

1 **AN INTERACTIVE COVID-19 MOBILITY IMPACT AND SOCIAL**
2 **DISTANCING ANALYSIS PLATFORM**

3
4 Lei Zhang, Herbert Rabin Distinguished Professor (Corresponding Author)
5 Director, Maryland Transportation Institute
6 Department of Civil and Environmental Engineering
7 1173 Glenn Martin Hall, University of Maryland
8 College Park, MD 20742, Email: lei@umd.edu

9
10 Sepehr Ghader, Research Scientist (co-first author)
11 Maryland Transportation Institute, Department of Civil and Environmental Engineering
12 1173 Glenn Martin Hall, University of Maryland
13 College Park, MD 20742, Email: sghader@umd.edu

14
15 Michael L. Pack
16 Director, Center for Advanced Transportation Technology Laboratory
17 University of Maryland
18 5000 College Ave., Suite 2206
19 College Park, MD 20740, Email: packml@umd.edu

20
21 Chenfeng Xiong, Assistant Research Professor
22 Maryland Transportation Institute, Department of Civil and Environmental Engineering
23 1173 Glenn Martin Hall, University of Maryland
24 College Park, MD 20742, Email: cxiong@umd.edu

25
26 Aref Darzi, Graduate Research Assistant
27 Maryland Transportation Institute, Department of Civil and Environmental Engineering
28 University of Maryland, Email: adarzi@umd.edu

29
30 Mofeng Yang, Graduate Research Assistant
31 Maryland Transportation Institute, Department of Civil and Environmental Engineering
32 University of Maryland, Email: mofeng@umd.edu

33
34 QianQian Sun, Graduate Research Assistant
35 Maryland Transportation Institute, Department of Civil and Environmental Engineering
36 University of Maryland, Email: qsun12@terpmail.umd.edu

37
38 AliAkbar Kabiri, Graduate Research Assistant
39 Maryland Transportation Institute, Department of Civil and Environmental Engineering
40 University of Maryland, E-mail: kabiri@umd.edu

41
42 Songhua Hu, Graduate Research Assistant
43 Maryland Transportation Institute, Department of Civil and Environmental Engineering
44 University of Maryland, Email: hsonghua@umd.edu

1 **ABSTRACT**

2
3 The research team has utilized privacy-protected mobile device location data, integrated with
4 COVID-19 case data and census population data, to produce a COVID-19 impact analysis platform
5 that can inform users about the effects of COVID-19 spread and government orders on mobility
6 and social distancing. The platform is being updated daily, to continuously inform decision-makers
7 about the impacts of COVID-19 on their communities using an interactive analytical tool. The
8 research team has processed anonymized mobile device location data to identify trips and produced
9 a set of variables including social distancing index, percentage of people staying at home, visits to
10 work and non-work locations, out-of-town trips, and trip distance. The results are aggregated to
11 county and state levels to protect privacy and scaled to the entire population of each county and
12 state. The research team are making their data and findings, which are updated daily and go back
13 to January 1, 2020, for benchmarking, available to the public in order to help public officials make
14 informed decisions. This paper presents a summary of the platform and describes the methodology
15 used to process data and produce the platform metrics.
16

1 1. INTRODUCTION

2
3 Informed decision-making requires data. In the case of COVID-19, no previous pandemic had
4 such a big universal impact on societies in the modern history, as a results historic data lacked key
5 information on how people react to such a universal pandemic and how the virus impacts
6 economies and societies. Data-driven decision-making becomes a challenge in such an
7 unprecedented event. Thanks to the technology, we now have an enormous amount of observed
8 data collected by mobile devices amid pandemic. We can now utilize this data to learn more about
9 the various impacts of a pandemic on our lives, make informed decisions to fight the current
10 invisible enemy, and be better prepared the next time such pandemics happen. Our research team
11 has utilized a national set of privacy-protected mobile device location data and produced a COVID-
12 19 Impact Analysis Platform to provide comprehensive data and insights on COVID-19's impact
13 on mobility, economy, and society.

14
15 Mobile device location data are becoming popular for studying human behavior, specially mobility
16 behavior. Earlier studies with mobile device location data were mainly using GPS technology,
17 which is capable of recording accurate information including, location, time, speed, and possibly
18 a measure of data quality ¹. Later, mobile phones and smartphones gained popularity, as they could
19 enable researchers to study individual-level mobility patterns ²⁻⁴. Other emerging mobile device
20 location data sources such as call detail record (CDR) ⁵⁻⁷, Cellular network data ⁸, and social media
21 location-based services ⁹⁻¹³ have also been used by the researchers to study mobility behavior.
22 Mobile device location data has proved to be a great asset for decision-makers amid the current
23 COVID-19 pandemic. Many companies such as Google, Apple, or Cuebiq have already utilized
24 location data to produce valuable information about mobility and economic trends ¹⁴⁻¹⁶.
25 Researchers have also utilized mobile device location data for studying COVID-19- related
26 behavior ^{17,18}.

