Understanding the Spatiotemporal Heterogeneities in the Associations between COVID-19 Infections and both Human Mobility and Close Contacts in the United States

Wen Ye
Geospatial Data Science Lab, University of Wisconsin
Madison, WI, USA
wye35@wisc.edu

Song Gao
Geospatial Data Science Lab, University of Wisconsin
Madison, WI, USA
song.gao@wisc.edu

ABSTRACT

It has been well-established that human mobility has an inseparable relationship with COVID-19 infections. As social-distancing and stay-at-home orders lifted and data availability increased, our knowledge on how human behaviors including mobility and close interpersonal contacts associate with the pandemic progression also needs to stay updated. In this paper, we examine the relationship of COVID-19 daily transmissibility measured by the total confirmed cases and the effective reproduction number (R_t) with the two indices that provide human behavior insights: Cuebiq Mobility Index (CMI) and Cuebiq Contact Index (CCI). The correlations between each index and COVID-19 infections are evaluated using the Maximal Information Coefficient (MIC) which is powerful in capturing complex relationships. Moreover, the study period is segmented into three periods by Bayesian Change Point Detection to examine temporal heterogeneity and the mainland US states are grouped into three distinct clusters using the KShape clustering algorithm to further examine spatial heterogeneity. The CCI and CMI exhibited very different patterns and we found significant temporal and spatial heterogeneities among the relationships between the two indices and COVID-19 infection rate. Although human mobility has demonstrated high correlation with COVID-19 infection rate in 2020, close contacts became much more correlated with COVID-19 infection than mobility in 2021. However, states in the Plains and Rocky Mountains area are exceptions to this observation. During the first wave in 2020, it is also shown that mobility has a high impact on states outside of Farwest and Southeast than those states within that region.

CCS CONCEPTS

• **Information systems → Spatial-temporal systems**; **Data analytics**; • **Mathematics of computing → Bayesian computation**.

KEYWORDS

time series analysis, clustering, human mobility, spatial epidemiology, bayesian inference, COVID-19, social computing

SpatialEpi '22, November 1, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9543-4/22/11…$15.00
https://doi.org/10.1145/3557995.3566117

ACM Reference Format:
Wen Ye and Song Gao. 2022. Understanding the Spatiotemporal Heterogeneities in the Associations between COVID-19 Infections and both Human Mobility and Close Contacts in the United States. In *The 3rd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi '22), November 1, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3557995.3566117

1 INTRODUCTION

Continuing to develop our knowledge on how human behaviors interact with the pandemic is critical for containment of the spread of COVID-19 virus and preparing better for future respiratory infectious diseases. It has been well-established that human behaviors have an inseparable relationship with COVID-19 infection rate [2]. However, most of the works were done only for certain periods in 2020 and was focused solely on the relationship between human mobility and COVID-19 infection [1, 14, 18, 40]. As social-distancing and stay-at-home orders lifted and data availability increased, the correlation between COVID-19 infections and human behaviors should be examined in more aspects together with the story of continued pandemic in 2021 and beyond. Existing approaches mainly look at human mobility through mobile device data from location-based services or business insight companies and calculate movement metrics (e.g., travel distance, number of trips or place visits, and dwell time) for geographic units of interest by looking at the origin and destination of user's trips [3, 11, 21, 25]. However, reduced mobility doesn't necessarily ensure that people follow the social (physical) distancing guidelines. Due to the mobile phone GPS location horizontal error and uncertainty, such social (physical) distancing patterns cannot be identified from the use of aggregated movement metrics. But individual-level interpersonal close contacts (i.e., spatiotemporal co-location patterns) can be estimated using mobile device data or through digital contact tracing [41] and have shown better association with the infection rates than the mobility metrics [8]. The spatiotemporal variations of such realtionships are still unknown though.

To fill such a research gap, in this paper, we aim to understand the temporal and spatial heterogeneities among the complex non-linear relationships between the two indices (mobility index and close contact index) and COVID-19 infections. Since the assumptions for linear regression models are extremely stringent, a lot of research then focused on non-linear models to capture the relationships between human behaviors and COVID-19 infections [43]. We will assess the relationships using the Maximal Information Coefficient (MIC) which is powerful in capturing non-linear relationships. This

will also provide valuable insights for relevant agencies when making decisions on future non-pharmaceutical interventions to a pandemic. The main contribution of this paper is providing an extended timely summary of the correlation between human behaviors and COVID-19 infections up to the end of 2021 and investigating the temporal and spatial heterogeneities in these relationships from looking at different phases of the COVID-19 pandemic and different regions of the United States.

The remainder of the paper is organized as follows. Section 2 provides a literature review on related work. We then introduce the data we used in this study in Section 3 and outline the detailed methods including Bayesian Change Point Detection and KShape clustering that are employed to examine temporal and spatial aspects of our data, and Maximal Information Coefficient in Section 4, respectively. Section 5 presents the experiments and results that we conduct in the US. Finally, Section 6 summarizes this work and shares our insights on future works.

2 RELATED WORK

2.1 Sources of Human Mobility Data in COVID-19 pandemic

For most of the studies that examine large-scale human mobility patterns, the most widely used are mobile device data given its large population coverage and low cost [15, 16, 36, 37]. SafeGraph (https://www.safegraph.com) and Google Mobility Reports (https://www.google.com/covid19/mobility) are the two most widely used sources in the COVID-19 pandemic due to their wide region coverage and open data policy to support research. Google Mobility Reports has the advantage of the massive amount of Google Map users and generate mobility insights from users' Location History [27]. SafeGraph generates mobility insights by estimating foot traffic at millions of points-of-interest in the US [3]. Others utilized public transit data, survey data, and census data to study human mobility patterns. In addition, index-based mobility measures also became popular because there is no need to calculate mobility inflow and outflow as an extra step and it's easy to compare between states or counties. The two widely used mobility indices were Tencent Mobility Index and Baidu Mobility Index, especially during the early stage of the pandemic. However, both indices only cover mainland China [43]. Cuebiq Mobility Index and Descartes Lab Mobility Index which cover the US were later employed by many studies as indicators of mobility including the New York Times [13, 19, 29]. In addition, some studies constructed their own index such as Social Contact Index (SCI) and Social Distancing Index (SDI). SCI is constructed by calculating the average number of users that a user is co-located with during a specified time period [42]. SDI looks at users' daily number of personal trips and the percentage of staying at home and out-of-county trips as indicators of social distancing in a geographic area of interest. A high value of SDI means that people are practicing social-distancing a lot while a low value indicates that people are not practicing social-distancing very well [30]. Overall, Cuebiq has the advantage that it is readily available and it has interpersonal close contacts insights in addition to the mobility insights.

