previously varies by the political affiliation and demographic makeup of each county. The MIT Election Data & Science Lab maintains a repository of publicly available data sets regarding various US elections. One of those data sets is a collection of county-level results for presidential elections from 1976 to 2020 [60]. We chose to focus on the results of the 2020 election, specifically the difference in the proportion of votes for the Democratic candidate, Joe Biden, and the proportion of votes for the Republican candidate, Donald Trump. Figure 9 displays the difference in voter proportion by county for each state. The MA counties clearly favored the Democratic candidate, while the MI counties exhibited a more even split.



(a) Massachusetts (b) Michigan

Figure 9: 2020 Presidential Election Voter Proportion by County

County-level demographic data is also made publicly available through the US Census Bureau. Our study makes use of 2019 estimates of the resident population for counties in MA and MI [61]. Specifically, we consider the total population per county as well as the proportion of individuals by race. The Census Bureau organizes individuals into the following six demographic groups: White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian, and Other. Figure 10 displays the proportion of individuals by race for each county within the two states. Looking across states, both MA and MI possess a few counties with large minority populations.



(a) Massachusetts (b) Michigan

Figure 10: Racial Population Estimates by County (2019)

Table 4 provides numerical summaries of the features from each of the data sources described above during the time period just prior to the beginning of the official declaration of COVID-19 as a national emergency (January 1st - March 12th, 2020) as well as the time period immediately after (March 13th - May 30th, 2020). Note that qualitative variables are summarized by frequency tables and quantitative tables are summarized by mean and standard deviation values. There are a few noteworthy observations found within Table 4. First, all of the NPIs included in the dataset were implemented at various stringency levels during the COVID-19 national emergency. Second, the average daily total POS transaction rate was quite lower during the national emergency in comparison to the period prior within both states. However, certain types of transactions such as grocery and convenience store-related purchases did not exhibit as pronounced of a decline during the national emergency in comparison to restaurant and health related purchases. Third and finally, daily COVID-19 case and testing rates following the declaration of COVID-19 as a national emergency suggest that the initial virus outbreak was more intense in MA than MI.

Variable (MA)	Before ($n = 360$)	During ($n = 395$)	Total ($n = 755$)	Variable (MI)	Before ($n = 504$)	During ($n = 553$)	Total ($n = 1057$)
NPIs				**NPIs**			
Workplace Closures				Workplace Closures			
-None	360 (100%)	55 (13.9%)	415 (55.0%)	*-None*	504 (100%)	21 (3.8%)	525 (49.7%)
-Recommended	0 (0%)	0 (0%)	0 (0%)	*-Recommended*	0 (0%)	0 (0%)	0 (0%)
-Required (some)	0 (0%)	0 (0%)	0 (0%)	*-Required (some)*	0 (0%)	315 (57.0%)	315 (29.8%)
-Required (all)	0 (0%)	340 (86.1%)	340 (45.0%)	*-Required (all)*	0 (0%)	217 (39.2%)	217 (20.5%)
Stay-at-Home Reqs.				Stay-at-Home Reqs.			
-None	360 (100%)	50 (12.7%)	410 (54.3%)	*-None*	490 (97.2%)	0 (0%)	490 (46.4%)
-Recommended	0 (0%)	345 (87.3%)	345 (45.7%)	*-Recommended*	14 (2.8%)	147 (26.6%)	161 (15.2%)
-Required (some ex.)	0 (0%)	0 (0%)	0 (0%)	*-Required (some ex.)*	0 (0%)	406 (73.4%)	406 (38.4%)
-Required (min ex.)	0 (0%)	0 (0%)	0 (0%)	*-Required (min ex.)*	0 (0%)	0 (0%)	0 (0%)
Gathering Restrictions				Gathering Restrictions			
-None	360 (100%)	0 (0%)	360 (47.7%)	*-None*	504 (100%)	0 (0%)	504 (47.7%)
-Restrict > 1,000 people	0 (0%)	0 (0%)	0 (0%)	*-Restrict > 1,000 people*	0 (0%)	0 (0%)	0 (0%)
-Restrict 101-1,000 people	0 (0%)	20 (5.1%)	20 (2.6%)	*-Restrict 101-1,000 people*	0 (0%)	28 (5.1%)	28 (2.6%)
-Restrict 11-100 people	0 (0%)	100 (25.3%)	100 (13.2%)	*-Restrict 11-100 people*	0 (0%)	49 (8.9%)	49 (4.6%)
-Restrict < 10 people	0 (0%)	275 (69.6%)	275 (36.4%)	*-Restrict < 10 people*	0 (0%)	476 (86.1%)	476 (45.0%)
Daily Transaction Rates (per 1,000 population)				**Daily Transaction Rates (per 1,000 population)**			
Total	387 ± 150	299 ± 96.5	341 ± 133	Total	243 ± 59.9	198 ± 57.1	219 ± 62.6
Grocery	6.70 ± 3.50	6.42 ± 3.68	6.56 ± 3.59	Grocery	6.06 ± 2.59	6.08 ± 2.87	6.07 ± 2.74
Convenience	7.16 ± 2.45	5.96 ± 2.02	6.53 ± 2.31	Convenience	5.37 ± 3.14	4.63 ± 2.96	4.98 ± 3.07
Restaurants	38.5 ± 19.0	24.3 ± 9.56	31.1 ± 16.4	Restaurants	31.30 ± 7.57	22.0 ± 6.76	26.4 ± 8.52
Health	38.2 ± 16.9	26.7 ± 8.67	32.2 ± 14.4	Health	29.30 ± 11.1	22.6 ± 8.63	25.8 ± 10.4
Hotels	15.1 ± 10.8	9.85 ± 4.81	12.3 ± 8.65	Hotels	6.51 ± 2.07	5.14 ± 1.65	5.79 ± 1.99
Daily COVID-19 Rates (per 100,000 population)				**Daily COVID-19 Rates (per 100,000 population)**			
Case	0.0265 ± 0.154	20.1 ± 16.4	10.5 ± 15.5	Case	0.108 ± 0.478	5.00 ± 4.21	2.67 ± 3.92
Test	0.225 ± 0.837	137 ± 77.5	71.6 ± 88.2	Test	0.000406 ± 0.0295	85.1 ± 59.3	44.5 ± 60.4

Table 4: NPIs, Transaction & COVID-19 Rates Before vs. During National Emergency

4.2 Modeling Approach

In this county-level study, we employ a two-step modeling approach centered around the use of mixed effects models with an autoregressive covariance structure (for the mathematical details underlying these types of models, see Section 3.1). First, we fit negative binomial mixed effects models to investigate the relationship between government-mandated NPIs and consumer mobility patterns, as indicated by daily point-of-sale (POS) credit card transaction data. Second, we fit linear mixed effects models to explore the relationship between changes in consumer mobility patterns compared to pre-pandemic baselines and the initial spread of COVID-19. In the following subsections, we describe each step in greater detail.

4.2.1 Mobility Modeling

Given the nature of our data (i.e., daily POS transaction counts clustered at the county-level), we fit negative binomial mixed effects models with an autoregressive covariance structure to investigate the relationship between NPIs and daily POS transaction rates. The explanatory variable NPI referenced in each model represents the different NPIs included in our data set (i.e., $NPI \in \{$national emergency, workplace closures, stay-at-home requirements, gathering restrictions$\}$). Each model also contains the variable *population* as an offset term, making our response variable the daily POS transaction rate (per 1,000 population) for observation i from county j. In each model, we adjust for seasonal patterns induced by the day of the week (see variable *weekend*, which takes on a value of zero on days Monday - Friday and a value of one on days Saturday - Sunday).

The first model,

$$\log\left(\frac{1000 \times transactions_{ij}}{population_{ij}}\right) = \beta_0 + \beta NPI_{ij} + \beta weekend_{ij} + \gamma_{0j}, \quad (1)$$

describes how daily POS transaction rates (per 1,000 population) change with the implementation of each NPI. Note that each NPI is represented as a categorical variable with multiple levels; therefore, fitting this model with each different NPI variable results in multiple β coefficients due to one hot encoding.

The second model,

$$\log(\frac{1000 \times transactions_{ij}}{population_{ij}}) = \beta_0 + \beta_1 NPI_{ij} + \beta_2 political_{ij} + \beta_3 NPI_{ij} political_{ij} + \beta_4 weekend_{ij} + \gamma_{0j}, \tag{2}$$

in which *political* represents the difference in the proportion of individuals who voted for the Democratic candidate and the Republican candidate in the 2020 US presidential election, describes how the association between daily POS transaction rates (per 1,000 population) and the implementation of each NPI varies by the political affiliation of the county. Note that in fitting this model one hot encoding is no longer performed, as doing so would require fitting an abundance of parameters and lead to less interpretable results. Instead, each NPI is represented as a continuous variable for which we are only concerned with one unit increases.

The third model,

$$\log(\frac{1000 \times transactions_{ij}}{population_{ij}}) = \beta_0 + \beta_1 NPI_{ij} + \beta_2 minority_{ij} + \beta_3 NPI_{ij} minority_{ij} + \beta_4 weekend_{ij} + \gamma_{0j}, \tag{3}$$

in which *minority* represents the minority population proportion, describes how the association between daily POS transaction rates (per 1,000 population) and the implementation of each NPI varies by the racial makeup of the county. Again, note that in fitting this model one hot encoding is not performed, as doing so would require fitting an abundance of parameters and lead to less interpretable results. Instead, each NPI is represented as a continuous variable for which we are only concerned with one unit increases.

The fourth and final model,

$$\log(\frac{1000 \times transactions_{ij}}{population_{ij}}) = \beta_0 + \beta_1 NPI_{ij} + \beta_2 state_{ij} + \beta_3 NPI_{ij} state_{ij} + \beta_4 weekend_{ij} + \gamma_{0j}, \tag{4}$$

is fit using data from both states, MA and MI, unlike the previous three models. This model, in which *state* represents whether the observation is a county within MA or MI, describes how the association between daily POS transaction rates (per 1,000 population) and the implementation of each NPI varies by state. Again, note that in fitting this model each NPI is represented as a continuous variable for which we are only concerned with one unit increases, as one hot encoding leads to an abundance of parameters and less interpretable results.

4.2.2 COVID-19 Modeling

In order to investigate the relationship between changes in consumer mobility patterns and the initial spread of COVID-19, we take a very similar approach to the one described in the previous section. We continue to use mixed-effects modeling, as our data is clustered at the county-level. However, linear models are now appropriate, as our new response variable is not a daily count. Instead, we model the daily exponential growth rate in confirmed COVID-19 cases, *case growth rate* (i.e., the natural logarithm of daily COVID-19 cases minus the logarithm of daily COVID-19 cases on the prior day, multiplied by 100 to be interpretable as percent). Note that we add one to each daily case count to prevent the natural logarithm of cases from being undefined when there are no cases in a county on a given day. Given the time-series nature of the data, we continue to fit these models with an autoregressive covariance structure.

The explanatory variable of interest in each COVID-19 model is changes in consumer mobility patterns. As indicated throughout our literature review, substantial efforts have been made to create data sets that reflect human mobility change since the outbreak of the COVID-19 pandemic, where starting points between January 2020 and February 2020 have been selected to provide a baseline. In this study, we attempt to do the same by creating a variable $\Delta transactions$, which represents the percentage of change between each county's daily POS transaction count and their pre-pandemic baseline day. Note that county pre-pandemic baseline days are median values of the daily POS transaction count for each day of the week during the period from January 1$^{\text{st}}$ - February 29$^{\text{th}}$, 2020.

A major challenge in modeling the relationship between changes in mobility patterns and the outbreak of COVID-19 is determining the lag in time between mobility data and COVID-19 cases. From the time an individual first makes contact with the virus and contracts the disease, confirming the diagnosis is a complex process. There are several factors that can affect the length of time necessary for a COVID-19 case to be confirmed–incubation period, testing speed, and reporting delay to name a few. Previous studies have found that the COVID-19 incubation period can vary, yet most cases manifest themselves between 3-7 days after initial infection [62]. The lag caused by testing speed and delays in reporting is much more variable, as it is largely dependent upon the region's medical facilities. Other statistical studies have attempted to quantify the optimal lag, most notably Badr et al. (2020) who found that a lag of 11 days achieved the highest correlation between mobility and COVID-19 growth rates for a single US all-county model [17].

In this study, we experiment with models that include mobility data lags ranging anywhere from 6-16 days. In each model, we control for the state-level COVID-19 testing rate and seasonal patterns induced by the day of the week (see variables $test\,rate$ and $weekend$). More formally, each model is defined mathematically as follows:

$$case\ growth\ rate_{ij} = \beta_0 + \beta_1 \Delta transactions_{ij} + \beta_2 test\ rate_{ij} + \beta_3 weekend_{ij} + \gamma_{0j}, \quad (5)$$

in which i represents the date of an observation and j represents the county. These models describe the association between lagged changes in POS transactions compared to pre-pandemic baselines and daily COVID-19 case growth rates.

4.3 Results

4.3.1 Mobility Modeling

Using the *glmmTMB* package in R [63, 64], we fit each of the negative binomial mixed effect models with an autoregressive covariance structure, as described in Section 4.2.1. In estimating the likelihood function of each model, we used Laplace Approximation with a non-linear optimizer to obtain best-fit model parameters [65]. Note that we fit the first three models on two separate subsets of the dataset by state (i.e., MA and MI) and the fourth and final model using the entire dataset.

First, we fit the class of models described by Equation 1 in Section 4.2.1:

$$\log(\frac{1000 \times transactions_{ij}}{population_{ij}}) = \beta_0 + \beta NPI_{ij} + \beta weekend_{ij} + \gamma_{0j}.$$

Tables 5 and 6 display the resulting fixed effect coefficients from these models for both MA and MI.

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	353.28	(279.49,446.56)	0.00***
	Effective	0.86	(0.80,0.93)	0.00***
	Weekend	0.87	(0.86,0.88)	0.00***
Workplace Closure	Intercept	345.31	(273.19,436.47)	0.00***
	Recommended	NA	NA	NA
	Required (some)	NA	NA	NA
	Required (all)	0.89	(0.82,0.96)	0.00**
	Weekend	0.87	(0.85,0.88)	0.00***
Stay-at-Home Reqs.	Intercept	347.44	(274.89,439.13)	0.00***
	Recommended	0.88	(0.81,0.95)	0.00**
	Required (some ex.)	NA	NA	NA
	Required (min ex.)	NA	NA	NA
	Weekend	0.87	(0.85,0.88)	0.00***
Gathering Restrictions	Intercept	371.96	(295.58,469.66)	0.00***
	Restrict > 1,000	NA	NA	NA
	Restrict 101-1,000	0.92	(0.86,0.99)	0.03*
	Restrict 11-100	0.80	(0.74,0.86)	0.00***
	Restrict < 10	0.77	(0.71,0.84)	0.00***
	Weekend	0.87	(0.85,0.88)	0.00***

Table 5: Model 1 Results (MA)

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	237.06	(205.53,273.43)	0.00***
	Effective	0.89	(0.83,0.96)	0.00**
	Weekend	0.82	(0.81,0.83)	0.00***
Workplace Closure	Intercept	254.76	(221.02,293.65)	0.00***
	Recommended	NA	NA	NA
	Required (some)	0.84	(0.79,0.90)	0.00***
	Required (all)	0.68	(0.64,0.73)	0.00***
	Weekend	0.81	(0.80,0.82)	0.00***
Stay-at-Home Reqs.	Intercept	239.41	(207.41,276.35)	0.00***
	Recommended	0.97	(0.90,1.03)	0.32
	Required (some ex.)	0.82	(0.76,0.89)	0.00***
	Required (min ex.)	NA	NA	NA
	Weekend	0.82	(0.80,0.83)	0.00***
Gathering Restrictions	Intercept	284.98	(242.80,334.49)	0.00***
	Restrict > 1,000	NA	NA	NA
	Restrict 101-1,000	0.94	(0.88,1.01)	0.08·
	Restrict 11-100	0.79	(0.72,0.86)	0.00***
	Restrict < 10	0.64	(0.57,0.71)	0.00***
	Weekend	0.82	(0.80,0.83)	0.00***

Table 6: Model 1 Results (MI)

The exponentiation of the intercept term in each model represents the daily POS transaction rate (per 1,000 population) without any NPI on a weekday within the average county.

Within each state, this estimate is very similar across the various NPIs, which makes sense considering that each NPI was instituted around the same date. However, looking across states, it is evident that the daily POS transaction rate was higher in the average MA county compared to the average MI county. The exponentiation of the NPI terms included in each model represents the multiplicative effect of instituting each NPI, controlling for the day of the week. Note that we were unable to obtain estimates for a few stringency levels of the various NPIs due to a lack of data. Almost every exponentiated NPI term is less than one and statistically significant at the $\alpha = 0.05$ level, indicating that each NPI was associated with a decrease in daily POS transaction rate in both states on average. For each NPI, as the stringency level increases, the coefficient estimate decreases, indicating that the counties within both states experienced less daily POS transactions on average as closures, requirements, and restrictions became more stringent.

Next, we fit the class of models described by Equation 2 in Section 4.2.1:

$$\log(\frac{1000 \times transactions_{ij}}{population_{ij}}) = \beta_0 + \beta_1 NPI_{ij} + \beta_2 political_{ij} + \beta_3 NPI_{ij} political_{ij} + \beta_4 weekend_{ij} + \gamma_{0j}.$$

Recall that the variable *political* represents the difference in the proportion of individuals who voted for the Democratic candidate and the Republican candidate in the 2020 US presidential election. Also recall that in fitting this model we did not perform one hot encoding, as doing so would require fitting an abundance of parameters and lead to less interpretable results. Instead, each NPI is represented as a continuous variable for which we are only concerned with one unit increases. Tables 7 and 8 display the resulting fixed effect coefficients from these models for both MA and MI.