27
28 Non-pharmaceutical interventions such as social distancing are important and effective tools for
29 preventing virus spread. One of the most recent studies projected that the recurrent outbreaks might
30 be observed this winter based on pharmaceutical estimates on COVID-19 and other coronaviruses,
31 so prolonged or intermittent social distancing may be required until 2022 without any interventions
32 ¹⁹, highlighting the importance of improving our understanding about individual's reaction to
33 social distancing. Researchers have highlighted the importance of social distancing in disease
34 prevention through modeling and simulation ²⁰⁻²³. The simulation models assume a level of
35 compliance, which can now be validated through observed data. Our current platform utilizes
36 mobile device location data to provide observed data and evidence on social distancing behavior
37 and the impact of COVID-19 on mobility. We used daily feeds of mobile device location data,
38 representing movements of more than 100 Million anonymized devices, integrated with COVID-
39 19 case data from John Hopkins University and census population data to monitor the mobility
40 trends in United States and study social distancing behavior ²⁴. In the next section we describe the
41 methodology used to process the anonymized location data and produce the metrics that are
42 available on the platform. The methodology section is followed by a brief overview of the
43 platform. The last section presents concluding remarks.

2. METHODOLOGY

The research team first integrated and cleaned location data from multiple sources representing person and vehicle movements in order to improve the quality of our mobile device location data panel. We then clustered the location points into activity locations and identified home and work locations at the census block group (CBG) level to protect privacy. We examined both temporal and spatial features for the entire activity location list to identify home CBGs and work CBGs for workers with a fixed work location. Next, we applied previously developed and validated algorithms²⁵ to identify all trips from the cleaned data panel, including trip origin, destination, departure time, and arrival time. Additional steps were taken to impute missing trip information for each trip, such as trip purpose (e.g., work, non-work), point-of-interest visited (restaurants, shops, etc.), travel mode (air, rail, bus, driving, biking, walking, and others), trip distance (airline and actual distance), and socio-demographics of the travelers (income, age, gender, race, etc.) using advanced artificial intelligence and machine learning algorithms. If an anonymized individual in the sample did not make any trip longer than one-mile in distance, this anonymized individual was considered as staying at home. A multi-level weighting procedure expanded the sample to the entire population, using device-level and trip-level weights, so the results are representative of the entire population in a nation, state, or county. The data sources and computational algorithms have been validated based on a variety of independent datasets such as the National Household Travel Survey and American Community Survey, and peer reviewed by an external expert panel in a U.S. Department of Transportation Federal Highway Administration's Exploratory Advanced Research Program project, titled "Data analytics and modeling methods for tracking and predicting origin-destination travel trends based on mobile device data"²⁵. Mobility metrics were then integrated with COVID-19 case data, population data, and other data sources. **Figure 1** shows a summary of the methodology.

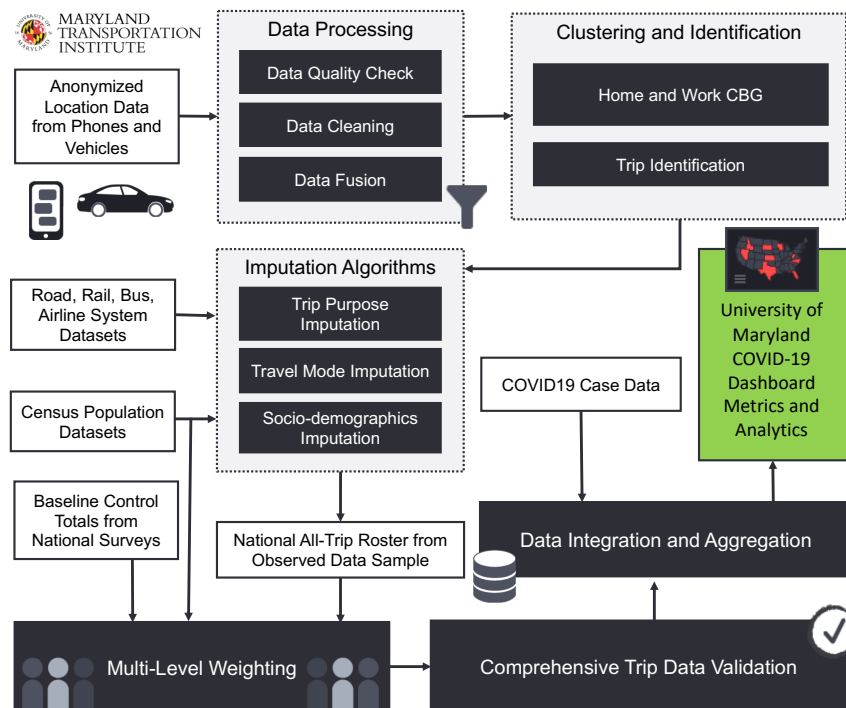


Figure 1. Methodology

26
27

2.1. Trip Identification

Trips are the unit of analysis for almost all transportation applications. Traditional data sources such as travel surveys include trip-level information. The mobile device location data, on the other hand, do not directly include trip information. Location sightings can be continuously recorded while a device moves, stops, stays static, or starts a new trip. These changes in status are not recorded in the raw data. As a result, researchers must rely on trip identification algorithms to extract trip information from the raw data. Basically, researchers must identify which locations form a trip together. The following subsections describe the steps our research team took to identify trips. The algorithm runs on the observations of each device separately.

