Proceedings of the IISE Annual Conference & Expo 2022
K. Ellis, W. Ferrell, J. Knapp, eds.

Interactive dashboards to study relations between early COVID-19 outbreaks and human mobility trends

Juan-Alberto Estrada-Garcia[a], Siqian Shen[a], and Wen Ye[a]

[a] *University of Michigan, Ann Arbor, Michigan, USA*

Abstract

Outbreaks of the COVID-19 pandemic, caused by the SARS-CoV-2 virus, have led to the creation of social distancing and lockdown policies to reduce the spread of the virus. Consequently, public/private transportation services, schools, workplaces, and retail stores' operations were disrupted. We gather user mobility reports worldwide to learn impacts of early COVID-19 outbreaks on human mobility patterns and trends. Building time series of six types of activities tracked in the Google Community Mobility Reports (CMR), we develop visualization tools and interactive dashboards for linking mobility and COVID-19 infection data at different levels (from county- and state-level in the US, to country level for the rest of the world). We show that the relationship between mobility and COVID-19 infection changes over time, and therefore the stage of the pandemic is essentially important for understanding how containment policies can affect infections and deaths caused by the COVID-19 pandemic.

Keywords
COVID-19 pandemic; mobility data analysis; time-series analysis; data visualization

1. Introduction
Since early 2020, the SARS-CoV-2 (COVID-19) pandemic has significantly affected human travel and other activities as disease control and intervention strategies, such as social distancing and government-mandated lockdowns, took place in many countries and regions. As the pandemic changes different aspects of human activities, the science community has developed methods and tools to understand key issues related to combating the disease spread. One important aspect is the mobility and its relationship with virus transmission. In this study, we aim to build real data-based online dashboards to link COVID-19 infections with travel behavior for different geographical zones.

Since the beginning of the pandemic, companies, institutions, and government agencies in different countries have released open data to better understand the trends of the pandemic. For example, the New York Times [1] releases daily updates about COVID-19 confirmed cases, deaths, and testing results in the United States (US), from national- to state- and county-levels. The daily infected cases and deaths for most countries are updated in [2], [3]. Meanwhile, Google has an online repository that updates daily mobility data for different regions in the world, called the Community Mobility Reports (CMR), specifying travel behavior and mobility results by activity type [4].

As most existing COVID-19 interactive dashboards report dynamic updates about COVID-19 infection and testing, few try to interpret them based on other variables such as mobility or economic activities [5], [6]. Some dashboards show the effects of COVID-19 infections on mobility, economy, health, and society, but only for certain areas in the US [3], [7], [8]. To our best knowledge, there have not been automated visualization dashboards that link mobility and COVID infections for different regions of the world using Google's CMR, which updates daily mobility data of different activities from a variety of countries.

Among the literature that draws relations between mobility and the COVID-19 spread, Warren & Skillman [9] utilized visualization tools to study mobility changes in response to disease spread and control policies for several US counties. They concluded that mobility patterns changed dramatically in US counties and were influenced by the urbanization degree in the region. Bonaccorsi et al. [10] analyzed mobility data before and after the early 2020 lockdown in Italy, to understand mobility restrictions' consequences to the society and economy, concluding that the lockdown affected poor populations the most. Similarly, Coven & Gupta [11] studied the impact of demographic differences on mobility response during the pandemic among New York City (NYC) neighborhoods. Using GPS data, they calculated the neighborhoods' propensity to leave the city, concluding that richer NYC residents were more likely to move out the city to less-populated zones, and lower-income areas exhibited more mobility and travel. One other study by Chang et al. [12] used mobility data from points of interest in the US to develop a Susceptible, Exposed, Infectious, Removed (SEIR) model to analyze geographic distribution of the infected cases in the US at the neighborhood level for ten metro areas. Their results suggested that infection differences in regions were mostly influenced by temporary short-term decisions rather than long-term preexisting conditions.

Sulyok & Walker [13] used data of the six activities reported in Google's CMR [4] to cluster countries based on the Kendall's tau rank correlation (described in Section 3.3.1) to study mobility's impacts on COVID-19 infections. They later used the best correlated data to fit generalized additive models [14] to predict worldwide infections. They concluded that mobility data can help to improve modelling of how infections spread through time, at least partially as a function of human daily activities. However, none of the studies propose interactive dashboards that can render their results accessible to users, nor make their analyses reproducible and available for replication. In this paper, we develop visualization dashboards using the Google's CMR data and plot 2D and 3D scatter diagrams that pair infection changes (considering lagged data), mobility changes and time to visualize the interaction of these variables for most counties in the US, that could also be expanded to the analysis of users' own data.

