Final Report to California Air Resources Board Research Division
CARB Agreement No. 20RD005

A Data Science Framework to Measure Vehicle Miles Traveled by Mode and Purpose

Principal Investigator: Marta C. González*
Shangqing Cao
Cristobal Pais
Violet Lingenfelter
Ruining Wang
Elouan Pulveric
Tomas Wenzel*
University of California, Berkeley
*Lawrence Berkeley National Laboratory

September 30th, 2023

Prepared for the California Air Resources Board and the California Environmental Protection Agency

Disclaimer

The statements and conclusions in this Report are those of the contractor and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as an actual or implied endorsement of such products.

Acknowledgements

The authors would like to thank the California Air Resources Board (CARB) for their close collaboration throughout the study, especially with Jose Lopez and Sarah Pittiglio. We would like to thank members of Spectus who provided data and support throughout the study, including Éadaoin Ilten and Brennan Lake. Thanks also extend to all the subject matter experts who agreed to participate in interviews with our team and whose valuable input strengthened the findings presented in this study. This Report was submitted in fulfillment of No. 20RD005: A Data Science Framework to Measure VMT by Mode and Purpose by University of California, Berkeley under the sponsorship of the California Air Resources Board. Work was completed as of 9/30/2023.

Contents

List of Figures

List of Tables

1 Abstract

On March 19^{th}, 2020, California implemented a Shelter-in-Place (SIP) order in response to the COVID-19 pandemic. This report delves into the alteration of human mobility patterns in California prompted by the pandemic and the corresponding response measures. Utilizing Location-Based Service (LBS) data derived from mobile phones, we introduce a mode detection algorithm to explore changes in vehicle miles traveled (VMT) statewide, categorized by trip purpose and origin tract within California. Our analysis uncovers both spatial and temporal disparities in how the COVID-19 pandemic and the SIP order affected VMT. The study investigates shifts in the number of commutes and the structure of the commuting flow network. Notably, we identify the emergence of two additional travel zones in the flow network, indicative of a growing separation among different regions in the wake of the pandemic. Furthermore, we compare the average travel distance, assessed through the radius of gyration (r_g), before and after the pandemic's onset, capturing shifts in individual travel behaviors. Our findings confirm that not only did people reduce their overall travel due to the pandemic, but they also began to travel shorter distances. Additionally, this report examines patterns of changes in residential locations during the COVID-19 pandemic using a two-step semi-supervised algorithm. Notably, we observe a higher frequency of home changes in March 2020. Users who relocated during this period tended to move over longer distances, suggesting a shift not only within the city but also within the broader region of residence. Finally, we evaluate the feasibility of employing LBS data to assess the effectiveness of various mobility-related interventions that took place in Sacramento in 2019, employing the aforementioned mode detection algorithm. While we do observe the potential to detect changes in motorized trips as a result of these interventions, it's important to note that the sample size for traveler detection is relatively small, limiting our ability to draw definitive conclusions.

2 Public outreach summary: People Movement through the Lens of Cell Phone Data in California during COVID-19

2.1 Issues

In 2018, California set an ambitious target: to reduce the state's greenhouse gas emissions, a major driver of global warming, to 40% below the 1990 level by 2030. One of the significant contributors to these emissions is the transportation sector, particularly the widespread use of motor vehicles. As a result, there is a pressing need for dependable data sources to quantify alterations in Vehicle Miles Traveled (VMTs).

The emergence of the COVID-19 pandemic brought about substantial limitations on people's mobility. A predominant shift towards remote work became the norm for many during this period, leading to a fundamental transformation in how and where people traverse within cities.

The COVID-19 pandemic has presented a unique opportunity to explore shifts in travel behavior through the lens of emerging technologies. Mobile phone data, in particular, offers a valuable means to investigate how people modified their vehicle usage and travel patterns during this transformative period.

2.2 Main Question

The central inquiry of the report revolves around the impact of the COVID-19 pandemic on individuals' mobility and vehicle utilization. Specifically, it explores the potential for utilizing location data generated by mobile phones to quantify these changes.

2.3 Key Results

The key results from this study are (1) Vehicle usage decreased up to 20% more in urban areas than in rural areas because of COVID-19. (2) The number of commutes decreased 30% more than the number of non-commute trips. The number of commuting travels we observed in 2022 had not returned to the respective 2019 level. (3) In the two-week period following the beginning of the lockdown in March 2020, many more people changed residential locations compared to other all other periods of observation in 2020. (4) Location based service data from mobile phones allow to measure the effects of transportation interventions on non-motorized trips.

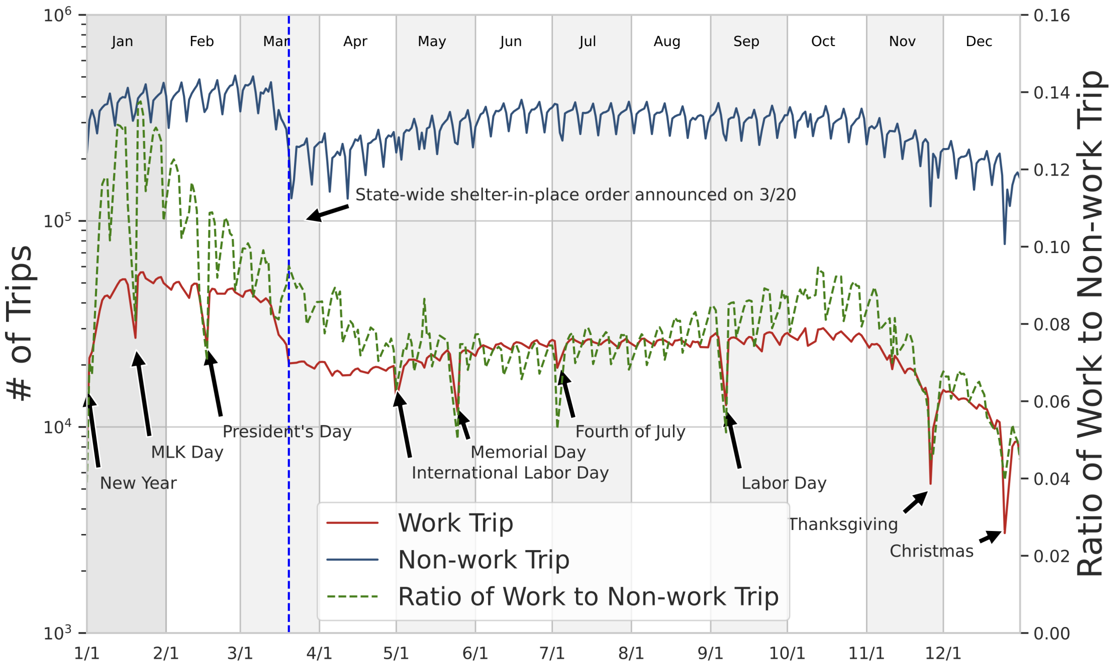
2.4 Conclusions

Our findings underscore the importance of tailoring policies aimed at reducing vehicle dependency to account for regional disparities. In general, urban areas reduced more their VMTs than rural areas during the same period of observation in 2020. Notably, the reduction in vehicle usage exhibits a significant variance, with some counties decreasing their Vehicle Miles Traveled only by 5% in 2020, while others achieved a more substantial reduction of 38%.

While Location-Based Service (LBS) data proves valuable in measuring travel mode and tracking changes in residential locations, it's worth noting that the sample sizes are relatively small. Consequently, drawing meaningful conclusions regarding the demographic characteristics of individuals based on this data remains challenging.

2.5 Additional Information

This study is funded under CARB grant No.20RD005 and completed by Shangqing Cao, Cristobal Pais, Violet Lingenfelter, Ruining Wang, Elouan Pulveric, Tom Wenzel and Marta C. González. The full title of the study is "A Data Science Framework to Measure Vehicle Miles Traveled by Mode and Purpose". The full paper can be found on the CARB website and the associated web application developed can be found here.



3 Executive Summary

3.1 *Background*

Human mobility science is the study of modeling and analyzing individuals' movements to extrapolate population-wide travel patterns and trends. These studies have been boosted by the new availability of mobile phones that track trips of large samples of the population. These data sources are passively collected in contrast with, and as a great complement to travel diary surveys. Mobile phone sources are divided into two types, call detailed records (CDRs) and location-based services (LBS). The former are collected by mobile phone providers and have spatial resolution given by the coverage areas of the antennas. The latter are collected by different applications installed in smartphones and they have spatial resolution of GPS and WiFi access points.

Characterizing human mobility patterns for a group of users, or in a given area, can assist policy makers in designing and implementing effective mobility-related and transportation-related policies. This report explores the potential of using mobile phone data and methods from mobility science to track changes in travel patterns in California. This helps the state of California to design policies that will lead to achieving the state's goal of reducing greenhouse gas emissions to 40% below the 1990 level by 2030. We focus on evaluating the impact of the COVID-19 pandemic on mobility patterns in California as detected via algorithms applied to mobile phone data.

