Assessing COVID-induced changes in spatiotemporal structure of mobility in the United States in 2020: A multi-source analytical framework.

Evgeny Noi, Alexander Rudolph, Somayeh Dodge

University of California Santa Barbara, Santa Barbara, CA 93117, USA

ARTICLE HISTORY
Compiled November 7, 2021

ABSTRACT
The COVID-19 pandemic resulted in profound changes in mobility patterns and altered travel behaviors locally and globally. As a result, movement metrics have widely been used by researchers and policy makers as indicators to study, model, and mitigate the impacts of the COVID-19 pandemic. However, the veracity and variability of these mobility metrics have not been studied. This paper provides a systematic review of mobility and social distancing metrics available to researchers during the pandemic in 2020 in the United States. Twenty-six indices across nine different sources are analyzed and assessed with respect to their spatial and temporal coverage as well as sample representativeness at the county-level. Finally global and local indicators of spatial association are computed to explore spatial and temporal heterogeneity in mobility patterns. The structure of underlying changes in mobility and social distancing is examined in different US counties and across different data sets. We argue that a single measure might not describe all aspects of mobility perfectly. A more comprehensive measure of mobility is required to model the complex effects of the epidemic. We urge that various limitations inherent in data sets from different providers should be acknowledged and these data must be used with caution.

KEYWORDS
COVID-19; SARS-CoV-2; pandemic; mobility patterns; mobility indices; social distancing; non-pharmaceutical interventions; spatial autocorrelation

1. Introduction

The COVID-19 pandemic led to major non-pharmaceutical interventions (NPIs) in the United States and worldwide to combat the spread of SARS-CoV-2. These NPIs are public policy measures such as mask mandates, travel restrictions, quarantines, stay-at-home orders, business shutdowns, social distancing, and contact tracing, which significantly impacted people's daily life, their travel behavior, and social interaction. Governments have relied heavily on digital technologies such as mobile phones, apps, and location-based services to monitor the implementation and assess the impacts of these unprecedented measures on the virus transmission (Kitchin 2020).

Mobility and close contacts between people are major contributing factors in the transmission of coronavirus infection across and within communities (Badr *et al.* 2020).

CONTACT Evgeny Noi. Email: noi@ucsb.edu

Therefore, it is essential that mobility insights are investigated and used to inform epidemiological models and to further amend interventions and health policies (Oliver *et al.* 2020). Since early 2020, we have witnessed a shift in the data sharing culture from big technology companies such as Google, Apple, and Facebook. Equally eager to join or spearhead data-sharing initiatives, are smaller location intelligence companies and educational institutions, such as Cuebiq, StreetLight, Descartes Labs, PlaceIQ, SafeGraph, Unacast and the Maryland Transportation Institute of the University of Maryland (UMD). The data sets published by these organizations have been widely used and made available through data dashboards to inform the general public and the scientific community in the United States (Cuebiq 2021, SafeGraph 2021, Maryland Transportation Institute 2020). These data have also enabled researchers and policy makers to model the impact of the pandemic and the associations between the transmission of COVID-19 and mobility through data-driven approaches (Dodge *et al.* 2020). Using these large aggregate mobility data, literature documented the profound impact of the pandemic on mobility and the effects of the NPIs on the containment of COVID-19 worldwide (McKenzie and Adams 2020) and across different regions and countries such as in the USA (Gao *et al.* 2020, Klein *et al.* 2020, Chang *et al.* 2020, Kim and Kwan 2021), Europe (Flaxman *et al.* 2020), China (Kraemer *et al.* 2020, Peng *et al.* 2020), Germany (Schlosser *et al.* 2020), Austria (Heiler *et al.* 2020), Italy (Gatto *et al.* 2020, Bonaccorsi *et al.* 2020), and elsewhere. In this paper, we mainly focus on the mobility data sets that are available for the United States.

Mobility data are complex and are commonly gathered from call data records (CDR), location data obtained from location-aware technologies such as Global Position Systems (GPS), Bluetooth, and mobile applications installed in smart phones and wearable devices (Grantz *et al.* 2020). Depending on the required granularity, modeling purposes and privacy guarding algorithms, raw measurements may be aggregated resulting in biases originating from ecological fallacy (Robinson 2009), the modifiable area unit problem (Yule and Kendall 1950), scaling problem, interpolation, misalignment, multi-scale and multi-resolution re-aggregation (Gotway and Young 2002). Various mobility metrics outlined in this paper have been designed to approximate and/or measure aggregate movement patterns. These indices may agree or disagree on the spatial and temporal specifics of the process in question (as demonstrated in the course of this paper), but at the very least, they further our understanding of the complexity of human mobility. Another data issue is missing data and gaps. Imperfect data can lead to major flaws in data-driven models and policies. While it is not always possible to assess data biases and deficiencies, as access to raw data is restricted due to privacy concerns or their proprietary nature, it is important to assess the validity of our inferences by utilizing multiple data sources and evaluate discrepancies in observed spatial and temporal patterns through comparative analysis. Especially during the turbulent pandemic times when the demands are high for fast pace flow of information and publication of scientific research, ignoring the inherent spatiotemporal biases and uncertainties in data sets may be inevitable, yet it can pose serious consequences when the data and research are used to inform governmental policies and the public.

Kishore *et al.* (2020) outlined the main characteristics and concerns regarding data pipelines for COVID-19 research, including data types, data-sharing policies, privacy safeguards, established baselines, data representativeness, and commonly used spatial and temporal aggregation methods to generate these data. Shuja *et al.* (2021) provided a comprehensive overview of open-source COVID-19 data sets by categorizing them as medical images, case reports, social media, lockdown and quarantine measures. Yet little attention is paid to inherent bias that could be introduced in analysis by

the choice of one particular data source or provider. Existing efforts for data-driven COVID-19 modeling, as well as research and media reports on the associations between mobility and the spread of the infection often rely on single-source or limited data for mobility insights. For example, McKenzie and Adams (2020) investigated place-based activity patterns in response to governments' mitigation actions across various nations using Google's community mobility data set. The New York Times mainly used Cuebiq Mobility Insight (CMI) data for their analyses (Gabriel Dance, Lazaro Gamio 2020). Arnal *et al.* (2020) examined COVID-related regional mobility trends in Spain using both Google and Facebook data sets, while Morita *et al.* (2020) utilized Apple and Google mobility data to analyze behavioral changes in Japan. Similarly, Cacciapaglia *et al.* (2020) utilized Google and Apple data to characterize social distancing patterns during the first wave of COVID-19 in Europe and in the USA. Fewer studies have focused on combining insights from several data sources. For instance, Huang *et al.* (2021) attempted to quantify mobility response to COVID-19 by combining data from Descartes Labs, Apple, Google, and Twitter to calculate the COVID responsive index in the United States.

In this study, we evaluate spatial and temporal structures of human mobility patterns recorded through multi-source mobility and social distancing indices. Specifically, through a comprehensive exploratory analysis of 26 mobility indices obtained from nine commercial and open data sources, this research provides unique insights on the spatial and temporal changes in mobility patterns across the US in response to the COVID-19 pandemic in 2020. The heterogeneity in data coverage, representativeness, and observed patterns are discussed. Using global and local Moran's I statistics on multi-source mobility data, we characterize and map spatial and temporal associations in the structure of daily mobility patterns at the county level. In doing so, we show how observed changes in mobility patterns vary across different data sources and different geographic regions in the United States. Compared to existing works, e.g. Huang *et al.* (2021), we draw from several more data sources and characterize mobility metrics in terms of their spatial and temporal coverage and by assessing the indicators of spatial association for different data providers.

2. Overview of Data Providers

As part of open research initiatives, private companies, universities and various organizations have made anonymized aggregate movement data available for scholars and policy-makers to help understand the impacts of COVID-19 NPI responses, and measure the influence of mobility on the overall infection rates in the communities. Table 1 summarizes major commercial and non-commercial providers used for the COVID-19 mobility response in the United States. As Table 1 shows, seven out of ten data sets are available at no cost for research purposes. Of the three commercial sources, we only had access to two: Cuebiq and Streetlight data. Hence, Unacast indicators are not analyzed in this paper.

This study investigates 26 indicators of mobility obtained from nine different data sources, including Apple, Cuebiq, Descartes Labs, Facebook, Google, PlaceIQ, Safegraph, Streetlight, and the University of Maryland (see Table 1 and Supplementary Material for a more detailed description of the indices). We focus mainly on four categories of indicators: traversed distance metrics, social distancing metrics, contact/exposure indices, and number of trips indices. Therefore, the network-based and origin-destination flow indices offered by these providers (such as PlaceIQ *LEX* and

Table 1. Major mobility data providers in COVID-19 studies for the United States, listed alphabetically.

	Data source	Access	Availability (since)	# of metrics	Coverage[b] (%)
1	Apple	Open	Jan 13, 2020	1	42.7
2	Cuebiq	Commercial	Jan 1, 2019	4	93.4
3	Descartes Labs	Open	Mar 1, 2020	2	65.3
4	Facebook	Open	Mar 1, 2020	2	62.6
5	Google	Open	Feb 15, 2020	6	35.0
6	PlaceIQ	Open	Jan 20, 2020	2	57.0
7	SafeGraph	Open	Jan 1, 2019	2	93.3
8	StreetLight	Commercial	Jan 1, 2019	1	98.5
9	University of Maryland	Open	Jan 1, 2020	6	97.8
10	Unacast[c]	Commercial	Feb, 24, 2020	3	-

[b]Coverage is defined as the percentage of complete (non-missing) records for all of the observed counties (n=3140) and days (n=366) in the observation period (January - December 2020) summarized for each data source.

[c]Unacast provides data on a commercial basis. Since we do not have access to this data, the coverage information is not provided and the readers are referred to their homepage (Engle *et al.* 2020). The number of metrics is calculated based on the number of products in *Unacast COVID-19 Location Data Toolkit.*

Cuebiq *Mobility Flows*) are not included in this study. The focus is narrowed to the metrics available at the county-level for the spatial unit and daily for the temporal granularity. Among the data providers, all except SafeGraph publish data at the county level. Safegraph provides raw data at the census block group (CBG[1]) and points-of-interest (POI[2]) levels. On the temporal scale, the metrics are tracked and made available on a daily resolution. The temporal extent considered in this study is January 1, 2020 – December 31, 2020.

3. Methodological Framework

3.1. *Overview*

Generally, assessing multi-source heterogeneous data that are measured and recorded using different units and baselines is cumbersome, particularly so for spatially and temporally correlated data. Here, we propose a multi-stage framework (summarized in Figure 1) which incorporates various geostatistics and exploratory data analysis tools and enables a more swift and comprehensive comparison of heterogeneous mobility data. Different processes involved in the proposed framework include:

(1) *Data collection and integration:* Various data sets are collected from different providers (as described in Section 2). The data are validated for errors and gaps and pre-processed for integration in a comprehensive database on a server. The database uses *county* as the spatial unit and a *week* for temporal granularity. We opted to analyze data at weekly intervals for two reasons: First, because raw data is available at the daily time scale, the next level of aggregation that is suitable for differentiating mobility trends and patterns is one week to account for the variability during weekdays and weekends. Second, the COVID-19 dynamics can be reasonably described at a weekly time scale as demonstrated by McAloon

[1]Census Block Groups - Census divisions containing Census blocks. Typically around 600-3000 people.

[2]Points of Interest - georeferenced point feature containing information on housing locations, gas stations, parking, tourist attractions, etc.

et al. (2020).

(2) *Exploratory Data Analysis (EDA):* A series of EDA is applied to gauge spatial sampling bias, assess spatial and temporal coverage (see Figures A1, A2), and temporal trends. The results for each outlined mobility index are discussed in Sections 4.1— 4.3. The spatial and temporal coverage of the data sets is detailed in Supplementary Material.

(3) *Spatial and temporal structure analysis:* Global and Local Moran's I are calculated to test for spatial associations across four groups of the mobility metrics and to pinpoint the locations of hotspots and coldspots in data (Sections 4.4—4.6). This stage further compares the trends for different mobility indices within the same data provider as well as across multiple providers:
   (a) *Cross-comparison of spatiotemporal structures:* Spatiotemporal structures for various mobility indices obtained from each data source are examined by summarizing the temporal trends and the consistency of the spatial clusters throughout 2020 for each individual county. These trends are then visualized to capture the clustering patterns on the national scale. The results are mapped and discussed in Section 4.5.
   (b) *Assessment of between-source consistency in patterns:* The variability in spatiotemporal structures of patterns across different data sources is compared by visualizing the agreement between consistency of hotspots and coldspots across mobility indices. The results are discussed in Section 4.6.

(4) *Mapping and data visualization:* At each stage, different data visualization techniques are offered to map the identified patterns in space and time and communicate the outcomes. The proposed visualizations are described and interpreted in light of the results in Section 4.

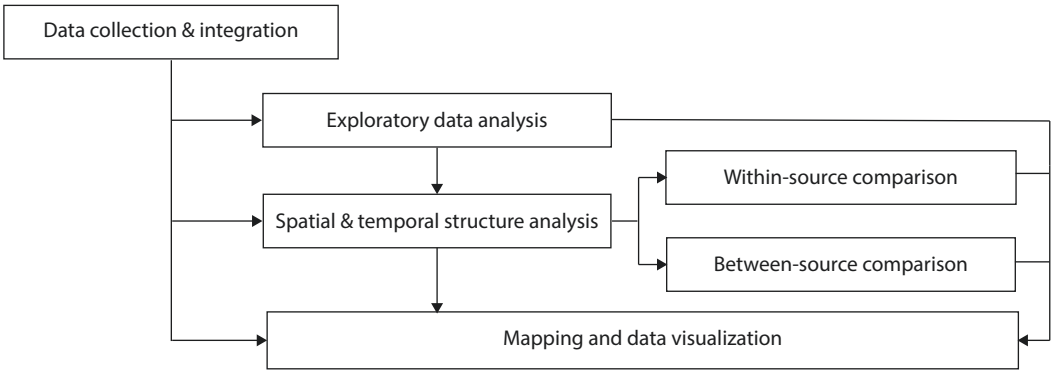

Figure 1. The proposed methodology to analyze spatial and temporal variability in mobility patterns within- and between-source.