2.2 Assessing Relationship between Human Behaviors and COVID-19 Pandemic

Numerous studies have put efforts into assessing relationship between human behaviors and COVID-19 infections and have pointed out a positive relationship between the increase of mobility and the growth of infection rate unanimously [14, 21, 40, 42]. From the modelling side, Xiong et al. used a simultaneous equations model (SEM) with time-varying coefficients and demonstrated a dynamic positive relationship between mobility inflow and new COVID-19 cases [40]. Kraemer et al. has shown that mobility could explain the spread distribution of COVID-19 infection in China fairly well at the early phases of COVID-19 onset adopting a generalized linear regression (GLR) model [24]. Wu et al. applied the Geographically Weighted Regression (GWR) and found static positive influence on COVID-19 infections from factors such as population and age structure [39]. Hou et al. [18] developed a human mobility-augmented stochastic SEIR Model for the intracounty modeling of COVID-19 infection and assessed the spatial heterogeneity of infections with business foot-traffic, age, and race. However, all studies mentioned above only modelled the relationships in 2020. Recently, Chen et al. accounted for neighborhood-level demographic and mobility differences to study different vaccine distribution strategies with consideration of heterogeneous COVID-19 risks in the US [4]. Despite successful epidemic modelling, correlation is a straightforward way to assess the relationship between human behaviors and COVID-19 infection rates that could provide timely insights. Most initial correlation analysis done so far applied the Pearson index, rank-sum test, and Kruskal-Wallis test with Dunn's post hoc analysis [42, 43]. Some also utilized the Spearman Ranked Correlation Coefficient and the Kendall-Tau Ranked Correlation Index [23, 30]. However, both of these correlation indices can only capture linear or monotomic relationships while the relationships between COVID-19 infection rates and human mobility & close contacts are not exactly the case. Moreover, ranked correlation coefficients is not necessarily applicable to assess the relationship between human behaviors and COVID-19 infections if we look at the global period as the global period time series can be viewed as different intervals that in a form is close to grouped frequency distribution. Then looking at pairs of two variables from different time intervals may not be the best way to assess correlation. Cheng et al. applied MIC to analyze the relationship between different categories of mobility (i.e. park visits, retail, workplaces, transit stations) and effective reproduction number and found close relationships among them [6]. This study is also done only for the period of March to September 2020 for regions in England.

3 DATA AND PREPROCESSING

The Cuebiq Mobility Index (CMI), Cuebiq Contact Index (CCI), and their Year-Over-Year YoY variations are collected from the Cuebiq dashboards at the state level daily for the two years of 2020 and 2021. According to Cuebiq, "The CMI measures the median aggregated movement (i.e distance traveled) by all devices in meters in each state on a log 10 scale. The CCI measures the number of devices with contact dividing by the number of devices seen at a place. A close contact is determined by whether two or more devices come within 50 feet of each other within a five minute time

period" (https://help.cuebiq.com/hc/en-us/articles/360041285051-Mobility-Insights-Mobility-Index-CMI). The mobility insights indices are all calculated by Cuebiq using their first-party data. We use the two indices as indicators for human mobility and close contacts between people, respectively. The YoY variation represents the percentage of change compared to the average value of the previous 365 days. The 2020 data is calculated as:

$$D_{2020} = \frac{D_{2021}}{YOY + 1} \tag{1}$$

The US state-level COVID-19 infection data including cumulative confirmed cases, hospitalized cases, and deaths are obtained from the New York Times (https://github.com/nytimes/covid-19-data). The daily confirmed cases are calculated as the difference between the cumulative confirmed cases and any negative values are replaced as a 0. The period we examined is from April 2020 to the end of 2021 as April 2020 is the first month with complete new cases reported for the whole month but a lot of cases were unreported or not tested in the US at the beginning of the pandemic (February and March, 2022) [5]. We then estimate the Effective Reproduction Number (R_t) from daily confirmed cases as R_t has been grounded to be an indicator of the trend and spread of a pandemic compared to Basic Reproduction Number R_0. R_t implies the average number of secondary infections caused by an infected individual at time t. Moreover, R_t is a measure that change with time which adapts to every stage of a pandemic while R_0 is only a static measure, which is not the best measure to determine the power of infection of a pandemic as there are diseases with $R_0 < 1$ that persisted and diseases with $R_0 > 1$ that vanished [6].

To compute the the Effective Reproduction Number (R_t), Bayesian inference assumes a Gamma Distribution as the prior distribution for R_t. It is calculated as the expectation E of the number of new infections I_t at time t over the infectiousness of the infected individual at time t. The infectiousness of the infected individual at time t is in turn determined as the sum of infected individuals up until time t-1 weighted by their infectivity function w_s. The infectivity profile w_s is estimated by the serial interval defined as the time from the onset of symptoms in an infected individual to the onset of symptoms in a secondary infected individual [7]. To avoid R_t being highly variable, the smoothing window τ is chosen to be the default value of 21 days in this analysis as it was demonstrated that 7 days lead to too much fluctuations and exaggerate the presence of weekly patterns which defeats the purpose of smoothing (according to the implementation package at https://github.com/lo-hfk/epyestim). R_t also has the advantage to assess the efficiency of control measures and interventions as it reflects changes very quickly [7].

$$R_t = \frac{E[I_t]}{\sum_{s=1}^{t} I_{t-s} w_s} \tag{2}$$

4 METHODS

4.1 Maximal Information Coefficient

The correlation analysis is broken down into temporal and spatial aspects respectively as our purpose is to understand the spatial-temporal heterogeneity among the relationships between the two indices and COVID-19 infections, which is not widely studied yet in the literature. For each period and each cluster of states, the MIC is applied to measure the non-linear correlation between each index and the COVID-19 infection rate including the daily cases and R_t. MIC is chosen because it is much more powerful in capturing complex non-linear relationships compared to other indices that measure correlation like Spearman[33].

MIC assigns scores nearly to 1 for all noiseless functional as well as non-functional relationships that are not constant; it assigns scores close to 0 to independent relationships [34]. In addition to the characteristics mentioned above, MIC has an outstanding advantage of having both generality and equitability, meaning that it could capture relationships that are not functional just as well as functional relationships and it does not bias toward certain types of functions at different noise levels [34]. Equitability is the concept that the correlation assessment method should give similar scores to relationships at similar noise levels regardless of the type of the relationship (i.e. different functional relationships or even non-functional relationships). And it was further shown that MIC is more equitable than the Distance Correlation and Spearman Correlation [33].