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	193.92	(131.34,286.31)	0.00****
	Emergency	0.97	(0.82,1.15)	0.73
	Political	7.41	(2.85,19.24)	0.00***
	Emergency × Political	0.72	(0.47,1.10)	0.13
	Weekend	0.87	(0.86,0.88)	0.00***
Workplace Closure	Intercept	203.09	(137.06,300.93)	0.00***
	Closure	0.94	(0.78,1.13)	0.50
	Political	5.70	(2.13,15.25)	0.00***
	Closure × Political	0.85	(0.53,1.36)	0.49
	Weekend	0.87	(0.85,0.88)	0.00***
Stay-at-Home Reqs.	Intercept	208.44	(140.87,308.44)	0.00***
	Requirements	0.92	(0.77,1.11)	0.38
	Political	5.52	(2.06,14.76)	0.00****
	Reqs. × Political	0.87	(0.55,1.38)	0.55
	Weekend	0.87	(0.85,0.88)	0.00***
Gathering Restrictions	Intercept	199.49	(144.48,275.45)	0.00***
	Restrictions	0.97	(0.91,1.03)	0.35
	Political	6.34	(2.88,13.95)	0.00***
	Restrict × Political	0.87	(0.74,1.01)	0.06·
	Weekend	0.87	(0.85,0.88)	0.00***

Table 7: Model 2 Results (MA)

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	265.92	(223.63,316.20)	0.00***
	Emergency	0.89	(0.83,0.96)	0.00**
	Political	1.25	(0.68,2.27)	0.47
	Emergency × Political	0.81	(0.65,1.03)	0.08·
	Weekend	0.82	(0.81,0.83)	0.00***
Workplace Closure	Intercept	311.65	(268.38,361.90)	0.00***
	Closure	0.83	(0.80,0.85)	0.00***
	Political	1.04	(0.62,1.77)	0.87
	Closure × Political	0.92	(0.82,1.04)	0.19
	Weekend	0.81	(0.80,0.82)	0.00***
Stay-at-Home Reqs.	Intercept	274.73	(235.69,320.23)	0.00***
	Requirements	0.90	(0.86,0.93)	0.00***
	Political	1.04	(0.60,1.79)	0.89
	Reqs. × Political	0.93	(0.82,1.07)	0.32
	Weekend	0.82	(0.80,0.83)	0.00***
Gathering Restrictions	Intercept	328.19	(276.71,389.25)	0.00***
	Restrictions	0.86	(0.83,0.90)	0.00***
	Political	1.07	(0.60,1.91)	0.82
	Restrict × Political	0.93	(0.83,1.04)	0.20
	Weekend	0.82	(0.80,0.83)	0.00***

Table 8: Model 2 Results (MI)

The exponentiation of the intercept term in each model represents the daily POS transaction rate (per 1,000 population) without any NPI on a weekday within the average county

in which the proportion of democratic and republican voters in the 2020 US election was equal. The exponentiation of the $NPI \times political$ term in each model represents the multiplicative effect of the interaction between these two variables. Note that both variables are included as main effects in each model as well in accordance with the hierarchy principle. The coefficient estimate of the interaction term in each model is not statistically significant at the $\alpha = 0.05$ level, suggesting that the association between each NPI and daily POS transaction rates does not differ by the political affiliation of the county.

Next, we fit the class of models described by Equation 3 in Section 4.2.1:

$$\log(\frac{1000 \times transactions_{ij}}{population_{ij}}) = \beta_0 + \beta_1 NPI_{ij} + \beta_2 minority_{ij} + \beta_3 NPI_{ij} minority_{ij} + \beta_4 weekend_{ij} + \gamma_{0j}.$$

Recall that the variable *minority* represents the minority population proportion. Also recall that in fitting this model we did not perform one hot encoding, as doing so would require fitting an abundance of parameters and lead to less interpretable results. Instead, each NPI is represented as a continuous variable for which we are only concerned with one unit increases. Tables 9 and 10 display the resulting fixed effect coefficients from these models for both MA and MI.

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	175.27	(121.62,252.59)	0.00***
	Emergency	0.97	(0.81,1.17)	0.79
	Minority	44.36	(9.70,202.79)	0.00***
	Emergency × Minority	0.57	(0.26,1.24)	0.16
	Weekend	0.87	(0.86,0.88)	0.00***
Workplace Closure	Intercept	184.67	(127.61,267.25)	0.00***
	Closure	0.94	(0.77,1.15)	0.53
	Minority	27.84	(5.76,134.64)	0.00***
	Closure × Minority	0.77	(0.32,1.81)	0.54
	Weekend	0.87	(0.85,0.88)	0.00***
Stay-at-Home Reqs.	Intercept	188.00	(130.20,271.46)	0.00***
	Requirements	0.93	(0.76,1.13)	0.46
	Minority	27.46	(5.73,131.75)	0.00***
	Reqs. × Minority	0.77	(0.33,1.81)	0.55
	Weekend	0.87	(0.85,0.88)	0.00***
Gathering Restrictions	Intercept	180.90	(136.90,239.05)	0.00***
	Restrictions	0.97	(0.91,1.04)	0.41
	Minority	33.89	(10.66,107.72)	0.00***
	Restrict × Minority	0.78	(0.59,1.04)	0.09·
	Weekend	0.87	(0.85,0.88)	0.00***

Table 9: Model 3 Results (MA)

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	270.51	(209.80,348.81)	0.00***
	Emergency	0.95	(0.85,1.05)	0.33
	Minority	0.93	(0.30,2.84)	0.89
	Emergency × Minority	0.70	(0.44,1.12)	0.14
	Weekend	0.82	(0.81,0.83)	0.00***
Workplace Closure	Intercept	333.46	(268.00,414.92)	0.00***
	Closure	0.85	(0.80.0.89)	0.00***
	Minority	0.69	(0.26,1.79)	0.45
	Closure × Minority	0.86	(0.68,1.09)	0.23
	Weekend	0.81	(0.80,0.82)	0.00***
Stay-at-Home Reqs.	Intercept	292.95	(233.96,366.83)	0.00***
	Requirements	0.92	(0.86,0.97)	0.00**
	Minority	0.70	(0.26,1.90)	0.48
	Reqs. × Minority	0.87	(0.67,1.13)	0.30
	Weekend	0.82	(0.80,0.83)	0.00***
Gathering Restrictions	Intercept	345.19	(268.93,443.07)	0.00***
	Restrictions	0.89	(0.84,0.94)	0.00***
	Minority	0.72	(0.25,2.14)	0.56
	Restrict × Minority	0.87	(0.69,1.10)	0.25
	Weekend	0.82	(0.80,0.83)	0.00****

Table 10: Model 3 Results (MI)

The exponentiation of the intercept term in each model represents the daily POS transaction rate (per 1,000 population) without any NPI on a weekday within the average county in which the minority population estimate was zero. The exponentiation of the $NPI \times minority$ term in each model represents the multiplicative effect of the interaction between these two variables. Note that both variables are included as main effects in each model as well in accordance with the hierarchy principle. The coefficient estimate of the interaction term in each model is not statistically significant at the $\alpha = 0.05$ level, suggesting that the association between each NPI and daily POS transaction rates does not differ by the racial makeup

of the county.

Finally, we fit the class of models described by Equation 4 in Section 4.2.1:

$$\log(\frac{1000 \times transactions_{ij}}{population_{ij}}) = \beta_0 + \beta_1 NPI_{ij} + \beta_2 state_{ij} + \beta_3 NPI_{ij} state_{ij} + \beta_4 weekend_{ij} + \gamma_{0j}.$$

Recall that the variable *state* represents whether the observation is a county within MA or MI. Also recall that in fitting this model we did not perform one hot encoding, as doing so would require fitting an abundance of parameters and lead to less interpretable results. Instead, each NPI is represented as a continuous variable for which we are only concerned with one unit increases. Table 11 displays the resulting fixed effect coefficients from each individual model.

NPI	**Fixed Effects**	e^{β}	**95% CI**	p**-value**
National Emergency	Intercept	406.69	(324.07,510.37)	0.00***
	Emergency	0.87	(0.81,0.94)	0.00***
	State (MI)	0.65	(0.49,0.88)	0.00**
	Emergency × State (MI)	1.02	(0.92,1.12)	0.71
	Weekend	0.84	(0.83,0.85)	0.00***
Workplace Closure	Intercept	409.90	(328.94,510.79)	0.00***
	Closure	0.87	(0.81,0.93)	0.00***
	State (MI)	0.73	(0.55,0.97)	0.03*
	Closure × State (MI)	0.96	(0.89,1.04)	0.37
	Weekend	0.83	(0.82,0.84)	0.00***
Stay-at-Home Reqs.	Intercept	410.42	(328.06,513.47)	0.00***
	Requirements	0.86	(0.80,0.93)	0.00***
	State (MI)	0.66	(0.50,0.88)	0.00**
	Reqs. × State (MI)	1.04	(0.96,1.13)	0.36
	Weekend	0.84	(0.83,0.85)	0.00***
Gathering Restrictions	Intercept	400.21	(325.29,492.39)	0.00***
	Restrictions	0.92	(0.89,0.95)	0.00***
	State (MI)	0.75	(0.56,1.00)	0.05·
	Restrict × State (MI)	0.97	(0.92,1.02)	0.20
	Weekend	0.84	(0.83,0.85)	0.00***

Table 11: Model 4 Results

The exponentiation of the intercept term in each model represents the daily transaction rate (per 1,000 population) without any NPI on a weekday within the average MA county. The exponentiation of the NPI main effect included in each model represents the multiplicative effect of instituting each NPI, controlling for the state of the observation and the day of the week. In each model, the coefficient associated with this term is both statistically significant at the $\alpha = 0.05$ level and less than one, indicating that each NPI is associated with a decrease in daily POS transaction rate on average within counties from both states. The exponentiation of the *state* main effect included in each model represents the multiplicative effect of an observation being a county within MI opposed to MA, controlling for the stringency level of each NPI and the day of the week. In each model, this coefficient is

also both statistically significant at the $\alpha = 0.05$ level and less than one, indicating that the daily POS transaction rate is less within MI counties than MA counties on average regardless of NPI stringency level and the day of the week. The exponentiation of the $NPI \times state$ term in each model represents the multiplicative effect of the interaction between these two variables. The coefficient estimate of this interaction term in each model is not statistically significant at the $\alpha = 0.05$ level, suggesting that the association between each NPI and daily POS transaction rates does not differ by state.

4.3.2 COVID-19 Modeling

Using the *nlme* package in R [66, 64], we fit each of the linear mixed effect models with an autoregressive covariance structure, as described in Section 4.2.2, under the restricted maximum likelihood (REML) approach. Note that we fit each model using data from during the national emergency (i.e., 13$^{\text{th}}$ March - 30$^{\text{th}}$ May, 2020) subset by state (i.e., MA and MI) and mobility lags ranging from 6-16 days:

$$case\ growth\ rate_{ij} = \beta_0 + \beta_1 \Delta transactions_{ij} + \beta_2 test\ rate_{ij} + \beta_3 weekend_{ij} + \gamma_{0j}.$$

Tables 12 and 13 display the resulting fixed effect coefficients from each model for both MA and MI.

Lag Period	Fixed Effects	β	95% CI	p-value
06 Days	Intercept	31.95	(22.66,41.24)	0.00***
	Δ Transactions	0.11	(-0.11,0.35)	0.33
	Test Rate	-11.29	(-16.22,-6.36)	0.00***
	Weekend	-33.78	(-42.97,-24.60)	0.00***
08 Days	Intercept	31.83	(22.87,40.79)	0.00***
	Δ Transactions	0.17	(-0.06,0.40)	0.15
	Test Rate	-10.57	(-15.67,-5.48)	0.00***
	Weekend	-33.10	(-42.35,-23.85)	0.00***
10 Days	Intercept	31.39	(22.61,40.17)	0.00***
	Δ Transactions	0.23	(-0.01,0.47)	0.06·
	Test Rate	-9.57	(-14.89,-4.26)	0.00***
	Weekend	-32.61	(-41.86,-23.36)	0.00***
12 Days	Intercept	30.46	(21.77,39.15)	0.00***
	Δ Transactions	0.30	(0.06,0.54)	0.02*
	Test Rate	-8.25	(-13.81,-2.68)	0.00***
	Weekend	-31.42	(-40.76,-22.07)	0.00***
14 Days	Intercept	28.47	(19.70,37.24)	0.00***
	Δ Transactions	0.39	(0.13,0.65)	0.00**
	Test Rate	-6.08	(-12.17,0.01)	0.05*
	Weekend	-29.49	(-39.06,-19.93)	0.00***
16 Days	Intercept	27.58	(18.53,36.63)	0.00***
	Δ Transactions	0.33	(0.05,0.60)	0.02*
	Test Rate	-6.38	(-12.97,0.21)	0.06·
	Weekend	-29.46	(-39.37,-19.55)	0.00***

Table 12: Model 5 Results (MA)

Lag Period	Fixed Effects	β	95% CI	p-value
06 Days	Intercept	17.08	(8.88,25.28)	0.00***
	Δ Transactions	0.13	(-0.05,0.31)	0.17
	Test Rate	-10.43	(-16.47,-4.40)	0.00***
	Weekend	-22.38	(-31.18,-13.58)	0.00***
08 Days	Intercept	18.22	(10.36,26.08)	0.00***
	Δ Transactions	0.19	(0.01,0.37)	0.03*
	Test Rate	-10.30	(-16.12,-4.47)	0.00***
	Weekend	-23.02	(-31.83,-14.21)	0.00***
10 Days	Intercept	16.54	(8.98,24.10)	0.00***
	Δ Transactions	0.14	(-0.04,0.32)	0.12
	Test Rate	-9.50	(-15.26,-3.75)	0.00***
	Weekend	-22.43	(-31.22,-13.64)	0.00***
12 Days	Intercept	15.94	(8.61,23.26)	0.00***
	Δ Transactions	0.13	(-0.05,0.31)	0.15
	Test Rate	-9.27	(-15.02,-3.52)	0.00***
	Weekend	-21.62	(-30.44,-12.80)	0.00***
14 Days	Intercept	15.32	(7.98,22.65)	0.00***
	Δ Transactions	0.09	(-0.08,0.27)	0.31
	Test Rate	-9.03	(-14.79,-3.27)	0.00***
	Weekend	-22.46	(-31.29,-13.64)	0.00***
16 Days	Intercept	15.81	(8.66,22.97)	0.00***
	Δ Transactions	0.15	(-0.03,0.33)	0.10*
	Test Rate	-8.62	(-14.39,-2.84)	0.00***
	Weekend	-22.14	(-30.92,-13.35)	0.00***

Table 13: Model 5 Results (MI)

The exponentiation of the intercept term in each model represents the daily COVID-19 case growth rate without any lagged changes in consumer mobility and COVID-19 testing on a weekday within the average county. Within each state, this estimate is very similar across

the various lag periods. However, looking across states, it is evident that the daily COVID-19 case growth rate was larger in the average MA county than the average MI county. The coefficient estimate of Δ transactions for each lag period represents the average change in the daily COVID-19 case growth rate associated with a 1% change in consumer mobility, as measured by POS transactions, compared to pre-pandemic baselines, controlling for the testing rate and the day of the week. Within MA counties, our results suggest that the strongest positive association between changes in consumer mobility and COVID-19 case growth rates appears at a lag of 14 days, as a 10% percent decrease in 14-day lagged change in consumer mobility was associated with a 3.9% decrease in the COVID-19 case growth rate. However, within MI counties, our results suggest that the strongest positive association between changes in consumer mobility and COVID-19 case growth rates appears at a lag of 8 days, as a 10% percent decrease in 8-day lagged change in consumer mobility was associated with a 1.9% decrease in the COVID-19 case growth rate. Within counties from both states, this association is not remarkably strong; however, it does appear stronger in MA counties than MI counties.

4.4 Discussion

Consumer mobility models detected the presence of a statistically significant negative association between all four NPIs of interest (i.e., national emergency, workplace closure, stay-at-home requirements, and gathering restrictions) and daily county-level POS transaction rates, adjusting for trends resulting from the time-series nature of the data. Stringent workplace closures and gathering restrictions exhibited the strongest negative association with daily POS transaction rates in both states. Specifically, required workplace closures were associated with a reduction of 11% in the daily POS transaction rate within the average MA county, and a reduction of 32% in the daily POS transaction rate within the average MI county. Similarly, restrictions on gatherings with greater than 10 people were associated with a reduction of 23% in the daily POS transaction rate within the average MA county, and a reduction of 36% in the daily POS transaction rate within the average MI county. The association between each NPI and daily POS transaction rates did not differ significantly by the political affiliation, racial makeup, or state of each county. Looking within both states, each NPI exhibited a dose-response relationship with daily POS transactions (i.e., incremental increases in the stringency of NPIs produced decreases in daily POS transactions), providing increased evidence of a causal relationship. These results support previous findings that the early implementation of a broad range of NPIs reduced human mobility.