3.2 *Objectives and Methods*

The goals of this report encompass evaluating the shifts in vehicle usage, trip purposes, commuting patterns, and residential relocations at a statewide level, all prompted by the COVID-19 pandemic, and relying on mobile phone data as our primary data source. Additionally, we assess the influence of multiple mobility-related initiatives implemented in Sacramento in 2019. To achieve these objectives, we employ a range of data science techniques and concepts from network science.

Specifically, we introduce innovative unsupervised learning methods designed for the identification of changes in residential locations and the classification of travel modes. Our analyses are driven by Location-Based Service (LBS) data for trip measurement, supplemented by demographic and socioeconomic information sourced from census data.

3.3 Results

We discover that mobile phone data enables the measurement of shifts in Vehicle Miles Traveled (VMTs) during the COVID-19 pandemic and its associated response measures in California. Notably, rural areas exhibited a lesser reduction in comparison to their urban and suburban counterparts, maintaining relatively high VMT levels even in March and April 2020. A more pronounced decline in VMT occurred on weekends, as non-essential travel bore a greater negative impact than essential travel. It's worth highlighting that the volume of work-related movements, particularly commutes, did not rebound to pre-pandemic levels as swiftly as non-commute trips. In fact, as indicated by network analysis, the statewide commuting structure has not fully returned to its pre-pandemic state as of 2022. An analysis of the radius of gyration (r_g) reinforces that people not only traveled less during the COVID-19 pandemic but also covered shorter distances.

A notable surge in the number of residential relocations is observed in March 2020, coinciding with the rapid spread of the virus in California. Furthermore, individuals relocated over greater distances during this period compared to usual times. The data also shows an increasing number of movements between Northern and Southern California during this period. Major urban centers such as Los Angeles and San Francisco experienced a substantial net outflow of residents at the pandemic's onset, while smaller cities in the central valley, like Fresno and Bakersfield, maintained relatively stable populations.

When evaluating mobility-related events in Sacramento in 2019 as captured by mobile phone data, the analysis reveals that the only event significantly impacting the vehicle usage rate was the expansion of the JUMP scooter fleet in June 2019.

3.4 Conclusion

In this report, we reveal the shifts in mobility patterns triggered by the COVID-19 pandemic. Our study delves into the utilization of mobility science and data science methodologies on mobile phone data to monitor vehicle usage and travel behaviors in California. As a crucial component of our research, we introduce and implement two innovative models crafted for pinpointing residential relocations and travel modes. Looking ahead, future research can be directed towards improving and devising supervised learning methods specifically designed to systematically detect modes and relocations, thereby shedding light on the implications of limited sample sizes for detecting these changes.

4 Introduction

On March 19^{th}, 2020, the state of California announced the Shelter-in-Place (SIP) order (also referred as ”lockdown” in parts of this report) as an effort to contain and mitigate the spread of Coronavirus Disease 2019 (COVID-19). The measure significantly limited people’s movements and introduced new mobility patterns. In order to meet the state’s goal of reducing greenhouse gas emissions to 40% below 1990 levels by 2030, understanding the impact of the COVID-19 pandemic and its associated responses is key to designing effective and targeted policies that reduce vehicle usage, and thus greenhouse gas emissions. The study sheds light on understanding the similarities and differences in human mobility patterns across different regions in California and during different periods through the lens of mobile phone data.

Understanding human mobility means understanding how and when people move from location to location in their everyday lives. In the past decade or so, advances in technology have opened up new opportunities for understanding human mobility through novel data sets. These data sets, such as Call Detail Records (CDR) from cell phone providers and location histories collected by smartphone apps, referred as Location-Based Services (LBS), allow researchers to observe individuals on a new scale [14][30]. The advent of these large data sets requires us to empirically derive new metrics, algorithms, and models to understand and capture how human behavior relates to human mobility. Improved models for human mobility have far reaching impacts, from models of disease spread [94] to transportation demand modeling [112], or to understanding behavior in natural disasters [55]. Section 15 provides a more comprehensive description of CDR and LBS data.

The objectives of this report are to assess the ability of mobile phone data to measure the statewide changes in Vehicle Miles Traveled (VMT) as a result of the SIP order, and analyze the changes to the temporal and spatial commute flow patterns. We also aim to uncover trends in home relocation during the SIP order and investigate the effect of mobility related initiatives on vehicle usage in selected area in the city of Sacramento. As a part of the study, We developed a novel unsupervised model for identifying vehicle trips to calculate VMT and a two-step semi-supervised scheme to identify changes in residential locations on a census block group level.

The rest of the report is organized as follows: In section 5, we present the data sets used in our study. In section 6, we discuss the mode detection algorithm and changes in VMT in 2020 due to the SIP order. In section 7, we dissect trips into two different categories, commute and non-commutes, and analyze changes in trends associated with each type of trip. In section 8, we showcase the web application that is developed as a part of this project to display Radius of Gyration (r_g) , a measure that captures the average distance covered by each individual. In section 9, we present the home detection algorithm and the patterns of home changes in March, 2020. Finally, in section 10, we evaluate the effectiveness of various mobility-related events and initiatives in changing vehicle use rate as detected via LBS data.

5 Material

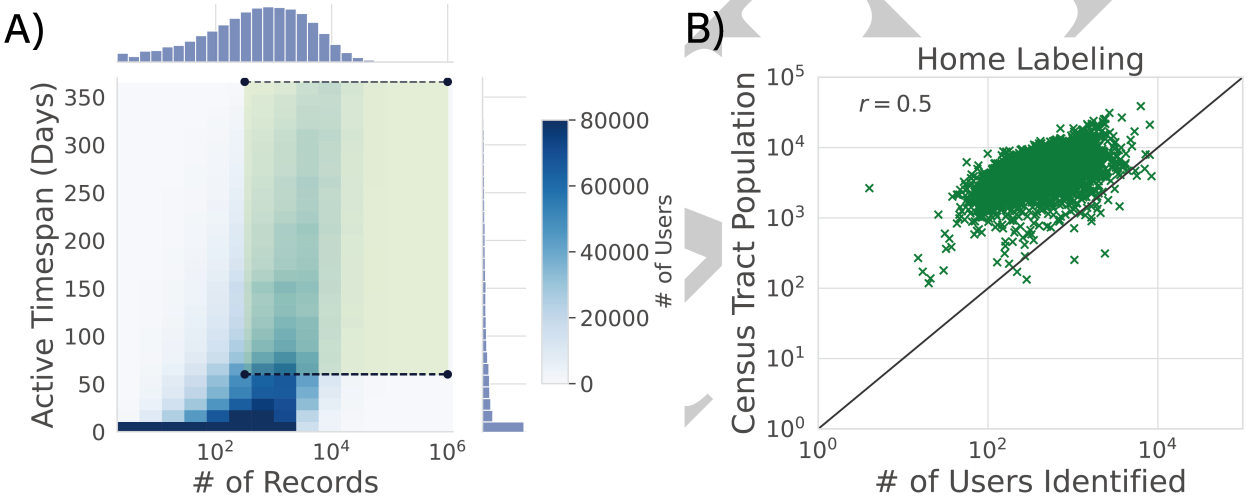
5.1 Location-based Service (LBS) Data

LBS data starting from January 1^{st}, 2019 is obtained from Spectus[1] and we use data up to the end of 2022 for this report. Raw spatial-temporal records (i.e. lat, long, and time) are transformed

[1] https://spectus.ai/

and processed by Spectus to construct trajectories [108]. The analysis is conducted using the trajectory data table built by Spectus. Because of privacy considerations, a device is only permitted to retain the same identification number for a duration of up to one year. This one-year limit represents the maximum period for which we can monitor a user. Therefore, we treat data in the years 2019, 2020, 2021, and 2022 as records collected from four distinct group of users. To obtain high-quality behavioral patterns, we adopt the user selection algorithm proposed by Xu et al. to select highly active users [112]. Section 15.5.2 explains the significance for using high-quality users in modeling human mobility patterns.

First, users who have more than 316 ($10^{2.5}$) pings and an active time-span longer than 60 days are deemed high-quality users. An active time-span is defined as the difference in number of days between the beginning of the last trajectory record of a user and the beginning of the first trajectory record of the user. Figure 1A) illustrates the number of users with particular lengths of active time-span and numbers of records in the 2020 dataset. We retain the highlighted users as high-quality users.



A) User selection criteria highlighted in the box. We select users with more than 60 days in their activity span and more than $10^{2.5}$ pings. B) Home detection validation with 2019 census data. We observe a good correlation between the number of users whose homes are found in particular census tracts and the tract population recorded by census.