For calculating the MIC, let G denote the x-by-y grid where the two variables are scattered. The MIC looks at the highest mutual information I obtained from G at any resolution and normalizes $I_G(x, y)$ from all resolutions. The MIC is taken as the maximum of $I_G(x, y)$ where $\{x^*, y^*\}$ is the optimal resolution that maximizes normalized I_G [33].

$$MIC = \max \frac{I_G(x^*, y^*)}{\log_2 \min\{x, y\}} \tag{3}$$

4.2 Temporal Analysis with Bayesian Change Point Detection

We applied the Bayesian change point detection approach, which is a Bayesian framework for detecting signals in the probability of a change point which suggests that the time series before and after it are come from different distributions or the same distribution with different parameters [12], to capture the temporal heterogeneity of the COVID-19 spread [9, 22]. It is especially powerful due to its flexibility with the number of change points and the positions of them [28]. The input of the algorithm is the daily average COVID-19 cases across all states as observations $y_1, \ldots, y_n$ where $n = 640$ is the number of data points (2020-04-01 to 2021-12-31) in our study. Denote $P_{a:b}$ as the probability that y_a and y_b are in the same period (there is no change point between y_a and y_b), Q_t as the likelihood of data after time t given a change point at t-1, $PCP_{i,t}$ as the likelihood that the i_{th} change point is at time t, and P_t as the probability of a change point of at time t which is the key output we need.

$$Q_t = P(y_{t:n} | t - 1 \textit{ is a change point}) \tag{4}$$

The algorithm examines two priors: i) the number of change points $\tau_1 \ldots \tau_m$ and conditioning on their positions $P(\tau_i | \tau_{i+1})$; ii) the point process that models the change points. Denote $g(t)$ as the probability mass function for the time between two successive points that specifies the point process, $G(t)$ as the distribution function of the distance between two successive points with $G_t = \sum_{s=1}^{t} g(t)$. We can derive recursion for Q_t as follows.

$$Q(1) = \sum_{s=1}^{n-1} P_{1:s} Q_{s+1} g(s) + P_{1:n}(1 - G(n-1)) \tag{5}$$

$$Q(t) = \sum_{s=t}^{n-1} P_{t:s} Q_{s+1} g(s+1-t) + P_{t:n}(1 - G(n-t)) \tag{6}$$

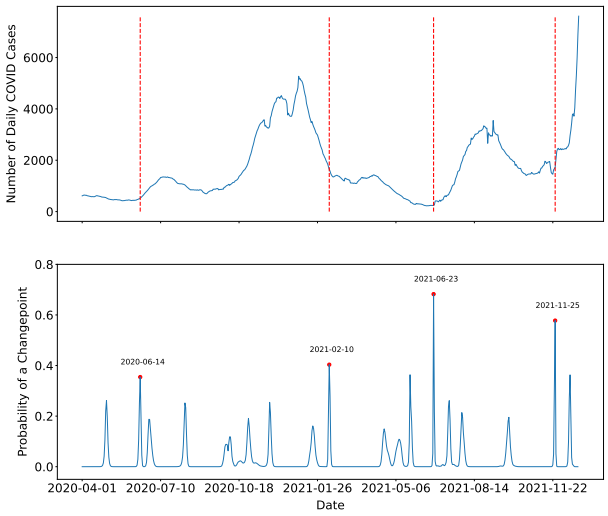

Figure 1: Bayesian Change Point Detection Result on the Average Daily COVID-19 Confirmed Cases of all States

We then simulate samples of change points from the posterior distribution of the next change point given the previous change-point at time t denoted as $P(\tau|y_{1:n,t})$ by simulating samples in the data. After conditioning on the number of change points and their positions, we can derive the likelihood of that the ith change point at time t. In this analysis, we assume constant prior for the change points as we do not have prior information on where a change point is more likely to occur. The constant prior used is the uniform distribution on the length of the data. To obtain the probability of a change point at time t P_t from the likelihood PCP, we have:

$$P_t = \sum_{i=1}^{t} PCP_i \tag{7}$$

For more details of the algorithm, please see [12]. We also acknowledge that there are other effective change point detection algorithm (e.g., TS-CP^2 via Contrastive Predictive Coding [10]). The daily COVID-19 cases time series is segmented into three periods by performing the Bayesian change point detection: 2020-06-14 to 2021-02-10, 2021-02-10 to 2021-06-23, and 2021-06-23 to 2021-11-25, as shown in Figure 1. The top plot in Figure 1 shows the average daily confirmed cases of COVID-19 cases of all states in the US and the bottom plot shows the likelihood of a change point. In general, tall and thin spikes indicate a change point with more confidence and short and wide spikes indicate a possible change point with more uncertainty [35]. The four change points we selected are based on this characteristic. Looking at the three periods diagnosed, the first period aligns well with the first wave in newly reported cases in 2020 [1, 32]; the second period corresponds to the downward trend followed by a small wave of a surge in Spring 2021; the third period corresponds to the peak surge in Fall of 2021. It is worth noting that the period after 2021-11-25 has a huge spike of COVID-19 infections which aligns with the discovery of the Omicron variant cases [17].

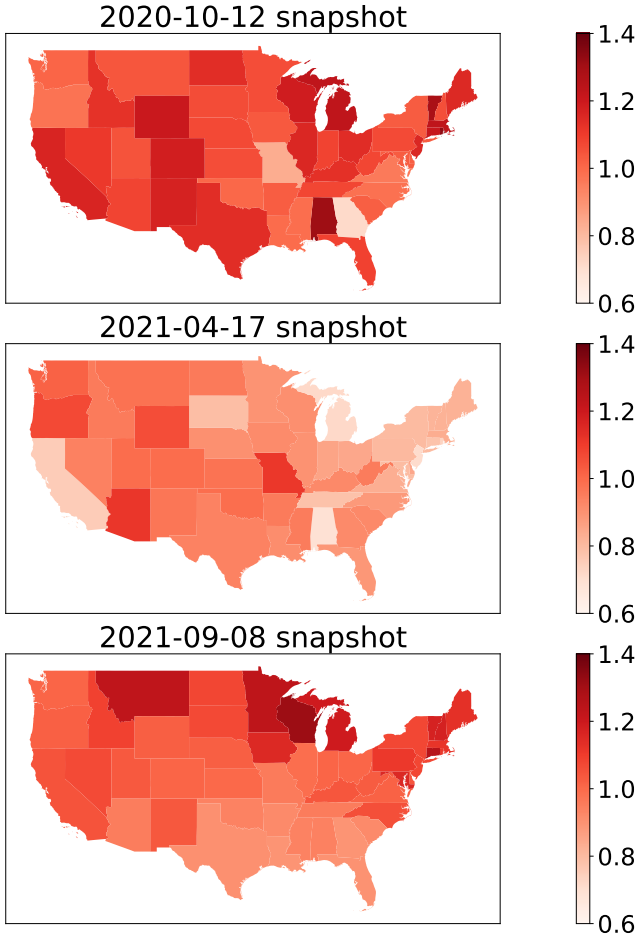

Figure 2: The spatial distribution snapshots of the effective reproduction number R_t in the middle of each period for US states

We also examined the snapshot distribution of R_t for US states in the middle of each period derived from the Bayesian Change Point Detection to get a sense of the spatial-temporal differences, as shown in Figure 2. The midpoints of each period derived are 2020-10-12, 2021-04-17, and 2021-09-08 which are at the rising uphill of the surge in Fall 2020, the small peak during the downward trend in Spring 2021, and the peak during the second largest wave since the pandemic in Fall 2021. Different geographic hotspots of infections can be found on the maps for each period (Figure 2), which is a motivation to perform spatial heterogeneity analysis in the following section. For each period, we apply MIC to measure the correlation between each index and COVID-19 daily cases

that are smoothed using a seven-day rolling average to take into account time delays. We also performed the same analysis on the seven-day rolling average of R_t which has been shown to be more representative of the spread of the pandemic at the moment in time.