2.1.1. Pre-Processing

First, all device observations are sorted by time. The trip identification algorithm assigns a hashed ID to every trip it identifies. The location dataset may include many points that do not belong to any trips. The algorithm assigns “0” as the trip ID to these points to identify them as static points. For every observation, we compute the distance, time, and speed between the point and its previous and next points if exist.

The trip identification algorithm has three hyper-parameters: distance threshold, time threshold, and speed threshold. The speed threshold is used to identify if an observation is recorded on the move. The distance and time thresholds are used to identify trip ends. At this step, the algorithm identifies the device’s first observation with $speed\ from \geq speed\ threshold$. This identified point is on the move, so a hashed trip ID is generated and assigned to this point. All points recorded before this point, if exist, are set to have “0” as their trip ID. Next, the recursive algorithm identifies if the next points are on the same trip and should have the same trip ID.

2.1.2. Recursive Algorithm

This algorithm checks every point to identify if they belong to the same trip as their previous point. If they do, they are assigned the same trip ID. If they do not, they are either assigned a new hashed trip id (when their $speed\ from \geq speed\ threshold$) or their trip ID is set to “0” (when their $speed\ from < speed\ threshold$). Identifying if a point belongs to the same trip as its previous point is based on the point’s “speed to”, “distance to” and “time to” attributes. If a device is seen in a point with $distance\ to \geq distance\ threshold$ but is not observed to move there ($speed\ to < speed\ threshold$), the point does not belong to the same trip as its previous point.

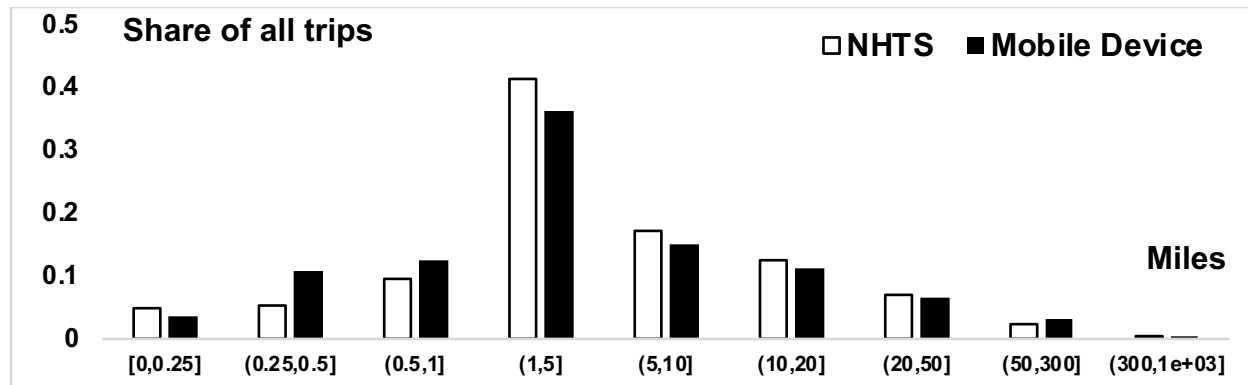
When the device is on the move at a point ($speed\ to \geq speed\ threshold$), the point belongs to the same trip as its previous point; but when the device stops, the algorithm checks the radius and dwell time to identify if the previous trip has ended. If the device stays at the stop (points should be closer than the distance threshold) for a period of time shorter than the time threshold, the points still belong to the previous trip. When the dwell time reaches above the time threshold, the trip ends, and the next points no longer belong to the same trip. The algorithm does this by updating “time from” to be measured from the first observation in the stop, not the point’s previous point. The algorithm may identify a local movement as a trip if the device moves within a stay location.

1 To filter out such trips, all trips that are within a static cluster and all trips that are shorter than 300
2 meters are removed.

3 4 2.1.3. Validation

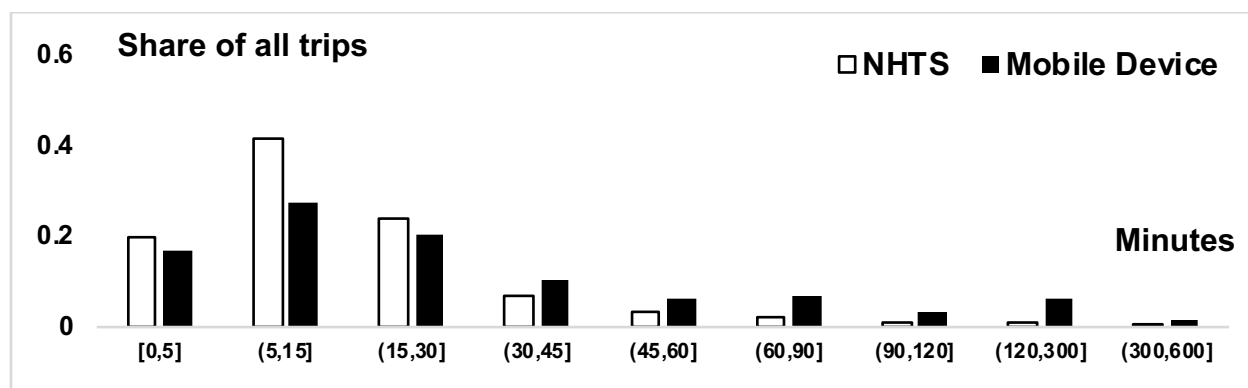
5
6 **Figure 2** and **Figure 3** show the validation of this algorithm by running the algorithm on a sample
7 of national mobile device location data and comparing the trip lengths and travel times with the
8 reported travel distances and travel times from the 2017 national household travel survey (NHTS
9 2017). A satisfactory match is observed between the two datasets.