Fig 1: User Selection and Validation

We identify home and work locations of a user to enable the analysis on change in commute patterns. As discussed in 15.2, both CDR and LBS data sets have been validated for capturing the commute flow that would otherwise be collected through traditional methods such as travel surveys. We define home location as the most frequently visited census block group between 7 pm and 7 am for each user with at least 10 visits. Work location is defined as the most frequently visited census block group between 7 am and 7 pm for each user with at least 10 visits. Users who have a home and a work location found in the same census block group are excluded in parts of the report that involve the comparative analysis on commutes and non-commutes [72]. Figure 1B) shows the correlation between the number of users with homes identified in a census tract and the respective census tract population from the 2019 census. We find a Pearson's correlation coefficient of 0.5, which serves as a validation of the home identification algorithm we implement.

Table 1 lists the number of users we retain for the analysis. The difference in the number of users between 2019 and the rest of the years is due to data availability.

Year	# Users	# High-Quality Users	# Users with Home and Work Found
2019	9,410,380	3,482,574	861,167
2020	5,912,373	2,396,990	431,190
2021	5,222,416	2,036,110	465,311
2022	5,618,760	2,582,405	702,847

Table 1: Number of Users

5.2 Definition of Regions

In parts of this report, we study regional differences in the impact of COVID-19 and the SIP order on mobility patterns. We include four metropolitan regions in the study: the Bay Area, greater Los Angeles, San Diego, and Sacramento. The Bay Area spans 9 counties, including San Francisco county, San Mateo county, Santa Clara county, Alameda county, Contra Coasta county, Solano county, Napa county, Sonoma county, and Marin county. Greater Los Angeles region includes Los Angeles county, Orange county, San Bernadino county, and Riverside county. The San Diego region and the Sacramento region refer to San Diego county and Sacramento county respectively.

5.3 Census Data

Census tract population and income data used in this report is obtained through the census Application Programming Interface (API) in python. We use the data of the American community survey of 2019.

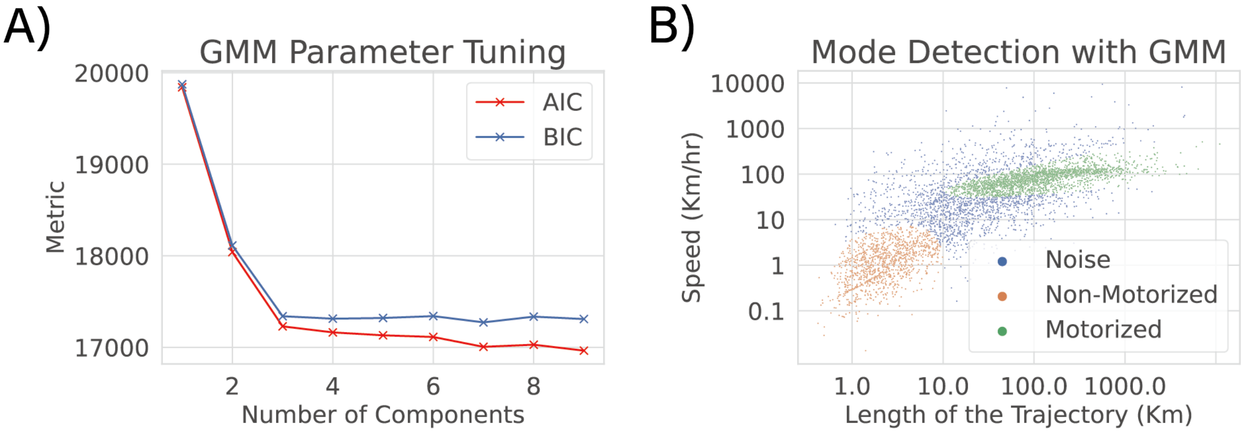
6 Statewide Changes in Vehicle Miles Traveled (VMT)

6.1 Mode Detection Algorithm

To distinguish motor vehicle trips from other trips made by users that are selected in section 5.1, we develop an unsupervised learning scheme to conduct mode detection. Existing literature relies heavily on GPS data, with features such as acceleration, heading, and jolt to train supervised learning models for mode detection with labels (i.e. trips with known modes) [86]. Other attempts have also been made using rule-based heuristics and geo-referencing dataset such as rail lines and highways to identify mode [43].

We develop a novel clustering approach to mode detection with Gaussian Mixture Model (GMM). GMM assumes that data in an n-dimensional feature space can be clustered into p mixtures using n-dimensional joint Gaussian distributions, characterized by the mean vector $u \in \mathbb{R}^n$ and the co-variance matrix $\Sigma \in \mathbb{R}^{n \times n}$ [49]. We use two features as inputs, the logged maximum speed (log kph) and the logged length of the trajectory (log km), to assign a trip one of three labels: motorized, non-motorized, or noise.

Using the elbow method, we determine from figure 2A) that the optimal number of mixtures, p, is 3. Figure 2B) illustrates a sample result of GMM. We conclude that trips labeled green must be vehicle trips, as the maximum speed that is observed, which is above $10^{1.6}$ (25 mph or 40 kph), is not feasible with walking or running. On the other hand, trips that have a maximum speed less than $10^{0.8}$ (6.3) kph are likely to be non-motorized. The region that these two mixtures occupy provides realistic insights into the mode of the trip. For all other trips, which are highlighted in blue, we assign noise as the label.



A) Parameter tuning. We select $p = 3$ as the optimal number of clusters. B) GMM clustering result

Fig 2: Gaussian Mixture Model (GMM) for Mode Detection

6.2 Results

We estimate VMT of users living in selected counties and the entire state of California by aggregating all motor vehicle trips identified using the method discussed in section 6.1. To compare the relative change in VMT across different counties, we use January 2020 VMT level as benchmark. Figure 3 shows the drastic decrease in VMT level from March to May, time in which the first wave of the pandemic arrived to California. We see it also in late 2020, when COVID-19 cases surged in the winter season. We observe a decrease in VMT starting in March, when the SIP order was announced by the State's government. The following month, April, sees the largest reduction in VMT levels, a change ranging from -40% to -60%.

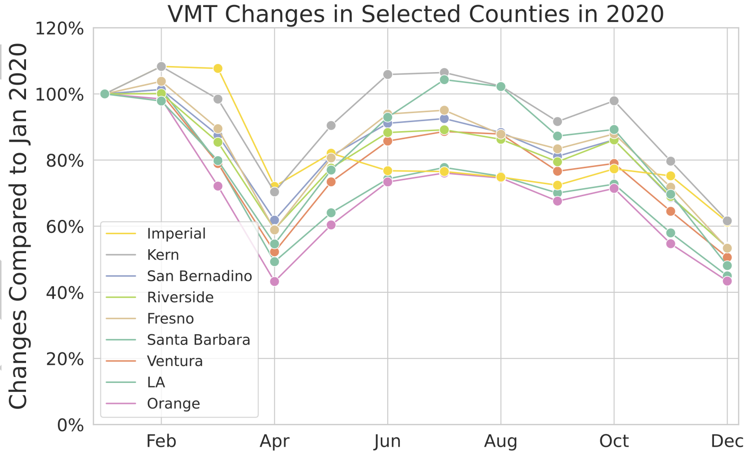
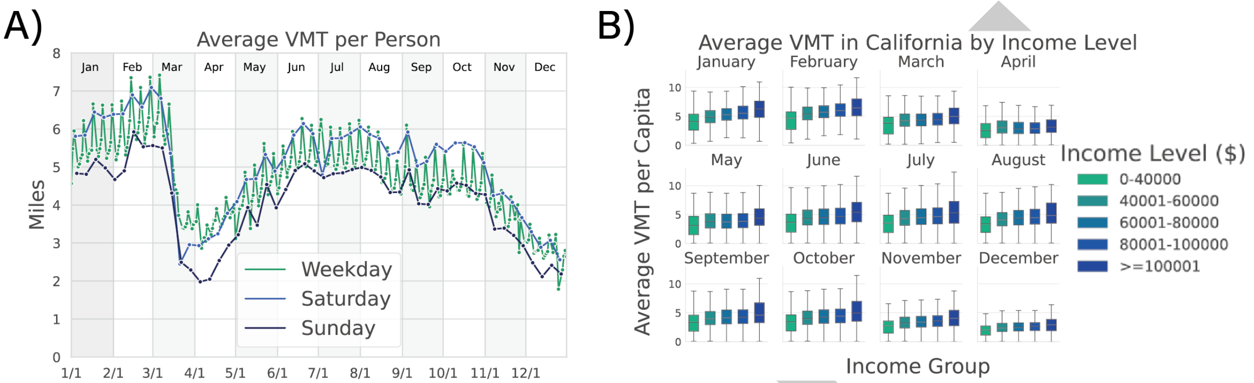


Fig 3: Vehicle miles travelled (VMT) changes in selected counties in California in 2020

Figure 4A) illustrates the average VMT changes, separated by day of the week in 2020. The SIP order did not change the relative VMT patterns across different days of the week. People travel consistently less with vehicles on Sundays and more on Saturdays. We also see the fluctuation in VMT within the weekdays always being distinguished from weekends. Still, the SIP order greatly reduced the average VMT, particularly on Sundays, during the first few weeks after its implementation. In the first two months of 2020, as well as in May and onward, per capita VMT level on Sundays is similar to the weekday with the lowest VMT. However, in the first few weeks after the implementation of the SIP order, VMT level on weekends decreased further compared to weekdays. Figure 4B) shows that across all months, census tracts of higher income levels observe a higher average VMT. Still, all census tracts, regardless of their income level, experience a decrease in VMT in March, April, November, and December 2020.