4.3 Spatial Analysis with KShape Clustering

To understand the spatial heterogeneity of COVID-19 infections across states and the regional homogeneity, spatial clustering is needed. KShape is a clustering algorithm specifically designed for clustering time series as the clustering process is based on the shape alignment. In particular, we chose KShape due to its characteristic of maintaining Scaling Invariance (being able to recognize time series' similarity despite scales), Shift Invariance (being able to recognize the similarity in time series despite a slight shift in phase), and Complexity Invariance (being able to recognize the similarity in time series when they have similar shape even if they have different degrees of complexities) during shape similarity detection of time series while performing the spatial clustering [31]. Hence, all the time series are z-normalized before undergoing the KShape clustering which is critical for maintaining the scaling invariance. To maintain the Shift Invariance, the shape-based distance compares time series similarity based on aligning the two time series with shift w that maximizes their cross-correlation CC_w. Since cross-correlation captures the similarity between time series itself, computing centroid is phrased as an optimization task that looks for the centroid time series ct that maximizes its cross-correlation with all other time series in the cluster. Figure 3 shows the centroids of the Kshape clustering of state-level COVID-19 daily cases in the US. Algorithm 1 below is an outline of the algorithm summarized from [31].

$$NCC_c = \left(\frac{CC_w(t, CEN(j))}{\sqrt{R_0(t,t) \cdot R_0(CEN(j), CEN(j))}}\right) \tag{8}$$

$$R_0(a,b) = \sum_{i=1}^{m} a_i \cdot b_i \tag{9}$$

In the Algorithm 1, Equation 10 calculates the centroid of cluster j as the maximizer of NCC_c with all time series in the cluster. $CEN(j)$ denotes the current centroid time series for cluster j. Let $D(j)$ denote the current shape-based distance that the time series t is deviated from the centroid of cluster j. Equation 11 calculates the shape-based distance between time series t and $CEN(j)$. As the coefficient normalization of cross correlation NCC_c defined in Equation 8 inherently measures the similarity between two time series, one minus the maximum NCC_c gives the minimum distance between t and $CEN(j)$. In other words, $D(j)$ is also a measure of dissimilarity and we want to assign the time series t to the cluster with the smallest dissimilarity. The algorithm will not stop until convergence, meaning the cluster membership CM does not change anymore or until the maximum number of iterations is reached.

The time series of daily COVID-19 cases for the Contiguous US including the 48 mainland states and Washington D.C. are clustered by KShape. The silhouette score of the z-normalized time series is computed for K ranging from 2 to 10 and we selected 3 as the optimal K with the largest silhouette score. As shown in Figure 4, the three clusters derived from Kshape largely correspond to

Algorithm 1 CM, CEN $\leftarrow KShape(X, k)$

Input X is an array of n time series each with length m; k is an integer that specifies the number of cluster

Output CM is an array of length n that specifies each time series' cluster membership, CEN is an $k \times n$ array with k centroids of length n

```
iteration ← 1
randomly assign clusters
while NOT converge or iteration<100 do
  for each cluster j do
      X' = []
      for each time series t in X do
      if t is in cluster j then
          X'.append(t)
      end if
      CEN(j) = arg max_ct Σ_{x_i ∈ X'} NCC_c(x_i, ct)^2        (10)
  end for
  for each time series t in X do
    for each cluster j do
      D(j) = 1 − max_w NCC_c(t, CEN(j))                        (11)
    end for
    assign t to the cluster with the minimum D
    update CM
  end for
  iteration+=1
end while
```

the eight economic regions in the US [26]: cluster 0 made of New England, Mideast, and Midwest; cluster 1 largely made up of Plains, Rocky Mountains, and Southwest; cluster 2 made up of mostly Far West and Southeast. Similarly, for each cluster of states, we apply MIC to measure the correlation between mobility index/close contact index and COVID-19 daily cases and R_t, respectively. Both the daily cases and R_t are smoothed using the seven-day rolling average to reduce noise due to the delay of reporting or other data issues [18, 32, 40].

We also further examined the spatial auto-correlation of the distribution MIC values for both indices and COVID-19 infections using the Global Moran's I index. It evaluates both the attribute location and attribute values at the same time. The Global Moran's I Index has range from 1 to -1 where a value of 0 indicates perfect spatial randomness, a value of 1 means perfect positive spatial auto-correlation and a value of -1 suggests perfect negative spatial auto-correlation [26]. A high value indicates a clustered pattern while a low value indicates a dispersed pattern. Let $w_{i,j}$ represent the i-by-j matrix of spatial weight between attribute x_i and x_j with all diagonal values $w_{i,i,}$ being 0; let n be the total number of features; let z_i denote the deviation of the ith attribute from its mean; let W denote the summation of all spatial weights; Using the geographic

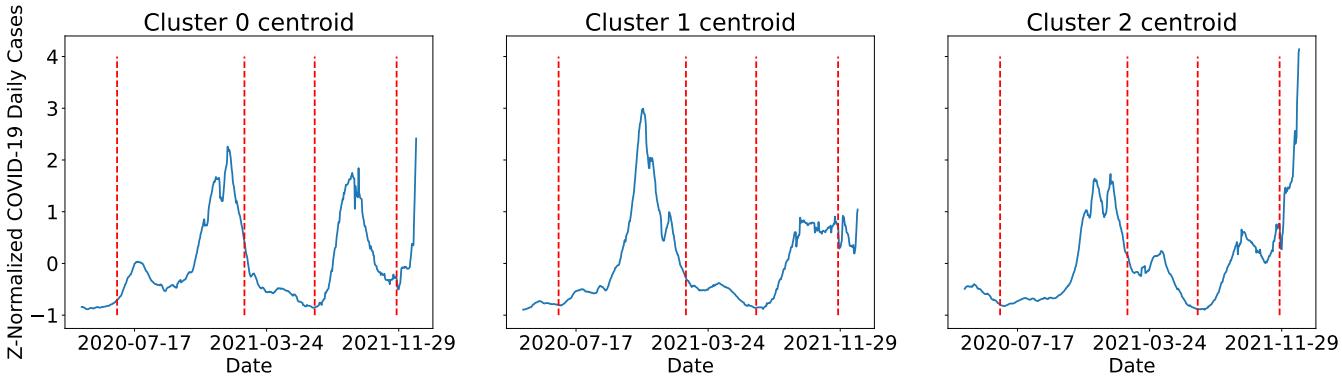

Figure 3: The z-normalized daily COVID-19 cases centroid time series of K Shape clustering result of US states