Below we expand on our approach to study variability in spatial and temporal structures of heterogeneous mobility data (i.e. Stage 3 of the methodology).

3.2. *Spatial and temporal structure analysis*

Since each metric quantifies mobility differently, in order to compare trends across data sets we make a comparative assessment of derived patterns representing spatial clusters instead of comparing raw values directly. To quantify the systemic spatial variation in mobility data, spatial autocorrelation is estimated. There are many global

measures that capture correlation of georeferenced data, including Getis and Ord's G (Getis and Ord 1992), Ripley's K (Ripley 1976), Geary's C (Geary 1954), Matheron's Semivariance $\gamma(h)$ (Matheron 1963), and Moran's I (Moran 1948). We use the latter, as it is well-documented and widely used in GIScience literature. Formally, Global Moran's I is calculated as follows:

$$I = \frac{\sum_i \sum_j w_{ij} z_i z_j \Big/ S_o}{\sum_i z_i^2 \Big/ n}, \tag{1}$$

where w_{ij} is the spatial-weight matrix, $S_o = \sum_i \sum_j w_{ij}$, $z_i = x_i - \bar{X}$ and $\bar{X}$ is the mean, and n is the number of observations. While there are many ways to model spatial effects (e.g. distance band, k-nearest neighbors, Thiessen tessellation), spatial models designed with the first order contiguity tend perform better on average, as demonstrated by Stakhovych and Bijmolt (2009). Following the scientific principle of parsimony and in line with Anselin and Rey (1991), Florax and Rey (1995), and Farber *et al.* (2009) low-order weight specification should be preferred to higher-order weight matrix specification, as high matrix connectivity and overspecification may lead to decreased probability of detecting the true model, introducing strong biases in estimation of spatial autocorrelation (Smith 2009). Therefore, Queen's contiguity was selected as a sufficient approximation of the true weight specification for the assessed mobility data. Because each spatial unit (i.e. county in this case) is assessed in relation to the global mean, larger deviations would produce larger cross-products. The resulting value of I is positive when the counties with high values are clustered with other counties of high values (i.e. hotspots), or when counties of low values are clustered with other counties of low values (i.e. coldspots). Conversely, Moran's I is negative when the high values are in close proximity to low values, or when low values are located near high values. Permutation inference is then used to test for spatial randomness and report simulated pseudo p-values:

$$p = \frac{R + 1}{M + 1}, \tag{2}$$

where R is the number of times when the simulated statistic is more extreme than the observed statistic, and M is the number of permutations. Since the null hypothesis for Global Moran's I is spatial randomness, the tests returning statistically significant simulated values would indicate clustering in spatial data (i.e. in this case across the entire United States). We can then pinpoint the location of these clusters by utilizing LISA–the Local Indicators of Spatial Association (Anselin 1995).

Geometrically, Global Moran's I is a slope of a fitted regression line, where original variable values (x-axis) are mapped against the spatially lagged variable values (y-axis), which denote weighted sum of the values observed at neighboring locations, where 'neighborhood' is defined by a spatial weights matrix drawn from Queen's contiguity. This effectively separates the values into four distinct types of spatial association, corresponding to four quadrants of Moran's plot: spatial clusters (high-high, low-low) and spatial outliers (high-low, low-high). Here, instead of a single value that quantifies global spatial autocorrelation, we calculate local Moran's I for each location i (i.e.

each county):

$$I_i = \frac{\sum_j w_{ij} z_i z_j}{\sum_i z_i^2}, \tag{3}$$

where $z_i = x_i - \bar{x}$ and $\bar{x}$ is the mean.

Following Benjamini and Hochberg (1995), we use the False Discovery Rate (FDR) to account for multiple testing in assessing the significance of local Moran's I values. Therefore, for each test, the metric is considered significant for counties, where $p_i \leq i/n \times \alpha$ (for $\alpha = 0.05$, simulated p-value (p_i), sample size n, sorted index i). Combining the notion of spatial clusters with the notion of significance allows us to pinpoint the core of clusters, including significant hotspots and coldspots for each mobility indicator. The word 'cluster' refers to the spatial arrangement when testing for complete spatial randomness (null hypothesis) and clustering / dispersion (alternative hypothesis) in the observed data. The core of clusters may appear as a single county or a group of counties on the map, depending on the set statistical significance threshold. The size of the clusters would depend on the adopted correction procedure for multiple testing. Inferences without such corrections would yield the least conservative estimates (producing clusters of larger size), and inferences utilizing Bonferroni corrections would lead to spatial arrangements of smaller sizes, most likely denoting 'cores' of positive spatial association (hotspots and coldspots). We selected FDR correction as it produces medium-sized spatial clusters, while also minimizing Type I error and decreasing the number of false positives (see Anselin *et al.* (2010) for discussion of statistical significance in assessing LISA clusters). In this context, a significant hotspot represents counties with higher mobility values that are spatially clustered with other neighboring counties exhibiting high values. In contrast, a significant coldspot represents counties with lower values that are spatially clustered with other neighboring counties exhibiting low values. Using this approach, the trends in the spatial structure of the identified hotspots and coldspots over time are analyzed and mapped as described in Section 4.5.

4. Results and Discussions

4.1. *Spatial sampling bias*

Sampling bias is a type of error that occurs when intended subgroups of the population (e.g. smartphone users) do not have proportional representation in the sample. If there is no sampling bias in our data set, we would expect the number of observations from each county to be proportional to the population of that county. Figure 2 uses scatter plots to map census population (x-axis) against the number of users (i.e. *userbase*) in each data set (y-axis), where such information is available. The regression lines in the scatter plots show the ideal situation where there is perfect correlation between the number of observations and the population. The counties above the regression lines are overrepresented in the data, while the counties below the regression lines are underrepresented in the data set. Among the data sets examined, the userbase information is only provided in four data sources: PlaceIQ, Descartes Labs, Cuebiq, and SafeGraph. All four data sources provide a high level of correlation, indicating that the sampling bias is fairly low. The estimated Pearson R is above 0.9 for all sources: SafeGraph (0.97), Cuebiq (0.94), Descartes Labs (0.90) and PlaceIQ (0.94). This demonstrates

a high correlation between the number of samples in the data and census population for most counties for all four data sources. The evident outliers are large metropolitan areas, where the captured smartphone users seem to underrepresent the actual population of these metro areas. Examples include: Los Angeles, CA; Harris, TX (Houston), Cook, IL (Chicago); Maricopa, AZ (Phoenix); Tarrant, TX (Fort Worth). This can be due to different demographic structures observed in these large and highly populated metropolitan areas. Also, these areas usually have different day-time and night-time populations that may affect the userbase counts.

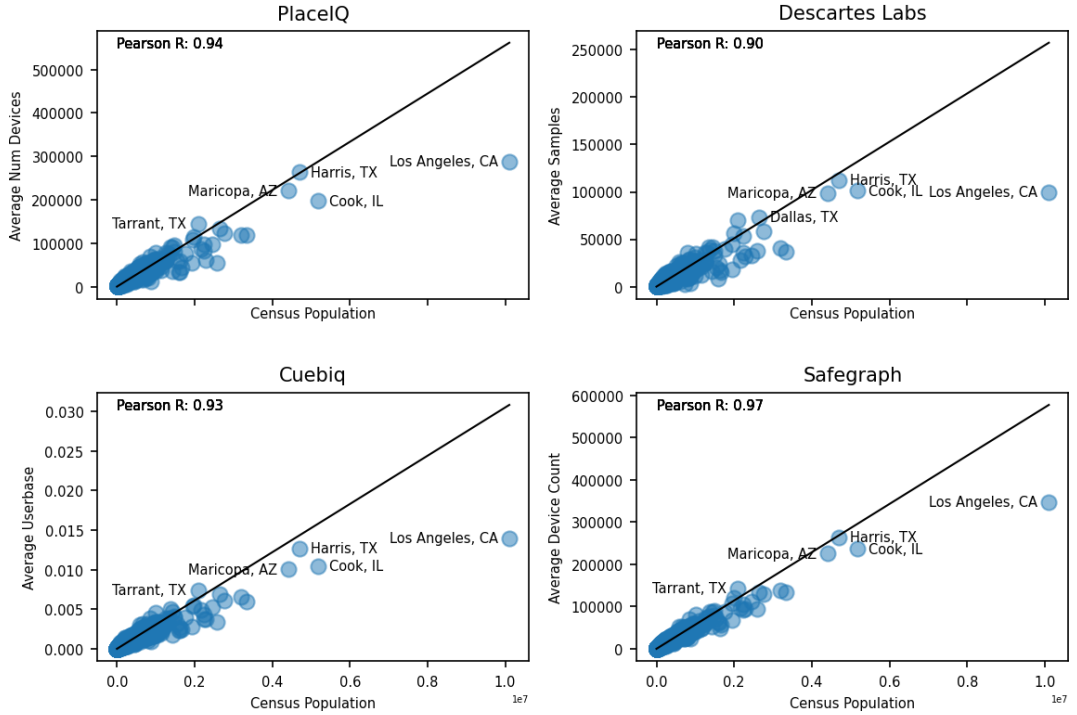

Figure 2. Sampling bias across four different data sources. The regression line corresponds to the expected counts for a truly random sample, where the userbase is representative of the population.

4.2. *Spatiotemporal coverage of mobility metrics*

We group mobility sources into three data-coverage Tiers (see Figures A1, A2): Tier 1 (complete coverage), Tier 2 (medium coverage), Tier 3 (heterogeneous coverage). There are four data sources in Tier 1 demonstrating the most complete spatial and temporal coverage: Cuebiq, University of Maryland, StreetLight and SafeGraph. Additionally, these four sources have more observation days in 2020 than any other data providers. Tier 2 includes three data sources: Descartes Labs, PlaceIQ, and Facebook. These data sets have a medium spatial and temporal coverage across the United States, but the coverage remains consistent within the observation period. The bulk of the missing values is in counties located in the Rockies and in some least populated counties. Tier 3 contains the data sources with patchier spatial and temporal coverage: Google and Apple mobility reports. Since Google publishes data for various category of trips, there is a lot of variation in the coverage. In particular, a highly heterogeneous coverage is observed for 'transit station %-change' focusing only on locations of transit stations, and 'parks %-change' concentrating on the locations of major national and state parks.

It is necessary to note that the userbase for both Google and Apple is dependent on the mobile operating systems (OS) and the portability of the various location-based applications across both OS. Therefore, it is logical to assume that the majority of users captured by Google come from Android devices, and the majority of users captured by Apple come from iOS devices[3]. A more detailed source-specific description of spatial and temporal coverage may be found in Supplementary Material.

4.3. *Temporal trends of mobility*

To evaluate the temporal trends of mobility across data sources, the indices are grouped into four categories based on the nature of mobility parameters that they approximate: (1) metrics quantifying social distancing, (2) metrics estimating the traversed distance, (3) metrics modeling contacts or exposure, and (4) metrics quantifying the number of trips or visit counts. To compare the trends, the time-series of these indices are then aggregated at Census Bureau-designated regions: South, Midwest, Northeast, and West (Figures 3–6). This delineation is often used for socioeconomic analysis with Census data, therefore it was logical to assess the temporal patterns of mobility at the regional level. While causal inferences on Census-designated regionalizations may be too broad to elicit meaningful insights, they are well-suited for descriptive and exploratory spatial data analysis. The time-series plotting is supplemented with corresponding violin plots (Figures B1, B2, B3, B4), and box plots (Figures B5, B6, B7, B8) in the Appendix. Violin plots are ideally suited for cross-comparison of distributions, as they allow to assess the number of modes, skewness, and overall dispersion of the distribution, whereas the box plots are better suited to denote median and interquartile range (IQR). The metrics are converted to standardized scores ($z = \frac{x-\mu}{\sigma}$) to ease the comparison, and kernel density estimation is used to plot and smooth distributions.

4.3.1. Metrics Quantifying Social Distancing

This category includes 7/26 metrics: *Ratio single tile users (Facebook)*, *Social distancing index (UMD)*, *% staying home (UMD)*, *% working from home (UMD)*, *% sheltered in place (Cuebiq)*, *% sheltered in place (SafeGraph), Residential %-change (Google).* In general, the indicators in this group approximate the number of people or proportion of population staying home (i.e. not moving away from home-base locations) on any given day. The general trends across different regions appear uniform in the observed period. As shown in Figure 3, there is a sharp increase across all of the metrics shortly after the start of the pandemic (March 11), which peaks around May and steadily decreases to slightly above the pre-pandemic levels. An exception is observed in *% working from home (UMD)*, which remains between 30%-40% until the end of 2020. There is less variability among different U.S. regions captured in these data sets compared to the *distance traversed metrics*, as demonstrated by the width of the confidence intervals and the spread of values in the time series plots (Figure 3 and 4). This can be attributed to the unit of measurements, standardized by the size of the population. As seen on Figure 3, according to Google, Cuebiq, Facebook and SafeGraph data sets, the highest percentage of sheltering at home is observed consistently in the Northeast, while the lowest is captured in the South. The violin plots and box plots (Figure B1, B5 in the Appendix) confirm that the distribution across regions and

[3]The market share of other OS, such as Windows and Linux, for tablets and smartphones is usually less than 0.5% for the United States

within these metrics are relatively similar and have unimodal distributions, with the exception of *% working from home (UMD)*, which has its second peak at two standard deviations away from the mean. This is most likely indicative of two distinct groups of counties in the sample centered at two observed modes. The number of outliers (i.e. $x = Q1 - 1.5 \times IQR$ or $x = Q3 + 1.5 \times IQR$) varies for all of the metrics across four regions from 5-15 standard deviations. Supporting our earlier claim, as seen on Figure B5, the median standardized values of these mobility metrics across all data sources are higher for the Northeast and the West, illustrating a higher percentage of sheltered population in these regions.