Our consumer mobility models possess a few limitations. First, several NPIs were implemented within a short period of time during the early stages of the COVID-19 pandemic, making it difficult to isolate the impact of each individual policy intervention. Second, the dataset used in our analysis is imperfect in that it represents daily POS transactions carried out by two large random samples of residents from two cities (i.e., Boston, MA, and

Ann Arbor, MI) aggregated at the county-level. This aggregation may introduce bias to our dataset; however, we limit this source of bias by restricting our analysis to the county that contains each city and its bordering counties. Third, although our models include several control variables and their results provide some evidence for causality, the possibility for omitted variable bias certainly exists, as is always the case in observational studies. Fourth, given the inclusion criteria of our study, the number of counties included in our analysis (i.e., 12 total) is relatively small; therefore, the generalizability of our results is concerning. Ideally, a future study would include counties from more states, yet this is dependent upon the willingness of BDEX to make such data available.

COVID-19 models identified the presence of a statistically significant positive association between changes in consumer mobility patterns and daily county-level COVID-19 case growth rates, adjusting for the COVID-19 testing rate and trends resulting from the time-series nature of the data. In a similar manner to other studies, we created a feature to represent changes in consumer mobility, specifically the percent change in daily POS transaction counts compared to pre-pandemic baseline days. Varying lags of this feature exhibited a statistically significant positive association with COVID-19 case growth rates. Within the average MA county, a 10% percent decrease in 14-day lagged change in consumer mobility was associated with a 3.9% decrease in the COVID-19 case growth rate. Within the average MI county, a 10% percent decrease in 8-day lagged change in consumer mobility was associated with a 1.9% decrease in the COVID-19 case growth rate. Contrary to previous findings, who found much stronger relationships between changes in human mobility and the COVID-19 spread, these results suggest that daily POS transactions are a limited mobility metric.

The COVID-19 modeling portion of our study also has a few limitations. First, as mentioned in the previous paragraph, POS credit card transactions are a limited measure of human mobility patterns. Additionally, there exist other potential COVID-19 mitigation factors beyond mobility reduction (e.g., mask-wearing, hand-washing, case tracing) that our analysis does not consider. Second, due to a lack of data availability, our models only represent the POS transactions of 2,000 randomly sampled residents from one city (i.e., Boston, MA, and Ann Arbor, MI) within each state. Ideally, our dataset would include all POS transactions that took place within each county. Third, the COVID-19 case data used in our analysis may contain errors due to reporting issues and limited testing capacity common among many medical facilities during the initial disease outbreak. Fourth, the COVID-19 testing data used in our analysis is representative of state-level totals, as both state governments failed to record county-level test counts during the early stages of the pandemic.

5 Individual-Level Study

5.1 Data

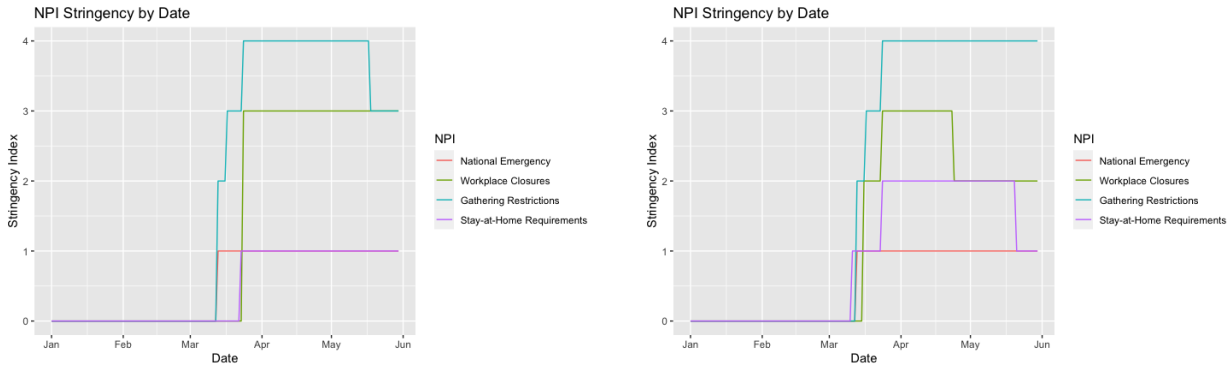
The population under observation in this study is 2,000 individuals from Boston, Massachusetts (MA), and 2,000 individuals from Ann Arbor, Michigan (MI), during the time period from January 1st, 2020, through May 30th, 2020. The final data set used in our statistical analysis is the aggregate of data from a few unique sources. In the following subsections, we describe each data source in detail.

5.1.1 Non-Pharmaceutical Intervention Data

The Oxford COVID-19 Government Response Tracker (OxCGRT) collected systematic information on the various policy measures implemented by government leaders to control the spread of COVID-19. For each US state, OxCGRT used publicly available sources such as news articles and government press releases to record data on indicators of government response to the outbreak of COVID-19 dating back to the beginning of 2020 [56]. In our analysis, we focus on four of the NPIs recorded by OxCGRT: national emergency, workplace closures, stay-at-home requirements, and gathering restrictions. Note that each of these NPIs is represented as a categorical variable with multiple stringency levels. Table 14 provides an in-depth description of each NPI variable. Figure 11 displays how the intensity of each NPI variable changed over time within each state. Both state governments instituted each NPI around the same time period (mid-March), yet the stringency level of each NPI varied across states. The MI government was more active in terms of altering the stringency of each NPI.

Policy	Description	Stringency Index
National Emergency	Record national emergency status	0 - not effective 1 - effective
Workplace Closures	Record closing of workplaces	0 - no measures 1 - recommend closing 2 - require closing (some sectors) 3 - require closing (all but essential)
Stay-at-Home Requirements	Record orders to "shelter in place"	0 - no measures 1 - recommend not leaving home 2 - require not leaving home (with exceptions) 3 - require not leaving home (minimal exceptions)
Gathering Restrictions	Record limits on gatherings	0 - no measures 1 - restrict gatherings $>$ 1000 people 2 - restrict gatherings 101-1000 people 3 - restrict gatherings 11-100 people 4 - restrict gatherings $\leq$ 10 people

Table 14: Non-Pharmaceutical Intervention Data Features



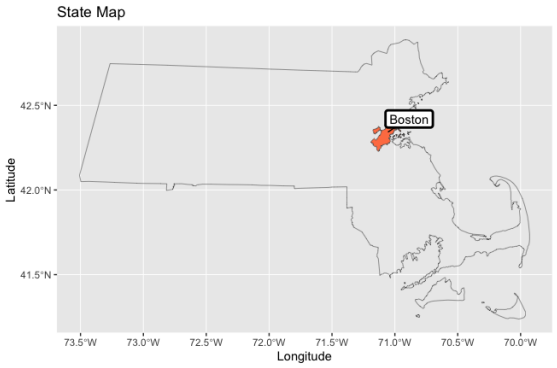
(a) Massachusetts (b) Michigan

Figure 11: Non-Pharmaceutical Intervention Stringency by Date

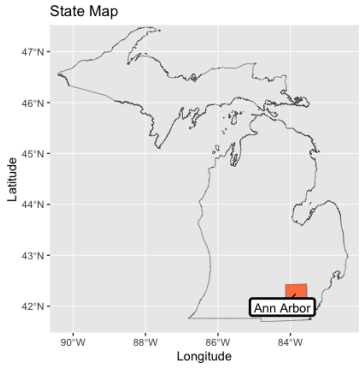
5.1.2 Point of Sale Transaction Data

BDEX, a popular data exchange platform, was established within the past decade in an effort to empower business-to-consumer companies to utilize the power of data to understand consumer behavior and intentions [57]. This platform privately made available point-of-sale (POS) transactions carried out by two large random samples of credit card accounts from Boston, MA, and Ann Arbor, MI, during the time period from January 1st, 2020, through May 30th, 2020. Figure 12 displays the geographical location of these two cities on state maps. Each observation within this collection of data represents an individual POS transaction, including a universally unique identifier of the transactor along with the date, time, type of business (i.e., grocery, convenience, restaurant, health, hotel), and location of the transaction. The *grocery* category covers businesses that retail food and household supplies (e.g., markets). The *convenience* category includes smaller retail stores that sell everyday items (e.g., mini marts, quick stops). The *restaurant* category encompasses businesses that prepare and sell food and drinks (e.g., fast food chains, pizzerias). The *health* category includes businesses that promote physical and/or mental well-being (e.g., counseling centers, rehabilitation facilities, dental clinics, fitness centers, physicians offices). The *hotel* category encompasses businesses that offer lodging to travelers (e.g., motels, inns, bed and breakfasts).

Given the sheer number of observations included in the original data set, we chose to aggregate the number of transactions by the date and unique individual identifier of the transaction, creating separate data sets for each sample (i.e., MA and MI). In the aggregation process we created several features to describe each observation, including the number of total transactions, the number of businesses involved in transactions, the number of zip codes involved in transactions, the number of transactions by time of day (i.e., morning, afternoon, evening, night), and the number of transactions by business type (i.e., grocery, convenience, restaurant, health, hotel). Note that morning transactions represents purchases made between 06:00am - 11:59am, afternoon transactions represents purchases made



(a) Massachusetts



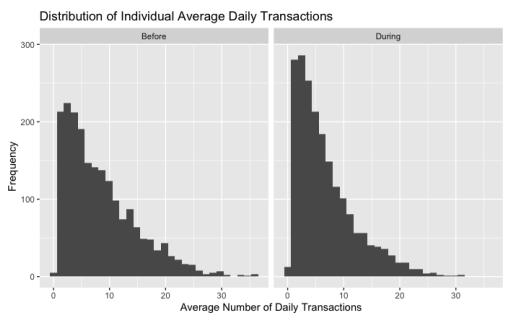
(b) Michigan

Figure 12: State Maps with Cities of Interest

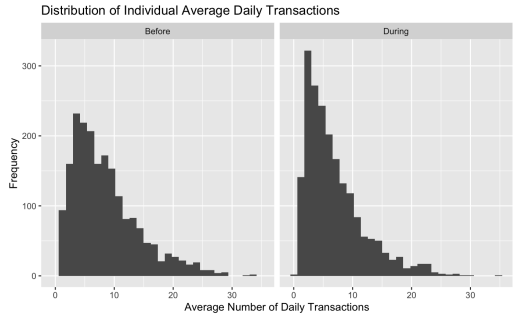
between 12:00pm - 5:59pm, evening transactions represents purchases made between 6:00pm - 11:59pm, and night transactions represents purchases made between 12:00am - 5:59am.

The observations included in our analysis met the following inclusion criteria. Some accounts appeared sparingly within the raw dataset; therefore, we chose to restrict our analysis to accounts that transacted on 25 or more days both prior to and during the national emergency. Some accounts also appeared to be business rather than personal credit lines; therefore, we restricted our analysis to unique individual identifiers without any daily observations of greater than fifty total daily transactions.

Figures 13, 14, and 15 display the distributions of individual average daily transactions, individual average daily businesses involved in transactions, and individual average daily zip codes involved in transactions before versus during the COVID-19 national emergency. Looking at the distribution of these three features before versus during the national emergency, each becomes denser at lower values during the national emergency, suggesting that on average more individuals were making less transactions at less different businesses within less different zip codes.

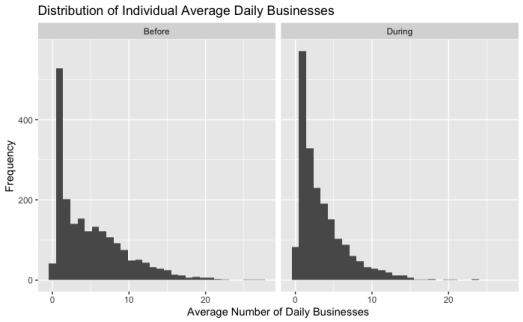


(a) Massachusetts

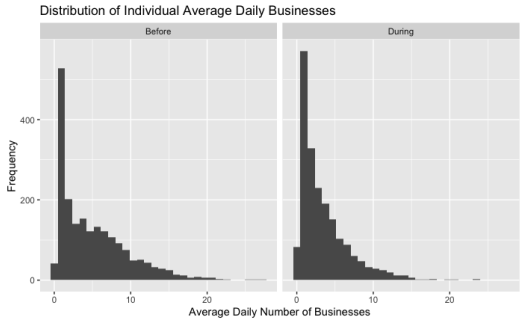


(b) Michigan

Figure 13: Distribution of Individual Average Daily Transactions

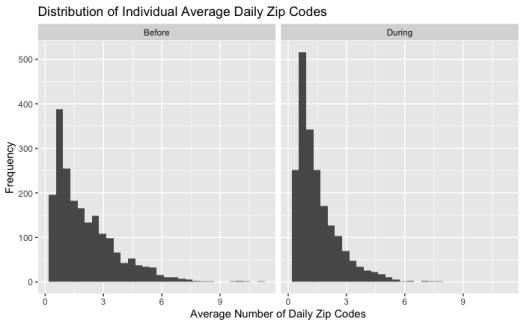


(a) Massachusetts

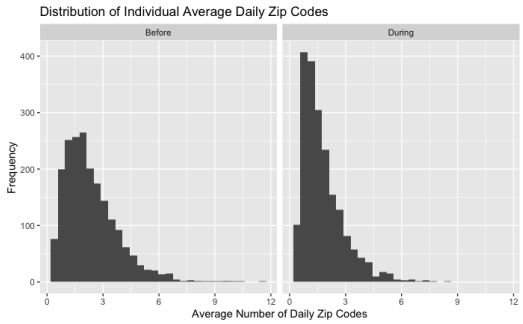


(b) Michigan

Figure 14: Distribution of Individual Average Daily Businesses



(a) Massachusetts



(b) Michigan

Figure 15: Distribution of Individual Average Daily Zip Codes

Table 15 summarizes how individuals' consumer mobility patterns changed across the national emergency within each sample. Note that each table is ordered by the percent change in average daily transactions before versus during the national emergency. Unexpectedly, some individuals within both samples exhibited a dramatic increase in consumer activity during the national emergency.

UUID (MA)	Avg. Daily Transactions			Avg. Daily Businesses			Avg. Daily Zip Codes		
	Before	*During*	*% Change*	*Before*	*During*	*% Change*	*Before*	*During*	*% Change*
MA1	27.44	1.01	-96.31	16.10	0.41	-97.48	3.03	0.34	-88.71
MA2	15.67	0.92	-94.10	11.06	0.67	-93.93	3.60	0.53	-85.22
MA3	11.82	0.75	-93.68	1.58	0.47	-70.42	0.88	0.42	-52.26
MA4	12.75	0.92	-92.75	1.00	0.35	-64.56	0.57	0.33	-42.20
MA5	10.18	0.77	-92.42	1.49	0.51	-65.93	1.40	0.46	-67.51
MA6	24.46	1.96	-91.98	18.63	1.51	-91.91	5.19	0.75	-85.62
MA7	2.32	15.72	+577.81	0.67	1.78	+167.72	0.47	1.22	+157.33
MA8	3.90	27.72	+610.30	1.93	5.05	+161.62	1.26	2.85	+125.34
MA9	1.53	12.33	+707.00	0.94	2.32	+145.27	0.89	1.49	+68.04
MA10	3.14	26.44	+742.43	1.19	3.06	+156.46	0.88	1.59	+82.28
MA11	1.29	17.01	+1217.11	0.60	12.63	+2015.28	0.60	3.08	+415.04
MA12	0.71	18.67	+2535.89	0.46	1.56	+217.61	0.42	0.42	+103.54

UUID (MI)	Avg. Daily Transactions			Avg. Daily Businesses			Avg. Daily Zip Codes		
	Before	*During*	*% Change*	*Before*	*During*	*% Change*	*Before*	*During*	*% Change*
MI1	14.17	0.84	-94.10	5.97	0.66	-88.98	0.97	0.56	-42.71
MI2	20.10	1.44	-92.81	15.28	1.32	-91.38	2.24	0.58	-73.96
MI3	9.83	0.76	-92.28	6.15	0.39	-93.62	1.78	0.39	-77.93
MI4	22.82	2.13	-90.68	14.75	1.62	-89.02	2.56	0.65	-74.74
MI5	23.87	2.35	-90.13	2.58	1.03	-60.31	1.86	0.84	-55.11
MI6	20.08	2.04	-89.85	17.33	1.67	-90.36	3.65	0.84	-77.13
MI7	2.08	16.09	+672.25	1.63	9.03	+455.40	0.83	1.71	+105.06
MI8	1.22	9.47	+674.68	0.96	3.90	+306.82	0.81	2.68	+233.13
MI9	1.19	10.34	+765.82	0.76	1.30	+70.68	0.64	1.25	+96.15
MI10	1.11	11.54	+938.99	1.04	2.87	+175.85	0.43	0.65	+49.94
MI11	0.78	11.86	+1424.95	0.47	1.09	+130.53	0.46	0.73	+60.18
MI12	1.49	26.92	+1711.71	1.42	17.32	1122.34	0.57	0.87	+53.38

Table 15: Daily Individual Consumer Behavior (Before vs. During National Emergency)

Figures 16, 17, and 18 display how individuals' consumer mobility patterns changed over time. Note that each variable is displayed as a 7-day moving average in order to smooth out short-term fluctuations caused by weekly trends. Looking across both samples, there was a major decline in the average number of daily transactions, the average number of businesses and the average number of zip codes involved in daily transactions following the declaration of COVID-19 as a national emergency during mid-March. The sample-level average of all three features remained exceptionally lower than normal until about late-April, when both samples started to become noticeably more active again.