We develop a tool that links mobility and COVID-19 data, therefore becoming a source for more detailed studies to analyze human activity patterns to better model their impact on COVID-19 infections. The main contributions of our paper include: (i) cross country dashboards organized by continent, (ii) two-level dashboards for regions and subregions in the US, (iii) infection-mobility correlation dashboard for eight selected countries, (iv) a tool to identify, for most US counties, if lagged infection data is better correlated with mobility, as well as (v) 2D and 3D scatter plots for average weekly aggregated changes in infections and mobility. The development of these tools requires matching and joining the records of different data sources. We develop algorithms to automate the update of the dashboards as new Google's CMR and COVID-19 infection data becomes available. Similarly, for the correlation analysis, we develop automation algorithms for the analysis and subsequent plotting of variables.

2. Data

2.1 Mobility data

We use Google Community Mobility Reports [4] as the main mobility data source. These reports are released as comma-separated values files that cover data on 135 countries. Historical data is available from February 15, 2020, to date. Three levels of granularity are reported if available: Level 0, corresponding to metrics aggregated by country or region (e.g., the US), Level 1, for metrics aggregated by geopolitical subdivisions (e.g., states in the US), and Level 2, for higher-resolution subdivisions (e.g., counties in the US). Levels 1 and 2 vary because different countries have different geopolitical subdivisions. No data is published for regions smaller than 3 km^2. The reports provide metrics on the daily relative change in mobility from Google applications' users for smartphones and other devices. Data is aggregated by region to ensure that anonymity is maintained from the records. When the number of visits does not meet the security threshold determined by Google, no value is reported for that record, and is presented as a gap in the dataset [15]. For each day of a week, a baseline value is calculated considering the mean value of mobility from records in the five-week period of January 3, 2020, to February 6, 2020. The report covers six different aggregated discrete categories (i) workplaces, (ii) residential, (iii) retail, (iv) grocery and pharmacy, (v) parks, and (vi) transit stations. For each category, several types of places are contained (e.g., for transit stations the considered places are subway stations, seaports, taxi stands, highway rest stops and car rental agencies).

2.2 COVID-19 infection cases

We obtain COVID-19 infection data from the Johns Hopkins GitHub Repository [2], using two available reports: (i) a worldwide dataset with country level confirmed cases time series and (ii) a US county-level confirmed cases time series. Data is aggregated from different sources, such as international and individual country health organizations depending on the aggregation level. For the US, Canada, and Australia, subregion-specific information is obtained from their specific government datasets. For China, the data is obtained from the Chinese CDC and is reported at the province level. For the rest of the world, data is reported at the country level.

3. Methods

3.1 Data processing

We first process Google's CMR by converting the integer to decimal (e.g., from 20 in the original dataset to 0.20). We then filter the data by geographic zones and compute weekly average. To compare the mobility relative change with the relative confirmed increased cases, we compute the daily difference of reported cases, and then scale the increase in confirmed cases as a number between 0 (i.e., the lowest increase) and 1 (i.e., the highest increase) on the dataset. We scale the time series using MinMaxScaler function in Python [16]. We then aggregate the daily infection increase by calculating the mean values over each period of seven days. We use an inner join between the mobility data and the infection data to match the county and date for each record in the database. We then generate 4-week lagged data from lag 1 to lag 4 to analyze cross correlation Kendall's tau (described in Section 3.3.1) between changes of mobility patterns and COVID-19 infections. Given these two main datasets, **Figure 1** shows the data management structure for the created dashboards and tools.

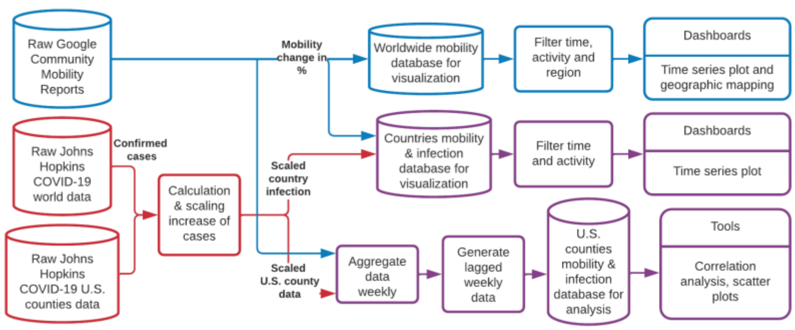


Figure 1: Data flow process. Using this basic process, we can generate datasets that serve different purposes. We develop two data structures: (i) time series of mobility and infection changes for different geographically-located regions and (ii) weekly aggregated mobility and infection time series for US counties.