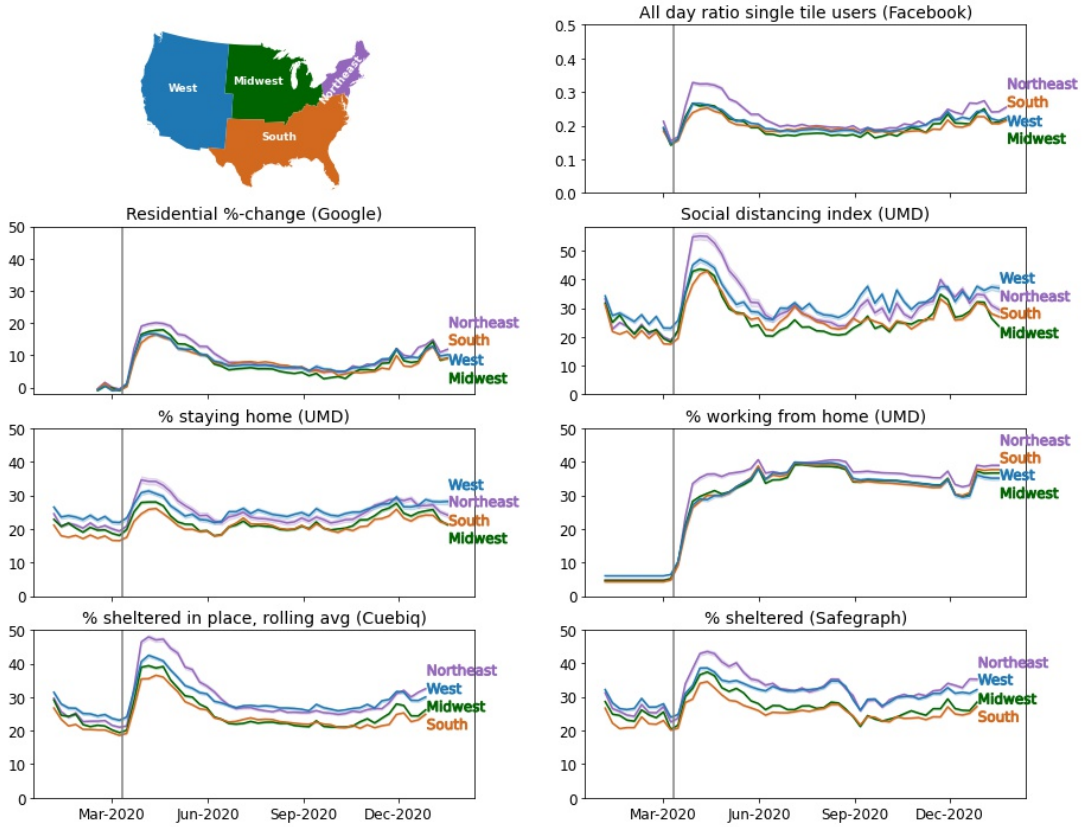

Figure 3. COVID-induced regional mobility trends in the U.S. in 2020 for *metrics quantifying the social distancing*. The gray vertical line denotes March 11, 2020, when COVID-19 was declared a pandemic. The shaded areas represent the confidence intervals of the average values for each time series.

4.3.2. Metrics Estimating the Traversed Distance

This category includes 8/26 metrics: *Bing tiles visited relative change (Facebook)*, *Median max-distance mobility (DL)*, *% of normal median max-distance (DL)*, *Miles per person (UMD)*, *Cuebiq mobility index (Cuebiq)*, *7d rolling average Cuebiq mobility index (Cuebiq)*, *Vehicles miles traveled (StreetLight)*, *median distance traveled (SafeGraph)*. These indicators estimate the average traversed distance by the recorded users. Overall, a similar dynamic is observed for most indices in this group (Figure 4): a sharp decline around the start of the pandemic (March 11) following various statewide stay-at-home orders, which reaches a global minima around April. From that point, the observed traveled distances steadily increase—with various degrees of variability across different data sets and regions—to either the pre-pandemic levels or to slightly

below the pre-pandemic levels. In spite of the similarity in the general trend, a high level of local variability is observed among different U.S. regions for these data sources as demonstrated by the width of the confidence intervals and the spread of values (Figure 4, B2, B6). One interesting observation is that Cuebiq points to the sharp decline in traversed distance for the Northeast compared to other regions. This trend, however, is not observed across other data sources (UMD, DL, Facebook and SafeGraph), where the value of indices remained lower than other regions. The confidence intervals calculated for StreetLight are very wide, especially in the Northeast and the western US, most likely indicating a higher variation in data. This is further supported by Figure B6, B2 in the Appendix. Compared to social distancing metrics, the number of outliers is much higher across this category for Facebook (25 standard deviations), Descartes Labs (from 50 to 155 standard deviations), and StreetLight (35 standard deviations). There is high variability in dispersion of the variables as illustrated by the shape of the distribution in violin plots and the size of boxes and whiskers on the box plots (Figures B6, B2). For example, Descartes Labs and StreetLight are much more centered around the median and have wider tales than the metrics in other sources.

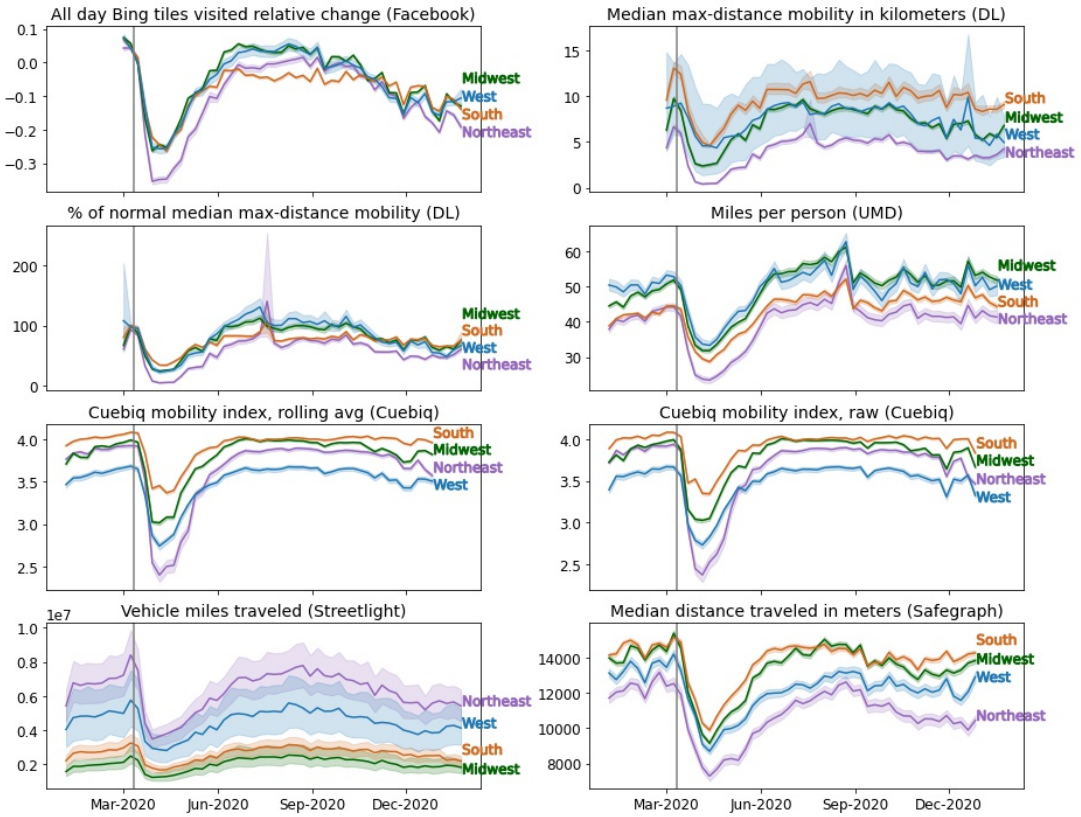

Figure 4. COVID-induced regional mobility trends in the U.S. in 2020 for *metrics estimating traversed distance*. The gray vertical line denotes March 11, 2020, when COVID-19 was declared a pandemic. The shaded areas represent the confidence intervals of the average values for each time series.

4.3.3. Metrics Modeling Contacts or Exposure

This category includes 3/26 metrics: *Device exposure index (PlaceIQ)*, *Adjusted device exposure index (PlaceIQ)*, *Cuebiq contact index, 7d rolling average (Cuebiq)*. These indicators approximate the number of users or contacts that the devices encounter

on a given day (i.e. the spatial proximity between users). Again, the trend is uniform across the three indices (Figure 5): an abrupt drop around the start of the pandemic (March 11), which plateaus around mid-April and then steadily increases to the pre-pandemic levels. The exposure appears to be highest in the South and the lowest in the West. The Northeast region also demonstrates a lower exposure, except for the Cuebiq Contact Index. This is supported by public health policies in the Pacific West states, which implemented rigorous stay-at-home orders to curb the possibilities for interaction by limiting or closing non-essential businesses such as restaurants, shopping malls, theaters, places of worships, salons, etc. It is important to note that although the West and the Northeast regions include densely populated areas such as the Californian coastal cities and the NYC metro, the exposure and contact indices generally capture a lower interaction level in these regions, perhaps highlighting the impact of the stay-at-home orders. The dispersion between metrics (Figure B3, B7 in the Appendix), as captured by IQR, is similar. The outliers cap at 20 standard deviations (for PlaceIQ) and 25 standard deviations for Cuebiq. The highest median and IQR is observed in the South across all the metrics.

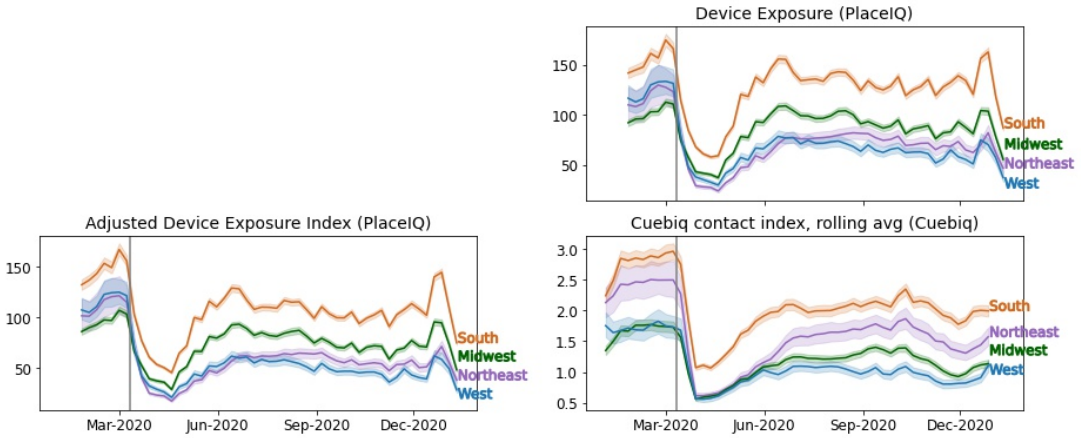

Figure 5. COVID-induced regional mobility trends in the U.S. in 2020 for *metrics modeling contacts or exposure*. The gray vertical line denotes March 11, 2020, when COVID-19 was declared a pandemic. The shaded areas represent the confidence intervals of the average values for each time series.

4.3.4. Metrics Quantifying the Number of Trips or Visit Counts

This category includes 8/26 metrics: *Retail and recreation %-change (Google)*, *Grocery and pharmacy %-change*, *Parks %-change (Google)*, *Transit stations %-change (Google)*, *Workplaces %-change (Google)*, *Non-work trips per person (UMD)*, *% change in consumption (UMD)*, *Mobility trend reports (Apple)*. The indicators in this group approximate mobility via the number of trips made per device. The behavior of different regions exhibits a sharp decline at the onset of the pandemic, but the 'recovery' phase captured in different indices appears to be quite different. For example, *Mobility trend reports (Apple)*, *Parks %-change (Google)* and *Non-working trips per person (UMD)* illustrate a more moderate decline, and a steeper rise peaking in the mid-summer, which is most likely indicative of the underlying connection between the normal travel season and the nature of these indices. That is, the users are more likely to use Google and Apple products to navigate for vacation/leisure and to travel to recreational places that are farther away from home and require routing help (as opposed to a memorized route to the grocery stores). It is necessary to remark that the

Apple Mobility Trends data does not capture the different phases of the pandemic in detail, as observed in the other distance or percentage based measures discussed earlier. The signal seems to only represent an oversimplified trend of mobility patterns with less granularity, and therefore it should be used as such. Apple data also indicate a higher mobility in the West and a lower mobility in the South, which is opposite of what was observed in other metrics as shown in Figures 4 and 3. This shows that the Apple Mobility Trends may not be a good measure of local mobility in general, and perhaps can mainly be used as an indicator for the trips made for non-commuting travel and vacation. Further examination (Figure B8, B4 in the Appendix) illustrates the differences between outlined regions. *Workplaces %-change (Google)* has a bimodal distribution with the first peak around the mean and the second around two standard deviations for all of the four regions, and *Transit stations %-change (Google)* has a bimodal distribution for the Northeast and the West. This means that two distinct workplace mobility patterns are latent in the data: one with a relatively high workplace %-change and another pattern with change values centered around the median. Similar trend is observed for %-change in transit stations mobility, suggesting two types of transit-dependent changes in the Northeast and the West, most likely attributed to changes in large metro areas (New York, Los Angeles, Portland, San Francisco) and less urbanized areas. The highest number of outliers is observed for *Non-working trips per person (UMD)* (60 standard deviations), and *% change in consumption (UMD)* (130 standard deviations), pointing to the highest spread in mobility metrics for data derived by the University of Maryland.