10



11
12
13

Figure 2. Distance validation of the trip identification algorithm against NHTS2017



14
15
16
17

Figure 3. Travel time validation of the trip identification algorithm against NHTS2017

18 2.2. Activity Identification

19 We first identify all activity points. Then, based on the temporal and spatial distribution of activity
20 points, we identify the home census block group (CBG) and the work CBG.

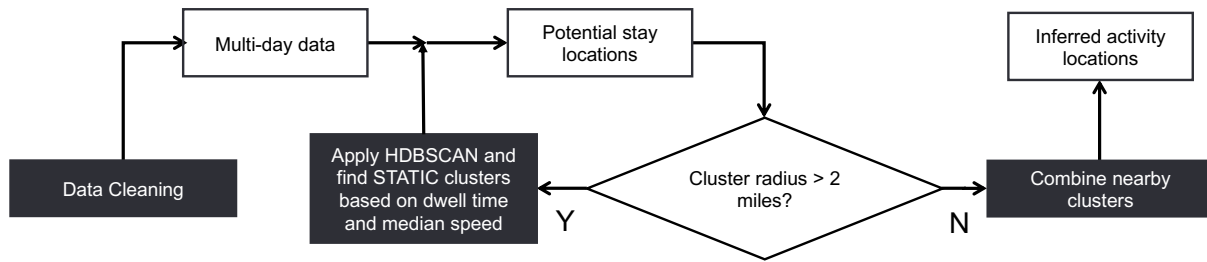
21

22 2.2.1. Activity Clustering

23

24 The algorithm starts by clustering all device observations into activity locations using HDBSCAN
25 ²⁶ clustering algorithm. This step takes the cleaned multi-day location data as input and applies an
26 iterative algorithm until no cluster has a radius larger than two miles. The iterative algorithm
27 consists of two parts: HDBSCAN based on a minimum number of point parameters and filtering

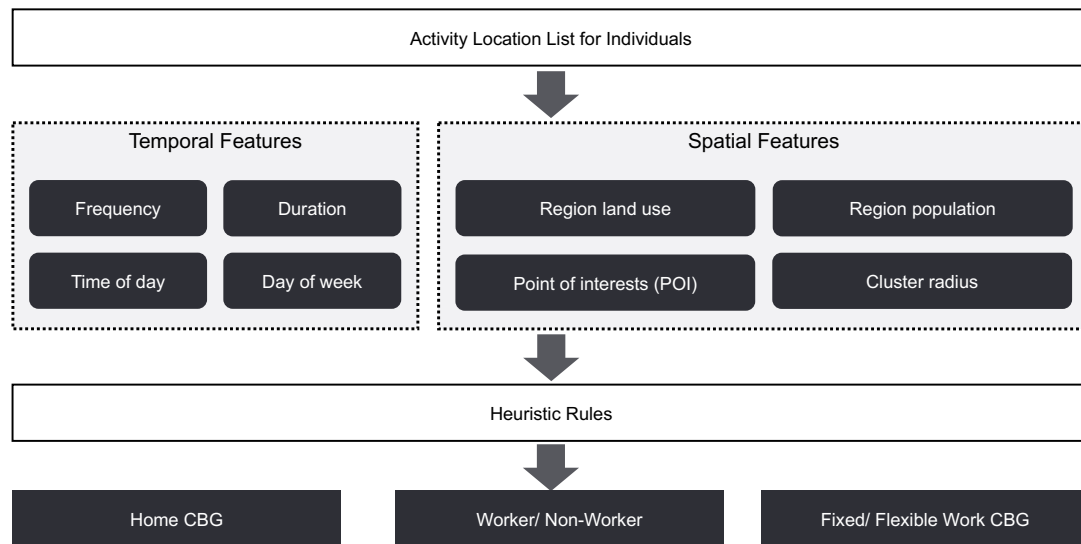
1 non-static clusters based on time and speed checks. After finalizing the potential stay clusters, the
 2 framework combines nearby clusters to avoid splitting a single activity (**Figure 4**).
 3



4 **Figure 4. Activity clustering methodology**

5
 6
 7 **2.2.2. Home and work CBG Identification**

8
 9 **Figure 5** shows the methodology for home and work CBG identification. Instead of setting a fixed
 10 time period for each type, e.g., 8pm to 8am as the study period for home CBG identification and
 11 the other half day for work CBG identification, the framework examines both temporal and spatial
 12 features for the entire activity location list. The benefits are two-fold: the results for workers with
 13 flexible or opposite work schedules would be more accurate and the employment type for each
 14 device could be detected simultaneously. **Figure 6** shows the validation of home and work location
 15 imputations, by comparing the distance from home to work between longitudinal employer-
 16 household dynamics (LEHD) data and the imputed locations for a set of mobile device location
 17 data for the Baltimore metropolitan area. We can observe a satisfactory match.
 18