A) Changes in average VMT per capita in California based on day of the week. A lower VMT level is consistently observed on Sundays and a higher VMT level is consisntly observed on Saturdays. B) Changes in VMT of census tracts of different income levels in California in 2020.

Fig 4: Changes in statewide vehicle miles travelled (VMT) in 2020

6.3 Discussion

It is worth noting that counties characterized by an urban settings such as Los Angeles county, Orange county, and Santa Barbara county experienced a larger reduction in VMT compared to their rural counterparts such as Imperial county and Kern county. The difference reveals that people living in remote, rural places are more dependent on vehicles as a means of transport. The greater reduction observed on weekends during the first few weeks after the implementation of the SIP order suggests that people cut more non-work related movement at the beginning of the pandemic. However, this divergence began to diminish starting in May and people's travel behaviors became more similar between weekends and weekdays even though the overall VMT was still much less than normal times. Additionally, in California we observe little correlation between income levels and the impact on VMT caused by the pandemic. Still, in periods when travel decreased drastically, such as in March and December, all census tracts, regardless of their income levels, have roughly the same VMT per capita. In comparison, during periods with fast recovery, such as the summer months, census tracts of high-income are more elastic in that they quickly returned to a relatively high level of VMT.

7 Change by Trip Purpose

7.1 Methodology

7.1.1 Trip Purpose Detection

With the home and work location identified using the process outlined in section 5.1, we define a work trip as a trajectory that starts in a user's home census block group and ends in the user's work census block group, or vice versa. Any trip that is not a work trip is defined as a non-work trip. The word commute, which by our definition is equivalent to a work-trip, is also used in the following sections.

7.1.2 Flow Network

We construct directed, weighted networks using the commutes to analyze the change in the structure of the commute network in California. The nodes in the commute flow network are census tracts. We define the weight on each directed edge between the origin node and a target node as commute flow, which is the number of people whose home is at the origin node and whose workplace is at the target node. We use the concept of communities, which are subsets of nodes that are densely connected [65], to characterize the network by showing the spatial distribution of connectivity. The Louvain method is used to identify communities in the flow network by maximizing modularity, which measures the degree to which a network can be split into distinct flow regions or communities. Therefore, a larger number of communities suggests that the network can be easily separated into small subsets as the flow within each community outnumbers the inter-community flow.

7.2 Results

Figure 5 shows the number of trips observed in 2020 separated by trip purpose. Although both the number of work and non-work trip decreased in March 2020 and experienced recovery in the summer months, work trip recovered at a slower pace compared to non-work trip. The ratio of work to non-work trip only started to rebound in September and October, before it quickly dropped again in November and December.

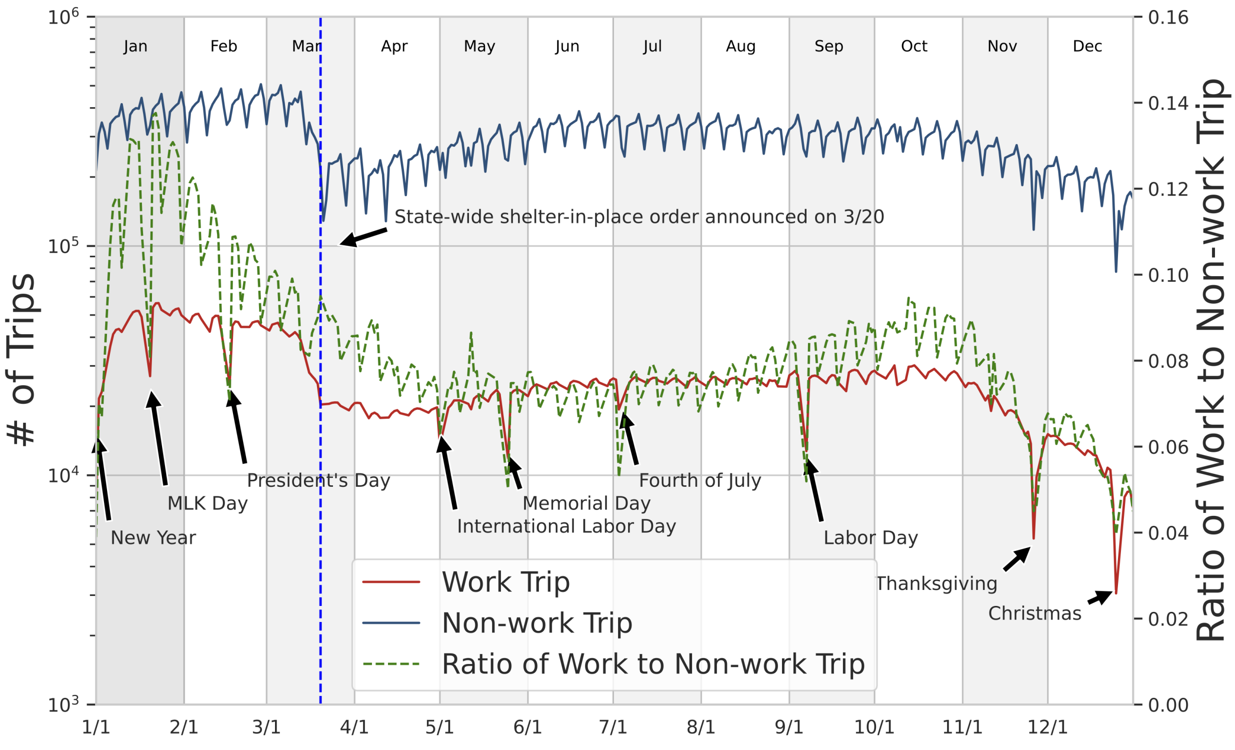
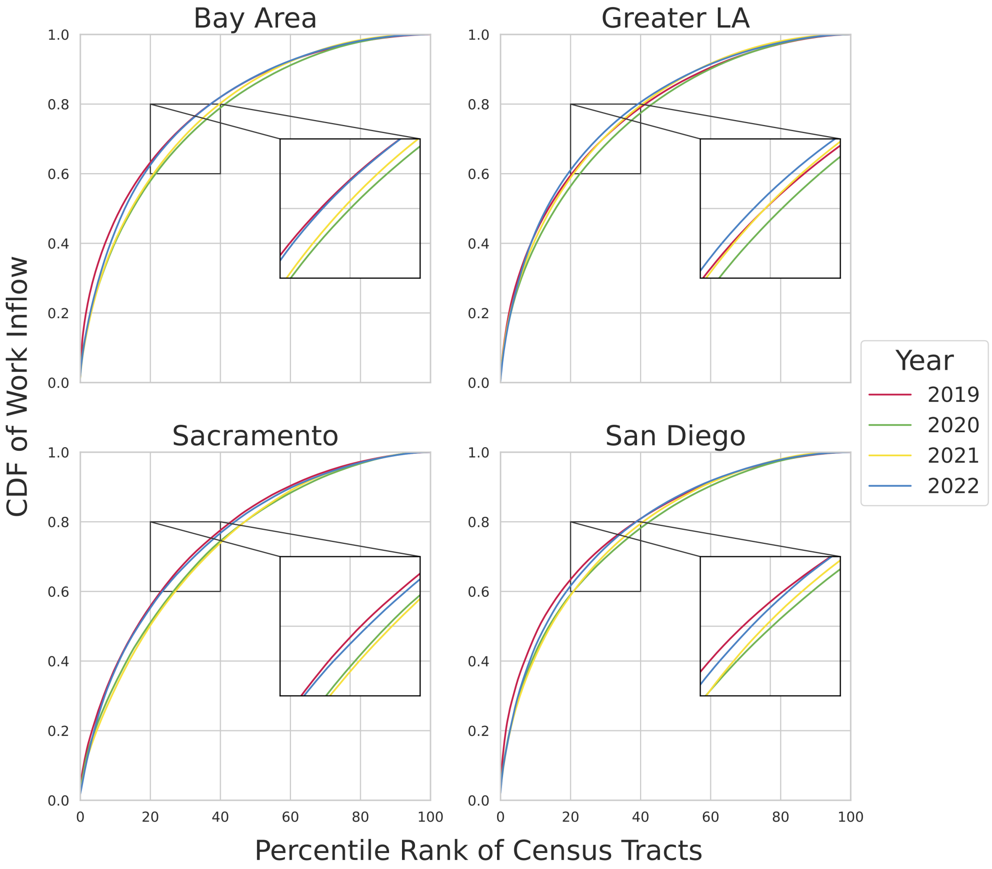


Fig 5: Number of trips made in 2020 by trip purpose

We also investigate regional differences in the impact of the COVID-19 pandemic on work trips and their recovery process. Figure 6 shows the cumulative distribution of work trip inflow in the four different regions across the four different years. The x-axis represents the rank of the census tract by the number of users that work in the respective census tract. The y-axis shows the cumulative percentage of the total work flow that go into these census tracts. We see that in all four regions, the curve for 2020 is shifted to the right, meaning that work locations of users are less concentrated in census tracts that previously drew a lot of users as work locations. Also, we observe that the distribution returned to 2019 level in 2022, showing a full recovery in the distribution of the ranks of work locations. The Greater LA metro region, in contrast to the other three regions, recovered more quickly as the distribution of flow in 2021 is already close to its 2019 level.