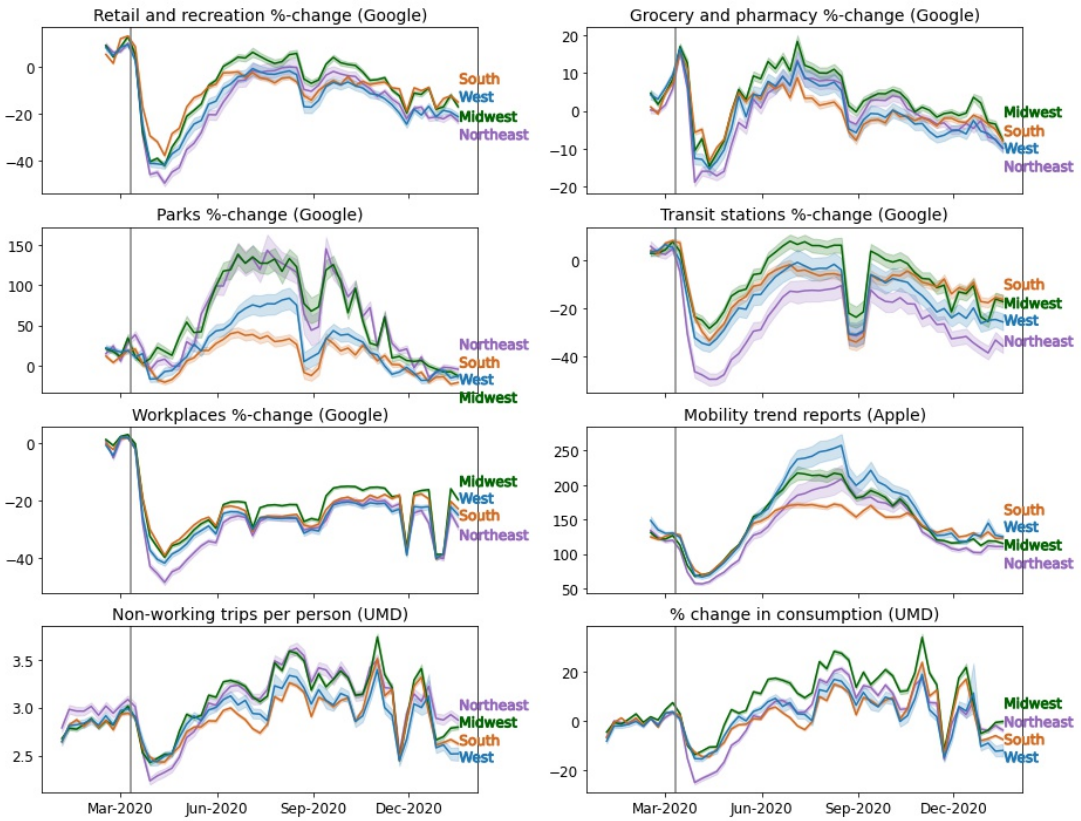

Figure 6. COVID-induced regional mobility trends in the U.S. in 2020 for *metrics quantifying the number of trips or visit counts*. The gray vertical line denotes March 11, 2020, when COVID-19 was declared a pandemic. The shaded areas represent the confidence intervals of the average values for each time series.

4.4. *Global spatial structure*

In order to measure spatial autocorrelation, the PySAL Python library (Rey and Anselin 2007) is used. Figure B9 in the Appendix plots Moran's I values for every week in 2020. Because Moran's I is highly dependent on spatial weights matrix, only mobility metrics (13 out of 26) that have complete spatial coverage are assessed, so that for each county there is at least one other neighboring county with shared edge and a corner (i.e. Queen's contiguity). These indices are metrics created for the UMD, Cuebiq, StreetLight and SafeGraph data sets. All of the Moran's I values are statistically significant ($\alpha = 0.05$) and positive, which implies clustering of high-high and low-low values for different metrics of mobility. The range of I values varies depending on the specific indicator of mobility: from 0.08 for Vehicle Miles Traveled (StreetLight) to 0.4 for social distancing index (UMD) and % change in consumption (UMD).

4.5. *Cross-comparison of spatiotemporal structures across metrics*

While the methodology described in Section 3.2 can be easily applied to any metric with high spatial and temporal coverage, we demonstrate this technique on four mobility metrics: *Cuebiq Mobility Index (Cuebiq)*, *Median Distance Traveled (SafeGraph), % sheltered in place, rolling average (Cuebiq)*, *% sheltered (SafeGraph)* due to the completeness of their coverage and comparability (see Section 2 and Supplementary Material for more details). The choice of these two sources is intentional: Cuebiq provides data by subscription, while SafeGraph provides data free of charge through an open data consortium. Both data sources are widely used by the media and the research community, and in COVID-19 related news and literature. They both contain indices that can be easily compared.

Local spatiotemporal structure of clusters for different mobility metrics is analyzed and mapped, as follows: (1) Local Moran's I is calculated for every county in the United States; (2) The corresponding statistically significant clusters (hotspots and coldspots) are extracted; (3) Cluster 'recency' (i.e. how recent a cluster appeared in 2020) is mapped to the color gradient, so that the hotspots/coldspots occurring in the beginning of 2020 are denoted in lighter shades of red/blue. Conversely, more recent clusters are denoted in darker shades; and (4) Cluster consistency (i.e. how frequently a cluster appeared in 2020) , defined as the number of weeks the county was consistently ranked as either a hotspot or coldspot, is mapped to the centroid marker size. Thus, larger consistency values correspond to proportionally larger marker sizes. Figure 7 illustrates the recency and consistency of the produced spatial clusters, assessed via LISA. In this figure, Cuebiq is depicted in the top row (a and b), and SafeGraph is plotted in the lower row (c and d). Distance-based metrics are plotted in the left half of the Figure (a and c), while social-distancing metrics are plotted in the right half of the Figure (b and d).

4.5.1. Metrics Estimating Traversed Distance

Because the number of counties is higher in the Eastern United States, there are a lot more markers in that part of the country. This is easily noticeable from Figure 7a (*Cuebiq Mobility Index*), where the highly attenuated hotspot cluster of counties is situated in the Southeastern states. This is the cluster of high-high values, associated with relatively high traversed distance in the neighboring counties. On the contrary, the same cluster of high-high values in the Southeast is not well pronounced in Safe-

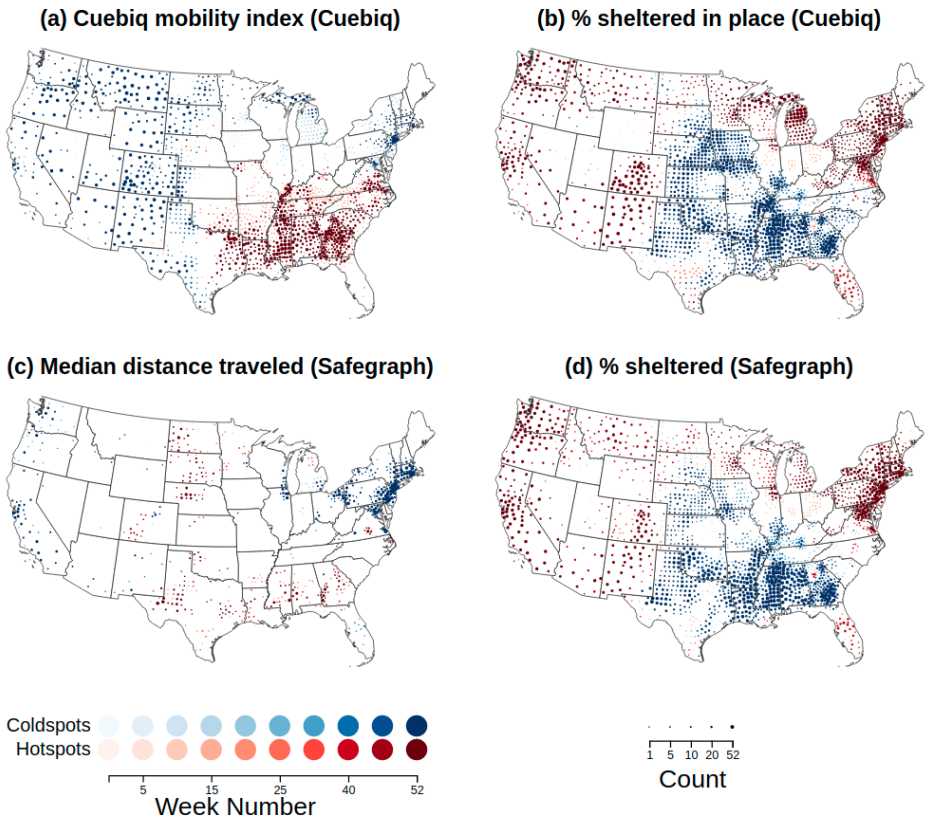

Figure 7. Spatial cluster recency and consistency for the United States. The markers of red color denote the counties that were classified as hotspots (high values co-located with other high values), the markers of blue color denote the counties that were classified as coldspots (low values co-located with other low values). Clustering consistency is illustrated by size. Clustering recency is depicted by color intensity.

Graph's *Median Distance Traveled* (Figure 7c). Using SafeGraph, fewer counties are classified as hotspots within these states. In contrast to hotspot delineation, the SafeGraph coldspot cluster located in the Northeast is more pronounced compared to Cuebiq. It is impossible to find out whether this is due to inherent methodological issues stemming from data aggregation or due to different algorithms used for privacy safeguarding. The original spatial unit of SafeGraph data is census block group, while for Cuebiq it is county. But under close examination, more differences can be found, particularly in the Dakotas, Nebraska and Colorado, which are classified as coldspots in Cuebiq and hotspots in SafeGraph. This might also be attributed to different userbase configurations of the two data sources.

4.5.2. Metrics Quantifying Social Distancing

Evidently there is a lot more agreement between *% sheltered (Cuebiq)* and *% sheltered (SafeGraph)* in location (see Figure 7b,d), but their recency and consistency varies from state to state. While the location of coldspots in both sources is centered in the South, with the core in Northern Texas, Mississippi, Alabama, Louisiana and Georgia, more cluster consistency (i.e. bigger markers) is observed in Cuebiq data for Nebraska, Iowa and Kansas. The location of hotspots is similar in the Northeastern states, Great Lakes region, and in Western United States, but the recency and consistency is higher

(i.e. darker and bigger markers) for Cuebiq data in Wisconsin and Michigan. In both sources, the states with some of the least recent hotspots are situated in Indiana and Ohio. It is necessary to remark that here, the hotspots represent lower mobility (higher percentage of sheltered-in-place population).

Some of the more interesting cases are found in the states that have both hotspots and coldspots, where the mobility patterns may be indicative of the differences in socioeconomic structure. For example, Northern Illinois hotspots with the core in Cook's County (Chicago) witnessed a consistently higher percentage of sheltered population together with its nearby counties, whereas Southern Illinois coldspot and its nearby counties witnessed a consistently lower percentage of sheltered population as compared to the national average. Another peculiar division found in spatiotemporal structures of mobility is between the East River and the West River South Dakota, separated by the Missouri River. The eastern part of the state exhibited consistent coldspots behavior, similar to other states with heavy agricultural production (Iowa, Nebraska), whereas the western part of the state exhibited consistent hotspots behavior (more people sheltering at home).

4.6. *Assessment of between-source consistency in patterns*

To supplement the assessment of trends in clustering consistency, a simple aggregate visualization is proposed that enables swift comparison over various mobility indices and data sources. The visualization is compiled as follows: (a) Identify hotspots and coldspots for each week and county, (b) Mark each county on the number of weeks a county remained significant as either a hotspot or a coldspot, and (c) Calculate the percentage of counties classified as either hotspots or coldspots for each state. Figure 8 illustrates the resulting summaries as a series of scatter plots, where each state is marked on the y-axis, and there is a separate column for each indicator. The value of each marker is mapped on the horizontal axis to denote the percentage of coldspots/hotspots (from the total number of counties within a state). The blue markers indicate coldspots and the red markers indicate hotspots. It is not uncommon for some states to include counties of both cluster types. In such cases, two markers (blue and red) per line are visualized. The primary goal of this visualization is to summarize and communicate the differences in spatial structure across various mobility indices. This works better when similar types of mobility indicators are assessed (e.g. metrics quantifying the social distancing). The visualization can be read vertically or horizontally. Vertically, it allows to assess the overall split between the hotspots and coldspots, and to identify the states which are primarily hotspots, coldspots or a combination of both. Horizontally, it allows a comparison in three dimensions: state-level, source-level, and indicator-level. For example, the first row in Figure 8 summarizes spatial clustering in Alabama over time. For distance-based metrics, only Cuebiq captures hotspots, which classifies 46.3% of counties as hotspots. For *%-sheltered at home*, the percentage of counties classified as coldspots is roughly the same. For metrics assessing traversed distance in Texas, both sources capture coldspots and hotspots, but the percentage of counties classified as hotspots/coldspots is higher for Cuebiq. In Massachusetts, no spatial clusters are identified when looking at *Cuebiq Mobility Index*, whereas *Safegraph Median Distance Traveled* identifies 23% of counties as coldspots. Overall, the figure reveals some similarities and discrepancies across the two different providers and mobility indices, as described in the following sections. These similarities and discrepancies also differ for various states. To learn more about the heterogeneity in the

Table 2. Major mobility data providers in COVID studies in the.