3.2 Visualization dashboards

We develop a web-based interactive dashboard using Google Data Studio (GDS) as the hosting platform. From the previously stated data structures, we develop an interactive dashboard for the mobility and infection data. As these dashboards serve a visualization and descriptive purpose, we identify the time series plot and choropleth map as useful graphics to be used for the dashboards. To enhance data visualization and take advantage of the different available region levels, we develop several variations of the dashboards: (i) With country-level data, a world wide dashboard and a dashboard for each continent, (ii) with US state-level data, a US national state dashboard, and (iii) with US county-level data, a US national county dashboard. The user has the ability to select a time frame, the type of activity to visualize, and the regions to be displayed. Additionally, the filtered datasets that are generated from these user selections can be downloaded from the dashboard.

3.3 County level analysis tool

Complementary to the visualization dashboards, we develop a county analysis tool, which focuses on the county-level (US subregions). Given a desired county, by providing the county's and the state it sits in, the tool filters the respective data, through a cross-correlation analysis determines the lag number that correlate the best with the infection increase time series, to then output the filtered data, the aggregated data, correlation results, and scatter plots. We select this approach to let users continue with further analysis with the data they are visualizing.

3.3.1 Cross-correlation analyses

Due to infection and mobility's weekly seasonality, we aggregate data weekly and calculate the mean value. Because of the infection's lagged nature, we consider the work by Sulyok & Walker [13], in which mobility data is considered non-parametric, and thus the rank correlation is selected to be the correlation test for the variables' ordinal association, in order to establish whether two datasets could be regarded as statistically dependent. We calculate Kendall's tau [17] as shown in **Equation 1**:

Given a set of observations $(x_1, y_1), \ldots, (x_n, y_n)$ of the joint random variables X and Y:

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} sgn(x_i - x_j)\, sgn(y_i - y_j) \tag{1}$$

The coefficient is in the range of $-1 \leq \tau \leq 1$. Here, 1 represents a perfect agreement between the two rankings and -1 is a perfect disagreement. If the coefficient is 0, the rankings are expected to be independent.

4. Results

4.1 Dashboards

We develop the dashboards on the GDS platform [18], each showing two-part mobility data visualizations: The left-most part plots the regions and mobility activities' time series, and the right-most maps the selected regions and colors them with a color gradient based on the comparison on the mean mobility changes in the selected time frame. **Figure 2** shows a dashboard that follows the described structure and is one of the developed dashboards. By selecting the regions from the drop down menu, the time span and the optional metrics for the selected activities, the time series are plotted and the corresponding data mapped.

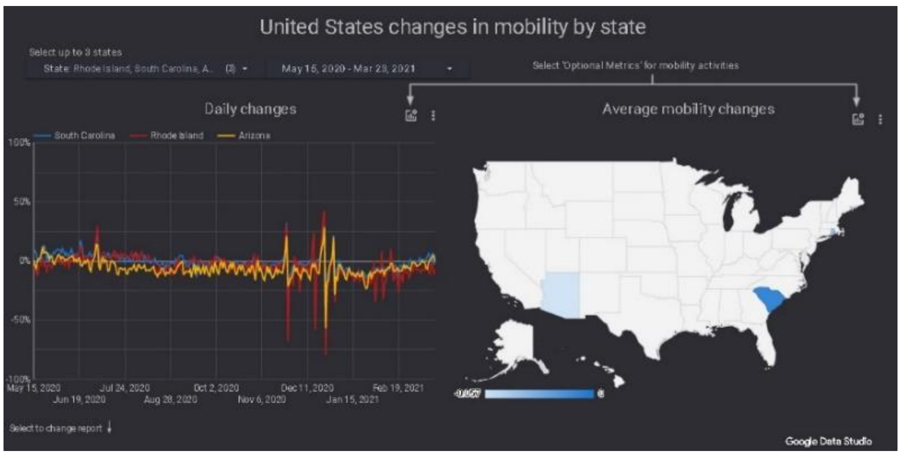


Figure 2: The user can download the data used for the current visualization, to get filtered datasets for further analysis.

4.2 County level analysis tool

4.2.1 Cross-correlation analyses

We consider the mobility change time series for one of the ranked sets (workplaces, residential, retail, grocery and pharmacy, parks, or transit stations) and the infection increase time series as the other ranked set (0, 1, 2, 3, or 4 week-lagged). We identify the lag in weeks that has the highest correlation (positive or negative) because of the virus' lagged nature (from the moment of infection to the moment of confirmed infection), and conduct the Kendall's correlation test for all activities and all lags. **Table 1** shows the comparison of lagged infection data for an example case using the data of New York, but can be generated for every county in the US.