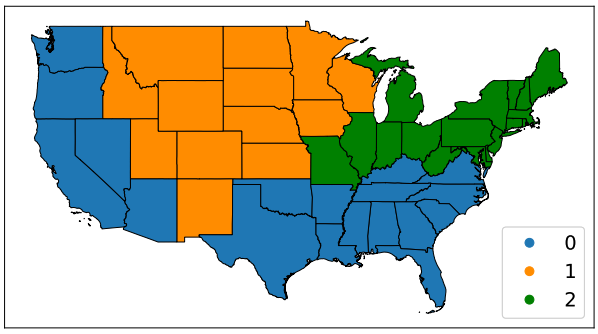

Figure 4: The K Shape clustering result of US states

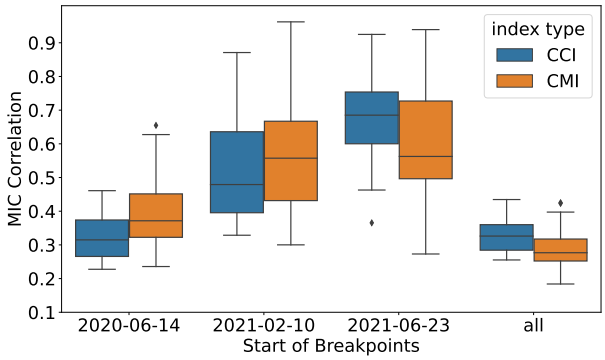

Figure 5: The MIC correlation between mobility and contact index with R_t for each time period

adjacency rule (sharing state boundary) in our study, the Global Moran's I Statistic is calculated as below:

$$I = \frac{n}{W} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} z_i z_j w_{i,j}}{\sum_{i=1}^{n} z_i^2} \tag{12}$$

$$W = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \tag{13}$$

$$z_i = x_i - \bar{X} \tag{14}$$

5 RESULTS

5.1 Temporal Heterogeneity

The experiment result using the abovementioned datasets in the US has shown significant temporal and spatial heterogeneities among the associations between the two indices (CMI and CCI) and the COVID-19 infections. As shown in Figure 5, in general, the positive correlation between both indices and R_t continues to grow as the pandemic progresses. This aligns with our intuition that at the beginning of the pandemic, the infected population was too small for mobility and close contacts to be highly correlated with the COVID-19 infection rate. In addition, with the stay-at-home and social-distancing orders, human mobility was at its lowest in March and April and did not see a rebound in mobility until May and June after partial reopening in 18 states on May 1st, 2020 [25, 29, 40]. Our observation that Summer 2020 has the lowest MIC correlation (0.228 between CCI and R_t and 0.236 between CMI and R_t) of all periods examined align with the knowledge that human mobility was at a rather idle and beginning to rebound phase during Summer 2020 [13]. For each period, the correlations are much higher than the global period (i.e. 2020-04-01 to 2021-12-31). This also demonstrated that it's critical to examine the role of human behaviors in the pandemic at different stages with consideration of the temporal heterogeneity. The first period (2020-06-14 to 2021-02-09) seems to be the time interval when the pandemic is most serious while the second period (2021-02-10 to 2021-06-22) seems to be a winding-down period as the first period has the highest average number of daily confirmed cases of all states as high as 6080 and the second period is in general a downward trend as shown in Figure 1 if we disregard the Omicron period after 2021-11-25.

From the temporal perspective, the average values of CCI and CMI exhibited very different patterns from April 2020 to the end of 2021 as shown in Figure 6. Close contacts and mobility are both rebounding prior to the first change point (2020-06-14) we detected. Stating from Summer 2020, human mobility level has remained at the same level with CMI values staying around 3.9 while the level of close contacts slowly rises during Spring 2021. But the level of close contacts had a drastic increase in Mid August and reached

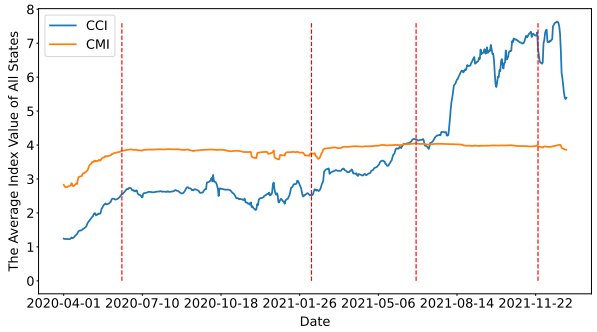

Figure 6: The Average CCI and CMI values of all states

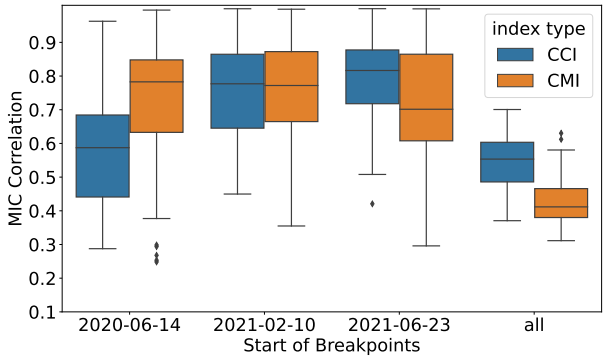

Figure 7: The MIC correlation between mobility and contact index with daily COVID-19 cases for each time period