	Distance-based metrics		Social distancing metrics	
	Cuebiq	SafeGraph	Cuebiq	SafeGraph
# of states with no counties classified as hot/coldspots	11	24	2	1
# of states with both hotspots and coldspots	8	1	16	12
# of states where # hotspots ≈ # coldspots	3	1	4	3

revealed patterns, our future research plan is to contextualize the identified hotspots and coldspots using sociodemographic information of the counties as well as statewide COVID-19 mitigation policies throughout 2020.

4.6.1. Metrics Estimating Traversed Distance

There is a lot of disagreement between the two sources, as shown in the scatter plot in Figure 8 and Table 2. For example, in Missouri, around 4% of counties are classified as hotspots in Cuebiq data, and roughly 2% of counties in SafeGraph data are classified as coldspots. This is indicative of the differences in the underlying spatiotemporal structure of mobility measured by different indices in certain states. Similar behavior is observed in Idaho and North Dakota: SafeGraph reveals co-location of high-high traversed distance values, whereas Cuebiq reveals co-location of low-low traversed distance values. The frequency and consistency of clusters vary widely for different states: the percentage of coldspots is very similar in Illinois, Maryland, Michigan, and Virginia. This indicates agreement in clustering delineation between the two sources. In other words, the spatiotemporal structure of mobility in these states is very similar. On the other hand, the 'disagreement' between the two sources for coldspots is highest in Arizona, Colorado, Nevada, and New Mexico and for hotspots is highest in Alabama, Georgia, Mississippi, North Carolina, and South Carolina. Some states appear to have no coldspots or hotspots whatsoever, indicating that counties in these states did not respond similarly to the intervention measures. For example, there are no cold/hotspots in Arizona, Arkansas, Alabama, Delaware, Florida, Georgia, Iowa, Louisiana, Maine, Montana, Mississippi, Minnesota, New Hampshire and several others in SafeGraph; and in Connecticut, Delaware, Iowa, Massachusetts and some others for Cuebiq. There are three states per each data source that have roughly the same number of hotspots and coldspots (Table 2): Maryland, New Jersey, and New York. Additionally, Cuebiq identifies eight states with both hotspots and coldspots, which is eight times as many as SafeGraph. This might be indicative of the fact that there is less smoothing in Cuebiq data and that it captures within-state variations more effectively. This claim is also supported by the higher number of states without statistically significant clusters for the SafeGraph measure (24 vs 11 states).

4.6.2. Metrics Quantifying Social Distancing

A lot more agreement is observed between the social distancing metrics as demonstrated in Table 2. Figure 8 illustrates similar percentage of coldspots in Alabama, Georgia, Indiana, Massachusetts and a few others, and similar percentage of hotpots in Georgia, Massachusetts, Montana, South Dakota and Utah, pointing to the similarities in spatial structure between mobility in these states. On other other hand, the disagreement is more pronounced in Iowa, Kansas and Nebraska, where Cuebiq differentiates higher number of coldspot counties; and in Connecticut, Delaware and Rhode Island where SafeGraph discerns higher number of hotspot counties. Overall,

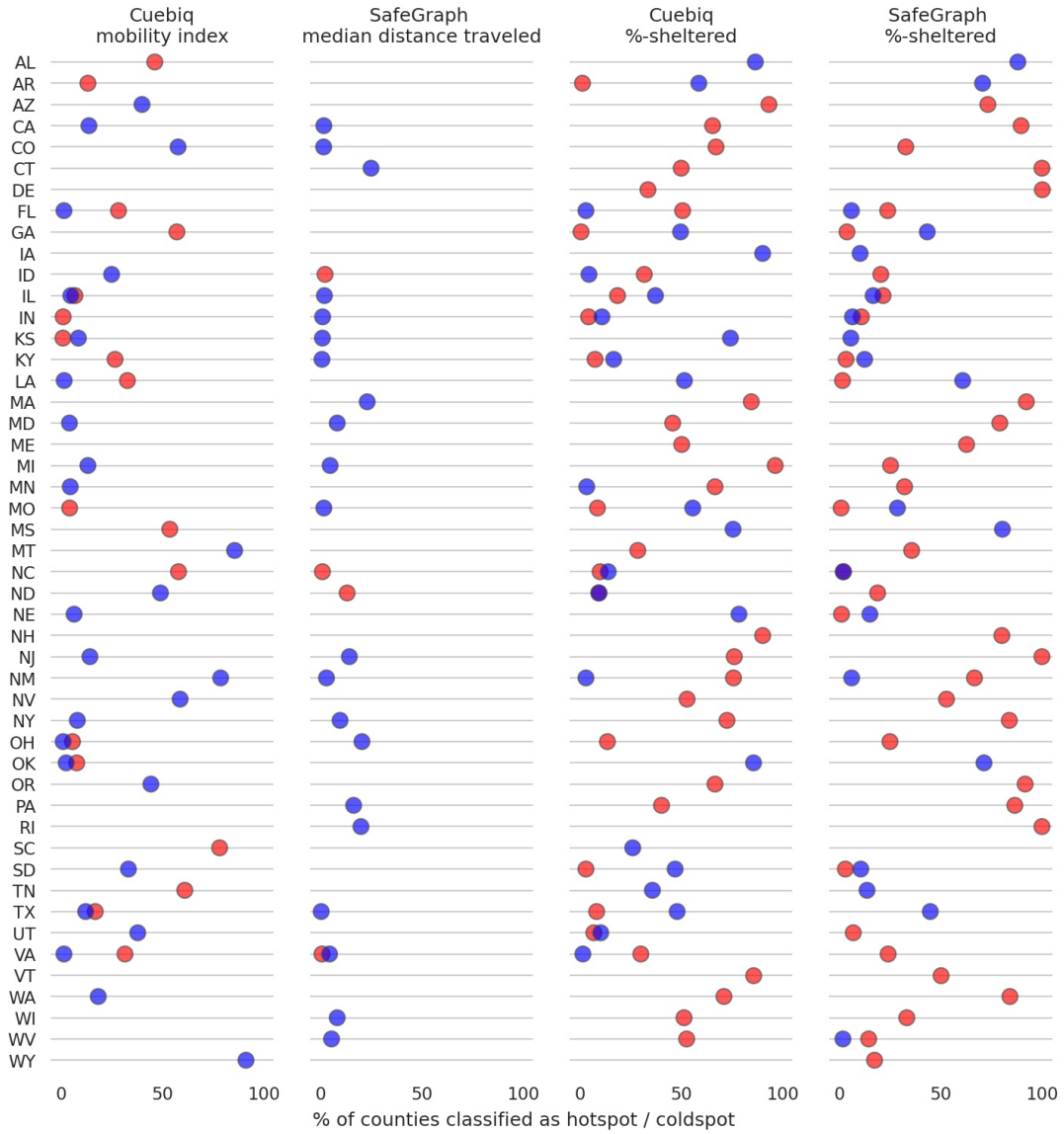

Figure 8. Number of counties classified as either hotspot (red) or coldspot (blue) per state across four indicators in 2020.

according to Table 2, we see a lot more states with both hotspots and coldspots (16 for Cuebiq and 12 for SafeGraph), and a lower number of states without counties classified as hotspot/coldspot, which is indicative of the higher degree of spatial heterogeneity in data, compared to metrics estimating traversed distance.

5. Conclusions

In this brief overview of available data, we identified and assessed the set of metrics that have been used to understand the impact of COVID-19 pandemic on human mobility. Since different metrics of mobility and social distancing have various measurement units and are created by different algorithms, it is extremely challenging to compare them. To do so, we proposed a framework to explore the spatial and temporal structure of these data sets using standard geostatistical tools, such as global and local autocorrelation. We further supplemented this analysis with multivariate cartographic visualizations to reveal the changes in spatial clusters and outliers over time. The developed visualizations enable us to detect the underlying spatiotemporal trends at a glance, and make visual comparison between the metrics more efficient. In general, since this is an exploratory data analysis, we used visualizations to illicit hypotheses, and not confirm them with inferential statistics.

While there are other methods that can be used to assess changes in spatiotemporal structure of mobility, such as Bivariate Spatial Association (Lee 2001), Differential Local Moran (Anselin and Rey 2014), and Local Neighbor Match Test (Anselin and Li 2020), we believe that the proposed methodology provides a robust framework for assessing both spatial and temporal structure of the data. It strikes a good balance between spatial data analysis and user-friendly visualization, that is easy to interpret. More importantly, this framework allows us to explore and examine each of the data source independently, by characterizing its spatial and temporal coverage, and contextualizing the metric in relation to other comparable and similar behaving metrics. Finally, it enables us to pinpoint general regional and national trends in clustering behavior in space and in time.

In conclusion, all reviewed data sets in this paper have limitations, which also limit the inferences that can be made on the data. With this paper, we aimed to shed light on inherent spatial and temporal heterogeneity in commonly used mobility indices, and to caution researchers to be mindful of biases that are present in data and can be exacerbated by re-aggregating, smoothing and interpolating. This paper might help researchers to better understand the differences in mobility data and utilize the metrics that are better suited for their research questions. However, relying solely on one data source may be misleading as demonstrated in this paper via discrepancies captured in recency and consistency of the clusters for various mobility metrics in space and time. Future research is needed to contextualize these similarities and differences in associations with sociodemographic and geographic heterogeneity of different states.

Further research should assess the effects of different spatial matrix specifications and spatial units (e.g. regular raster grids, heterogeneous polygons) on the delineation of hotspots and coldspots, and investigate the issues of analytical scale on visualizations of patterns. Additionally, standardizing and summarizing reviewed mobility metrics should yield a more universal mobility metric, allowing more formal analysis with inferential statistics.

6. Acknowledgment

The authors gratefully acknowledge the support from the National Science Foundation Award BCS #2043202, and the UCSB Vice Chancellor for Research COVID-19 Seed Program to conduct this research.

Data and Codes Availability Statement

Code

To supplement reusability and replicability in research and to allow reproduction of maps, figures and tables included in the manuscript, we shared our code and a data sample from SafeGraph via Figshare: `https://doi.org/10.6084/m9.figshare.16892320.v1`. Please see the *README* file in the root directory for more details.

Data

There are two types of availability for the data used in the paper: open-source data (Group 1) and commercial data (Group 2).

Group 1: the data that support the findings of this study are openly available or available upon request at no charge:

(1) Apple COVID-19 Mobility Trend Reports at `https://covid19.apple.com/mobility`
(2) Facebook Data for Good at `https://dataforgood.fb.com/tools/movement-range-maps/`
(3) Google COVID-19 Community Mobility Reports at `https://www.google.com/covid19/mobility/`
(4) Exposure indices derived from PlaceIQ movement data at `https://github.com/COVIDExposureIndices/COVIDExposureIndices`
(5) SafeGraph Data for Academics at `https://www.safegraph.com/academics` (upon request, for academic use)
(6) Data for Mobility Changes in Response to COVID-19 (Descartes Labs) at `https://github.com/descarteslabs/DL-COVID-19`
(7) University of Maryland COVID-19 Impact Analysis Platform at `https://data.covid.umd.edu/`

Group 2: Restrictions apply to the availability of these data, which were used under license for this study. Data are available with the permission of third party:

(1) Cuebiq Mobility Insights at `https://www.cuebiq.com/visitation-insights-covid19/`
(2) StreetLight Big Data for Mobility at `https://www.streetlightdata.com/covid-transportation-metrics/`

References

Anselin, L., 1995. Local indicators of spatial association—LISA. *Geographical Analysis*, 27 (2), 93–115.

Anselin, L. and Li, X., 2020. Tobler's Law in a Multivariate World. *Geographical Analysis*, 52 (4), 494–510.

Anselin, L. and Rey, S., 1991. Properties of tests for spatial dependence in linear regression models. *Geographical analysis*, 23 (2), 112–131.

Anselin, L. and Rey, S.J., 2014. *Modern spatial econometrics in practice: A guide to GeoDa, GeoDaSpace and PySAL*. GeoDa Press LLC.

Anselin, L., Syabri, I., and Kho, Y., 2010. Geoda: an introduction to spatial data analysis. *In*: *Handbook of applied spatial analysis*. Springer, 73–89.

Arnal, R.P., *et al.*, 2020. Private sources of mobility data under COVID-19. *arXiv preprint arXiv:2007.07095*.

Badr, H.S., *et al.*, 2020. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*, 20 (11), 1247–1254.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57 (1), 289–300.

Bonaccorsi, G., *et al.*, 2020. Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences*, 117 (27), 15530–15535. Available from: `https://www.pnas.org/content/117/27/15530`.

Cacciapaglia, G., Cot, C., and Sannino, F., 2020. Mining Google and Apple mobility data: Twenty-one shades of European social distancing measures for COVID-19. *arXiv preprint arXiv:2008.02117*.

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

Cuebiq, 2021. Cuebiq mobility insights. Accessed: 2021-01-19, Available from: `https://www.cuebiq.com/visitation-insights-covid19/`.