Table 1: New York County, Cross-correlation Kendall's tau comparison

Week of lag	Workplaces	Residential	Retail	Grocery & pharmacy	Parks	Transit stations
No lag	-0.1346	0.4433	-0.3206	-0.3723	-0.6719	-0.2807
1	-0.1346	0.4433	-0.3206	-0.3723	-0.6719	-0.2807
2	-0.0121	0.3413	-0.1790	-0.2328	-0.5368	-0.1617
3	0.1038	0.2003	-0.0471	-0.1014	-0.3879	-0.0267
4	0.2243	0.0711	0.0916	0.0400	-0.2517	0.1128

Given the results shown in **Table 1**, we consider the best lag for each activity, as the one with highest correlation 2D and 3D scatter plots. From the results in **Table 1**, we select lag 1.

4.2.2 Scatter plots

Given the selected lag to study, to visualize the relationship between time, the change in infected cases, and the change in mobility for different activities, we use scatter plots to visualize the relationship of these variables. All scatter plots follow a temporal gradient color scheme, in which light-red points represent the first weeks of data (February 21, 2020) and the light blue points represent the last weeks (February 21, 2021).

The first scatter plot consists of mobility change in the x-axis and the infection increase data for the y-axis. Continuing the exploration of New York County, New York example, **Figure 3** shows the relationship between mobility changes to workplaces and the infection increase.

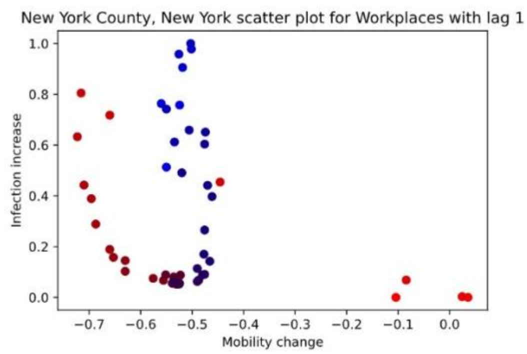
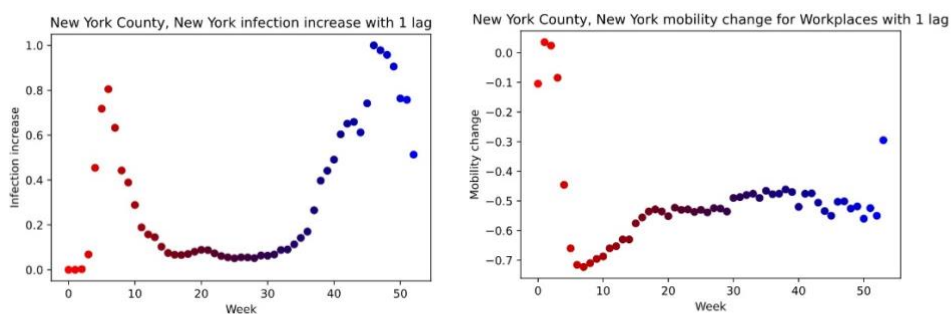


Figure 3: Mobility changes and infection increase scatter plots

Considering that the time component can have an impact on how mobility and infection change, we generate two more visualizations for the time series of infection change in **Figure 4a** and the time series of mobility change in **Figure 4b**, respectively. We plot the time component expressed in weeks on the x-axis and the infection or mobility in the y-axis.



(a) (b)

Figure 4: Time series scatter plots showing **(a)** infection change time series and **(b)** mobility change time series.

5. Conclusions

5.1 Discussion

The interactive tool we develop is an addition to the current dashboards and tools available in the web. As new data is generated, our tool will incorporate it, accordingly, to help in the tracking, understanding of the pandemic, and solutions for new similar events. It can serve as a visualization tool and source for further analysis related to mobility and disease spread and infections. This tool can be helpful to the analysis of differences in the effects of the pandemic between regions (from countries comparisons, in-country state-level comparisons, up to in-state county-level comparisons for the US, where data proliferates) and the influence of different variables on infection and mobility.

5.2 Limitations

Since only time series and choropleth maps are available, our dashboards could include more types of graphs to create a more comprehensive visualization of the mobility and infection changes for different regions. Additionally, the data sources are limited in our tool, since only data from Google's CMR is used, and using Apple's mobility data to verify and enrich the datasets could be an improvement to the mobility data quality. For the trends of COVID-19 infection and disease spread, this same opportunity exists to enhance the quality of infection data from other sources.