19 **Figure 5. Home/work CBG imputation methodology**

20
 21

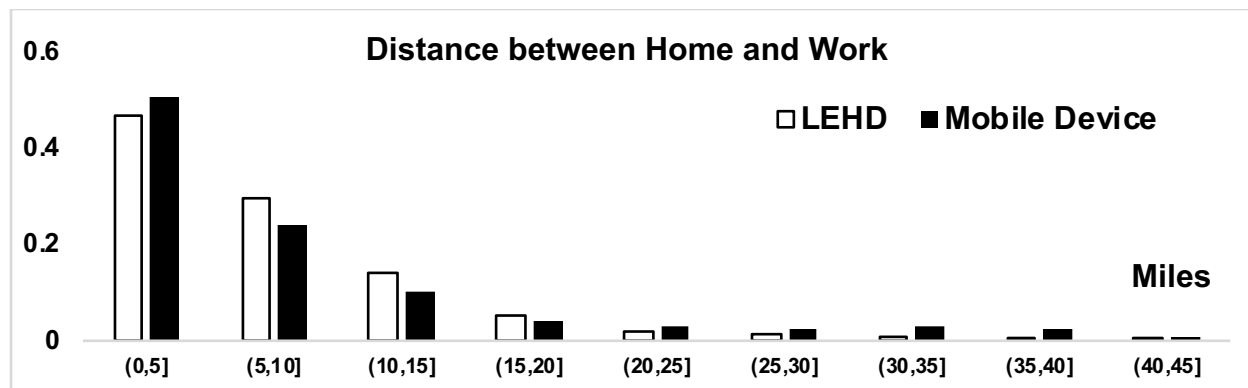


Figure 6. Validation of home and work imputation against LEHD

2.3. Imputation

2.3.1. Mode Imputation

Our research team developed a jointly trained single-layer model and deep neural network²⁷ for travel mode detection of this project. This model combines the advantages of both types of models to be able to make sufficient generalizations using a multi-layer DNN and capture the exceptions using the wide single-layer model. The datasets used for training the model were collected from the incenTrip mobile phone app²⁸, developed by the authors, where the ground truth information for car, bus, rail, bike, walk, and air trips was collected. To effectively detect the travel mode for each trip, feature construction is critical in providing useful information. Travel mode-specific knowledge is needed to improve the detection accuracy. In addition to the traditional features used in the literature (e.g. average speed, maximum speed, trip distance, etc.), we also integrated the multi-modal transportation network data to construct innovative features in order to improve the detection accuracy based on network data integration. The wide and deep learning method utilized in this study achieved over 95% prediction accuracy for drive, rail, air, and non-motorized, and over 90% for bus modes on test data. We have applied the trained algorithms on the location dataset to obtain multimodal trip rosters (see **Figure 7** that shows raw location data points by different travel modes). The resulting mode shares show a decent match with the available travel surveys at both national and metropolitan levels.