Dodge, S., *et al.*, 2020. Progress in computational movement analysis – towards movement data science. *International Journal of Geographical Information Science*, 34 (12), 1–6. Available from: `https://doi.org/10.1080/13658816.2020.1784425`.

Engle, S., Stromme, J., and Zhou, A., 2020. Staying at home: mobility effects of COVID-19. *Available at SSRN https://ssrn.com/abstract=3565703 or http://dx.doi.org/10.2139/ssrn.3565703*.

Farber, S., Páez, A., and Volz, E., 2009. Topology and dependency tests in spatial and network autoregressive models. *Geographical Analysis*, 41 (2), 158–180.

Flaxman, S., *et al.*, 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*.

Florax, R.J. and Rey, S., 1995. The impacts of misspecified spatial interaction in linear regression models. *In*: *New directions in spatial econometrics*. Springer, 111–135.

Gabriel Dance, Lazaro Gamio, 2020. Millions in U.S. are leaving home again. May. Available from: `https://www.nytimes.com/interactive/2020/05/12/us/coronavirus-reopening-shutdown.html`.

Gao, S., *et al.*, 2020. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SIGSPATIAL Special*, 12 (1), 16–26. Available from: `https://doi.org/10.1145/3404820.3404824`.

Gatto, M., *et al.*, 2020. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences*, 117 (19), 10484–10491. Available from: `https://www.pnas.org/content/117/19/10484`.

Geary, R.C., 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5 (3), 115–146.

Getis, A. and Ord, J.K., 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24 (3), 189–206. Available from: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4632.1992.tb00261.x`.

Gotway, C.A. and Young, L.J., 2002. Combining incompatible spatial data. *Journal of the American Statistical Association*, 97 (458), 632–648.

Grantz, K.H., *et al.*, 2020. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature Communications*, 11 (4961), 1–8.

Heiler, G., *et al.*, 2020. Country-wide mobility changes observed using mobile phone data during COVID-19 pandemic. *arXiv preprint arXiv:2008.10064*.

Huang, X., *et al.*, 2021. The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the US during the COVID-19 pandemic. *International Journal of Digital Earth*, 1–19.

Kim, J. and Kwan, M.p., 2021. The impact of the COVID-19 pandemic on people's mobility: A longitudinal study of the U.S. from March to September of 2020. *Journal of Transport Geography*, 93 (December 2020), 103039.

Kishore, N., *et al.*, 2020. Measuring mobility to monitor travel and physical distancing interventions: A common framework for mobile phone data analysis. *The Lancet Digital Health*.

Kitchin, R., 2020. Using digital technologies to tackle the spread of the coronavirus: Panacea or folly? *The Programmable City Working Paper*, 44 (April), 1–24. Available from: `http://progcity.maynoothuniversity.ie/wp-content/uploads/2020/04/Digital-tech-spread-of-coronavirus-Rob-Kitchin-PC-WP44.pdf`.

Klein, B., *et al.*, 2020. Assessing changes in commuting and individual mobility in major metropolitan areas in the United States during the COVID-19 outbreak. Available from: `https://www.networkscienceinstitute.org/publications/assessing-changes-in-commuting-and-individual-mobility-in-major-metropolitan-areas-in-the-united-states-during-the-covid-19-outbreak`.

Kraemer, M.U., *et al.*, 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368 (6490), 493–497.

Lee, S.I., 2001. Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I. *Journal of Geographical Systems*, 3 (4), 369–385.

Maryland Transportation Institute, 2020. University of Maryland COVID-19 Impact Analysis Platform. Accessed: 2020-01-19, Available from: `https://data.covid.umd.edu/`.

Matheron, G., 1963. Principles of geostatistics. *Economic Geology*, 58 (8), 1246–1266.

McAloon, C., *et al.*, 2020. Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research. *BMJ open*, 10 (8), e039652.

McKenzie, G. and Adams, B., 2020. A country comparison of place-based activity response to COVID-19 policies. *Applied Geography*, 125, 102363. Available from: `https://www.sciencedirect.com/science/article/pii/S0143622820306238`.

Moran, P.A.P., 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10 (2), 243–251. Available from: `http://www.jstor.org/stable/2983777`.

Morita, H., Nakamura, S., and Hayashi, Y., 2020. Changes of urban activities and behaviors due to COVID-19 in Japan. *Available at SSRN: https://ssrn.com/abstract=3594054 or http://dx.doi.org/10.2139/ssrn.3594054*.

Oliver, N., *et al.*, 2020. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*, 6 (23). Available from: `https://advances.sciencemag.org/content/6/23/eabc0764`.

Peng, Z., *et al.*, 2020. Exploring urban spatial features of COVID-19 transmission in Wuhan based on social media data. *ISPRS International Journal of Geo-Information*, 9 (6), 402.

Rey, S.J. and Anselin, L., 2007. PySAL: A Python library of spatial analytical methods. *The Review of Regional Studies*, 37 (1), 5–27.

Ripley, B.D., 1976. The second-order analysis of stationary point processes. *Journal of applied probability*, 13 (2), 255–266.

Robinson, W.S., 2009. Ecological correlations and the behavior of individuals. *International Journal of Epidemiology*, 38 (2), 337–341.

SafeGraph, 2021. Safegraph data examples. Accessed: 2021-01-19, Available from: `https://www.safegraph.com/data-examples`.

Schlosser, F., *et al.*, 2020. COVID-19 lockdown induces structural changes in mobility networks–Implication for mitigating disease dynamics. *arXiv preprint arXiv:2007.01583*.

Shuja, J., *et al.*, 2021. Covid-19 open source data sets: A comprehensive survey. *Applied Intelligence*, 51, 1296–1325.

Smith, T.E., 2009. Estimation bias in spatial models with strongly connected weight matrices. *Geographical Analysis*, 41 (3), 307–332.

Stakhovych, S. and Bijmolt, T.H., 2009. Specification of spatial models: A simulation study on weights matrices. *Papers in Regional Science*, 88 (2), 389–408.

Yule, G. and Kendall, M., 1950. An introduction to the theory of statistics. *London: Griffin*.

Appendix A. Spatial and temporal coverage of mobility indices

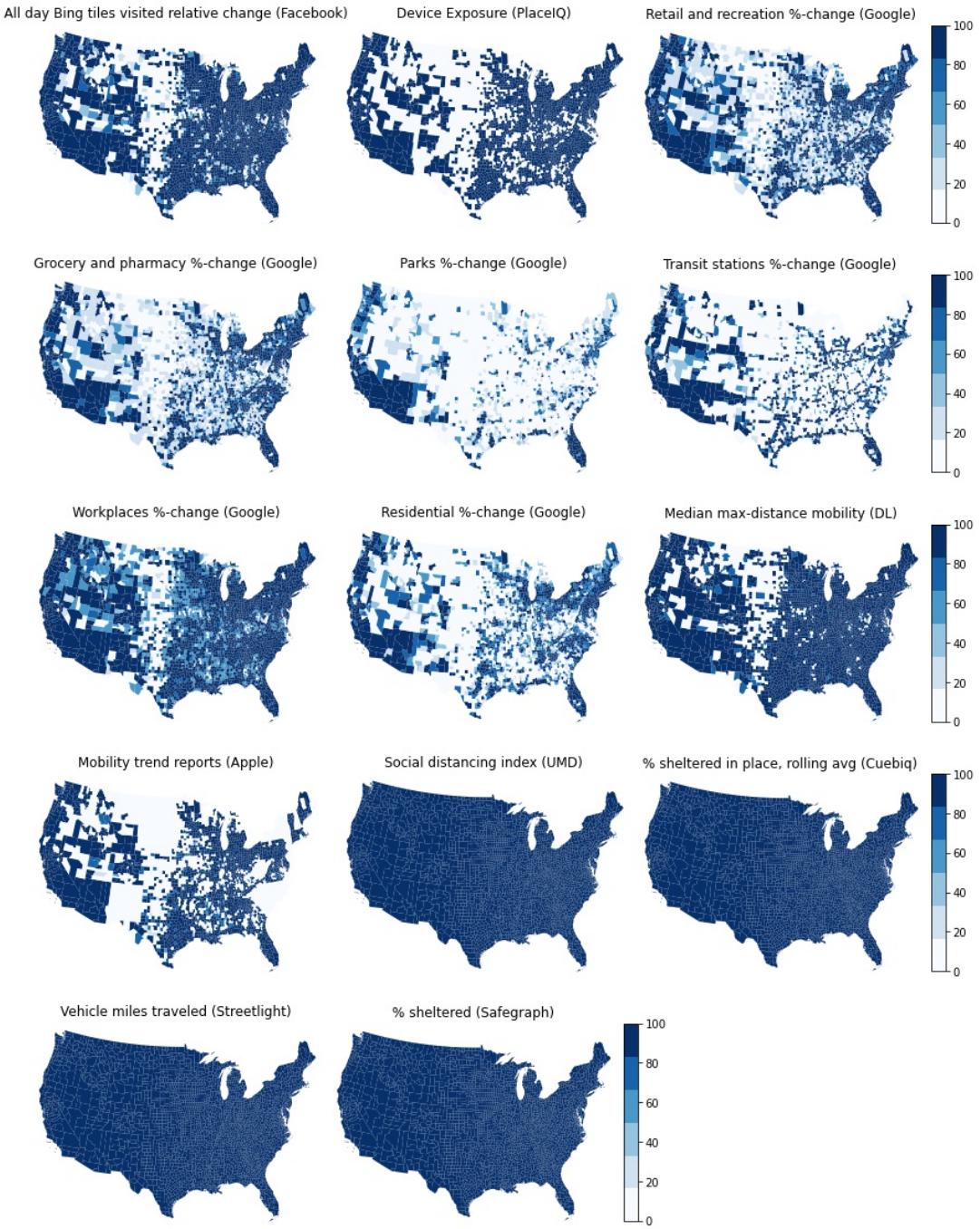

Figure A1. Spatial coverage of mobility sources. Darker shade indicates higher percentage of complete observations per county for spatial coverage.

Appendix B. Box plots of standardized mobility metrics

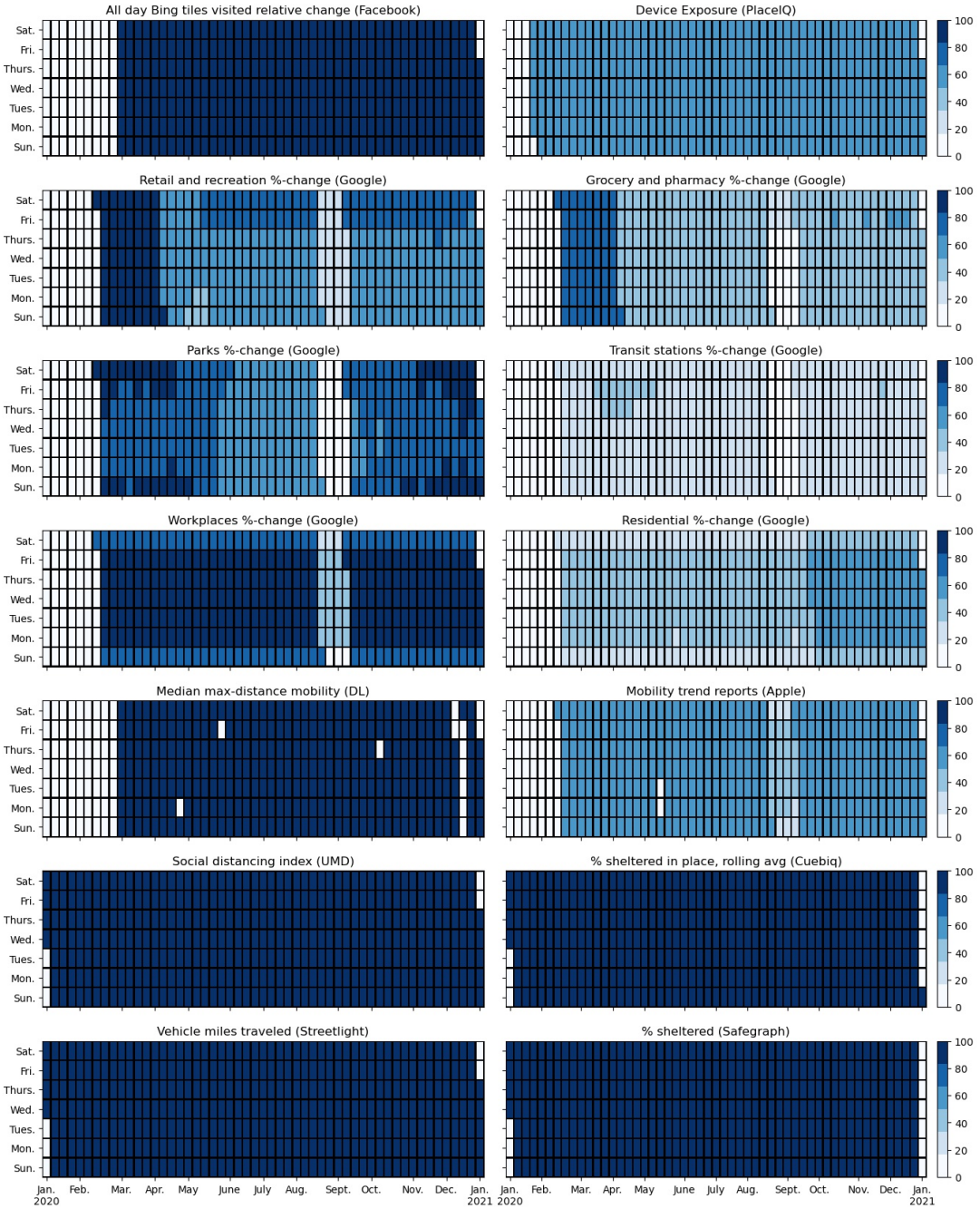

Figure A2. Temporal coverage of mobility sources. Darker shade indicates higher percentage of complete observations per day.