The distributions reveal the degree to which work locations are concentrated in a group of census tracts. We observe work locations are less concentrated during the COVID-19 pandemic. All four regions eventually returned to the 2019 level in 2022.

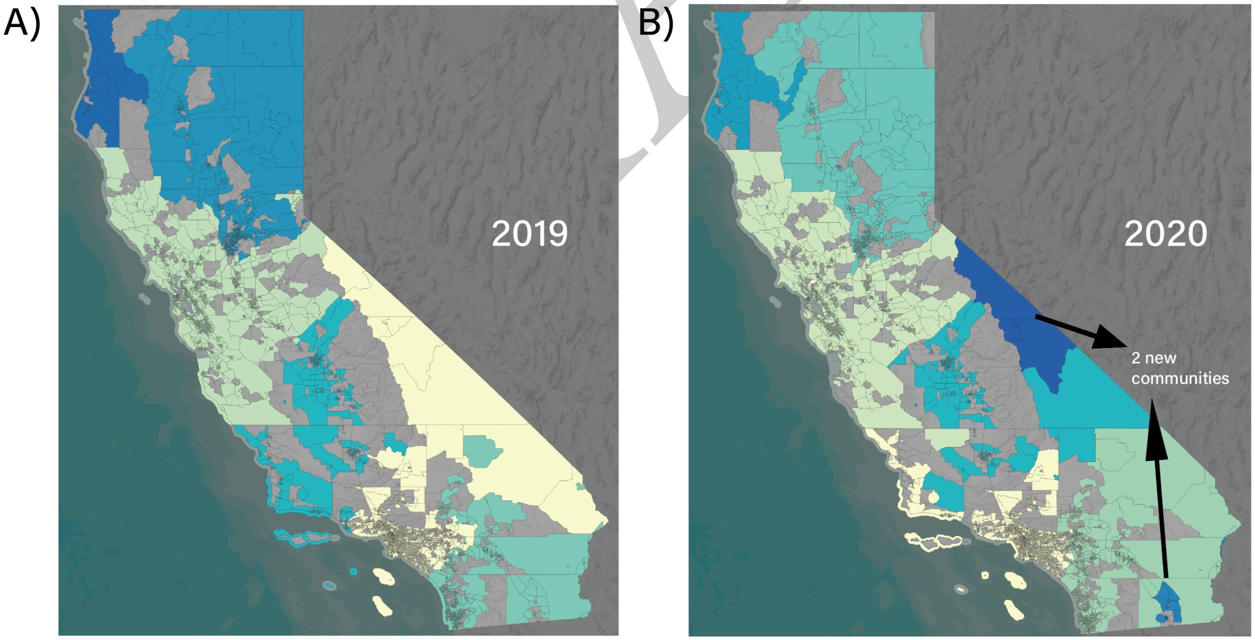
Fig 6: Concentration of work locations

As discussed in section 7.1.2, we construct commute flow networks and use network communities to study the differences across the four-year period. Table 2 lists the characteristics of the flow networks in each year. The 2020 and 2021 commute networks possess a much smaller number of edges compared to 2019 and 2022, which shows the reduction in the number of census tract pairs that have users commuting between them. While the number of edges began to return to 2019 level in 2022, the number of communities remained at 8 and the modularity remained high.

This reveals that the commute flow network in California stayed fragmented. Figure 7 shows the difference in the community structure between 2019 and 2020. In both years, we observe separations among different regions. Each of the large urban areas, including Los Angeles, the Bay Area, San Diego, Bakersfield, and Sacramento act as centers of a wider region, demonstrating the gravitational pull these urban centers have on the surrounding region in terms of employment opportunities. It is interesting to observe the overwhelming dominance of the Bay Area over its surrounding communities. Cities that are geographically closer to Sacramento and Bakersfield also see more commutes to the Bay Area rather than to Sacramento or Bakersfield. In 2020, two more communities came into existence: one located in El Centro, Southern California, and the other situated on the eastern side of the upper Sierras. These locations are isolated from major urban centers due to natural obstacles like mountains and deserts. This highlights the heightened impact of the pandemic on geographically remote regions.

	# Nodes	# Edges	# Communities	Modularity
2019	8,033	145,838	6	0.628
2020	8,026	89,382	8	0.653
2021	8,009	73,401	8	0.648
2022	8,019	111,652	8	0.666

Table 2: Commute Flow Networks



A) Communities in California commute network in 2019. B) Communities in California commute network in 2020.
Fig 7: Spatial Distribution of Communities in California

7.3 Discussion

Our findings reveal that the COVID-19 pandemic and associated lockdown measures had varying effects on trips based on their purposes and geographical regions. Commuting was more adversely impacted compared to non-work-related travel. Even as the overall number of trips began to recover in the summer of 2020, the ratio of commutes to non-work-related trips did not return to pre-pandemic levels. This suggests that the lockdown measures didn't necessarily deter people from traveling but rather the shift to remote work reduced the necessity for travel.

The policy measures had a comparatively smaller impact on personal travel. Additionally, our analysis of four different regions showed that the top 20% of census tracts, in terms of work inflow, accounted for over half of the users. However, in 2020 and 2021, we observed a more dispersed distribution of work locations. Census tracts that previously attracted a significant number of users for work played a less dominant role, as the same percentage of users were distributed across a greater number of census tracts for work.

Through the construction of four flow networks spanning a four-year period, we observe a significant reduction in the number of edges, suggesting a decline in commutes between numerous pairs of census tracts. The heightened modularity indicates a decrease in inter-community commutes post-COVID-19 compared to pre-pandemic times. Notably, the number of communities remained constant at 8 even in 2022, and the count of census tract pairs with commutes did not revert to 2019 levels in 2022. This underscores the enduring and lasting impact of the pandemic on commuting patterns in California

8 Radius of Gyration (r_g)

8.1 Methodology

Radius of Gyration (r_g) measures the spatial spread of a user's activity, it is an easy-to-compute metric that provides direct yet valuable insights on the travel behaviors of individual users. Higher r_g suggests more vehicle use and long-distance travel, and a lower r_g indicates less vehicle use and more local travel. Section 15.1.1 provides a detailed discussion on the implication of r_g in human mobility. For the following analysis, r_g of an individual u is defined as:

$$r_g(u) = \sqrt{\frac{1}{n_u}\sum_{i=1}^{n_u} dist(r_i(u) - r_{cm}(u))^2}$$

where n_u is the number of records of an individual and the $dist$ operator calculates the distance between the location of a record and the center of mass of all records [41]. To compare different geographical areas in California, we further define the radius of gyration of an area as the average radius of gyration of all users that live in the given area. The home detection process is outlined in section 5.1.

We built an interactive web application that allows the user to retrieve radius of gyration statistics for various counties and regions. The web application user can not only select different counties in California but also periods, in weeks, enabling a quick comparison of radius of gyration from both a spatial and a temporal perspective.

The web application provides easy-to-access informational visualizations on mobility trends and behaviors with a database that can be refreshed daily.

Figure 8 shows an example of a map of radius of gyration in California. We see a clear distinction between urban or suburban regions and rural regions. The Bay area, Los Angeles and San Diego metro regions have much smaller radius of gyration compared to the Central Valley, which is more rural. We also see larger radius of gyration in census tracts along the interstate 5, where exists a combination of low density urban development and convenient access to freeways.