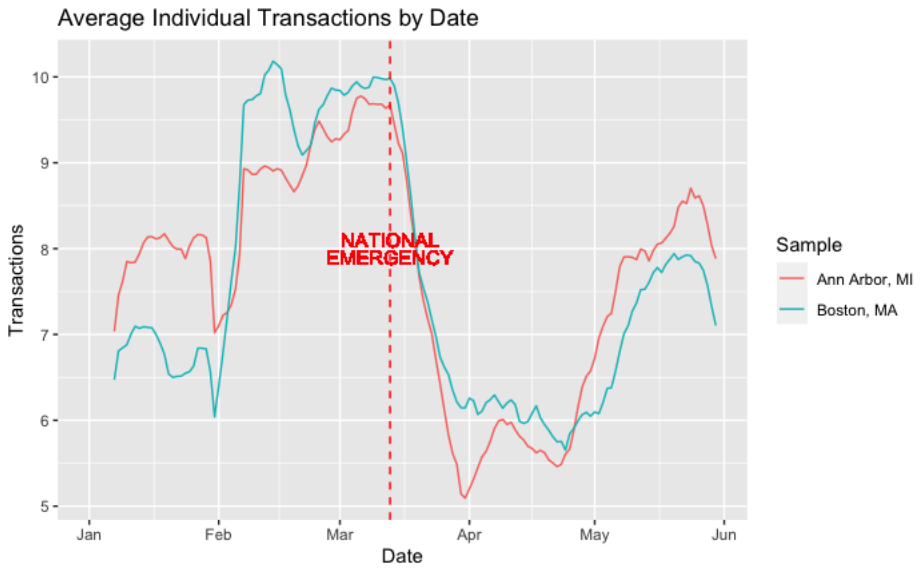


Figure 16: Average Individual Transactions by Date

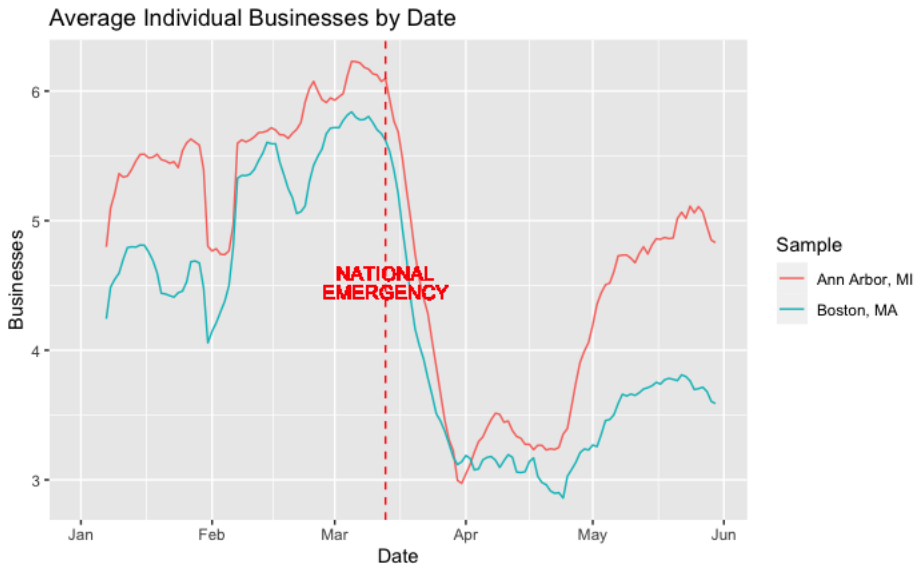


Figure 17: Average Individual Businesses by Date

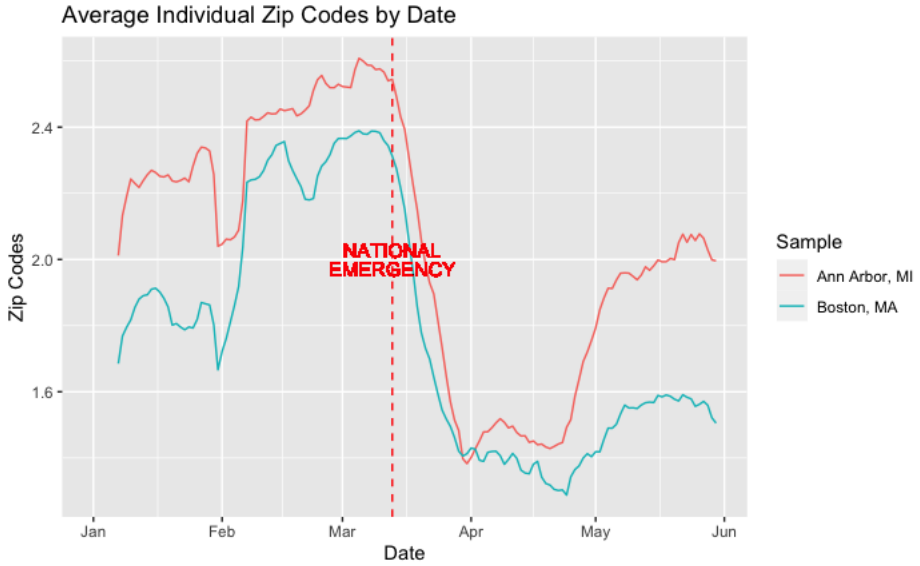
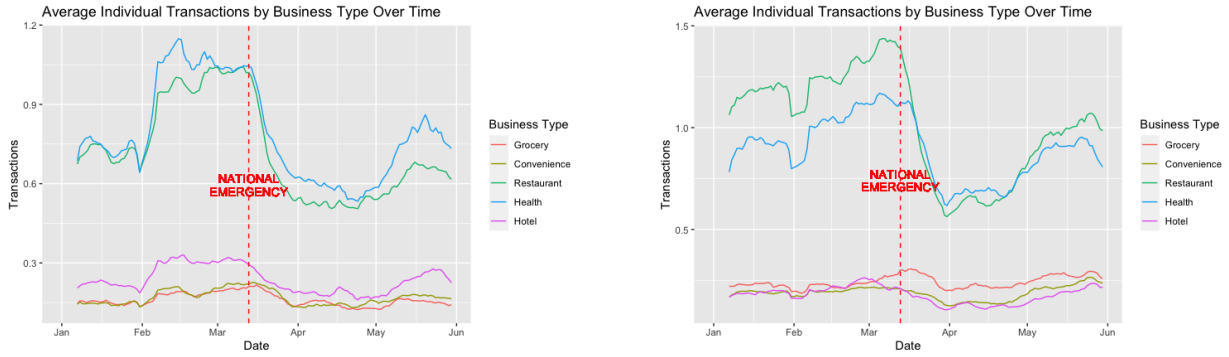


Figure 18: Average Individual Zip Codes by Date

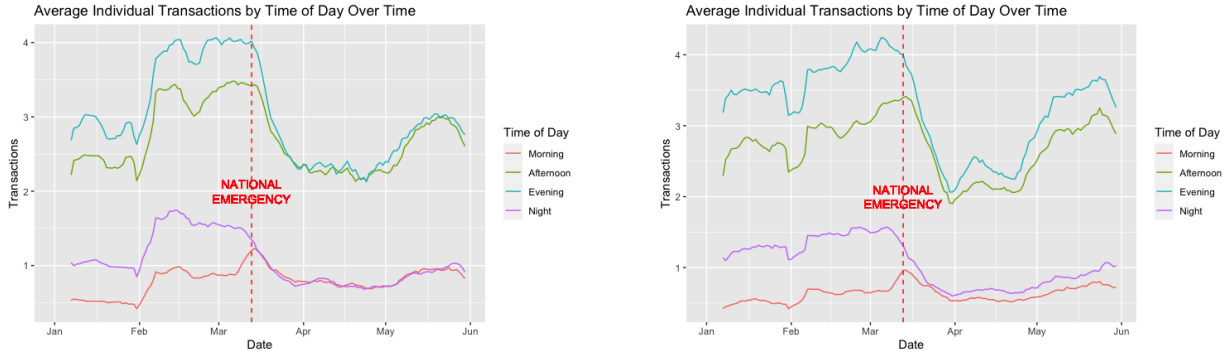
Figures 19 and 20 display how average individual daily transactions by business type and time of day changed over time within each sample. Note that each variable is displayed as a 7-day moving average in order to smooth out short-term fluctuations caused by weekly trends. In terms of average transactions by business type, restaurant and health-related transactions were noticeably higher than grocery, convenience, and hotel-related transactions among both samples. The decline in individual average restaurant and health-related transactions following the declaration of COVID-19 as a national emergency was also much

more pronounced than that of grocery, convenience, and hotel-related transactions. In terms of consumer behavior by time of day, individuals from both samples were considerably more active during the afternoon and evening periods.



(a) Massachusetts (b) Michigan

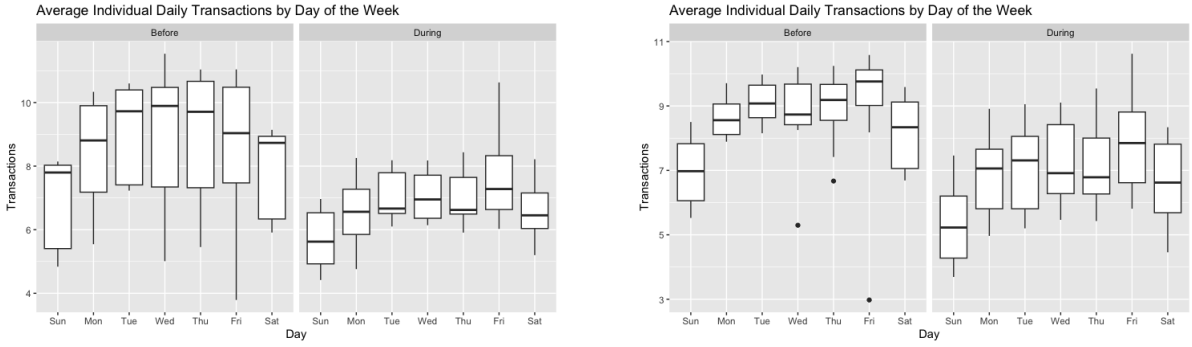
Figure 19: Average Individual Transactions by Business Type Over Time



(a) Massachusetts (b) Michigan

Figure 20: Average Individual Transactions by Time of Day Over Time

Figure 21 displays the distribution of average individual daily transactions by the day of the week before versus during the national emergency for each sample. Both plots suggest that there exists a seasonal trend, as individuals from both samples were more active on weekdays (i.e., Monday-Friday) than weekends (i.e., Saturday-Sunday) both before and during the national emergency. However, this seasonal trend appears to become less pronounced among both samples following the COVID-19 outbreak.



(a) Massachusetts (b) Michigan

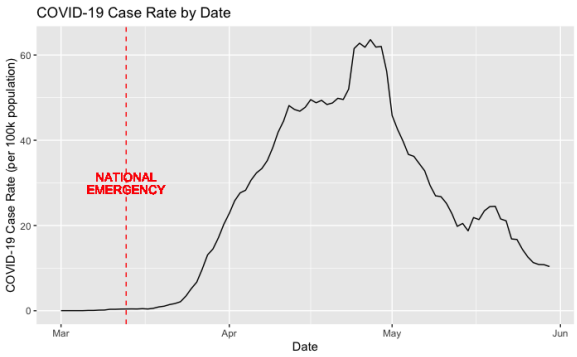
Figure 21: Distribution of Average Individual Daily Transactions by Day of the Week

5.1.3 COVID-19 Case and Testing Data

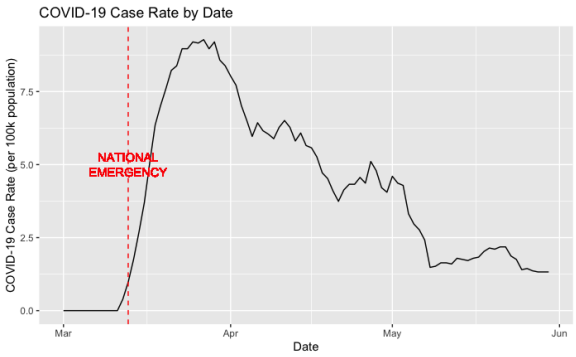
Both the MA and MI state governments kept well-organized COVID-19 databases dating all the way back to the beginning of the outbreak in early 2020 [58, 59]. These two databases contain daily and cumulative confirmed case and testing counts at the county-level. Recall that our two random consumer samples contain residents of Boston, Suffolk County (MA), and Ann Arbor, Washtenaw County (MI). Therefore, the COVID-19 case and testing data observed within these two counties is of particular interest.

Figures 22 and 23 display the COVID-19 case rate (per 100k population) by date and day of the week within our two regions of interest (Suffolk County, MA, and Washtenaw County, MI). Note that in Figure 22 case rate is displayed as a 7-day moving average in order to smooth-out short term fluctuations caused by weekly patterns. Looking across counties, both experienced dramatic spikes in the COVID-19 case rate. The spike within Washtenaw County, MI, occurred immediately following the declaration of COVID-19 as a national emergency. However, the spike within Suffolk County occurred closer to the end of April and was greater in terms of both duration and magnitude. Figure 23 suggests that case rates did not follow a major seasonal trend.

Figures 24 and 25 display the COVID-19 test rate (per 100k population) by date and day of the week within our two regions of interest (Suffolk County, MA, and Washtenaw County, MI). Note that MA COVID-19 testing data is representative of state-level totals, as testing estimates were not available at the county-level. Also note that in Figure 24 testing rate is displayed as a 7-day moving average in order to smooth-out short term fluctuations caused by weekly patterns. Looking across counties, both experienced a relatively steady increase in testing from March - June 2020, with the exception of a short period of decline during April in MI. Figure 25 suggests that testing rates followed a slight seasonal trend, as testing appeared to be more popular on weekdays than weekends.

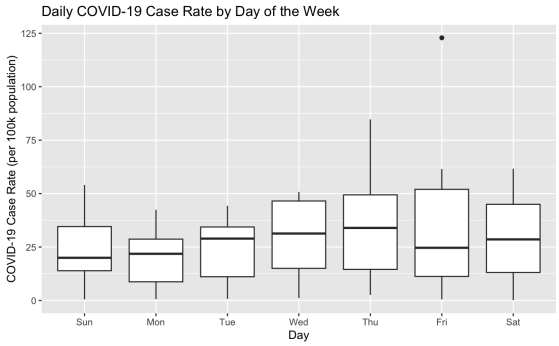


(a) Suffolk County, Massachusetts

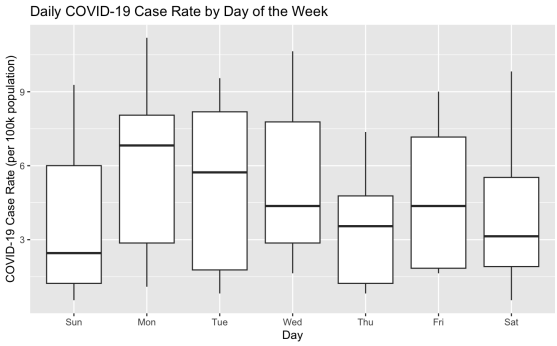


(b) Washtenaw County, Michigan

Figure 22: COVID-19 Case Rate (per 100,000 population) by Date

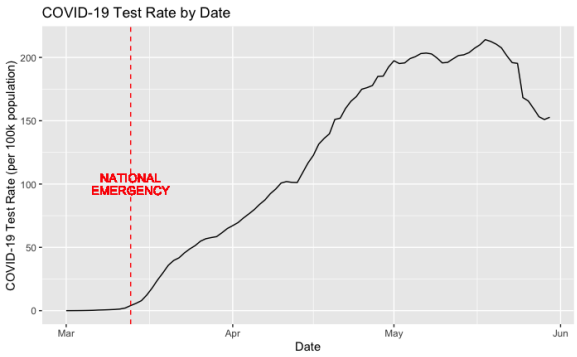


(a) Suffolk County, Massachusetts

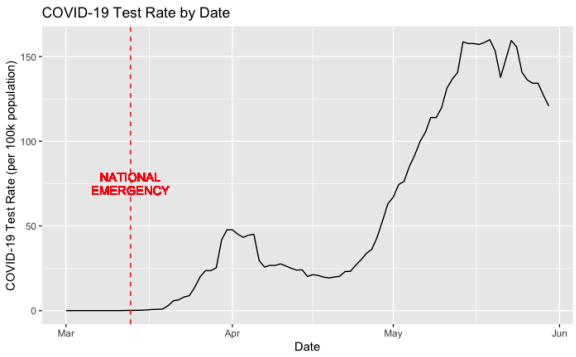


(b) Washtenaw County, Michigan

Figure 23: Daily COVID-19 Case Rate (per 100,000 population) by Day of the Week

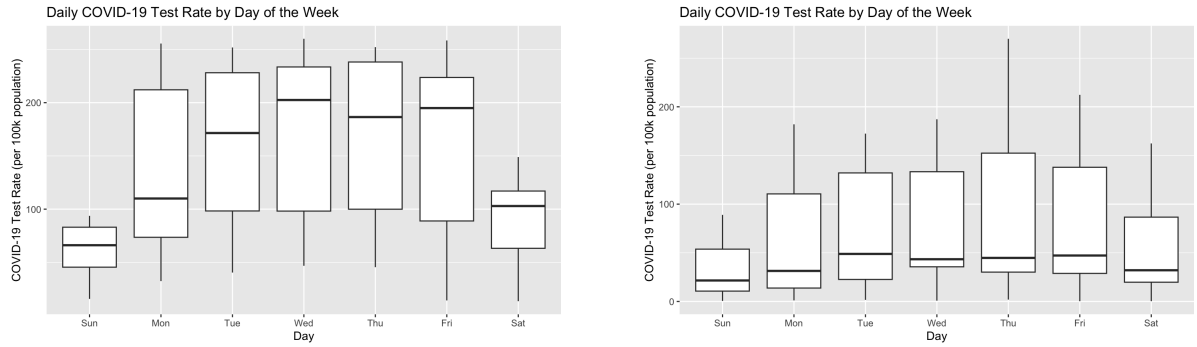


(a) Suffolk County, Massachusetts



(b) Washtenaw County, Michigan

Figure 24: Daily COVID-19 Testing Rate (per 100,000 population) by Date



(a) Suffolk County, Massachusetts (b) Washtenaw County, Michigan

Figure 25: Daily COVID-19 Testing Rate (per 100,000 population) by Day of the Week

Table 16 provides numerical summaries of the features from each of the data sources described above during the time period just prior to the beginning of the official declaration of COVID-19 as a national emergency (January 1st - March 12th, 2020) as well as the time period immediately after (March 13th - May 30th, 2020). Note that qualitative variables are summarized by frequency tables and quantitative variables are summarized by mean and standard deviation values.