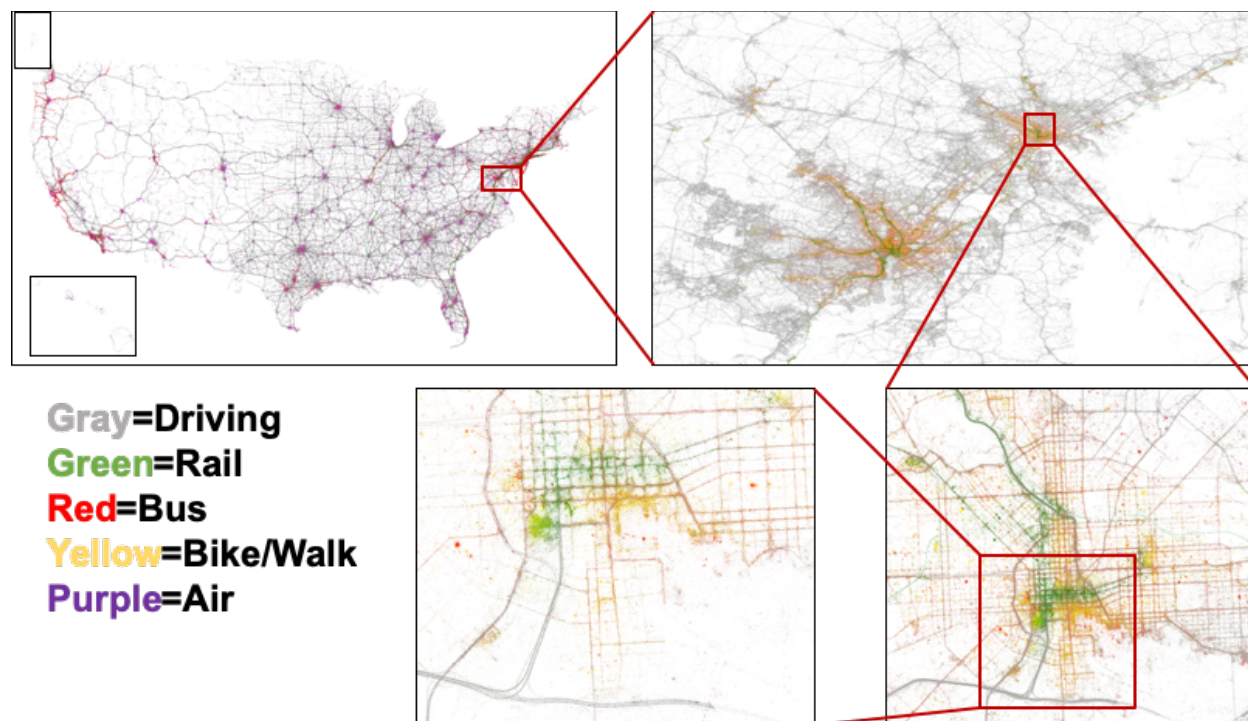


Figure 7. Demonstration of the multi-modal travel patterns

1
2
3
4 2.3.2. Purpose Imputation
5

6 In the Home and work CBG Identification, we described how home and work CBGs can be
7 identified. Other purposes can be directly identified through spatial joint of trip end locations and
8 point of interest (POI) data. We have used a popular commercial POI dataset that includes more
9 than forty million records for the U.S. For each trip end, we first filter all POIs that are located
10 within a 200-meter radius of the trip-end. Next, we identify the trip purpose by the POI type of the
11 closest POI.

12
13 2.3.3. Socio-Demographic Imputation
14

15 Due to privacy concerns, mobile device location data contain very little ground truth information
16 about the device owners. However, it is essential to understand how representative the sample is
17 and how different segments of the population travel. The state-of-the-practice method is to assign
18 either the census population socio-demographic distribution or the public use microdata sample
19 (PUMS) units to the sample devices within the same geographic area based on the imputed home
20 locations. More advanced socio-demographic imputation methods utilize travel patterns and
21 visited POI types to impute the socio-demographics. These methods require a significant amount
22 of computation, as various features from different databases should be calculated and used. In
23 order to balance the computations and conduct a timely analysis for the pandemic, we have used
24 the state-of-the-practice method and assigned socio-demographic information to the anonymized
25 devices based on the census socio-demographic distribution of their imputed CBG. Five-year
26 American Community Survey (ACS) estimates for 2014 to 2018 from the U.S. Census Bureau can
27 be used to obtain median income, age distribution, gender distribution, and race distribution for
28 each U.S. CBG²⁹. For each device, we used Monte-Carlo simulation³⁰ to draw from the age,

1 gender, and race distribution at the device's imputed home CBG. We also assigned the CBG's
2 median income to the device.

3 4 **2.4. Weighting**

5
6 The sample data needs to be weighted to represent population-level statistics. First, the devices
7 available in our dataset are a sample of all individuals in the population, so we need to apply
8 device-level weights. Second, for an observed device, only a sample of all trips may be recorded,
9 so trip-level weights are also needed. For the sake of timeliness, we have applied simple weighting
10 methods to obtain county-level device weights and state-level trip weights. In order to obtain
11 device-level weights, we have used the home county, obtained from the imputed home CBG
12 information. The weight for each device is equal to the number of devices observed in the device's
13 imputed home county divided by the population of the county, so all devices residing in a county
14 would have the same device-level weight. For instance, if our sample includes 100 devices in a
15 county with a population of 2,000, each device would be assigned a weight of 20. Population of
16 each county can be obtained from the U.S. Census Bureau. For trip-level weights, we have
17 calculated number of trips per person (trip rate) for each state during an average weekday in the
18 first two weeks of February from our sample. We have also calculated this trip rate number for
19 each state from NHTS2017. We have used a single trip rate for all trips generated from each state,
20 equal to the NHTS trip rate divided by our observed trip rate.

21 22 **3. PLATFORM OVERVIEW**

23
24 The COVID-19 Impact Analysis Platform, available at data.covid.umd.edu provides data and
25 insights on COVID-19's impact with daily data updates. The research team are exploring how
26 social distancing and stay-at-home orders are affecting travel behavior, spread of the coronavirus,
27 and local economies. Through this interactive analytics platform, we are making our data and
28 research findings available to other researchers, agencies, non-profits, media, and the general
29 public. The platform will evolve and expand over time as new data and impact metrics are
30 computed and additional visualizations are developed. **Table 1** shows the current metrics available
31 in the platform at the national, state, and county levels in the United States with daily updates.
32 **Figure 8** illustrates the platform.

33

1

Table 1. List of metrics available on the COVID-19 impact analysis platform

Current Metrics	Description
social distancing index	<p>An integer from 0~100 that represents the extent residents and visitors are practicing social distancing. “0” indicates no social distancing is observed in the community, while “100” indicates all residents are staying at home and no visitors are entering the county.</p> <p>It is computed by this equation:</p> $\text{social distancing index} = 0.8 * [\% \text{staying home} + 0.01 * (100 - \% \text{staying home}) * (0.1 * \% \text{reduction of all trips compared to pre - COVID - 19 benchmark} + 0.2 * \% \text{reduction of work trips} + 0.4 * \% \text{reduction of non - work trips} + 0.3 * \% \text{reduction of travel distance})] + 0.2 * \% \text{reduction of out - of - county trips}$
% staying home	Percentage of residents staying at home (i.e., no trips more than one mile away from home)
#trips/person	Average number of trips taken per person.
% out-of-county trips	The percent of all trips taken that travel out of a county. Additional information on the origins and destinations of these trips at the county-to-county level is available, but not currently shown on the platform.
miles traveled/person	Average person-miles traveled on all modes (car, train, bus, plane, bike, walk, etc.)
#work trips/person	Number of work trips per person (where a “work trip” is defined as going to or coming home from work)
#non-work trips/person	Number of non-work trips per person. (e.g. grocery, restaurant, park, etc.). Additional information on trip purpose (restaurant, shops, etc.) is available, but not currently shown on the platform.
#COVID-19 cases	Number of new confirmed COVID-19 cases from the Johns Hopkins University’s GitHub repository.
population	Number of residents in a nation, state, or county as reported from the national Census database.