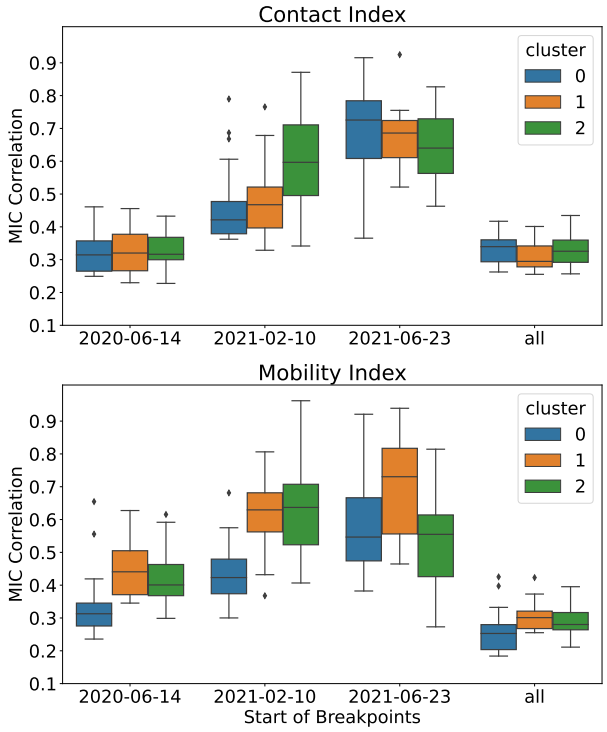

Figure 8: The MIC correlation between each index with R_t for each cluster

an unprecedented level in Fall 2021 which matches the second peak in COVID-19 cases in Fall 2021. This leads to the temporal heterogeneity in the relationships between COVID-19 infections and the two indices. The CMI is much more correlated with R_t compared to CCI in 2020 as demonstrated in Figure 5. However, when it comes to the Fall surge in 2021, the CCI is much strongly correlated with R_t. This suggests that travelling is an important factor that contribute to infections during the first wave but close contacts between people have more impact during the Fall peak in COVID-19 cases. This pattern would not have been visible if we look at the global period only. This matches our knowledge that as vaccination rates went up and mask requirement, human mobility would not be as influential as close contacts in terms of contributing to COVID-19 infections [8, 20]. Looking at the MIC correlation between both indices and COVID-19 daily cases in Figure 7, the temporal heterogeneity is in general smaller between each period in time. The distribution of MIC values for the daily COVID-19 cases are also in general higher than the MIC values of R_t. But it still showed a similar pattern as the correlation between the two indices and R_t.

5.2 Spatial Heterogeneity

From the spatial perspective, the spatial heterogeneity which can lead to substantial local and regional variations in COVID-19 timing and severity [18, 38] is already visible from looking at the centroids of COVID-19 infection rates in Figure 3. States in cluster 1 are impacted by the first peak in Fall 2020 much more than all the surges coming afterwards. On the other hand, states in cluster 2 has a much more visible small surge during the downward trend in Spring 2021 and an incredibly sharp increase during the Omicron period at the end of 2021. States in cluster 0 experienced a rather similar impact from each surge of COVID-19. The values of centroids can not be used for comparison of severity of infections between clusters as each state's time series are z-normalized toward its own mean and the centroids are extracted in a shape-based manner. The centroids are only meaningful in understanding the overall trend and the impact of each surge on all states within a cluster.

Figure 2 shows the spatial distribution of R_t in US states as a snapshot at the middle of each period we examined. The New England, Mideast, and Midwest regions were shown to be a rather homogeneous area with similar pandemic situation with smaller R_t values than the rest of the states looking at the snapshot from Spring 2021. This group of states actually aligns very well with cluster 2 (in green) we derived from the KShape Clustering result in Figure 4. It's also noticeable that Southeast and some Southwest states have similar R_t values smaller than the rest of the states during Fall 2021. This observation also matches with the cluster 0 (in blue) delineated from the KShape Clustering. Another interesting observation is that Alabama has a significantly higher R_t value during the Summer-Winter 2020 period but later demonstrated a much smaller R_t value in Spring 2021. This may be due to outstanding policies that successfully contained COVID-19 spread at that time.

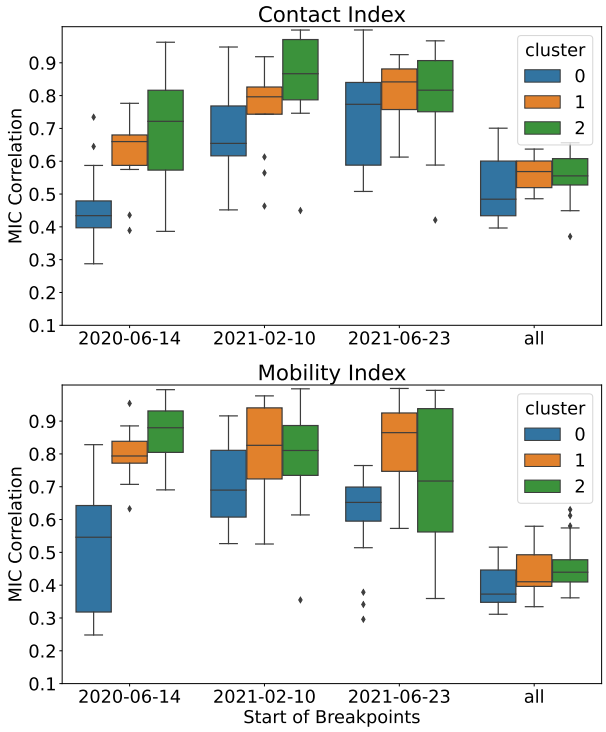

Figure 9: The MIC correlation between each index with daily COVID-19 cases for each cluster

Looking at the MIC correlation in Figure 8, although mobility was highly associated with COVID-19 infection rates R_t during the first two periods compared to close contacts, there is noticeable spatial heterogeneity within it. On the mobility side, it has a smaller impact on states in cluster 0 (i.e. Far West and Southeast) compared to states in the other two clusters. Moreover, states in cluster 2 (New England, Mideast, and Midwest region) is the only cluster that experienced a decrease in the correlation between mobility and COVID-19 infection. In terms of close contact, the trend of the association with R_t is similar, showing a steady upward trend in each cluster of states. However, states in cluster 2 (New England, Mideast, and Midwest areas) saw a bump in the correlation between close contacts and COVID-19 infections as early as Spring 2021 while other states did not see this bump until Summer 2021. The difference in median MIC correlation values for states in cluster 2 (in green) between Spring 2021 (2021-02-10 to 2021-06-23) and Summer 2020 (2020-06-14 to 2021-02-10) is as high as 0.28 while the increase in MIC correlation values for the other two clusters were only 0.15 and 0.11. When it comes to Summer 2021, the increase in median MIC values from the previous period for cluster 0 and 1 reached 0.3 and 0.22, respectively. Although we know the correlation between close contacts and R_t exceeded the correlation between human mobility and R_t in fall 2021, we can see that it's mostly due to the increasing correlation with close contacts in states in cluster 0 and 2. States in cluster 1's (Plains, Rocky Mountains, and Southwest) correlation between R_t and close contacts actually never exceeded that with mobility. This observation may be related with the fact that the population density for states in cluster 1 are the lowest region in the US. Even so, our result still showed a noticeable increase in the MIC correlation between close contacts and COVID-19 infections in fall 2021 for states in cluster 1 (in orange).

On the other hand, when it comes to the MIC correlation between the two indices and daily COVID-19 cases as shown in Figure 9, it has also shown a rather similar pattern with the MIC correlation with R_t. Similarly, the MIC values for daily cases are in general much higher than that for R_t. Looking at the global spatial auto-correlation analysis results in Table 1, we have all positive values for each period's MIC correlation values, indicating neighboring states tend to have similar values of MIC correlation as well. The MIC correlation for both indices with R_t realized the highest spatial dependency at different periods. The correlation between CCI and R_t showed the highest spatial dependency during Spring 2021 (Moran's I: 0.561) while the correlation between CMI and R_t exhibited the highest spatial dependency during Summer 2020 (Moran's I: 0.675). The spatial dependency between human mobility and R_t is slowly becoming weaker over time.