	Before	**During**	**Total**
Variable (MA)	$(n = 144,000)$	$(n = 158,000)$	$(n = 302,000)$
NPIs			
Workplace Closures			
-None	144,000 (100%)	22,000 (13.9%)	166,000 (55.0%)
-Recommended	0 (0%)	0 (0%)	0 (0%)
-Required (some)	0 (0%)	0 (0%)	0 (0%)
-Required (all)	0 (0%)	136,000 (86.1%)	136,000 (45.0%)
Stay-at-Home Reqs.			
-None	144,000 (100%)	20,000 (12.7%)	164,000 (54.3%)
-Recommended	0 (0%)	138,000 (87.3%)	138,000 (45.7%)
-Required (some ex.)	0 (0%)	0 (0%)	0 (0%)
-Required (min ex.)	0 (0%)	0 (0%)	0 (0%)
Gathering Restrictions			
-None	144,000 (100%)	0 (0%)	144,000 (47.7%)
-Restrict > 1,000 people	0 (0%)	0 (0%)	0 (0%)
-Restrict 101-1,000 people	0 (0%)	8,000 (5.1%)	8,000 (2.6%)
-Restrict 11-100 people	0 (0%)	40,000 (25.3%)	40,000 (13.2%)
-Restrict < 10 people	0 (0%)	110,000 (69.6%)	110,000 (36.4%)
Daily Transaction Counts			
Total	8.40 ± 12.3	6.80 ± 10.9	7.57 ± 11.6
Unique Businesses	5.05 ± 7.31	3.47 ± 5.53	4.22 ± 6.48
Unique Zip Codes	2.08 ± 2.65	1.50 ± 1.99	1.78 ± 2.35
Grocery	0.17 ± 0.88	0.15 ± 0.83	0.16 ± 0.86
Convenience	0.18 ± 0.81	0.16 ± 0.97	0.17 ± 0.891
Restaurants	0.86 ± 2.28	0.60 ± 2.06	0.72 ± 2.17
Health	0.91 ± 2.98	0.68 ± 2.79	0.79 ± 2.89
Hotels	0.27 ± 1.30	0.21 ± 1.46	0.24 ± 1.39
Morning	0.75 ± 2.95	0.83 ± 2.97	0.79 ± 2.96
Afternoon	2.89 ± 5.49	2.53 ± 5.37	2.70 ± 5.43
Evening	3.44 ± 6.02	2.61 ± 5.32	3.01 ± 5.68
Night	1.32 ± 3.78	0.83 ± 2.88	1.06 ± 3.35
Daily COVID-19 Rates (per 100,000 population)			
Case	0.040 ± 0.172	28.2 ± 21.6	14.8 ± 21.0
Test	0.221 ± 0.837	137 ± 77.4	71.6 ± 88.2

	Before	**During**	**Total**
Variable (MI)	$(n = 144,000)$	$(n = 158,000)$	$(n = 302,000)$
NPIs			
Workplace Closures			
-None	144,000 (100%)	6,000 (3.8%)	150,000 (49.7%)
-Recommended	0 (0%)	0 (0%)	0 (0%)
-Required (some)	0 (0%)	90,000 (57.0%)	90,000 (29.8%)
-Required (all)	0 (0%)	62,000 (39.2%)	62,000 (20.5%)
Stay-at-Home Reqs.			
-None	140,000 (97.2%)	0 (0%)	140,000 (46.4%)
-Recommended	4,000 (2.8%)	42,000 (26.6%)	46,000 (15.2%)
-Required (some ex.)	0 (0%)	116,000 (73.4%)	116,000 (38.4%)
-Required (min ex.)	0 (0%)	0 (0%)	0 (0%)
Gathering Restrictions			
-None	144,000 (100%)	0 (0%)	144,000 (47.7%)
-Restrict > 1,000 people	0 (0%)	0 (0%)	0 (0%)
-Restrict 101-1,000 people	0 (0%)	8,000 (5.1%)	8,000 (2.6%)
-Restrict 11-100 people	0 (0%)	14,000 (8.9%)	14,000 (4.6%)
-Restrict < 10 people	0 (0%)	136,000 (86.1%)	136,000 (45.0%)
Daily Transaction Counts			
Total	8.53 ± 11.4	6.90 ± 10.7	7.67 ± 11.1
Unique Businesses	5.58 ± 7.08	4.14 ± 6.01	4.83 ± 6.58
Unique Zip Codes	2.35 ± 2.66	1.77 ± 2.27	2.05 ± 2.48
Grocery	0.23 ± 0.97	0.25 ± 1.10	0.24 ± 1.04
Convenience	0.20 ± 0.74	0.18 ± 0.92	0.19 ± 0.84
Restaurants	1.23 ± 2.65	0.83 ± 2.14	1.02 ± 2.41
Health	0.99 ± 2.75	0.80 ± 2.75	0.89 ± 2.75
Hotels	0.20 ± 1.12	0.16 ± 1.06	0.18 ± 1.09
Morning	0.61 ± 2.67	0.64 ± 2.68	0.63 ± 2.67
Afternoon	2.86 ± 5.24	2.56 ± 5.23	2.71 ± 5.24
Evening	3.70 ± 5.70	2.90 ± 5.39	3.28 ± 5.56
Night	1.36 ± 3.53	0.80 ± 2.73	1.06 ± 3.15
DailyCOVID-19 Rates (per 100,000 population)			
Case	0.038 ± 0.319	4.58 ± 3.01	2.42 ± 3.15
Test	0.008 ± 0.064	68.8 ± 62.9	36.0 ± 57.0

Table 16: NPIs, Transaction & COVID-19 Rates Before vs. During National Emergency

There are a few noteworthy observations found within Table 16. First, all of the NPIs included in the dataset were implemented at various stringency levels during the COVID-19 national emergency. Second, the total number of transactions, the number of unique businesses involved in transactions, and the number of unique zip codes involved in transactions was lower during the national emergency in comparison to the period prior among

both samples of individuals. However, certain types of transactions such as grocery and convenience store-related purchases did not exhibit as pronounced of a decline during the national emergency in comparison to restaurant and health related purchases. Third and finally, COVID-19 case and testing rates following the declaration of COVID-19 as a national emergency suggest that the initial virus outbreak was more intense in MA than MI.

5.2 Modeling Approach

In this individual-level study, we employ a three-step modeling approach. First, we experiment with hierarchical and partitional clustering methods in order to detect subgroups among each sample of consumers based on their pre-pandemic spending behavior. Second, we employ negative binomial mixed effects models to explore how the mobility patterns of consumers changed across various NPIs. Third, we use linear mixed effects models to investigate the relationship between changes in consumer mobility patterns compared to pre-pandemic baselines and the initial spread of COVID-19. In the following subsections, we describe each step in greater detail.

5.2.1 Clustering

Given that we possess no identifying information about the individuals included in our dataset, we chose to perform consumer segmentation using clustering. Cluster analysis involves four distinct steps: feature selection, algorithm selection, cluster validation, and results interpretation. In the paragraphs to follow, we outline our approach step-by-step.

Our final dataset represents the spending patterns of 2,000 unique individuals from Boston, MA, and 2,000 unique individuals from Ann Arbor, MI, over the period from January 1$^{\text{st}}$ - May 30$^{\text{th}}$, 2020. Using observations from January 1$^{\text{st}}$ - February 31$^{\text{st}}$, we create a dataset of features that characterizes each individual consumer's behavior prior to the COVID-19 pandemic. These features include average daily transactions, average daily AM (i.e., night and morning) transactions, average daily PM (i.e., afternoon and evening) transactions, a ratio of average daily AM versus PM transactions, average daily weekend transactions (i.e., Saturday - Sunday), average daily weekday transactions (i.e., Monday - Friday), and a ratio of average daily weekend versus weekday transactions.

Using the dataset of features described in the previous paragraph, we perform cluster analysis with three different algorithms: agglomerative clustering (i.e., complete linkage method, single linkage method, average linkage method, and Ward's method), divisive clustering, and K-means clustering. We use Euclidean distance for calculating the dissimilarities between observations. In order to identify the optimal number of clusters for each algorithm, we employ the elbow method, the silhouette method, and the gap statistic method. The mathematical details behind these algorithms and methods can be found within Section 3.2.

In order to evaluate each clustering algorithm, we rely on several criterion. For each hierarchical method (i.e., agglomerative and divisive clustering), we assess performance using the agglomerative coefficient, divisive coefficient, and dendrograms. For our sole partitional method (i.e., K-means clustering), we assess performance using the within-cluster sum of squares. For all methods, we also assess performance using principal component analysis (PCA) plots colored by cluster.

Finally, after using the set of criterion outlined in the previous paragraph to identify the most appropriate clustering of individuals, we interpret our results. Specifically, we create graphical and numerical summaries of various individual-level transaction-related features by cluster. Using these summaries, we assign intuitive labels to each sub-group.

5.2.2 Mobility Modeling

Given the nature of our data (i.e., daily POS transaction counts clustered at the individual-level), we fit negative binomial mixed effects models with an autoregressive covariance structure to investigate the relationship between government-mandated NPIs and daily POS transactions. The explanatory variable NPI referenced in the models below represents the different NPIs included in our dataset (i.e., $NPI \in \{$national emergency, workplace closures, stay-at-home requirements, gathering restrictions$\}$). The response variable in each model is the daily POS transaction count on date i for individual j. In each model, we adjust for seasonal patterns induced by the day of the week (see variable *weekend*, which takes on a value of zero on days Monday - Friday and a value of one on days Saturday-Sunday).

The first model,

$$\log(transactions_{ij}) = \beta_0 + \beta NPI_{ij} + \beta weekend_{ij} + \gamma_{0j}, \tag{6}$$

describes the association between daily POS transaction counts and the implementation of each NPI. Note that each NPI is represented as a categorical variable with multiple levels; therefore, fitting this model with each different NPI variable results in multiple β coefficients due to one hot encoding.

The second model,

$$\begin{aligned} \log(transactions_{ij}) = \beta_0 + \beta_1 NPI_{ij} + \beta_2 consumer_{ij} \\ + \beta_3 NPI_{ij} consumer_{ij} + \beta_4 weekend_{ij} + \gamma_{0j}, \end{aligned} \tag{7}$$

in which *consumer* represents the subgroups identified through K-means clustering (see Section 5.3.1), describes how the association between daily POS transaction counts and the implementation of each NPI varies by consumer type. Note that in fitting this model one hot encoding is no longer performed, as doing so would require fitting an abundance of parameters and lead to less interpretable results. Instead, each NPI is represented as a continuous

variable for which we are only concerned with one unit increases.

The third and final model,

$$\begin{aligned}\log(transactions_{ij}) = \beta_0 + \beta_1 NPI_{ij} + \beta_2 state_{ij} \\ + \beta_3 NPI_{ij} state_{ij} + \beta_4 weekend_{ij} + \gamma_{0j},\end{aligned} \tag{8}$$

is fit using data from both states, MA and MI, unlike the previous three models. This model, in which *state* represents whether the observation is an individual from MA or MI, describes how the association between daily POS transaction counts and the implementation of each NPI varies by state. Again, note that in fitting this model each NPI is represented as a continuous variable for which we are only concerned with one unit increases, as one hot encoding leads to an abundance of parameters and less interpretable results.

5.2.3 COVID-19 Modeling

In order to investigate the relationship between changes in consumer mobility patterns and the spread of COVID-19, we take a slightly different approach than the one described in the previous section. We group all individual consumer data by date, creating sample-level transaction features that are representative of a large random sample of residents from two counties (Suffolk, MA, and and Washtenaw, Michigan) for which COVID-19 case and test counts are publicly available. This eliminates the need for mixed-effects modeling. Negative binomial models are also no longer necessary, as our new response variable is not a daily count. Instead, we model the daily exponential growth rate in confirmed COVID-19 cases, *case growth rate* (i.e., the natural logarithm of daily COVID-19 cases minus the logarithm of daily COVID-19 cases on the prior day, multiplied by 100 to be interpretable as percent), making linear models a more appropriate approach. Note that we add one to each daily case count to prevent the natural logarithm of cases from being undefined if there are no cases on a given day. Given the time-series nature of the data, we continue to fit these models with an autoregressive covariance structure.

The explanatory variable of interest in these COVID-19 models is changes in consumer mobility patterns. As indicated throughout our literature review, substantial efforts have been made to create datasets that reflect human mobility change since the outbreak of the COVID-19 pandemic, where starting points between January 2020 and February 2020 have been selected to provide a baseline. In this study, we attempt to do the same by creating a variable $\Delta transactions$, which represents the percentage of change between each sample's total daily POS transaction count and the pre-pandemic baseline day. Note that these pre-pandemic baseline days are median values of the total daily POS transaction count for each day of the week during the period from January 1$^{\text{st}}$ - February 29$^{\text{th}}$, 2020.

A major challenge in modeling the relationship between changes in mobility patterns and the outbreak of COVID-19 is determining the lag in time between the mobility data and

COVID-19 cases. From the time an individual makes initial contact with the virus and contracts the disease, confirming the diagnosis is a complex process. There are several factors that can affect the length of time necessary for a COVID-19 case to be confirmed–incubation period, testing speed, and reporting delay to name a few. Previous studies have found that the COVID-19 incubation period can vary, yet most cases manifest themselves between 3-7 days after initial infection [62]. The lag caused by testing speed and delays in reporting is much more variable, as it is largely dependent upon the medical facilities within a region. Other statistical studies have attempted to quantify the optimal lag, most notably Badr et al. (2020) who found that a lag of 11 days achieved the highest correlation between mobility and COVID-19 growth rate ratios for a single US all-county model [17].

In this study, we experiment with models that include mobility data lags ranging anywhere from 6-16 days. In each model, we control for the state-level COVID-19 testing rate (per 100,000 population) and seasonal patterns induced by the day of the week (see variables *test rate* and *weekend*). More formally, we mathematically define each model as follows:

$$\begin{aligned} case\ growth\ rate_i = \beta_0 + \beta_1 \Delta transactions_i + \beta_2 test\ rate_i \\ + \beta_3 weekend_i, \end{aligned} \tag{9}$$

in which i represents the date of an observation. These models describe the association between COVID-19 case growth rates and lagged changes in POS transactions compared to pre-pandemic baselines.

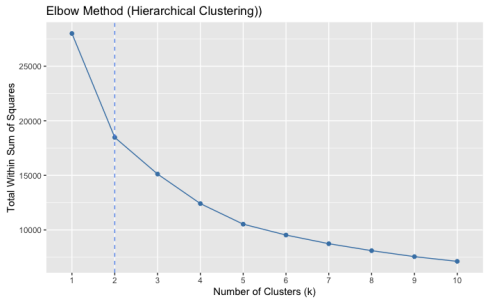
5.3 Results

5.3.1 Clustering

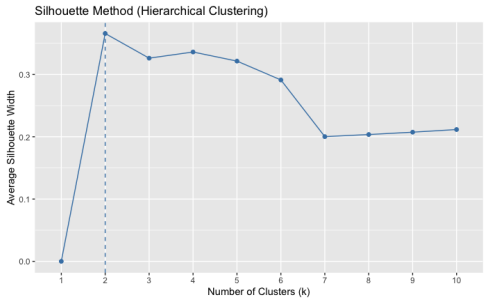
Using the *cluster* and *factoextra* packages in R [64, 67, 68], we performed three different clustering algorithms (i.e., agglomerative, divisive, and K-means) on the 4,000 individuals in our final dataset (i.e., 2,000 MA individuals and 2,000 MI individuals) using the following features: average daily transactions, average daily weekend transactions, average daily weekday transactions, the ratio of average daily weekend versus weekday transactions, average daily AM transactions, average daily PM transactions, and the ratio of average daily AM versus PM transactions. Note that for each clustering algorithm, we calculated dissimilarity values using Euclidean distance. Prior to performing any clustering, we used the elbow, silhouette, and gap statistic methods to determine the optimal number of clusters in our dataset.

Figure 26 displays the elbow, silhouette, and gap statistic methods for identifying the appropriate number of clusters to use in performing hierarchical clustering. Both the elbow and silhouette methods suggest extracting two clusters, while the gap statistic method suggests extracting three clusters. Given the conflicting results obtained from these methods,

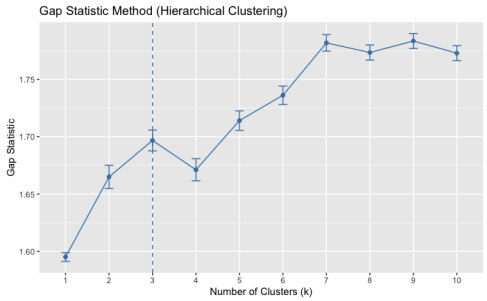
we opted to experiment with the extraction of both two and three clusters from each hierarchical clustering algorithm.