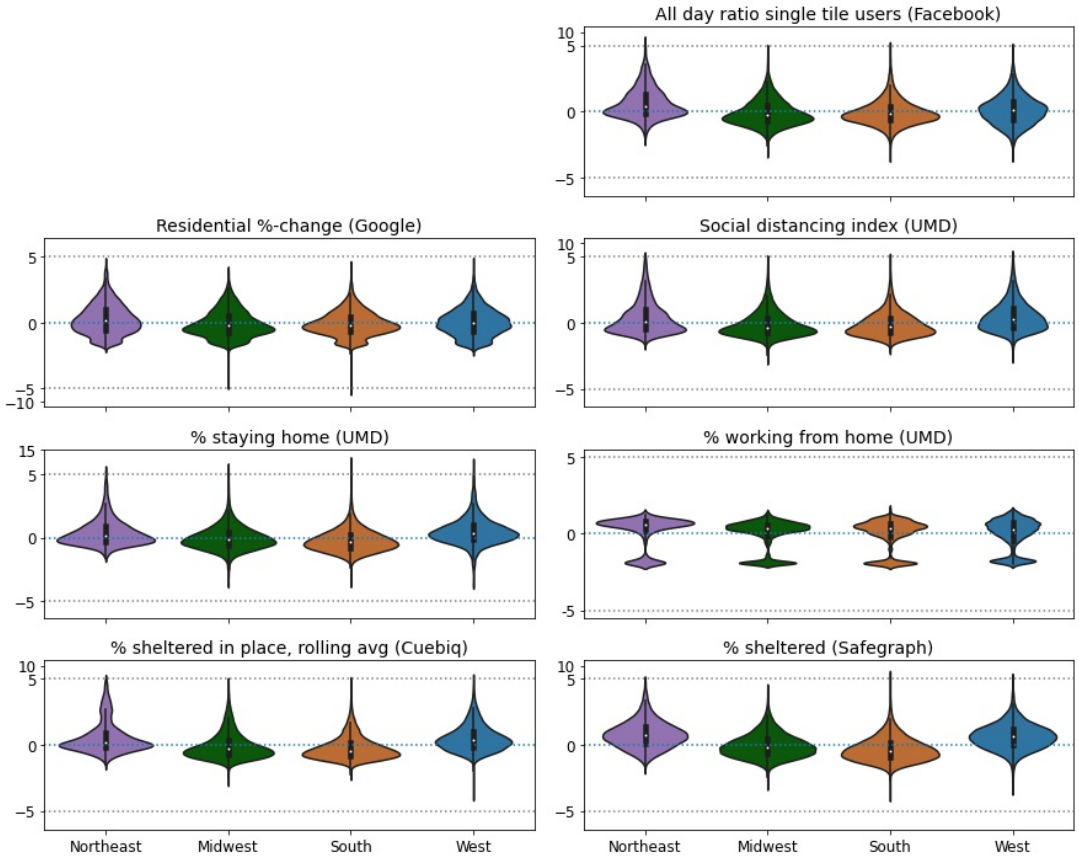

Figure B1. The violin plots of standardized mobility metrics quantifying social distancing by US region (blue - West, green - Midwest, purple - Northeast, tawny - South). Custom linear scaling on the y-axis is used for values exceeding 5 standard deviations. Inside of each violin plot is a box plot. Each distribution is scaled to be of the same area.

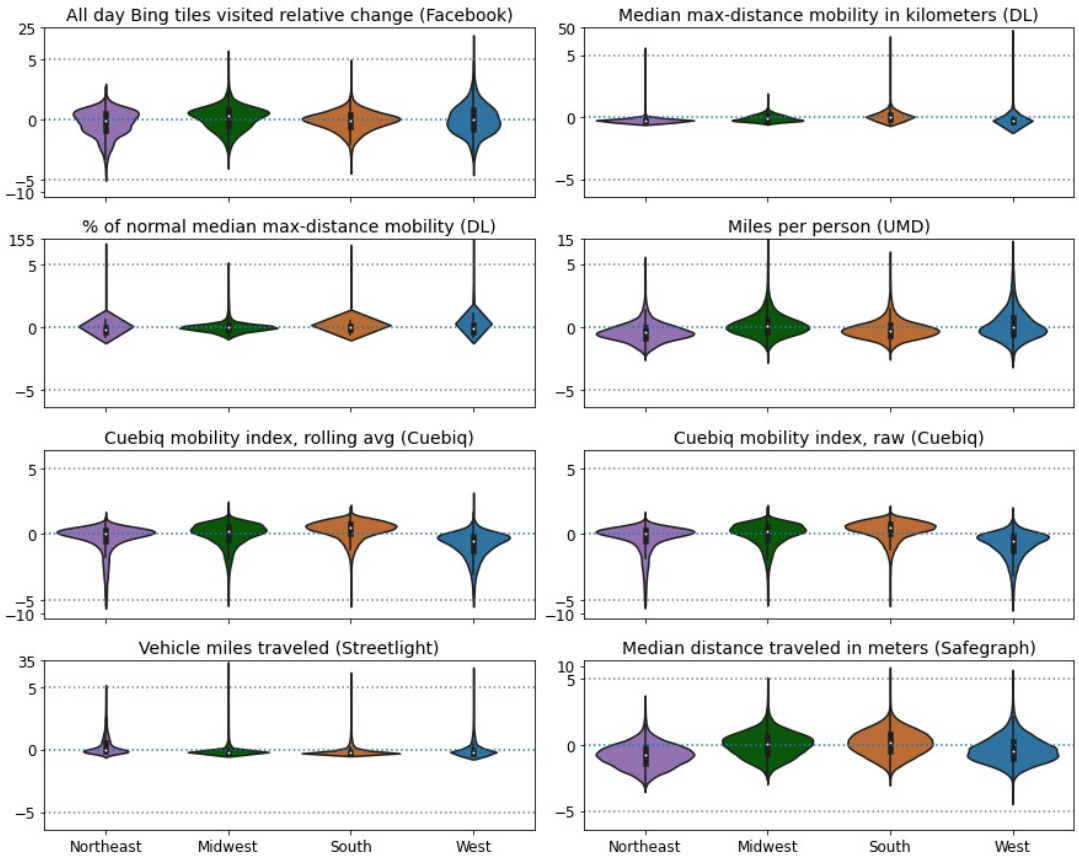

Figure B2. The violin plots of standardized mobility metrics estimating traversed distance by US region (blue - West, green - Midwest, purple - Northeast, tawny - South). Custom linear scaling on the y-axis is used for values exceeding 5 standard deviations. Inside of each violin plot is a box plot. Each distribution is scaled to be of the same area.

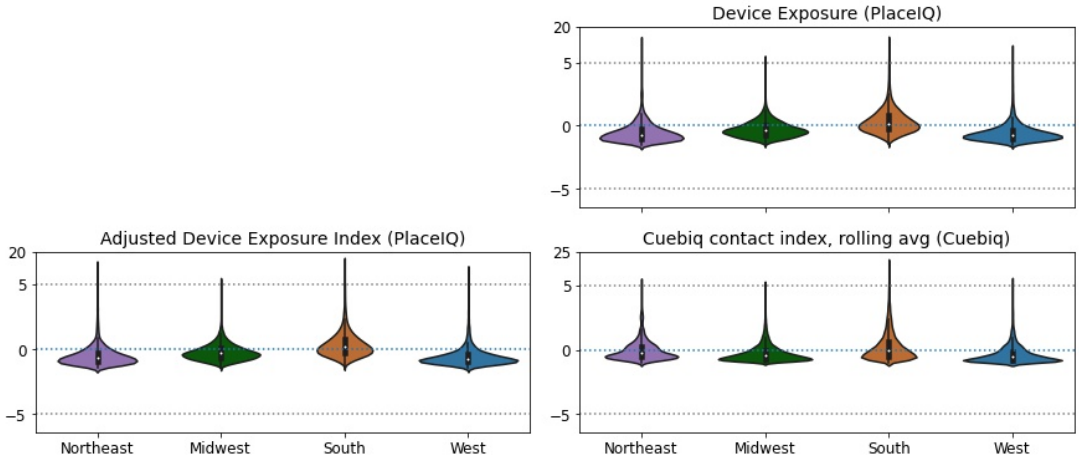

Figure B3. The violin plots of standardized mobility metrics modeling contacts or exposure by US region (blue - West, green - Midwest, purple - Northeast, tawny - South). Custom linear scaling on the y-axis is used for values exceeding 5 standard deviations. Inside of each violin plot is a box plot. Each distribution is scaled to be of the same area.

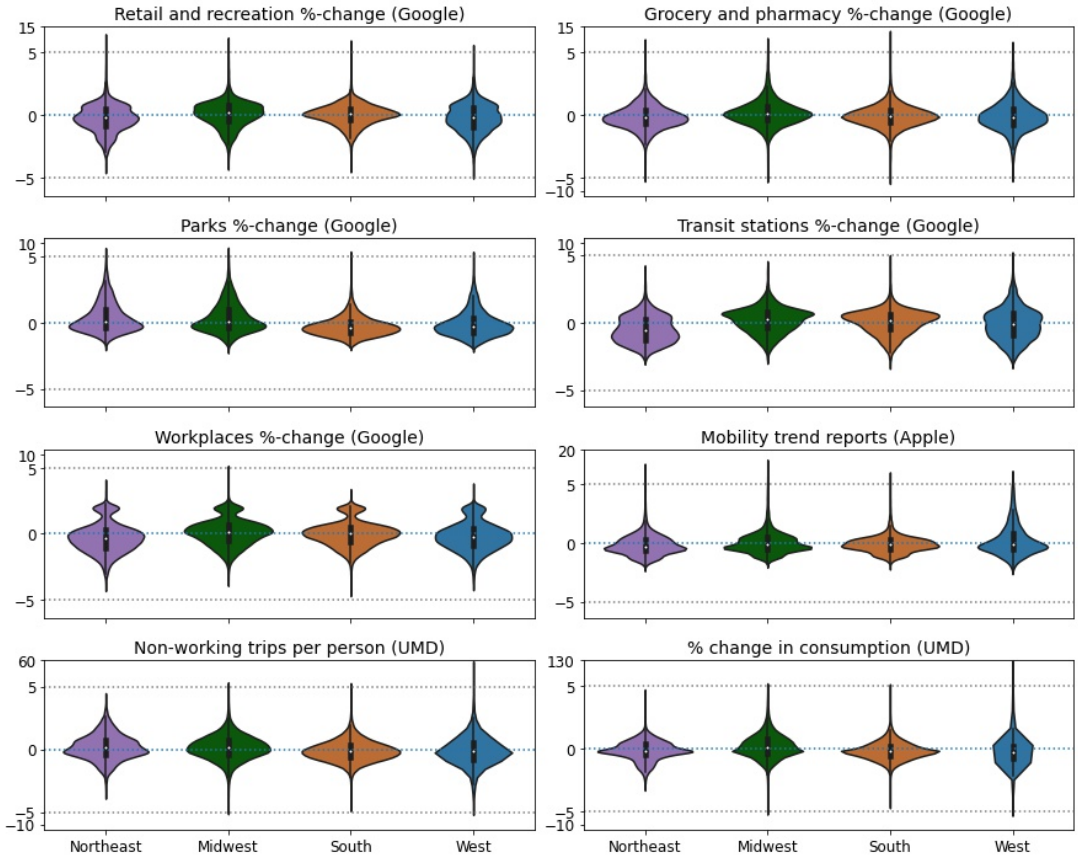

Figure B4. The violin plots of standardized mobility metrics quantifying the number of trips or visit counts by US region (blue - West, green - Midwest, purple - Northeast, tawny - South). Custom linear scaling on the y-axis is used for values exceeding 5 standard deviations. Inside of each violin plot is a box plot. Each distribution is scaled to be of the same area.

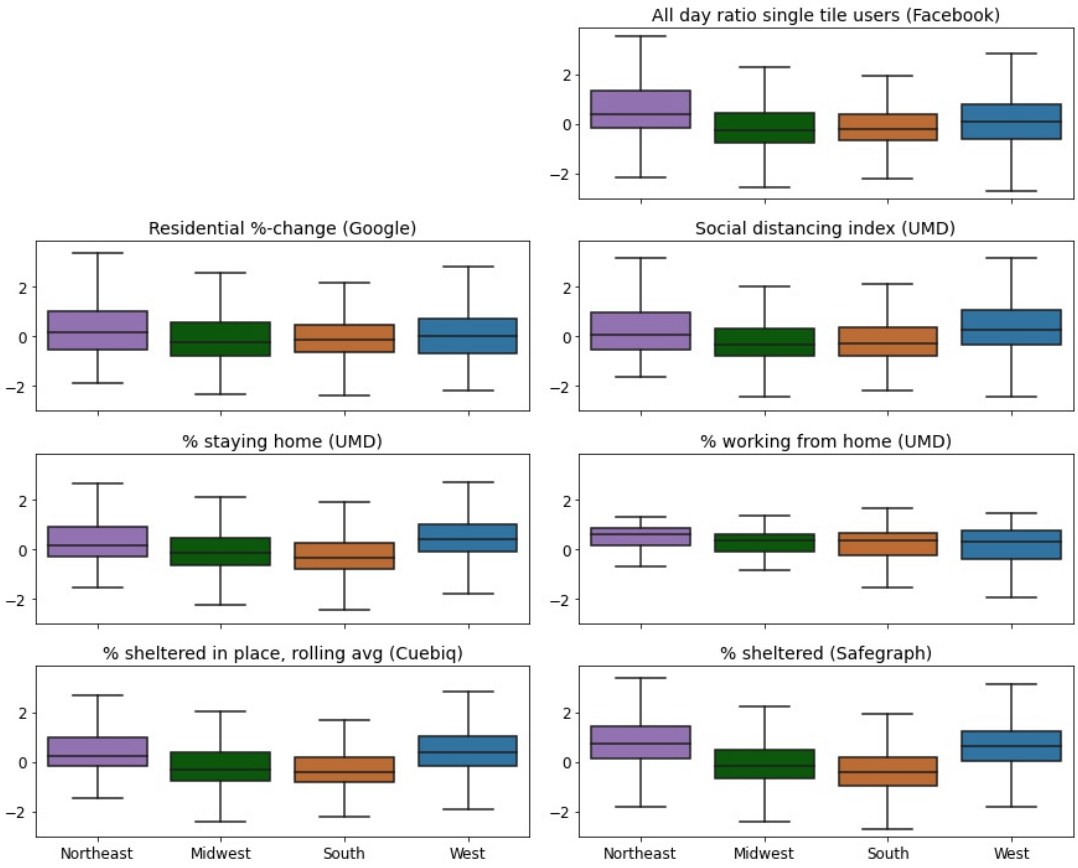

Figure B5. The box plots of standardized mobility metrics quantifying social distancing by US region (blue - West, green - Midwest, purple - Northeast, tawny - South).