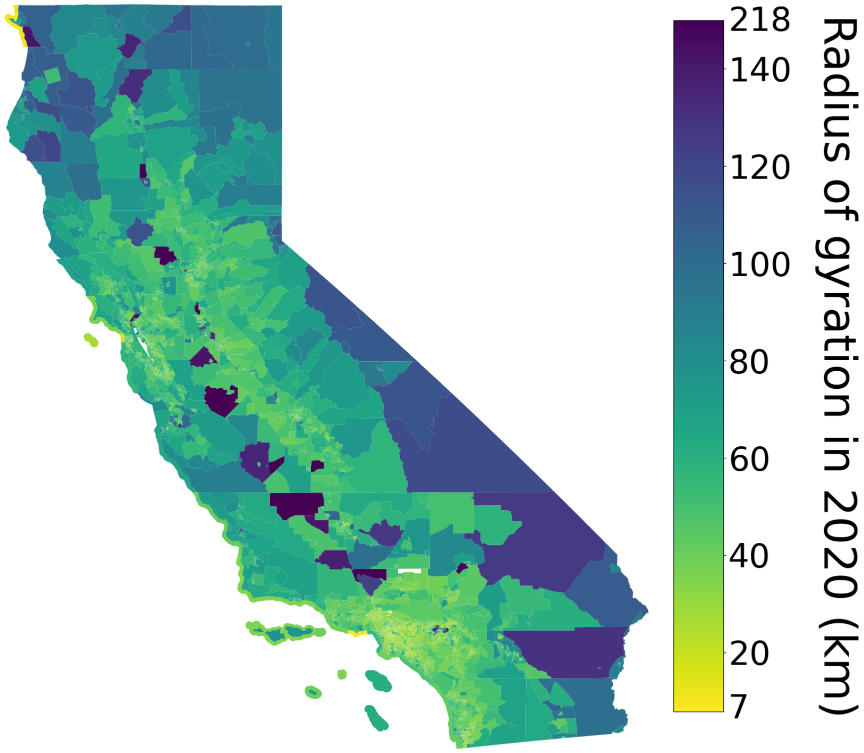
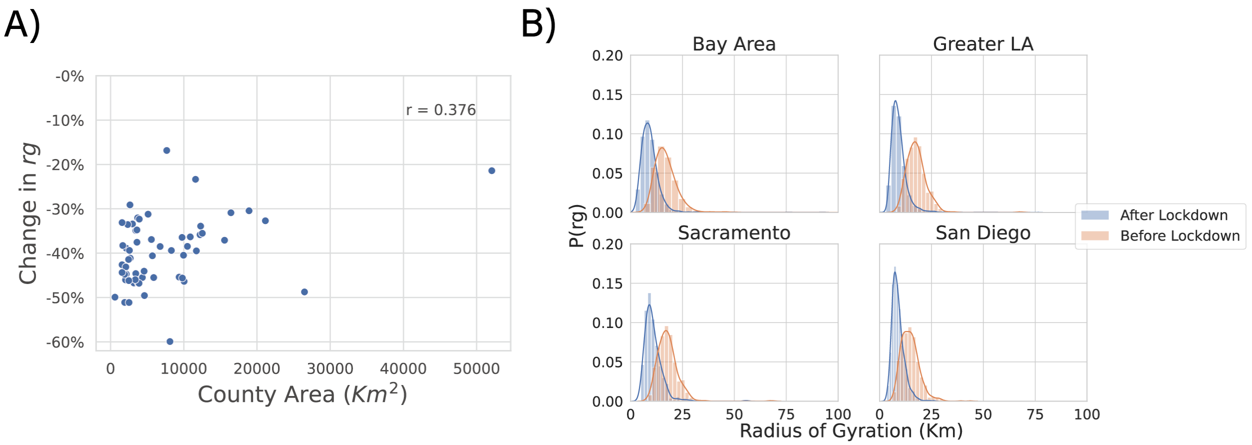


Fig 8: Radius of gyration distribution in California in 2020 by census tract

8.2 Results and Discussion

We compute the change in r_g by comparing r_g in the week after and prior to the implementation of the SIP order. The average reduction in r_g of a given area, as shown in figure 9A), is positively correlated with the area of a county, where a large size indicates a more rural environment. Although all counties experienced decrease in their average r_g, we see disparity in impacts of the SIP on different regions. We observe smaller decrease in travel in rural counties, which are larger in size. In metro regions, as shown in figure 9B), the shift in the distribution of r_g appears more homogeneous.



A) Correlation between the change in average r_g and the area of county after the announcement of the SIP order. B) Distribution of r_g before and after SIP in four metro regions in California.

Fig 9: Regional r_g trends

9 Changes in Residential Locations

9.1 *Methodology*

9.1.1 *Home Change Detection Algorithm*

As 15.1.3 examines, LBS and CDR records have proven to be effective in capturing human migration patterns on a city-level scale. However, challenges remain in using LBS and CDR data to model and detect home changes over short distances. To examine changes in residential locations (intra-state moves) on a census tract level during the COVID-19 pandemic in California, we devised a two-step semi-supervised learning algorithm. Our assumption is that a user demonstrates two clusters of spatial-temporal travel behaviors, with each cluster corresponding to the residential location before and after the move, provided that the user has relocated once. We first detect the two clusters using k-means algorithm, which is the unsupervised step. We then use the assigned cluster labels of the observations as pseudo-labels. We use these pseudo-labels to train a Support Vector Machine (SVM) model, which returns the separating hyperplane in the input space, completing the supervised step. Lastly, we impose heuristics to determine if the two clusters indeed resemble two distinct home locations in order to conclude home changes.

K-means algorithm is a clustering technique that updates the location of k centroids by minimizing the sum of squared distance between each observation and its closest centroids, thus partitioning the n-dimensional population in to k sets [61]. SVM is a supervised classification algorithm that identifies labels of observations by finding the best separating hyperplane that maximizes the margins between the two groups [23].

For each user, we use the latitude and longitude of the end point of the user's trajectories (trajectories defined by Xiang et al. [108]), as well as the start time (number of days since 1/1/2020) of all trajectories that end between 7pm and 7am as features in the input space. Each observation in the input space, therefore, is a set of features associated with a trajectory that a user makes. We use SVM to prevent potential over-fitting in results from k-means clustering, which can lead to observations being falsely classified due to their geographical proximity to the centroid of the wrong cluster. For example, a user might still visit places near their original home after changing their residential location. Because k-means treats all three features, latitude, longitude, and time with equal weight, the algorithm can still assign these visits after the home change to the cluster that represents the original home location. We set parameter $c = 0.01$, which is the penalty attached to misclassifying observations to create a wider separating hyperplane in SVM, to a very small number. This helps correct the mistakes that are made in k-means to construct the clusters of behaviors before and after the change in residential location. After identifying the two clusters, we determine the date of home change of a user as *Move Date* $= min(max(c_{bh}, c_{ac}))$, where c_{bh} is a set of dates of the observations that belong to the cluster before the home change and c_{ac} is a set of dates of observations that belong to the cluster after the home change. Using the date of home change, we apply the home detection algorithm outlined in section 5.1 to the trajectory observations before and after the *Move Date* to determine the two home locations.

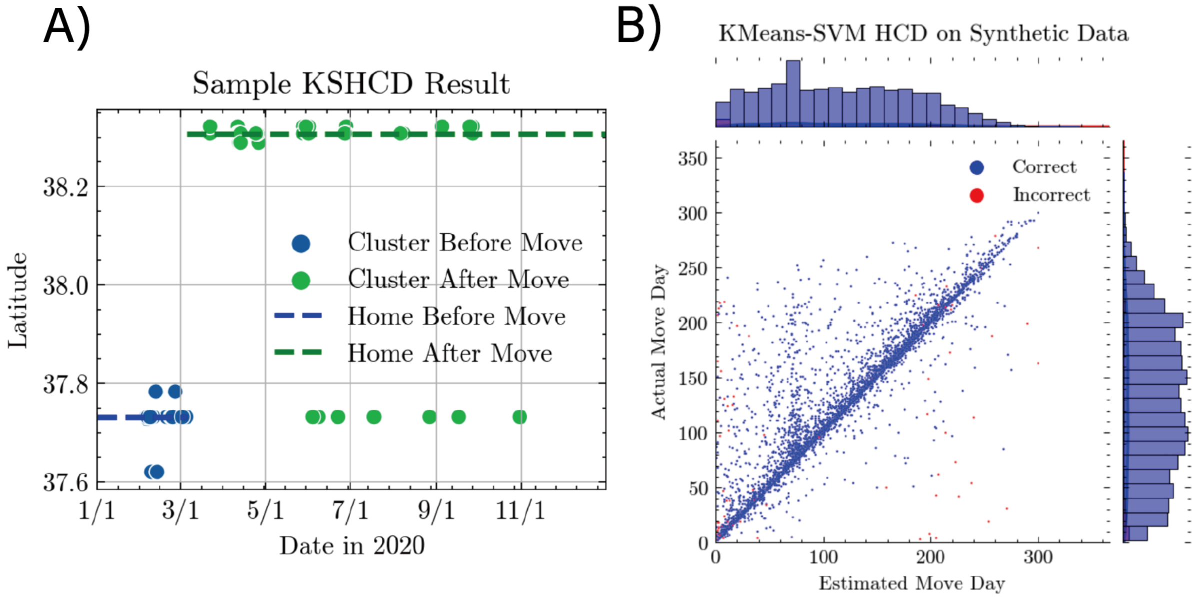
We impose additional conditions to select users who changed their residential locations. Only users whose two home locations are at least 5 miles (8km) apart from each other and that are observed for at least 20 times (trajectories) in each of the two clusters are deemed as users who moved. Algorithm 1 details the home change detection process.