(a) Elbow Method



(b) Silhouette Method



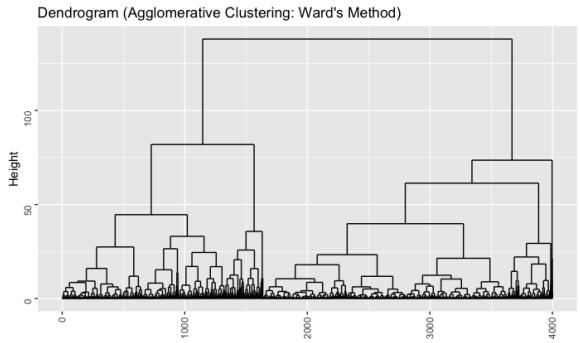
(c) Gap Statistic Method

Figure 26: Optimal Number of Clusters (Hierarchical Clustering)

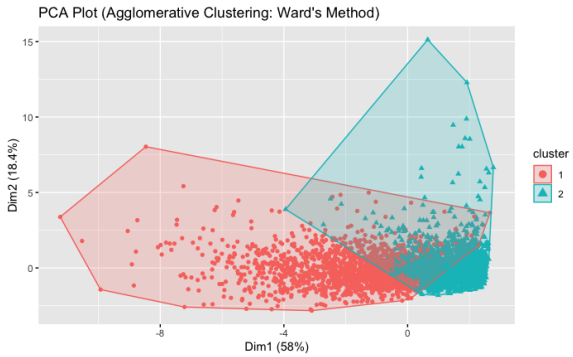
First, we performed agglomerative clustering. Recall that there are several different agglomeration methods (i.e., complete linkage, single linkage, average linkage, Ward's method). Table 17 displays the agglomerative coefficient obtained using each different method. Although all four methods obtained extremely high agglomerative coefficients, Ward's method identified the strongest clustering structure. Figure 27 displays the resulting dendrogram and PCA plot obtained from agglomerative clustering using Ward's method. The PCA plot suggests that this clustering is not effective, as there is a large amount of overlap between the two clusters.

Method	**AC**
Complete	0.9928
Single	0.9772
Average	0.9873
Ward	0.9985

Table 17: Agglomerative Clustering Results



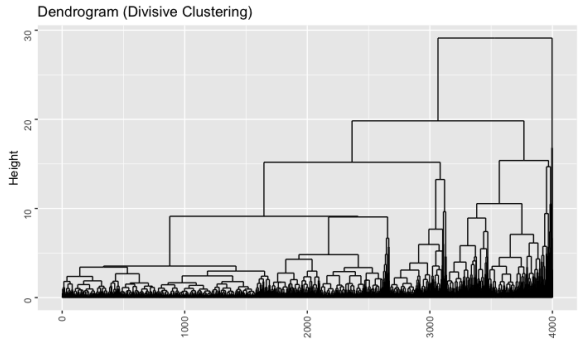
(a) Dendrogram



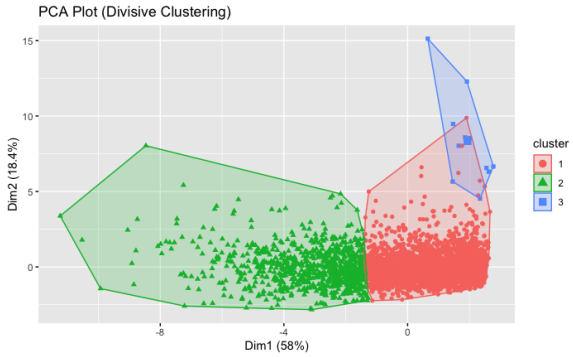
(b) PCA Plot

Figure 27: Agglomerative Clustering Results (Ward's Method)

Next, we performed divisive clustering. The divisive coefficient obtained using this method is 0.9917, suggesting that the algorithm found a strong clustering structure within the data selected to represent the individuals in our dataset. Figure 28 displays the resulting dendrogram and PCA plot obtained from divisive clustering. The PCA plot indicates that divisive clustering was a bit more effective than agglmoerative clustering; however, there is clearly still overlap between clusters one and three.



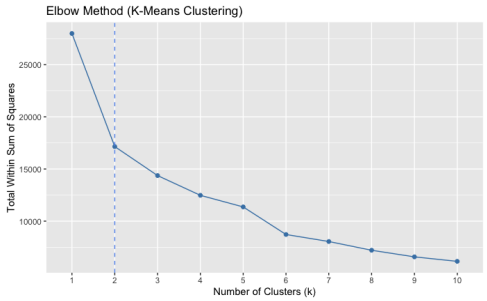
(a) Dendrogram



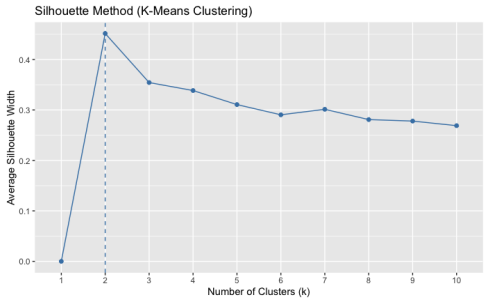
(b) PCA Plot

Figure 28: Divisive Clustering Results

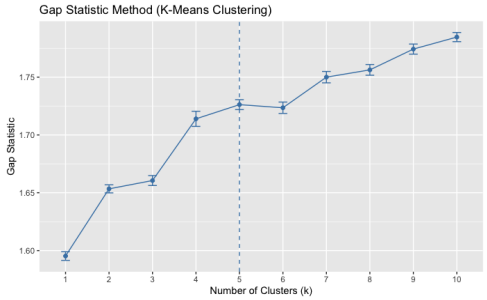
Lastly, we performed k-means clustering. Figure 29 displays the elbow, silhouette, and gap statistic methods for identifying the appropriate number of clusters to use in performing k-means clustering. Both the elbow and silhouette methods suggest using $k = 2$ clusters, while the gap statistic method suggests using $k = 5$ clusters. Given the conflicting suggestions provided by these methods, we chose to meet in the middle, opting to perform 3-means clustering.



(a) Elbow Method



(b) Silhouette Method



(c) Gap Statistic Method

Figure 29: Optimal Number of Clusters (K-Means Clustering)

Table 18 displays the 3-means clustering results. Note that the total within cluster sum of squares is 14370.43, indicating that this clustering method explains approximately 48.7% of the variance in the data selected to represent the individuals in our dataset. Figure 30 displays the PCA plot obtained using 3-means clustering colored by cluster. This plot highlights the effectiveness of 3-means clustering, as all three groups are well-separated.

Measure	**Cluster 1**	**Cluster 2**	**Cluster 3**
Size	2157	1385	458
Within Cluster SS	7869.647	3320.855	3179.931

Table 18: K-Means Clustering Results

Based on the results described throughout this section, 3-means clustering provides the most appropriate clustering of the individuals within our dataset. However, assigning intuitive labels to the subgroups identified by this clustering requires investigating the characteristics of each cluster in greater depth. Table 19 provides a numerical summary of the features selected to represent the individuals in our dataset by cluster. Figure 31 provides the distributions of individual average daily transactions by cluster. Figure 32 displays each individual's average number of daily transactions during weekends versus weekdays and PM versus AM colored by cluster. All three suggest that each cluster best represents consumer

Figure 30: PCA Plot (K-Means Clustering)

activity level, with cluster one containing low-frequency consumers, cluster two containing medium-frequency consumers, and cluster three containing high-frequency consumers.

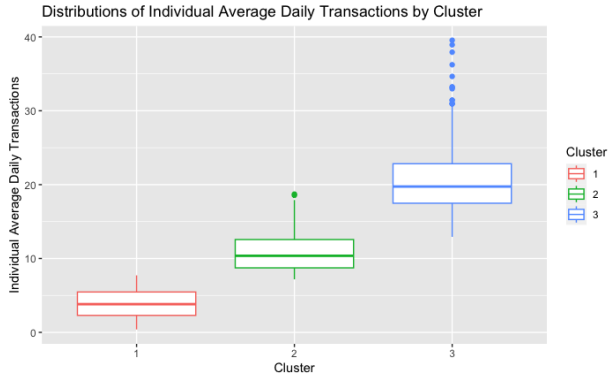
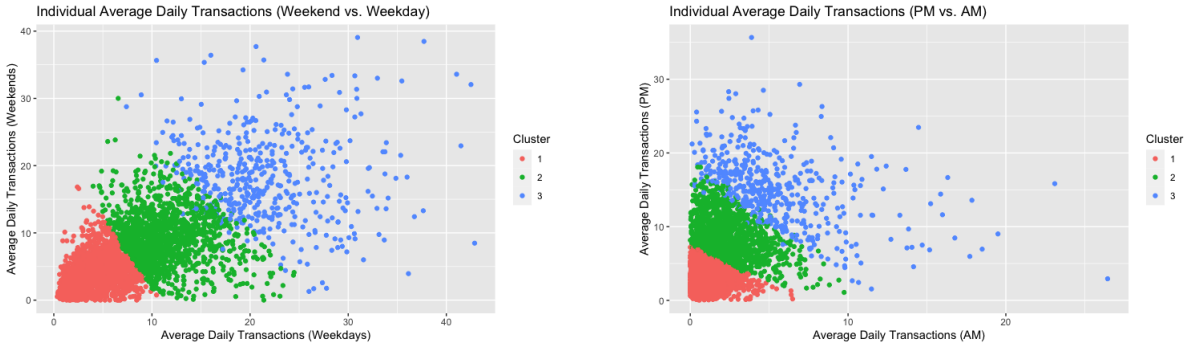


Figure 31: Distributions of Individual Average Transactions by Cluster



(a) Weekend vs. Weekday

(b) PM vs. AM

Figure 32: Individual Average Daily Transactions by Cluster

Variable	Cluster 1 ($n = 2157$)	Cluster 2 ($n = 1385$)	Cluster 3 ($n = 458$)	Overall ($n = 4000$)
Avg. Daily Total Transactions				
-Mean (± SD)	3.87 (± 1.85)	10.80 (± 2.52)	20.60 (± 4.42)	8.20 (± 6.04)
-Median (Min,Max)	3.82 (0.40,7.72)	10.40 (7.18,18.70)	19.70 (12.90,39.60)	6.65 (0.40,39.60)
Avg. Daily AM Transactions				
-Mean (± SD)	0.95 (± 0.84)	2.33 (± 1.52)	5.51 (± 3.54)	1.95 (± 2.16)
-Median (Min,Max)	0.73 (0.00,6.48)	2.08 (0.00,9.75)	4.72 (0.12,26.50)	1.28 (0.00,26.50)
Avg. Daily PM Transactions				
-Mean (± SD)	2.92 (± 1.62)	8.51 (± 2.70)	15.10 (± 4.77)	6.25 (± 4.82)
-Median (Min,Max)	2.75 (0.03,7.37)	8.10 (1.08,18.10)	15.20 (1.55,35.70)	5.03 (0.03,35.70)
Avg. Daily Ratio (AM vs. PM) Transactions				
-Mean (± SD)	0.56 (± 1.58)	0.35 (± 0.46)	0.50 (± 0.72)	0.48 (± 1.22)
-Median (Min,Max)	0.28 (0.00,32.40)	0.24 (0.00,9.00)	0.33 (0.01,9.02)	0.27 (0.00,32.4)
Avg. Daily Weekend Transactions				
-Mean (± SD)	3.62 (± 2.39)	9.36 (± 3.88)	18.30 (± 6.42)	7.29 (± 5.97)
-Median (Min,Max)	3.18 (0.00,16.80)	9.06 (0.00,30.00)	17.90 (1.29,39.10)	5.76 (0.00,39.10)
Avg. Daily Weekday Transactions				
-Mean (± SD)	3.96 (± 2.05)	11.40 (± 3.37)	21.50 (± 5.82)	8.56 (± 6.60)
-Median (Min,Max)	3.79 (0.35,10.50)	10.80 (4.00,24.50)	20.70 (7.40,42.90)	6.81 (0.35,42.90)
Avg. Daily Ratio (Weekend vs. Weekday) Transactions				
-Mean (± SD)	1.05 (± 0.78)	0.91 (± 0.52)	0.93 (± 0.46)	0.98 (± 0.67)
-Median (Min,Max)	0.89 (0.00,9.73)	0.82 (0.00,4.57)	0.88 (0.05,3.89)	0.86 (0.00,9.73)

Table 19: Consumer Behavior by Cluster

5.3.2 Mobility Modeling

Using the *glmmTMB* package in R [63, 64], we fit each of the negative binomial mixed effect models with an autoregressive covariance structure, as described Section 5.2.2. In order to estimate the likelihood function of each model, we employed Laplace Approximation with a non-linear optimizer to obtain best-fit model parameters [65]. Note that we fit the first two models on two separate subsets of the dataset by state (i.e., MA and MI) and the third and final model using the entire dataset.

First, we fit the class of models described by Equation 6 in Section 5.2.2:

$$\log(transactions_{ij}) = \beta_0 + \beta NPI_{ij} + \beta weekend_{ij} + \gamma_{0j}.$$

Tables 20 and 21 display the resulting fixed effect coefficients from these models for both MA and MI.

The exponentiation of the intercept term included in each model represents the number of daily POS transactions made by the average individual without any NPI on a weekday. Within each state, this estimate is very similar across the various NPIs, which makes sense considering that each NPI was instituted around the same date. However, looking across

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	3.30	(3.14,3.47)	0.00***
	Effective	0.88	(0.85,0.91)	0.00***
	Weekend	0.86	(0.85,0.87)	0.00***
Workplace Closure	Intercept	3.29	(3.13,3.46)	0.00***
	Recommended	NA	NA	NA
	Required (some)	NA	NA	NA
	Required (all)	0.87	(0.84,0.91)	0.00***
	Weekend	0.85	(0.85,0.86)	0.00***
Stay-at-Home Reqs.	Intercept	3.30	(3.14,3.47)	0.00***
	Recommended	0.87	(0.84,0.91)	0.00***
	Required (some ex.)	NA	NA	NA
	Required (min ex.)	NA	NA	NA
	Weekend	0.85	(0.84,0.86)	0.00***
Gathering Restrictions	Intercept	3.42	(3.25,3.60)	0.00***
	Restrict > 1,000	NA	NA	NA
	Restrict 101-1,000	0.93	(0.89,0.97)	0.00***
	Restrict 11-100	0.82	(0.78,0.86)	0.00***
	Restrict < 10	0.82	(0.77,0.86)	0.00***
	Weekend	0.86	(0.85,0.87)	0.00***

Table 20: Model 6 Results (MA)

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	4.80	(4.62,4.99)	0.00***
	Effective	0.87	(0.84,0.90)	0.00***
	Weekend	0.84	(0.83,0.85)	0.00***
Workplace Closure	Intercept	5.13	(4.94,5.34)	0.00***
	Recommended	NA	NA	NA
	Required (some)	0.84	(0.82,0.87)	0.00***
	Required (all)	0.62	(0.59,0.65)	0.00***
	Weekend	0.83	(0.82,0.84)	0.00***
Stay-at-Home Reqs.	Intercept	4.91	(4.72,5.11)	0.00***
	Recommended	0.94	(0.91,0.97)	0.00***
	Required (some ex.)	0.77	(0.74,0.80)	0.00***
	Required (min ex.)	NA	NA	NA
	Weekend	0.83	(0.83,0.84)	0.00***
Gathering Restrictions	Intercept	5.40	(5.19,5.63)	0.00***
	Restrict > 1,000	NA	NA	NA
	Restrict 101-1,000	1.03	(0.99,1.07)	0.15
	Restrict 11-100	0.87	(0.83,0.91)	0.00***
	Restrict < 10	0.67	(0.64,0.70)	0.00***
	Weekend	0.83	(0.82,0.84)	0.00***

Table 21: Model 6 Results (MI)

states, it is evident that the average individual from MI made more daily POS transactions than the average individual from MA. The exponentiation of the NPI coefficient terms included in each model represents the multiplicative effect of instituting each NPI, controlling for the day of the week. Note that we were unable to obtain estimates for a few stringency levels of the various NPIs due to a lack of data. Almost every exponentiated NPI term is less than one and statistically significant at the $\alpha = 0.05$ level, indicating that each NPI was associated with a decrease in the average individual's number of daily POS transactions among both samples. For each NPI, as the stringency level increases, the NPI coefficient estimate decreases, indicating that both samples of individuals engaged in less daily POS transactions on average as closures, requirements, and restrictions became more stringent.

Next, we fit the class of models described by Equation 7 in Section 5.2.2:

$$\begin{aligned}\log(transactions_{ij}) = \beta_0 + \beta_1 NPI_{ij} + \beta_2 consumer_{ij} \\ + \beta_3 NPI_{ij} consumer_{ij} + \beta_4 weekend_{ij} + \gamma_{0j}.\end{aligned}$$

Recall that the variable *consumer* represents the type of individual consumer in terms of activity level (i.e., *consumer* $\in$ {low, medium, high}), as identified through 3-means clustering (see Section 5.3.1). Also recall that in fitting this model we did not perform one hot encoding, as doing so would require fitting an abundance of parameters and lead to less interpretable results. Instead, each NPI is represented as a continuous variable for which we are only concerned with one unit increases. Tables 22 and 23 display the resulting fixed effect coefficients from these models for both MA and MI.