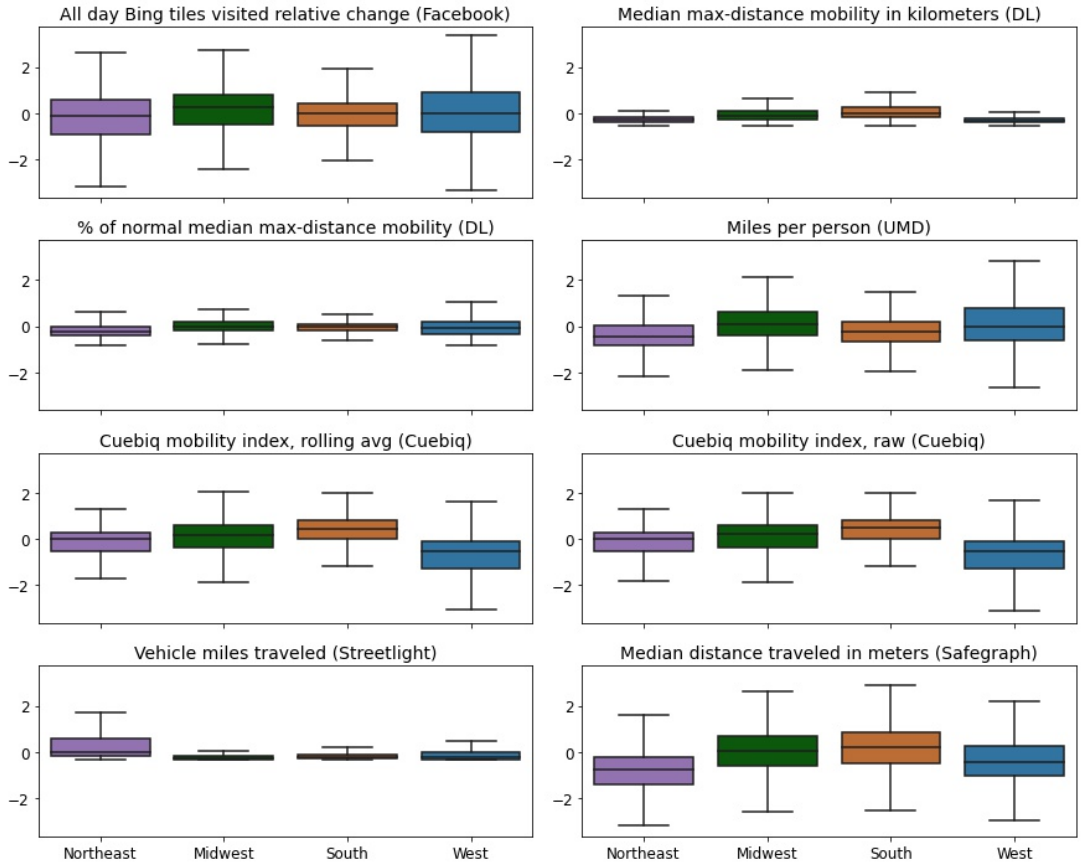

Figure B6. The box plots of standardized mobility metrics estimating traversed distance by US region (blue - West, green - Midwest, purple - Northeast, tawny - South).

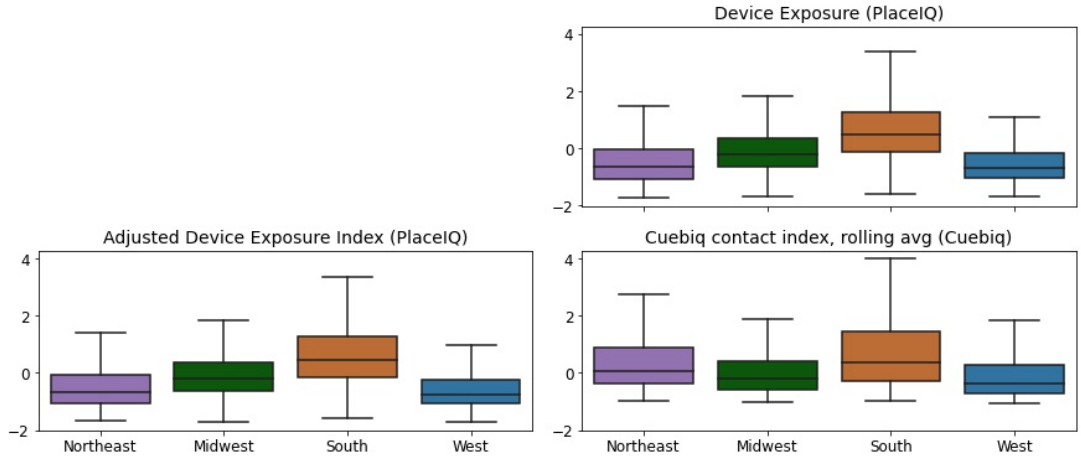

Figure B7. The box plots of standardized mobility metrics modeling contacts or exposure by US region (blue - West, green - Midwest, purple - Northeast, tawny - South).

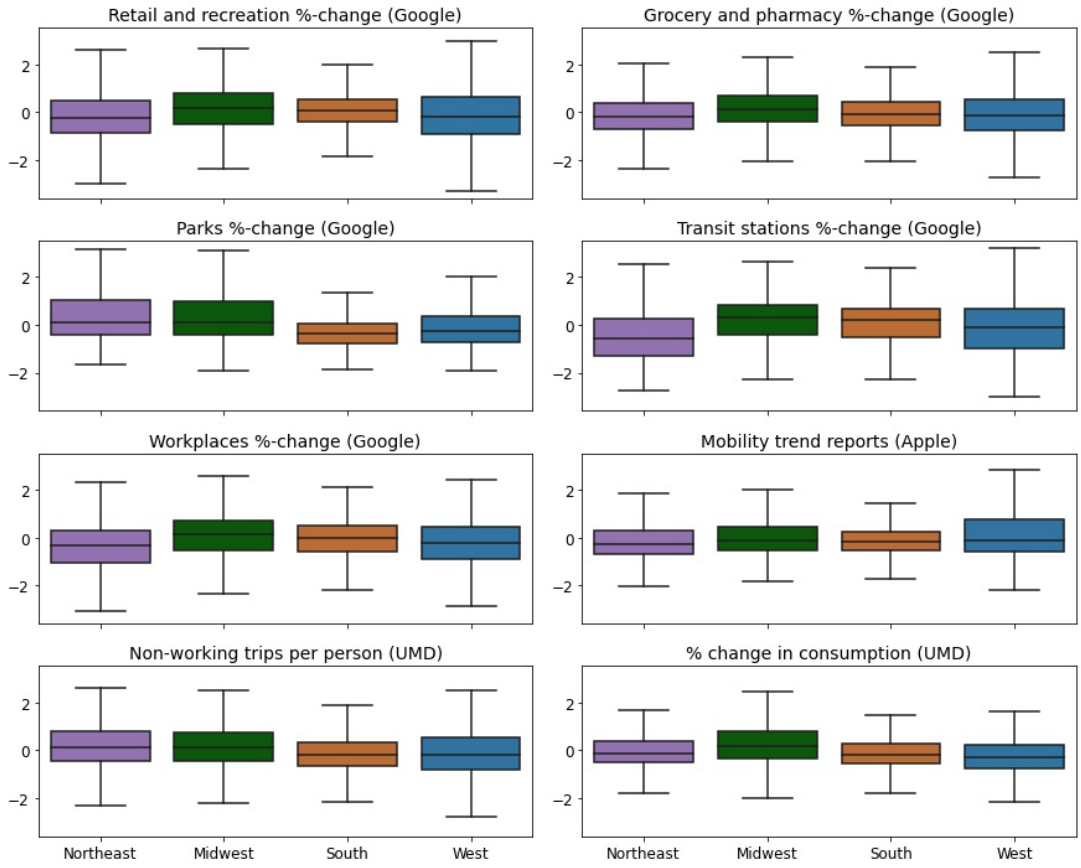

Figure B8. The boxplot of standardized mobility metrics quantifying the number of trips or visit counts by US region (blue - West, green - Midwest, purple - Northeast, tawny - South).

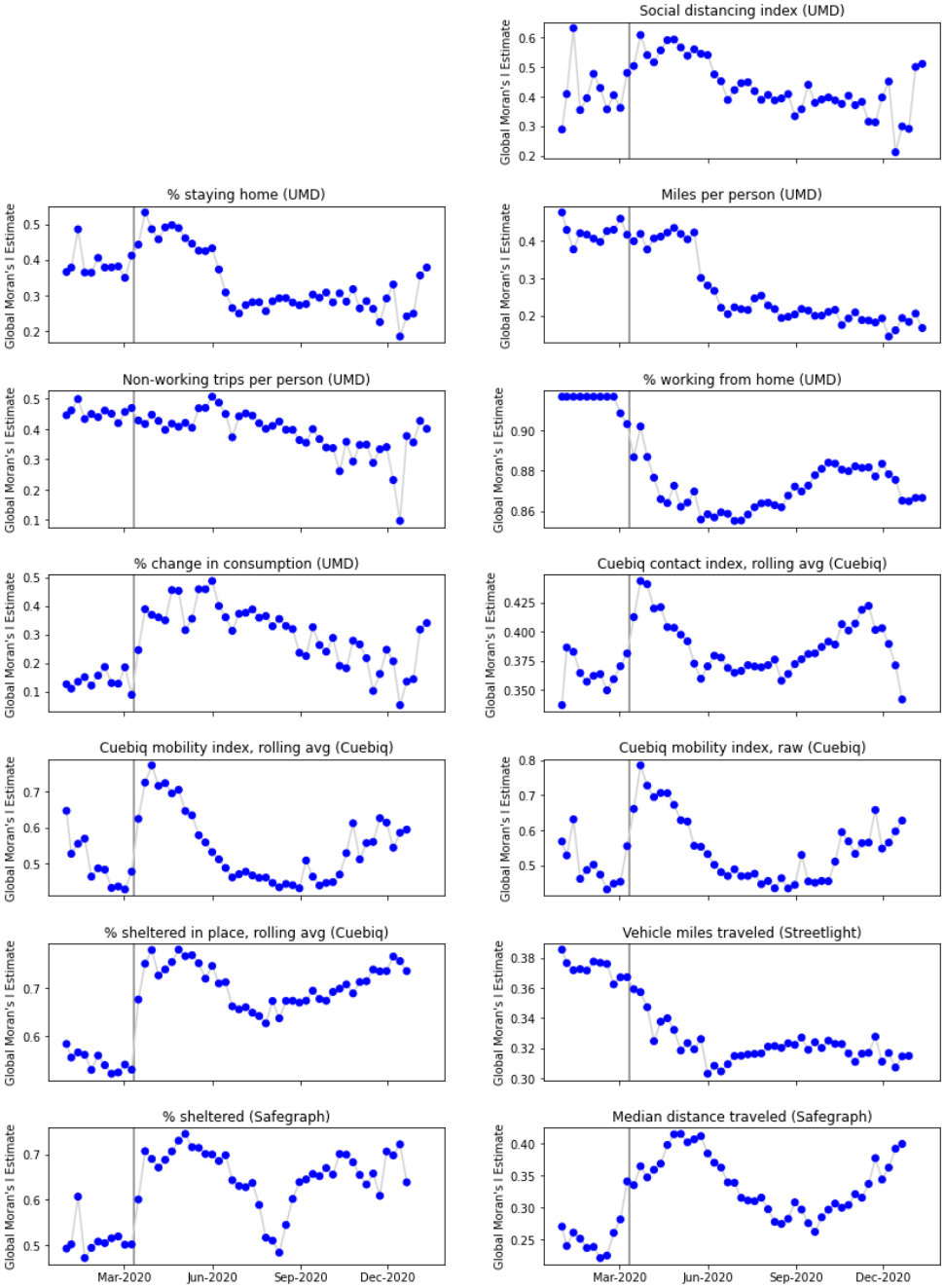

Figure B9. Weekly estimated Global Moran's I for the United States. Blue points denote statistically significant Moran's I values. Grey line denotes the date when COVID-19 was declared a pandemic by the World Health Organization.