Algorithm 1 Home Change Detection

```
$\alpha = 20$
$d = 5$ miles or $8$ km
for every user do
    Select all trajectories that end between 7pm and 7am
    Fit a $k$-means model with $k$=2 using 3 features: lat, lon, and day of visit
    Use the fitted $k$-means model to assign each observation to one of one labels
    Fit SVM with the pseudo-labels to cluster the observations
    Compute the move date as $min(max(c_{bh}, c_{ac}))$
    Conduct frequentist home detection before and after the move date
    if # of records both before and after home change $> \alpha$ then
        if Distance between two home locations $> d$ then
            Conclude home change
        end if
    end if
end for
```

9.1.2 Validation

Figure 10A) shows a user who changed home in 2020 and was successfully detected by algorithm 1. The figure shows two distinct clusters of behaviors in which the algorithm is able to label the most frequently visited location as the home location before and after the home change. Figure 10B) is the result of applying the algorithm to a group of synthetic users that are created using the procedures described in 16.1. Since the synthetic users are created artificially, we know the home locations of these users before and after the move. A correct home change labeling is defined as accurately identifying the home location both before and after the home change. We achieve an accuracy of 82% on the synthetic user data set.



A) Sample home change detection on a real user. B) Home change detection validation on synthetic users.
Fig 10: Home change detection algorithm validation

9.2 Results

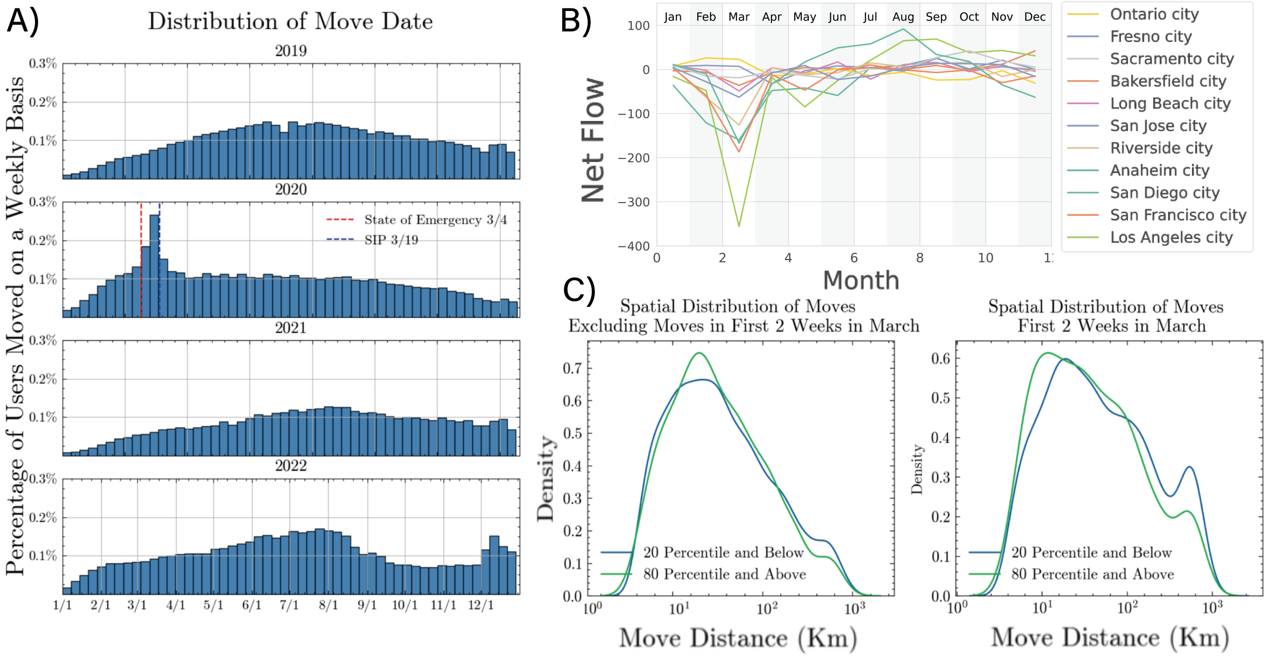
Figure 11A) shows the temporal distribution of home changes in each year between 2019 and 2022 and table 3 shows the overall percentage of users that moved within California each year. Overall, a smaller percentage of users moved in 2020, possibly due to the travel restrictions and the closing of businesses during the SIP order.

	# High-Quality Users	# Home Changes	% Moved
2019	3,482,574	199,286	5.72%
2020	2,396,990	125,929	5.25%
2021	2,036,110	118,446	5.82%
2022	2,582,405	164,927	6.39%

Table 3: Number of home changes detected in each year between 2019 and 2022

We observe significantly higher number of moves within the two-week period between when the state announced the state of Emergency on 3/4/2020 and when the SIP order was implemented on 3/19/2020. The results indicate that the SIP order only had short-term impact on people's moving behaviors as the percentage of moves quickly returned to normal levels in the weeks following the SIP. The gradual climb in the percentage of moves at the beginning of each year and the decrease in the percentage of moves towards the end of each year are attributed to the nature of the dataset and the algorithm. Because of the fact that algorithm 1 only considers users whose two clusters before and after the home change contain more than 20 observations as users who moved, it is more difficult to conclude home changes at the beginning and at the end of a given time window. Additionally, since a user identification number is only present in the database up to one year due to privacy reasons, not all observations of a user is used in home change detection.

Figure 11B) illustrates the the net flow of moves in the largest cities in California. The net flow of a city is defined as the difference in the number of people who moved into and out of



A) Temporal distribution of move dates in 2019, 2020, 2021, and 2022. B) Monthly net flow of migration in selected cities in California C) Distribution of move distance in (excluding) the two week period between 3/4/2020 and 3/19/2020.

Fig 11: Home change behaviors in 2020

the city. A negative net flow means that more resident left the area than those moved in. We observe negative net flow across almost all cities in California in March 2020. Although more populous cities experienced proportionally more outflow compared to smaller cities, it is important to note that neither Ontario nor Fresno experienced a negative net flow in March. Figure 11C) shows the distribution of move distance of users in two groups, which are those living in the top 20 percentile census tracts by income level, and those that are living in the lower 20 percentile census tracts. Move distance is defined as the euclidean distance between the two centroids of the census tracts in which a user's two homes before and after the move exist. The disparity in move distance distribution between these two demographic groups is minimal when we exclude the moves that took place between 3/4/2020 and 3/19/2020. Nevertheless, during this specific time frame, individuals residing in the poorest census tracts relocated to greater distances compared to those in other areas. Importantly, the overall move distance within this two-week period saw a significant increase, resulting in a bimodal distribution. The second mode, shown on the right, correspond to a move distance of $10^{2.7}(\,313mi/\,500km)$ to $10^{2.8}(\,394mi/\,630km)$, which translate to the distance between the Bay Area and Southern California.

9.3 Discussion

A noticeable surge in residential relocations occurred in early March 2020, in response to the array of measures implemented by the state of California to mitigate the spread of COVID-19. The immediate effects of the pandemic and associated policies persisted for roughly two weeks. Notably, the moves that transpired during this timeframe were distinguished by their longer distances, especially among individuals residing in economically disadvantaged census tracts. The uptick in move distances suggests that individuals making these relocations underwent sudden shifts in their work or study circumstances, necessitating or enabling inter-regional moves. Intriguingly, this surge in relocations was relatively brief, lasting for only about two weeks.

One plausible explanation is that, following the implementation of Shelter-in-Place (SIP) measures, logistical challenges associated with moving increased substantially. Additionally, regional disparities are evident. Cities in the Central Valley and suburban areas surrounding population centers witnessed minimal to no significant net flow changes in March 2020. This highlights the greater stability of suburban locations as compared to urban areas for establishing a residence.

10 Detecting Event-Induced VMT Changes

Our task in this section involves assessing the effects of four mobility-related events that took place in Sacramento in 2019. Section 16.2 has the complete list of census tracts in which these mobility-related programs were launched and figure 12 shows the locations of these census tracts, which are all in downtown Sacramento. Table 4 presents a comprehensive list detailing the events and the expected impact on vehicle usage. Our hypothesis is that electric scooters, given their classification as non-motor vehicles and a speed limitation of approximately 15 mph, would contribute to a reduction in vehicle use. On the other hand, since car share services and rapid transit all provide mobility services with motor vehicles, the introduction of such programs would lead to an increase in motorized trips.