The exponentiation of the intercept term included in each model represents the number of daily POS transactions made by the average high-frequency consumer without any NPI on a weekday. The exponentiation of the two *consumer* main effects (i.e., low and medium) included in each model represents the multiplicative effect of being classified as a low-frequency or high-frequency consumer, controlling for the stringency level of each NPI

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	12.89	(11.52,14.41)	0.00***
	Emergency	0.66	(0.60,0.74)	0.00***
	Consumer (Low)	0.13	(0.12,0.15)	0.00***
	Consumer (Medium)	0.50	(0.43,0.56)	0.00***
	Emergency × Cons. (Low)	1.59	(1.42,1.78)	0.00***
	Emergency × Cons. (Med)	1.15	(1.02,1.30)	0.02*
	Weekend	0.86	(0.85,0.87)	0.00***
Workplace Closure	Intercept	18.27	(15.21,21.96)	0.00***
	Closure	0.68	(0.61,0.76)	0.00***
	Consumer (Low)	0.09	(0.08,0.12)	0.00***
	Consumer (Medium)	0.45	(0.36,0.55)	0.00***
	Closure × Cons. (Low)	1.50	(1.33,1.68)	0.00***
	Closure × Cons. (Med)	1.13	(1.00,1.28)	0.05*
	Weekend	0.85	(0.85,0.86)	0.00***
Stay-at-Home Reqs.	Intercept	20.18	(16.79,24.26)	0.00***
	Reqs.	0.64	(0.57,0.71)	0.00***
	Consumer (Low)	0.09	(0.07,0.10)	0.00***
	Consumer (Medium)	0.39	(0.32,0.49)	0.00***
	Reqs. × Cons. (Low)	1.60	(1.42,1.80)	0.00***
	Reqs. × Cons. (Med)	1.23	(1.09,1.39)	0.00***
	Weekend	0.85	(0.84,0.86)	0.00***
Gathering Restrictions	Intercept	16.87	(14.62,19.47)	0.00***
	Restrict	0.81	(0.78,0.85)	0.00***
	Consumer (Low)	0.10	(0.08,0.12)	0.00***
	Consumer (Medium)	0.45	(0.38,0.53)	0.00***
	Restrict × Cons. (Low)	1.26	(1.20,1.33)	0.00***
	Restrict × Cons. (Med)	1.08	(1.02,1.14)	0.02*
	Weekend	0.86	(0.85,0.87)	0.00***

Table 22: Model 7 Results (MA)

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	14.68	(13.36,16.14)	0.00***
	Emergency	0.61	(0.56,0.68)	0.00***
	Consumer (Low)	0.19	(0.17,0.21)	0.00***
	Consumer (Medium)	0.52	(0.47,0.58)	0.00***
	Emergency × Cons. (Low)	1.66	(1.50,1.85)	0.00***
	Emergency × Cons. (Med)	1.25	(1.12,1.40)	0.00***
	Weekend	0.84	(0.83,0.85)	0.00***
Workplace Closure	Intercept	20.81	(18.34,23.62)	0.00***
	Closure	0.70	(0.66,0.74)	0.00***
	Consumer (Low)	0.18	(0.16,0.21)	0.00***
	Consumer (Medium)	0.55	(0.47,0.63)	0.00***
	Closure × Cons. (Low)	1.20	(1.13,1.28)	0.00***
	Closure × Cons. (Med)	1.05	(0.98,1.12)	0.17
	Weekend	0.83	(0.82,0.84)	0.00***
Stay-at-Home Reqs.	Intercept	19.52	(17.12,22.25)	0.00***
	Reqs.	0.75	(0.71,0.79)	0.00***
	Consumer (Low)	0.16	(0.14,0.18)	0.00***
	Consumer (Medium)	0.52	(0.45,0.60)	0.00***
	Reqs. × Cons. (Low)	1.27	(1.19,1.35)	0.00***
	Reqs. × Cons. (Med)	1.07	(1.01,1.15)	0.03*
	Weekend	0.83	(0.83,0.84)	0.00***
Gathering Restrictions	Intercept	23.04	(20.25,26.22)	0.00***
	Restrict	0.75	(0.72,0.78)	0.00***
	Consumer (Low)	0.13	(0.12,0.15)	0.00***
	Consumer (Medium)	0.47	(0.41,0.54)	0.00***
	Restrict × Cons. (Low)	1.28	(1.22,1.34)	0.00***
	Restrict × Cons. (Med)	1.10	(1.05,1.15)	0.00***
	Weekend	0.84	(0.83,0.84)	0.00***

Table 23: Model 7 Results (MI)

and the day of the week. In each model, both coefficients are statistically significant and less than one, indicating that the average low-frequency and medium-frequency consumer completes less daily POS transactions than the average high-frequency consumer, regardless of NPI stringency level and the day of the week. The exponentiation of the NPI main effect included in each model represents the multiplicative effect of instituting each NPI on high-frequency consumers, controlling for the day of the week. In each model, the coefficient associated with this term is both statistically significant and less than one, indicating that each NPI is associated with a decrease in the average high-frequency consumer's number of daily POS transactions. The exponentiation of the $NPI \times consumer$ terms (i.e., low and medium) in each model represents the multiplicative effect of the interaction between these two variables. In most models, both terms are statistically significant and greater than one, indicating that the association between each NPI and daily POS transactions was more positive among the average medium and low-frequency consumer.

Finally, we fit the class of models described by Equation 8 in Section 5.2.2:

$$\begin{aligned}\log(transactions_{ij}) = \beta_0 + \beta_1 NPI_{ij} + \beta_2 state_{ij} \\ + \beta_3 NPI_{ij} state_{ij} + \beta_4 weekend_{ij} + \gamma_{0j}.\end{aligned}$$

Recall that the variable *state* represents whether the observation is a county within MA or MI. Also recall that in fitting this model we did not perform one hot encoding, as doing so would require fitting an abundance of parameters and lead to less interpretable results. Instead, each NPI is represented as a continuous variable for which we are only concerned with one unit increases. Table 24 displays the resulting fixed effect coefficients from each individual model.

NPI	Fixed Effects	e^{β}	95% CI	p-value
National Emergency	Intercept	3.50	(3.35,3.66)	0.00***
	Emergency	0.88	(0.85,0.91)	0.00***
	State (MI)	1.30	(1.23,1.39)	0.00***
	Emergency × State (MI)	1.00	(0.95,1.06)	0.89
	Weekend	0.85	(0.84,0.85)	0.00***
Workplace Closure	Intercept	4.04	(3.78,4.32)	0.00***
	Closure	0.87	(0.84,0.90)	0.00***
	State (MI)	1.60	(1.47,1.74)	0.00***
	Closure × State (MI)	0.89	(0.86,0.94)	0.00***
	Weekend	0.84	(0.83,0.85)	0.00***
Stay-at-Home Reqs.	Intercept	4.07	(3.81,4.36)	0.00***
	Requirements	0.86	(0.83,0.90)	0.00***
	State (MI)	1.36	(1.24,1.48)	0.00***
	Reqs. × State (MI)	1.01	(0.97,1.05)	0.74
	Weekend	0.84	(0.84,0.85)	0.00***
Gathering Restrictions	Intercept	3.79	(3.59,4.00)	0.00***
	Restrictions	0.94	(0.92,0.95)	0.00***
	State (MI)	1.57	(1.45,1.69)	0.00***
	Restrict × State (MI)	0.93	(0.91,0.95)	0.00***
	Weekend	0.85	(0.84,0.85)	0.00***

Table 24: Model 8 Results

The exponentiation of the intercept term included in each model represents the number of daily POS transactions made by the average MA individual without any NPI on a weekday. The exponentiation of the NPI main effect included in each model represents the multiplicative effect of instituting each NPI, controlling for the resident state of the individual and the day of the week. In each model, the coefficient associated with this term is both statistically significant and less than one, indicating that each NPI is associated with a decrease in daily POS transactions on average within the two samples of individuals from each state. The exponentiation of the *state* main effect included in each model represents the multiplicative effect of an an individual being a resident of MI opposed to MA, controlling for the stringency level of each NPI and the day of the week. In each model, this coefficient is both statistically significant and greater than one, indicating that the average individual from MI completes more daily POS transactions than the average individual from MS, regardless of NPI stringency level and the day of the week. The exponentiation of the $NPI \times state$ term in each model represents the multiplicative effect of the interaction between these two variables. The coefficient estimate of this interaction term for workplace closures and gathering restrictions is statistically significant and less than one, suggesting that the association between these two NPIs and daily POS transactions was more negative within MI.

5.3.3 COVID-19 Modeling

Using the *nlme* package in R [66, 64], we fit each of the linear models with an autoregressive covariance structure, as described in Section 5.2.3, under the restricted maximum

likelihood (REML) approach. Note that we fit each model using data from during the national emergency (i.e., 13$^{\text{th}}$ March - 30$^{\text{th}}$ May, 2020) subset by state (i.e., MA and MI) and mobility lags ranging from 6-16 days:

$$\begin{aligned} case\ growth\ rate_i = \beta_0 + \beta_1 \Delta transactions_i + \beta_2 test\ rate_i \\ + \beta_3 weekend_i, \end{aligned}$$

Tables 25 and 26 display the resulting fixed effect coefficients from each model for both MA and MI.

Lag Period	Fixed Effects	β	95% CI	p-value
06 Days	Intercept	28.99	(12.05,45.93)	0.00***
	Δ Transactions	0.38	(0.00,0.75)	0.05*
	Test Rate	-0.09	(-0.19,0.01)	0.08·
	Weekend	-33.82	(-52.33,-15.31)	0.00***
08 Days	Intercept	30.36	(12.89,47.83)	0.00***
	Δ Transactions	0.16	(-0.20,0.52)	0.39
	Test Rate	-0.10	(-0.21,-0.00)	0.05*
	Weekend	-38.10	(-56.45,-19.76)	0.00***
10 Days	Intercept	29.40	(11.58,47.22)	0.00***
	Δ Transactions	0.20	(-0.20,0.59)	0.33
	Test Rate	-0.09	(-0.20,0.02)	0.12
	Weekend	-40.68	(-58.35,-23.02)	0.00***
12 Days	Intercept	25.40	(5.97,44.83)	0.01**
	Δ Transactions	0.27	(-0.11,0.64)	0.17
	Test Rate	-0.07	(-0.19,0.05)	0.24
	Weekend	-34.71	(-53.95,-15.47)	0.00***
14 Days	Intercept	17.32	(-4.14,38.79)	0.11
	Δ Transactions	0.45	(0.03,0.86)	0.04*
	Test Rate	-0.02	(-0.15,0.10)	0.71
	Weekend	-25.63	(-47.56,-3.69)	0.02*
16 Days	Intercept	24.51	(4.22,44.81)	0.02*
	Δ Transactions	0.28	(-0.13,0.68)	0.18
	Test Rate	-0.06	(-0.19,0.07)	0.38
	Weekend	-38.61	(-56.27,-20.94)	0.00***

Table 25: Model 9 Results (MA)

Lag Period	Fixed Effects	β	95% CI	p-value
06 Days	Intercept	18.15	(1.33,34.97)	0.04*
	Δ Transactions	0.32	(-0.12,0.76)	0.16
	Test Rate	-0.12	(-0.25,0.01)	0.07
	Weekend	-20.40	(-39.75,-1.04)	0.04*
08 Days	Intercept	13.77	(-2.42,29.95)	0.09·
	Δ Transactions	0.15	(-0.28,0.58)	0.50
	Test Rate	-0.10	(-0.23,0.02)	0.11
	Weekend	-19.91	(-39.51,-0.30)	0.05*
10 Days	Intercept	13.80	(-2.04,29.65)	0.09·
	Δ Transactions	0.15	(-0.26,0.57)	0.47
	Test Rate	-0.10	(-0.22,0.03)	0.13
	Weekend	-20.83	(-40.63,-1.04)	0.04*
12 Days	Intercept	25.40	(5.97,44.83)	0.01*
	Δ Transactions	0.27	(-0.11,0.64)	0.17
	Test Rate	-0.07	(-0.19,0.05)	0.24
	Weekend	-34.71	(-53.95,-15.47)	0.00***
14 Days	Intercept	12.21	(-2.05,26.48)	0.09·
	Δ Transactions	0.12	(-0.27,0.51)	0.54
	Test Rate	-0.09	(-0.21,0.03)	0.16
	Weekend	-19.57	(-39.22,0.08)	0.05*
16 Days	Intercept	11.36	(-2.85,25.57)	0.12
	Δ Transactions	0.06	(-0.33,0.44)	0.77
	Test Rate	-0.09	(-0.21,0.03)	0.16
	Weekend	-19.92	(-39.64,-0.20)	0.05*

Table 26: Model 9 Results (MI)

The exponentiation of the intercept term included in each model represents the average COVID-19 case growth rate without any changes in consumer mobility and COVID-19 testing on a weekday. Within each sample, this estimate is very similar across the various lag periods. However, looking across samples, it is evident that on average the COVID-19 case growth rate was larger in Suffolk, MA, than Washtenaw, MI. The coefficient estimate of Δ transactions for each lag period represents the average change in the daily COVID-19 case growth rate associated with a 1% change in consumer mobility, as measured by POS transactions, compared to pre-pandemic baselines, controlling for the testing rate and the day of the week. Within the sample of Suffolk (MA) residents, our results suggest that the strongest positive association between changes in consumer mobility and COVID-19 case growth rates appears at lags of six and fourteen days. Specifically, a 10% percent decrease in the 6-day lagged change in consumer mobility was associated with a 3.8% decrease in the daily COVID-19 case growth rate. Similarly, a 10% percent decrease in the 6-day lagged change in consumer mobility was associated with a 4.5% decrease in the COVID-19 case growth rate.

Within the sample of MI residents, our results indicate that there was not a significant association between changes in consumer mobility and COVID-19 case growth rates at any lag.

5.4 Discussion

Cluster analysis indicated that the individuals within our final dataset fell into three distinct sub-groups based on their pre-pandemic spending behavior: low, medium, and high-frequency consumers. Low-level consumers carried out anywhere from (0.40,7.72) average daily total POS transactions prior to the pandemic. Medium-level consumers fell between (7.18,18.70) average daily total POS transactions prior to the pandemic. High-level consumers carried out anywhere from (12.90,39.60) average daily total POS transactions prior to the pandemic. These ranges suggest that our labeling scheme is not perfect, as there is some overlap between the medium and high-level consumer groups. However, this overlap primarily occurs between outliers within each cluster (see Figure 31). Perhaps a better clustering would also distinguish consumers by the time of day they typically shop as well as the day of the week, yet our clustering algorithm was unable to accurately detect this.

Consumer mobility models detected the presence of a statistically significant negative association between all four NPIs of interest (i.e., national emergency, workplace closure, stay-at-home requirements, and gathering restrictions) and daily individual-level POS transaction counts, adjusting for trends resulting from the time-series nature of the data. Stringent workplace closures and gathering restrictions exhibited the strongest negative associations with daily POS transaction counts in both states. Specifically, required workplace closures were associated with a reduction of 13% in the average MA individual's daily POS transaction count, and a reduction of 38% in the average MI individual's daily POS transaction count. Similarly, restrictions on gatherings of greater than 10 people were associated with a reduction of 18% in the average MA individual's daily POS transaction count, and a reduction of 33% in the average MI individual's daily POS transaction count. Looking across states, the association between these two NPIs (i.e., workplace closures and gathering restrictions) and individual-level daily POS transactions was more negative within MI than MA. Looking within both states, each NPI exhibited a dose-response relationship with daily POS transactions (i.e., incremental increases in the stringency of NPIs produced decreases in daily POS transactions), providing increased evidence of a causal relationship. These results support previous findings that the early implementation of a broad range of NPIs reduced human mobility.

Our consumer mobility models possess a few limitations. First, several NPIs were implemented within a short period of time during the early stages of the COVID-19 pandemic, making it difficult to isolate the impact of each individual policy intervention. Second, although we attempted to avoid this issue by excluding user identities with more than fifty transactions on any given day, the data used in our analysis likely contained both personal

and business credit lines, as evidenced by multiple user identities with an abnormally large average daily transaction count (see Table 15). This aspect of the dataset creates a bit of uncertainty concerning our study population. Third, although our models include several control variables and their results provide some evidence of causality, the possibility for omitted variable bias certainly exists, as is always the case in observational studies.

COVID-19 models identified the presence of a statistically significant positive association between changes in residents' consumer mobility patterns and daily COVID-19 case growth rates within Suffolk County, MA, adjusting for the COVID-19 testing rate and trends resulting from the time-series nature of the data. In a similar manner to prior studies, we created a feature to represent changes in residents' consumer mobility patterns, specifically the percent change in aggregated daily POS transaction counts compared to pre-pandemic baseline days. Using data from Suffolk County, MA, we uncovered a statistically significant positive association between both 6 and 14 day lagged changes in consumer mobility and COVID-19 case growth rates. Specifically, a 10% percent decrease in the 6-day lagged change in consumer mobility was associated with a 3.8% decrease in the daily COVID-19 case growth rate. Similarly, a 10% percent decrease in the 6-day lagged change in consumer mobility was associated with a 4.5% decrease in the COVID-19 case growth rate. The relationship between lagged changes in consumer mobility and COVID-19 case growth rates was not significant at any lag within Washtenaw County, Michigan. Contrary to previous studies, who found much stronger relationships between changes in human mobility and the COVID-19 spread, these results suggest that daily POS transactions are a limited mobility metric.

The COVID-19 modeling portion of our study also has a few limitations. First, as mentioned in the previous paragraph, daily POS credit card transactions are a limited measure of human mobility patterns. Additionally, there exist other potential COVID-19 mitigation factors beyond mobility reduction (e.g., mask-wearing, hand-washing, case tracing) that our analysis does not consider due to a lack of data. Second, the COVID-19 case data used in our analysis may contain errors due to reporting issues and limited testing capacity common among many medical facilities during the initial disease outbreak. Third, the COVID-19 testing data used in our analysis is representative of state-level totals, as both state governments failed to record county-level test counts during the early stages of the pandemic.