References

[1] The New York Times, *Coronavirus COVID-19 Data in the United States,* 2021. Available: https://github.com/nytimes/covid-19-data, [Accessed January 4, 2022]

[2] Johns Hopkins Center for Systems Science and Engineering, *COVID-19 Data Repository,* 2020. Available: https://github.com/CSSEGISandData/COVID-19, [Accessed January 4, 2022]

[3] T. Yang, K. Shen, S. He, E. Li, P. Sun, P. Chen, L. Zuo, J. Hu, Y. Mo, W. Zhang, H. Zhang, J. Chen and Y. Guo, *CovidNet: To bring data transparency in the era of COVID-19,* 2020. Available: https://arxiv.org/abs/2005.10948, [Accessed January 4, 2022]

[4] Google LLC, *Google COVID-19 Community Mobility Reports,* 2021. Available: https://www.google.com/covid19/mobility/, [Accessed January 4, 2022]

[5] E. Dong, H. Du and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *The Lancet,* vol. 20, p. 533–534, February 2020.

[6] World Health Organization, *WHO Coronavirus disease (COVID-19) dashboard,* 2021. Available: https://covid19.who.int/, [Accessed January 4, 2022]

[7] L. Zhang, S. Ghader, M. L. Pack, C. Xiong, A. Darzi, M. Yang, Q. Sun, A. Kabiri and S. Hu, "An interactive COVID-19 mobility impact and social distancing analysis platform," *medRxiv,* 2020. Available: https://www.medrxiv.org/content/early/2020/05/05/2020.04.29.20085472, [Accessed January 4, 2022]

[8] F. Zou, J. Wang, J. Gao, K. Ozbay, X. Ban, Y. Shen, H. Yang and S. Iyer, "An interactive data visualization and analytics tool to evaluate mobility and sociability trends during COVID-19," *UrbComp 2020: The 9th SIGKDD international workshop on urban computing, San Diego, California, USA,* 2020.

[9] M. S. Warren and S. W. Skillman, "Mobility Changes in Response to COVID-19," *arXiv preprint arXiv:2003.14228v1,* 2020.

[10] G. Bonaccorsi, F. Pierri, M. Cinelli, A. Flori, A. Galeazzi, F. Porcelli, A. L. Schmidt, C. M. Valensise, A. Scala, W. Quattrociocchi and F. Pammolli, "Economic and social consequences of human mobility restrictions under COVID-19," *Proceedings of the National Academy of Sciences,* vol. 117, p. 15530–15535, 2020.

[11] J. Coven and A. Gupta, "Disparities in Mobility Responses to COVID-19," *Preprint,* 2020. Available: https://static1.squarespace.com/static/56086d00e4b0fb7874bc2d42/t/5ebf201183c6f016ca3abd91/1589583893816/DemographicCovid.pdf, [Accessed January 4, 2022]

[12] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky and J. Leskovec, "Mobility network models of COVID-19 explain inequities and inform reopening," *Nature,* vol. 589, p. 82–87, 2021.

[13] M. Sulyok and M. Walker, "Community Movement and COVID-19: A global study using Google's Community Mobility Reports," *Epidemiology and Infection,* vol. 148, November 2020.

[14] T. Hastie and R. Tibshirani, Generalized additive models, Chapman & Hall, 1999.

[15] A. Aktay, S. Bavadekar, G. Cossoul, J. Davis, D. Desfontaines, A. Fabrikant, E. Gabrilovich, K. Gadepalli, B. Gipson, M. Guevara, C. Kamath, M. Kansal, A. Lange, C. Mandayam, A. Oplinger, C. Pluntke, T. Roessler, A. Schlosberg, T. Shekel, S. Vispute, M. Vu, G. Wellenius, B. Williams and R. J. Wilson, *Google COVID-19 Community Mobility Reports: Anonymization Process Description (version 1.1),* 2020. Available: https://arxiv.org/abs/2004.04145, [Accessed January 4, 2022]

[16] G. Van Rossum and F. L. Drake, Python 3 Reference Manual, Scotts, Valley: CreateSpace, 2009. Available: https://docs.python.org/3/reference/index.html, [Accessed January 4, 2022]

[17] M. G. Kendall, "A new measure of rank correlation," *Biometrika,* vol. 30, pp. 81-93, June 1938.

[18] J. Estrada-Garcia and S. Shen, *COVID-19 Effects on Mobility and Transportation,* 2021. Available: https://sites.google.com/umich.edu/decision-tools-for-covid19/data-analysis/covid-19-and-mobility, [Accessed January 4, 2022]