2

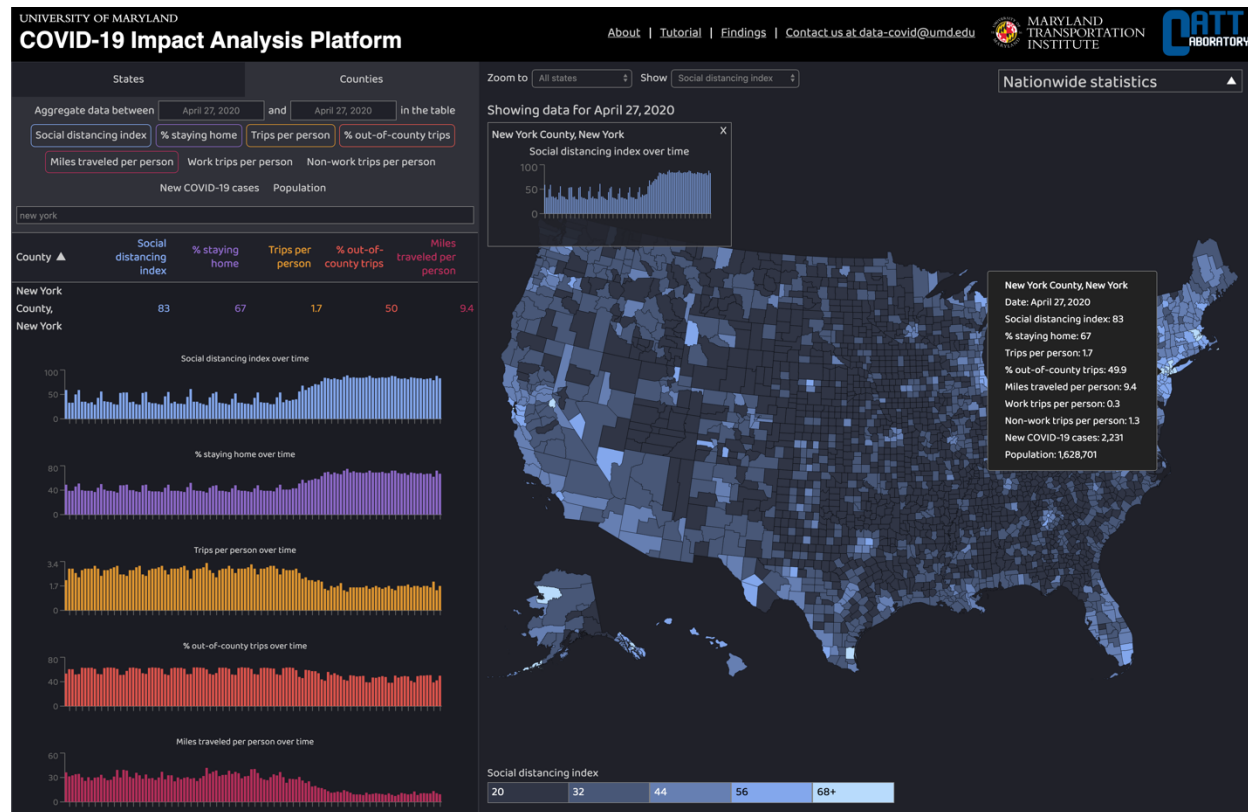


Figure 8. Platform illustration

4. CONCLUSION

The Integrated dataset compiled by our research team shows how the nation and different states and counties are impacted by the COVID-19 and how the communities are conforming with the social distancing and stay-at-home orders issued to prevent the spread of the virus. The platform utilizes privacy-protected anonymized mobile device location data integrated with healthcare system data and population data to assign a social distancing score to each state and county based on derived information such as percentage of people who are staying home, average number of trips per person and average distance traveled by each person. As the next steps, the research team is integrating socio-demographic and economic data into the platform to study the multifaceted impact of COVID-19 on our mobility, health, economy, and society.

ACKNOWLEDGMENT

We would like to thank and acknowledge our partners and data sources in this effort: (1) Amazon Web Service and its Senior Solutions Architect, Jianjun Xu, for providing cloud computing and technical support; (2) computational algorithms developed and validated in a previous USDOT Federal Highway Administration's Exploratory Advanced Research Program project; (3) mobile device location data provider partners; and (4) partial financial support from the U.S. Department of Transportation's Bureau of Transportation Statistics and the National Science Foundation's RAPID Program.