Table 1: Global Moran's I index of the MIC values in the US for each phase of the pandemic

Spatial Auto-Correlation (by Moran's I)		
Time Period	CCI & R_t	CMI & R_t
2020-06-14 to 2021-02-09	0.358	**0.675**
2021-02-10 to 2021-06-22	**0.561**	0.410
2021-06-23 to 2021-11-23	0.258	0.376
Global Period	0.235	0.266

6 CONCLUSION

In this study, we looked at the non-linear correlations between COVID-19 infections and human mobility as well as close contacts and revealed significant spatiotemporal heterogeneities. We reviewed effective analytical methods that examine spatial and temporal characteristics of the COVID-19 infection in the US. Specifically, we adopted a Bayesian inference framework to determine a meaningful segmentation of the pandemic into rather independent time periods. We applied a time series specific shape-based clustering algorithm (KShape) to group states that experience similar pandemic situations into one cluster. To capture the complex relationships, we exploited the Maximal Information Coefficient to measure correlations. Our results showed that close contacts between individuals starts to play a more influential role compared to mobility as the pandemic progresses into 2021. However, states in the Plains, Rocky Mountains, and Southwest regions never saw the correlation of close contacts exceeding mobility. We also found that the MIC correlation values are in general higher for both indices (CMI and CCI) with daily COVID-19 cases and has less heterogeneity than with Effective Reproduction Number R_t. The spatial dependency for the correlation between close contacts and R_t and the correlation between mobility and R_t each reached the maximum during Spring 2021 and Summer 2020 respectively.

Our results demonstrated the value of mobile device location big data and the importance of spatiotemporal heterogeneity of human

behaviors in understanding the spread of the virus. Understanding the spatiotemporal heterogeneity in how human behaviors correlates with the pandemic progression is critical for government officials to make decisions and plan ahead Non-Pharmaceutical Interventions (NPIs) as of when to lift travel restrictions, when to impose loose or strict social distancing order, and for which regions, especially as the pandemic continues to develop and resurge, which essentially require coordination efforts [36]. With the enrichment of spatial-temporal data, our work provided a timely summary of the pandemic situation since the beginning of the pandemic. A key contribution of this study is to fill the gap of measuring the complex relationships between the pandemic spread and human behaviors all the way up to the end of 2021. We have investigated human behaviors from both a human mobility perspective and a close contacts perspective. Some further research directions include advancing the correlation analysis to higher geographic resolutions such as county or census tracts and investigate any changed or hidden behaviors at county or census tract level.

ACKNOWLEDGMENTS

We acknowledge the funding support from the County Health Rankings and Roadmaps program of the University of Wisconsin Population Health Institute, Wisconsin Department of Health Services, and the National Science Foundation [Grant No. BCS-2027375]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funders.

REFERENCES

[1] Hamada S Badr, Hongru Du, Maximilian Marshall, Ensheng Dong, Marietta M Squire, and Lauren M Gardner. 2020. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases* 20, 11 (2020), 1247–1254.

[2] Jay J Van Bavel, Katherine Baicker, Paulo S Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J Crockett, Alia J Crum, Karen M Douglas, James N Druckman, et al. 2020. Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour* 4, 5 (2020), 460–471.

[3] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589, 7840 (2021), 82–87.

[4] Lin Chen, Fengli Xu, Zhenyu Han, Kun Tang, Pan Hui, James Evans, and Yong Li. 2022. Strategic COVID-19 vaccine distribution can simultaneously elevate social utility and equity. *Nature Human Behaviour* (2022), 1–12.

[5] Shi Chen, Qin Li, Song Gao, Yuhao Kang, and Xun Shi. 2020. State-specific projection of COVID-19 infection in the United States and evaluation of three major control measures. *Scientific Reports* 10, 1 (2020), 1–9.

[6] Tao Cheng, Xinchen Zhong, Yunzhe Liu, Yang Zhang, and Guangsheng Dong. 2021. Dynamic Spreading of COVID-19 Versus Community Mobility in Regions of England. In *Mapping COVID-19 in Space and Time*. Springer, 233–251.

[7] Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. 2013. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology* 178, 9 (2013), 1505–1512.

[8] Forrest W Crawford, Sydney A Jones, Matthew Cartter, Samantha G Dean, Joshua L Warren, Zehang Richard Li, Jacqueline Barbieri, Jared Campbell, Patrick Kenney, Thomas Valleau, et al. 2022. Impact of close interpersonal contact on COVID-19 incidence: Evidence from 1 year of mobile device data. *Science Advances* 8, 1 (2022), eabi5499.

[9] Jonas Dehning, Johannes Zierenberg, F Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann. 2020. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* 369, 6500 (2020), eabb9789.

[10] Shohreh Deldari, Daniel V. Smith, Hao Xue, and Flora D. Salim. 2021. Time Series Change Point Detection with Self-Supervised Contrastive Predictive Coding. In *Proceedings of The Web Conference 2021 (WWW '21)*. Association for Computing Machinery. https://doi.org/10.1145/3442381.3449903

[11] Justin Elarde, Joon-Seok Kim, Hamdi Kavak, Andreas Züfle, and Taylor Anderson. 2021. Change of human mobility during COVID-19: A United States case study. *PloS one* 16, 11 (2021), e0259031.

[12] Paul Fearnhead. 2006. Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and computing* 16, 2 (2006), 203–213.

[13] Song Gao, Jinmeng Rao, Yuhao Kang, Yunlei Liang, and Jake Kruse. 2020. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SIGSpatial Special* 12, 1 (2020), 16–26.

[14] Song Gao, Jinmeng Rao, Yuhao Kang, Yunlei Liang, Jake Kruse, Dorte Dopfer, Ajay K Sethi, Juan Francisco Mandujano Reyes, Brian S Yandell, and Jonathan A Patz. 2020. Association of mobile phone location data indications of travel and stay-at-home mandates with COVID-19 infection rates in the US. *JAMA network open* 3, 9 (2020), e2020485–e2020485.

[15] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779–782.

[16] Kyra H Grantz, Hannah R Meredith, Derek AT Cummings, C Jessica E Metcalf, Bryan T Grenfell, John R Giles, Shruti Mehta, Sunil Solomon, Alain Labrique, Nishant Kishore, et al. 2020. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature communications* 11, 1 (2020), 1–8.

[17] Xuemei He, Weiqi Hong, Xiangyu Pan, Guangwen Lu, and Xiawei Wei. 2021. SARS-CoV-2 Omicron variant: characteristics and prevention. *MedComm* 2, 4 (2021), 838–845.

[18] Xiao Hou, Song Gao, Qin Li, Yuhao Kang, Nan Chen, Kaiping Chen, Jinmeng Rao, Jordan S Ellenberg, and Jonathan A Patz. 2021. Intracounty modeling of COVID-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proceedings of the National Academy of Sciences* 118, 24 (2021), e2020524118.