6 Conclusion

US state governments relied on a complex combination of non-pharmaceutical interventions (NPIs) in order to control the initial outbreak of COVID-19. NPIs led to varied patterns of human movement and behavioural changes throughout the country. Simultaneously, the progression and intensity of the COVID-19 pandemic fluctuated considerably by location. In conjunction, these two factors make measuring the effect of NPI's on con-

sumer mobility patterns and the spread of COVID-19 a non-trivial task, which explains the popularity of studies aimed at achieving this end. Our work is unique in comparison to these other studies primarily for one reason–we conducted our analysis using a novel dataset made privately available by BDEX that reflected consumer mobility patterns of individuals from urban regions within two states, Massachusetts (MA) and Michigan (MI). Most other analyses were carried out using aggregated nation-wide mobility reports made publicly available by Google, SafeGraph, Teralytics, and Unacast. Using the dataset provided by BDEX in conjunction with published NPI and COVID-19 datasets, we quantified the relationship between government-mandated NPIs and consumer mobility patterns (as represented by point-of-sale (POS) credit card transactions) and the relationship between subsequent changes in consumer mobility patterns compared to pre-pandemic baselines and the COVID-19 case growth rate at both the county and individual-level.

Our county-level study uncovered a statistically significant negative association between four NPIs (i.e., national emergency, workplace closures, stay-at-home-requirements, and gathering restrictions) and daily POS transaction rates. Specifically, stringent workplace closures and gathering restrictions displayed the strongest negative association with daily county-level POS transaction rates. Each association did not differ by the political affiliation, demographic makeup, or state of the county. Within both states, each NPI exhibited a dose-response relationship with daily POS transaction rates (i.e., incremental increases in the stringency of NPIs produced decreases in daily POS transaction rates), providing increased evidence of a causal relationship. Subsequent lagged changes in consumer mobility patterns (as represented by POS transaction rates) compared to pre-pandemic baselines exhibited a statistically significant positive association with daily COVID-19 case growth rates. The optimal lag period varied, as MI counties exhibited their strongest association at 8 days and MA counties exhibited their strongest association at 12 days.

Our individual-level study also detected a statistically significant negative association between the four NPIs referenced previously and daily POS transaction counts. Likewise, workplace closures and gathering restrictions exhibited the strongest negative association with daily individual POS transaction counts. Looking across states, the association between these two NPIs and POS transactions was more negative within our sample of MI residents than MA residents. Looking within both states, each NPI displayed a dose-response relationship with transactions, again providing increased evidence of a causal relationship. Subsequent lagged changes in residents' consumer mobility patterns (as represented by POS transaction counts) compared to pre-pandemic baselines exhibited a statistically significant positive association with daily COVID-19 case growth rates within Suffolk County, MA. The optimal lag period was unclear, as both 6 and 14 day lagged changes in consumer mobility patterns were significantly associated with decreased COVID-19 case growth rates.

Overall, the results of both our county and individual-level analyses support previous findings that NPIs, especially workplace closures and gathering restrictions, reduced human

mobility during the initial COVID-19 outbreak. However, unlike previous studies, which often uncovered stronger relationships between changes in human mobility patterns and the spread of COVID-19, our results indicate that daily POS transactions are a limited mobility metric. Note that these results are tied to a study of large random samples of individuals from two cities (Boston, MA, and Ann Arbor, MI) during roughly the first three months following the declaration of COVID-19 as a national emergency. Future studies should seek to analyze more complete measures of consumer mobility patterns across various states over a longer duration.

References

[1] J. Michael Ryan. "Timeline of COVID-19". In: *COVID-19: Global Pandemic, Societal Responses, Ideological Solutions* (2021), pp. xiii–xxxii.

[2] AJMC Staff. *A Timeline of COVID-19 Developments in 2020*. 2021. URL: https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020.

[3] H. Dele Davies. "Unlike COVID-19, Ebola Outbreak in the DRC Disproportionately Affected Children". In: *AAP News* (2020). URL: https://publications.aap.org/aapnews/news/14545?autologincheck=redirected?nfToken=00000000-0000-0000-0000-000000000000.

[4] Brianna Abbott and Jason Douglas. "How Deadly Is Covid-19? Researchers Are Getting Closer to an Answer". In: *Wall Street Journal* (2020). URL: https://www.wsj.com/articles/how-deadly-is-covid-19-researchers-are-getting-closer-to-an-answer-11595323801.

[5] Katie Pearce. "What is Social Distancing and How Can it Slow the Spread of COVID-19". In: *Johns Hopkins Magazine* (2020). URL: https://hub.jhu.edu/2020/03/13/what-is-social-distancing/#:~:text=The%5C%20CDC%5C%20defines%5C%20social%5C%20distancing,no%5C%20hugs%5C%2C%5C%20no%5C%20handshakes.%5C%22.

[6] Laura Matrajt and Tiffany Leung. "Evaluating the Effectiveness of Social Distancing Interventions to Delay or Flatten the Epidemic Curve of Coronavirus Disease". In: *Emerging Infectious Diseases* 26.8 (2020), pp. 1740–1748. URL: https://stacks.cdc.gov/view/cdc/91562.

[7] Ann O'M Bowman and James H McKenzie. "Managing a Pandemic At a Less Than Global Scale: Governors Take the Lead". In: *The American Review of Public Administration* 50.6-7 (2020), pp. 551–559. URL: https://doi.org/10.1177/0275074020941700.

[8] Alba Mendez-Brito, Charbel El Bcheraoui, and Francisco Pozo-Martin. "Systematic Review of Empirical Studies Comparing the Effectiveness of Non-pharmaceutical Interventions Against COVID-19". In: *Journal of Infection* 83.3 (2021), pp. 281–293. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8214911/.

[9] Xue Zhang and Mildred E Warner. "COVID-19 Policy Differences Across US states: Shutdowns, Reopening, and Mask Mandates". In: *International Journal of Environmental Research and Public Health* 17.24 (2020), p. 9520. URL: https://pubmed.ncbi.nlm.nih.gov/33353158/.

[10] Mark J Siedner et al. "Social Distancing to Slow the US COVID-19 Epidemic: Longitudinal Pretest–Posttest Comparison Group Study". In: *PLoS Medicine* 17.8 (2020). URL: https://doi.org/10.1371/journal.pmed.1003244.

[11] Nickolas Dreher et al. "Policy Interventions, Social Distancing, and SARS-CoV-2 Transmission in the United States: A Retrospective State-level Analysis". In: *The American Journal of the Medical Sciences* 361.5 (2021), pp. 575–584. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7833753/.

[12] Yun Li et al. "The Impact of Policy Measures on Human Mobility, COVID-19 cases, and Mortality in the US: A Spatiotemporal Perspective". In: *International Journal of Environmental Research and Public Health* 18.3 (2021). URL: https://www.mdpi.com/1660-4601/18/3/996.

[13] Andrew M Olney et al. "Estimating the Effect of Social Distancing Interventions on COVID-19 in the United States". In: *American Journal of Epidemiology* 190.8 (2021), pp. 1504–1509. URL: https://pubmed.ncbi.nlm.nih.gov/33406533/.

[14] Charles Courtemanche et al. "Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate". In: *Health Affairs* 39.7 (2020), pp. 1237–1246. URL: https://doi.org/10.1377/hlthaff.2020.00608.

[15] Senan Ebrahim et al. "Reduction of COVID-19 Incidence and Nonpharmacologic Interventions: Analysis Using a US County–level Policy Data Set". In: *Journal of Medical Internet Research* 22.12 (2020). URL: https://www.jmir.org/2020/12/e24614.

[16] Aliea M Jalali et al. "Delayed Interventions, Low Compliance, and Health Disparities Amplified the Early Spread of COVID-19". In: *medRxiv* (2020). URL: https://www.medrxiv.org/content/early/2020/08/04/2020.07.31.20165654.

[17] Hamada S Badr et al. "Association Between Mobility Patterns and COVID-19 Transmission in the USA: A Mathematical Modelling Study". In: *The Lancet Infectious Diseases* 20.11 (2020), pp. 1247–1254. DOI: https://doi.org/10.1016/S1473-3099(20)30553-3.

[18] Oliver Gatalo et al. "Associations Between Phone Mobility Data and COVID-19 Cases". In: *The Lancet Infectious Diseases* 21.5 (2020), e111. DOI: https://doi.org/10.1016/S1473-3099(20)30725-8.

[19] Jonathan I Dingel and Brent Neiman. "How Many Jobs Can Be Done at Home?" In: *Journal of Public Economics* 189 (2020). URL: https://doi.org/10.1016/j.jpubeco.2020.104235.

[20] John McLaren and Su Wang. *Effects of Reduced Workplace Presence on COVID-19 Deaths: An Instrumental-Variables Approach*. Tech. rep. National Bureau of Economic Research, 2020. URL: https://www.nber.org/papers/w28275.

[21] David Welsch. "The Impact of Mask Usage on COVID-19 Deaths: Evidence from US Counties Using a Quasi-Experimental Approach". In: *The BE Journal of Economic Analysis & Policy* 22.1 (2022), pp. 1–28. URL: https://doi.org/10.1515/bejeap-2021-0157.

[22] Gareth James et al. *Statistical Learning*. Vol. 112. Springer, 2013.

[23] Julian J Faraway. "Preface". In: *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman and Hall/CRC, 2016.

[24] G. Rodriguez. "Poisson Models for Count Data". In: *Lecture Notes on Generalized Linear Models*. 2007. URL: https://data.princeton.edu/wws509/notes/.

[25] Julian J Faraway. "Count Regression". In: *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman and Hall/CRC, 2016.

[26] Stephen Parry. *To Offset or Not: Using Offsets in Count Models*. Tech. rep. Cornell Statistical Consulting Unit, 2020. URL: https://cscu.cornell.edu/wp-content/uploads/94_offsets.pdf.

[27] Edward E Gbur et al. "Generalized Linear Mixed Models". In: *Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences*. Vol. 156. John Wiley & Sons, 2020.

[28] Joseph M Hilbe. "Negative Binomial Regression". In: *Modeling Count Data*. Cambridge University Press, 2014.

[29] Andrew Gelman and Jennifer Hill. "Multilevel Structures". In: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.

[30] Jon Starkweather. "Linear Mixed Effects Modelling Using R". In: *Unpublished Manuscript* (2010). URL: http://bayes.acs.unt.edu:8083/BayesContent/class/Jon/Benchmarks/LinearMixedModels_JDS_Dec2010.pdf.

[31] Robert R Corbeil and Shayle R Searle. "Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model". In: *Technometrics* 18.1 (1976), pp. 31–38. URL: https://www.stat.cmu.edu/technometrics/70-79/VOL-18-01/v1801031.pdf.

[32] Ashenafi A Yirga et al. "Negative Binomial Mixed Models for Analyzing Longitudinal CD4 Count Data". In: *Scientific reports* 10.1 (2020). DOI: https://doi.org/10.1038/s41598-020-73883-7.

[33] Andrew Gelman and Jennifer Hill. "Multilevel Generalized Linear Models". In: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.

[34] José C Pinheiro and Douglas M Bates. "Approximations to the log-likelihood function in the nonlinear mixed-effects model". In: *Journal of computational and Graphical Statistics* 4.1 (1995), pp. 12–35. URL: https://www.tandfonline.com/doi/abs/10.1080/10618600.1995.10474663.

[35] Ke Ju et al. "Laplace approximation, penalized quasi-likelihood, and adaptive Gauss–Hermite quadrature for generalized linear mixed models: towards meta-analysis of binary outcome with sparse data". In: *BMC medical research methodology* 20.1 (2020), pp. 1–11. URL: https://doi.org/10.1186/s12874-020-01035-6.

[36] C Kincaid. "Guidelines for Selecting the Covariance Structure in Mixed Model Analysis". In: *Proceedings of the Thirtieth Annual SAS Users Group International Conference*. Vol. 30. 8. SAS Institute Inc Cary NC. 2005. URL: https://support.sas.com/resources/papers/proceedings/proceedings/sugi30/198-30.pdf.

[37] Ramon C Littell, Jane Pendergast, and Ranjini Natarajan. "Modelling Covariance Structure in the Analysis of Repeated Measures Data". In: *Statistics in Medicine* 19.13 (2000), pp. 1793–1819. URL: https://doi.org/10.1002/1097-0258(20000715)19:13%3C1793::AID-SIM482%3E3.0.CO;2-Q.

[38] José Pinheiro and Douglas Bates. "Extending the Basic Linear Mixed-Effects Model". In: *Mixed-Effects Models in S and S-PLUS*. Springer Science & Business Media, 2006.

[39] Rui Xu and Donald Wunsch. "Survey of Clustering Algorithms". In: *IEEE Transactions on Neural Networks* 16.3 (2005), pp. 645–678. DOI: 10.1109/TNN.2005.845141..

[40] Charu C. Aggarwal and Chandan K. Reddy. *A Survey of Partitional and Hierarchical Clustering Algorithms*. CRC Press, 2014.

[41] Sabine Landau et al. "Measurement of Proximity". In: *Cluster Analysis*. John Wiley & Sons, 2011.

[42] Alboukadel Kassambara. *Determining the Optimal Number of Clusters*. Vol. 1. Sthda, 2017.

[43] Robert Thorndike. "Who Belongs in the Family?" In: *Psychometrika* 18.4 (1953), pp. 267–276. DOI: 10.1007/BF02289263.

[44] Peter J Rousseeuw. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis". In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65. URL: https://doi.org/10.1016/0377-0427(87)90125-7.

[45] Robert Tibshirani, Guenther Walther, and Trevor Hastie. "Estimating the Number of Clusters in a Data Set Via the Gap Statistic". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.2 (2001), pp. 411–423. URL: https://doi.org/10.1111/1467-9868.00293.

[46] Leonard Kaufman and Peter J Rousseeuw. *Agglomerative Nesting (Program AGNES)*. John Wiley & Sons, 2009.

[47] Alboukadel Kassambara. "Hierarchical Clustering". In: *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. Vol. 1. Sthda, 2017.

[48] Sabine Landau et al. "Hierarchical Clustering". In: *Cluster Analysis*. John Wiley & Sons, 2011.

[49] Fionn Murtagh and Pedro Contreras. "Algorithms for Hierarchical Clustering: An Overview". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2.1 (2012), pp. 86–97. URL: http://i2pc.es/coss/Docencia/SignalProcessingReviews/Murtagh2012.pdf.

[50] T Soni Madhulatha. "An Overview on Clustering Methods". In: *arXiv* (2012). DOI: https://doi.org/10.48550/arXiv.1205.1117.

[51] Joe H Ward Jr. "Hierarchical Grouping to Optimize an Objective Function". In: *Journal of the American statistical association* 58.301 (1963), pp. 236–244. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845.

[52] Leonard Kaufman and Peter J Rousseeuw. *Divisive Analysis (Program DIANA)*. John Wiley & Sons, 2009.

[53] Alboukadel Kassambara. "K-Means Clustering". In: *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. Vol. 1. Sthda, 2017.

[54] Sabine Landau et al. "Optimization Clustering Techniques". In: *Cluster Analysis*. John Wiley & Sons, 2011.

[55] J MacQueen. "Some Methods for Classification and Analysis of Multivariate Observations". In: *Proc. 5th Berkeley Symposium on Math., Stat., and Prob.* 1965, p. 281. URL: https://digitalassets.lib.berkeley.edu/math/ucb/text/math_s5_v1_article-17.pdf.

[56] Laura Hallas et al. *Variation in US States' COVID-19 Policy Responses 3.0*. Tech. rep. Blavatnik School of Government–University of Oxford, 2022. URL: https://www.bsg.ox.ac.uk/sites/default/files/2021-05/BSG-WP-2020-034-v3.pdf.

[57] *The Story of BDEX*. Accessed: 2022-08-01. URL: https://www.bdex.com/about/.

[58] Massachusetts Government. *COVID-19 Response Reporting*. 2022. URL: https://www.mass.gov/info-details/covid-19-response-reporting.

[59] Michigan Government. *Coronavirus*. 2022. URL: https://www.michigan.gov/coronavirus/stats.

[60] MIT Election Data and Science Lab. *U.S. President 1976–2020*. Version DRAFT VERSION. 2017. URL: https://doi.org/10.7910/DVN/42MVDX.

[61] US Census Bureau. *Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2010 to July 1, 2019*. 2020. URL: https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html.

[62] Bo Wang et al. "Airborne Particulate Matter, Population Mobility and COVID-19: A Multi-city Study in China". In: *BMC public health* 20.1 (2020), pp. 1–10. URL: https://doi.org/10.1186/s12889-020-09669-3.

[63] Mollie E. Brooks et al. "glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling". In: *The R Journal* 9.2 (2017), pp. 378–400. URL: https://journal.r-project.org/archive/2017/RJ-2017-066/index.html.

[64] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2022. URL: https://www.R-project.org/.

[65] Kasper Kristensen et al. “TMB: Automatic Differentiation and Laplace Approximation”. In: *Journal of Statistical Software* 70.5 (2016). DOI: 10.18637/jss.v070.i05.

[66] Jose Pinheiro et al. “Linear and Nonlinear Mixed Effects Models”. In: *R package version* 3.57 (2007), pp. 1–89.

[67] Martin Maechler et al. *cluster: Cluster Analysis Basics and Extensions*. 2022. URL: https://CRAN.R-project.org/package=cluster.

[68] Alboukadel Kassambara and Fabian Mundt. “Package ‘factoextra’”. In: *Extract and Visualize the Results of Multivariate Data Analyses* 76.2 (2017).