1 REFERENCES

- 2
- 3 1 Stopher, P., Clifford, E., Zhang, J. & FitzGerald, C. Deducing mode and purpose from GPS
4 data. *Institute of Transport and Logistics Studies*, 1-13 (2008).
- 5 2 Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M. & Puchinger, J. Inferring
6 dynamic origin-destination flows by transport mode using mobile phone data.
7 *Transportation Research Part C: Emerging Technologies* **101**, 254-275 (2019).
- 8 3 Huang, Z. *et al.* Modeling real-time human mobility based on mobile phone and
9 transportation data fusion. *Transportation research part C: emerging technologies* **96**, 251-
10 269 (2018).
- 11 4 Cui, Y., Meng, C., He, Q. & Gao, J. Forecasting current and next trip purpose with social
12 media data and Google Places. *Transportation Research Part C: Emerging Technologies*
13 **97**, 159-174 (2018).
- 14 5 Pappalardo, L. *et al.* Returners and explorers dichotomy in human mobility. *Nature*
15 *communications* **6**, 8166 (2015).
- 16 6 Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human
17 mobility. *Nature Physics* **6**, 818 (2010).
- 18 7 Çolak, S., Lima, A. & González, M. C. Understanding congested travel in urban areas.
19 *Nature communications* **7**, 10793 (2016).
- 20 8 Nyhan, M., Kloog, I., Britter, R., Ratti, C. & Koutrakis, P. Quantifying population exposure
21 to air pollution using individual mobility patterns inferred from mobile phone data. *Journal*
22 *of exposure science & environmental epidemiology* **29**, 238 (2019).
- 23 9 Bao, J., Lian, D., Zhang, F. & Yuan, N. J. Geo-social media data analytic for user modeling
24 and location-based services. *SIGSPATIAL Special* **7**, 11-18 (2016).
- 25 10 Liu, Q., Wu, S., Wang, L. & Tan, T. in *AAAI*. 194-200.
- 26 11 Riederer, C. J., Zimmeck, S., Phanord, C., Chaintreau, A. & Bellovin, S. M. in *Proceedings*
27 *of the 2015 ACM on Conference on Online Social Networks*. 185-195 (ACM).
- 28 12 Gu, Y., Yao, Y., Liu, W. & Song, J. in *Computer Communication and Networks (ICCCN),*
29 *2016 25th International Conference on*. 1-9 (IEEE).
- 30 13 Liao, Y., Lam, W., Jameel, S., Schockaert, S. & Xie, X. in *Proceedings of the 2016 ACM*
31 *International Conference on the Theory of Information Retrieval*. 271-280 (ACM).
- 32 14 Google. *See how your community is moving around differently due to COVID-19,*
33 [<https://www.google.com/covid19/mobility/>](https://www.google.com/covid19/mobility/) (2020).
- 34 15 Apple. *Mobility Trends Reports,* [<https://www.apple.com/covid19/mobility/>](https://www.apple.com/covid19/mobility/) (2020).
- 35 16 Cuebiq. *COVID-19 Mobility Insights,* [<https://www.cuebiq.com/visitation-insights-](https://www.cuebiq.com/visitation-insights-covid19/)
36 [covid19/>](https://www.cuebiq.com/visitation-insights-covid19/) (2020).
- 37 17 Gao, S., Rao, J., Kang, Y., Liang, Y. & Kruse, J. Mapping county-level mobility pattern
38 changes in the United States in response to COVID-19. *Available at SSRN 3570145* (2020).
- 39 18 Warren, M. S. & Skillman, S. W. Mobility Changes in Response to COVID-19. *arXiv*
40 *preprint arXiv:2003.14228* (2020).
- 41 19 Kissler, S., Tedijanto, C., Goldstein, E., Grad, Y. & Lipsitch, M. Projecting the
42 transmission dynamics of SARS-CoV-2 through the post-pandemic period. (2020).
- 43 20 Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel
44 coronavirus (COVID-19) outbreak. *Science* (2020).

1 21 Kelso, J. K., Milne, G. J. & Kelly, H. Simulation suggests that rapid activation of social
2 distancing can arrest epidemic development due to a novel strain of influenza. *BMC public*
3 *health* **9**, 117 (2009).

4 22 Greenstone, M. & Nigam, V. Does Social Distancing Matter? *University of Chicago,*
5 *Becker Friedman Institute for Economics Working Paper* (2020).

6 23 Li, D. *et al.* Estimating the scale of COVID-19 Epidemic in the United States: Simulations
7 Based on Air Traffic directly from Wuhan, China. *medRxiv* (2020).

8 24 Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19
9 in real time. *The Lancet infectious diseases* (2020).

10 25 Zhang, L. & Ghader, S. Data Analytics and Modeling Methods for Tracking and Predicting
11 Origin-Destination Travel Trends Based on Mobile Device Data. (Federal Highway
12 Administration Exploratory Advanced Research Program, 2020).

13 26 Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. in *Kdd.* 226-231.

14 27 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444 (2015).

15 28 Xiong, C. *et al.* An integrated and personalized traveler information and incentive scheme
16 for energy efficient mobility systems. *Transportation Research Part C: Emerging*
17 *Technologies* (2019).

18 29 McMaster, R. B., Lindberg, M. & Van Riper, D. in *Proceedings 21st International*
19 *Cartographic Conference.* 821-828.

20 30 Hammersley, J. *Monte carlo methods.* (Springer Science & Business Media, 2013).

21