[19] Tao Hu, Siqin Wang, Bing She, Mengxi Zhang, Xiao Huang, Yunhe Cui, Jacob Khuri, Yaxin Hu, Xiaokang Fu, Xiaoyue Wang, et al. 2021. Human mobility data in the COVID-19 pandemic: characteristics, applications, and challenges. *International Journal of Digital Earth* 14, 9 (2021), 1126–1147.

[20] Bo Huang, Jionghua Wang, Jixuan Cai, Shiqi Yao, Paul Kay Sheung Chan, Tony Hong-wing Tam, Ying-Yi Hong, Corrine W Ruktanonchai, Alessandra Carioli, Jessica R Floyd, et al. 2021. Integrated vaccination and physical distancing interventions to prevent future COVID-19 waves in Chinese cities. *Nature Human Behaviour* 5, 6 (2021), 695–705.

[21] Xiao Huang, Junyu Lu, Song Gao, Sicheng Wang, Zhewei Liu, and Hanxue Wei. 2022. Staying at home is a privilege: Evidence from fine-grained mobile phone location data in the United States during the COVID-19 pandemic. *Annals of the American Association of Geographers* 112, 1 (2022), 286–305.

[22] Xiao Huang, Yang Xu, Rui Liu, Siqin Wang, Sicheng Wang, Mengxi Zhang, Yuhao Kang, Zhe Zhang, Song Gao, Bo Zhao, et al. 2022. Exploring the spatial disparity of home-dwelling time patterns in the USA during the COVID-19 pandemic via Bayesian inference. *Transactions in GIS* 26, 4 (2022), 1939–1961.

[23] Kamal Kishore, Vidushi Jaswal, Madhur Verma, Vipin Koushal, et al. 2021. Exploring the utility of Google mobility data during the COVID-19 pandemic in India: digital epidemiological analysis. *JMIR public health and surveillance* 7, 8 (2021), e29957.

[24] Moritz UG Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Open COVID-19 Data Working Group†, Louis Du Plessis, Nuno R Faria, Ruoran Li, et al. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368, 6490 (2020), 493–497.

[25] Minha Lee, Jun Zhao, Qianqian Sun, Yixuan Pan, Weiyi Zhou, Chenfeng Xiong, and Lei Zhang. 2020. Human mobility trends during the early stage of the COVID-19 pandemic in the United States. *PLoS One* 15, 11 (2020), e0241468.

[26] Kangli Li, Natasha Zhang Foutz, Yuxin Cai, Yunlei Liang, and Song Gao. 2021. Impacts of COVID-19 lockdowns and stimulus payments on low-income population's spending in the United States. *PloS one* 16, 9 (2021), e0256407.

[27] Grant McKenzie and Benjamin Adams. 2020. A country comparison of place-based activity response to COVID-19 policies. *Applied geography* 125 (2020), 102363.

[28] Octavian Niculita, Zakwan Skaf, and Ian K Jennions. 2014. The application of Bayesian change point detection in UAV fuel systems. *Procedia CIRP* 22 (2014), 115–121.

[29] Evgeny Noi, Alexander Rudolph, and Somayeh Dodge. 2022. Assessing COVID-induced changes in spatiotemporal structure of mobility in the United States in 2020: a multi-source analytical framework. *International Journal of Geographical Information Science* 36, 3 (2022), 585–616.

[30] Yixuan Pan, Aref Darzi, Aliakbar Kabiri, Guangchen Zhao, Weiyu Luo, Chenfeng Xiong, and Lei Zhang. 2020. Quantifying human mobility behaviour changes during the COVID-19 outbreak in the United States. *Scientific Reports* 10, 1 (2020), 1–9.

[31] John Paparrizos and Luis Gravano. 2015. k-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*. 1855–1870.

[32] Sen Pei, Teresa K Yamana, Sasikiran Kandula, Marta Galanti, and Jeffrey Shaman. 2021. Burden and characteristics of COVID-19 in the United States during 2020. *Nature* 598, 7880 (2021), 338–341.

[33] David Reshef, Yakir Reshef, Michael Mitzenmacher, and Pardis Sabeti. 2013. Equitability analysis of the maximal information coefficient, with comparisons. *arXiv preprint arXiv:1301.6314* (2013).

[34] David N Reshef, Yakir A Reshef, Hilary K Finucane, Sharon R Grossman, Gilean McVean, Peter J Turnbaugh, Eric S Lander, Michael Mitzenmacher, and Pardis C Sabeti. 2011. Detecting novel associations in large data sets. *Science* 334, 6062 (2011), 1518–1524.

[35] Eric Ruggieri. 2013. A Bayesian approach to detecting change points in climatic records. *International Journal of Climatology* 33, 2 (2013), 520–528.

[36] Nick Warren Ruktanonchai, JR Floyd, Shengjie Lai, Corrine Warren Ruktanonchai, Adam Sadilek, Pedro Rente-Lourenco, Xue Ben, Alessandra Carioli, Joshua Gwinn, JE Steele, et al. 2020. Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science* 369, 6510 (2020), 1465–1470.

[37] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of predictability in human mobility. *Science* 327, 5968 (2010), 1018–1021.

[38] Loring J Thomas, Peng Huang, Fan Yin, Xiaoshuang Iris Luo, Zack W Almquist, John R Hipp, and Carter T Butts. 2020. Spatial heterogeneity can lead to substantial local variations in COVID-19 timing and severity. *Proceedings of the National Academy of Sciences* 117, 39 (2020), 24180–24187.

[39] Xiu Wu and Jinting Zhang. 2021. Exploration of spatial-temporal varying impacts on COVID-19 cumulative case in Texas using geographically weighted regression (GWR). *Environmental Science and Pollution Research* 28, 32 (2021), 43732–43746.

[40] Chenfeng Xiong, Songhua Hu, Mofeng Yang, Weiyu Luo, and Lei Zhang. 2020. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. *Proceedings of the National Academy of Sciences* 117, 44 (2020), 27087–27089.

[41] Li Xiong, Cyrus Shahabi, Yanan Da, Ritesh Ahuja, Vicki Hertzberg, Lance Waller, Xiaoqian Jiang, and Amy Franklin. 2020. REACT: Real-time contact tracing and risk monitoring using privacy-enhanced mobile tracking. *SIGSPATIAL Special* 12, 2 (2020), 3–14.

[42] Takahiro Yabe, Kota Tsubouchi, Naoya Fujiwara, Takayuki Wada, Yoshihide Sekimoto, and Satish V Ukkusuri. 2020. Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic. *Scientific Reports* 10, 1 (2020), 1–9.

[43] Mengxi Zhang, Siqin Wang, Tao Hu, Xiaokang Fu, Xiaoyue Wang, Yaxin Hu, Briana Halloran, Zhenlong Li, Yunhe Cui, Haokun Liu, et al. 2022. Human mobility and COVID-19 transmission: a systematic review and future directions. *Annals of GIS* (2022), 1–14.