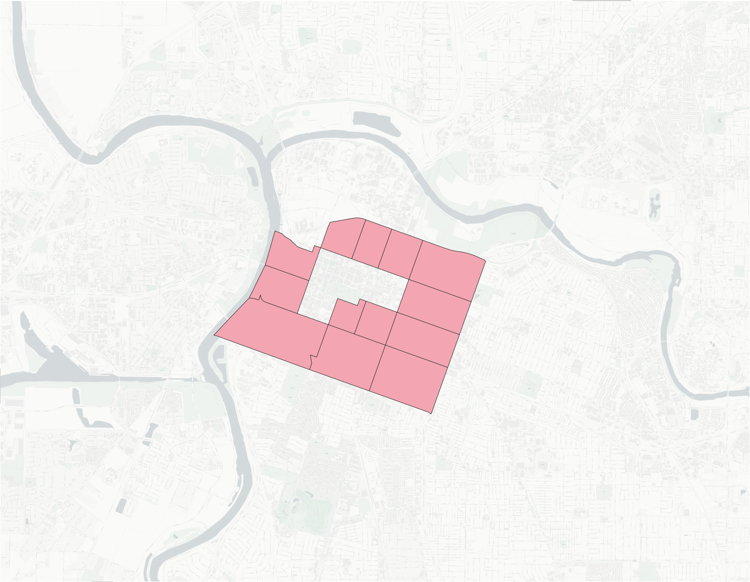


Fig 12: Selected census tracts in Sacramento

10.1 Methodology

To investigate the impact of these events on motorized travel, we use the mode detection algorithm presented in section 6.1 to calculate the percentage of trips that are made using motor vehicles. Bootstrap sampling is a sampling technique that calculates a target statistics by repeatedly drawing samples of observations from the same population. After each draw, the selected observations are returned to the population from which subsequent samples are collected. For each event, we create 100 bootstrap samples of trajectories that begin in the selected census tracts in the month prior to the event and in the month of the event, each of a sample size of 50,000. We compare the distribution of the percentage of vehicles trips between these two months.

Time	Event	Hypothesized Impact
Feburary, 2019	JUMP released electric scooters	Decrease in vehicle usage
March, 2019	GIG Car Share released shared-vehicles	Increase in vehicle usage
June, 2019	JUMP increased its electric bike fleet	Decrease in vehicle usage
September, 2019	Sacramento Rapid Transit launched a new transit program SacRT Forward	Increase in vehicle usage

Table 4: Mobility-related events in 2019

10.2 Results and Discussion

Figure 13 illustrates the impact of the events listed in table 4 and table 5 shows the sample size in terms of both the number of users and the number of trajectories used in this analysis.

	# Trips	# Users
January, 2019	192,426	34,636
Feburary, 2019	190,643	36,203
March, 2019	248,462	46,121
May, 2019	270,871	50,633
June, 2019	277,004	36,203
August, 2019	256,785	45,665
September 2019	257,476	46,948

Table 5: Sample size in different months in selected census tracts in Sacramento

Among the four events, only the increase in fleet size of JUMP scooters leads to distributional changes in vehicle usage rate, shown in C). We have noted a substantial decrease in vehicle utilization in the month subsequent to the release. It’s crucial to note that vehicle usage is influenced by numerous variables, such as weather and seasonal variations. In the absence of a controlled environment, these findings should not be construed as conclusions derived from an experiment. Nevertheless, the analysis of Location-Based Services (LBS) data provides valuable insights into the effects of these initiatives.

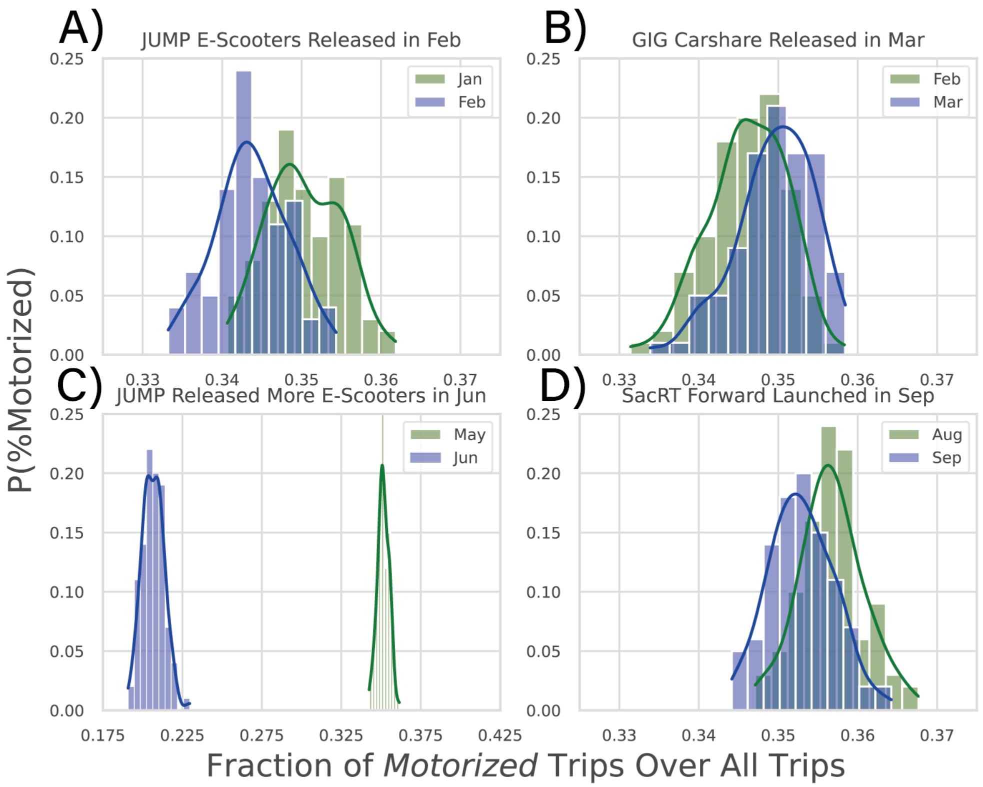


Fig 13: Bootstrapped distribution of vehicle trip percentage

11 Summary and Conclusions

In this research, we leverage Location-Based Services (LBS) data to examine shifts in human mobility patterns in California caused by the COVID-19 pandemic and the corresponding measures enforced by the state. Below, we provide a summary of our key findings and achievements:

- We have identified a significant decline in overall vehicle miles traveled (VMT) across the state during March and April of 2020. However, notable regional disparities emerged. Urban counties, like Los Angeles County, experienced a VMT reduction of up to 55%, while rural counties, such as Imperial and Kern County, saw VMT decrease by approximately 20-30%

- We observe variation in trips of different purposes. The number of non-commute trips recovered much faster compared to commutes. The analysis of the network of commuting flows shows that commutes did not return to pre-pandemic levels when measured in 2022.

- We use r_g to measure the spread of a user's activity space. We find that the reduction in r_g resulted from the SIP order is positively correlated with the area of a county. Rural counties experienced little to no reduction in r_g.

- We developed a home change detection algorithm, and we find that people moved over a much longer distance within the two week period between 3/4/20, when the state of emergency was announced, and 3/19/20, when the SIP order was announced. It shows the impact of the COVID-19 pandemic had on disrupting people's ties to work or study locations.

- We assess the effectiveness of selected transportation projects in changing mode preferences in selected census tracts in the city of Sacramento. We find that JUMP's increase in fleet size in June, 2019 decreased the overall vehicle usage rate.
- We evaluate and validate the use of CDR and LBS data in modelling human mobility patterns and review their applications in public health and urban planning. The results can be found in the white paper under section 15.

The main contribution of this report is that we unfold the reactions to the COVID-19 pandemic in California through the lens of mobile phone data. Not only did we observe a reduction in people's movement but also spatial and temporal disparities in their impact, as well as disparities in trips of different purposes. We demonstrate the viability of using LBS data to monitor human mobility patterns in California. We develop novel unsupervised algorithms to conduct mode detection and home change detection, overcoming the need for labeled datasets for training purposes. The developed algorithms allow us to observe that work travel and work-induced VMT were more susceptible to exogenous disruptions, such as the COVID-19 pandemic, when compared to non-work travel. We also measured changes from home locations exacerbated during the SIP order.

12 Recommendations

Future efforts can be focused on advancing and fine-tuning unsupervised algorithms for detecting changes in residential settings and identifying activity modes. These algorithms empower the utilization of Location-Based Services (LBS) data for critical inferences without relying on costly and challenging-to-acquire labels.

13 References

[1] Armin Akhavan et al. “Accessibility Inequality in Houston”. In: *IEEE Sensors Letters* 3.1 (Jan. 2019). Conference Name: IEEE Sensors Letters, pp. 1–4. ISSN: 2475-1472. DOI: 10.1109/LSENS.2018.2882806.

[2] Laura Alessandretti, Ulf Aslak, and Sune Lehmann. “The scales of human mobility”. en. In: *Nature* 587.7834 (Nov. 2020). Number: 7834 Publisher: Nature Publishing Group, pp. 402–407. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2909-1. URL: https://www.nature.com/articles/s41586-020-2909-1 (visited on 06/24/2022).

[3] Lauren Alexander et al. “Origin–destination trips by purpose and time of day inferred from mobile phone data”. en. In: *Transportation Research Part C: Emerging Technologies*. Big Data in Transportation and Traffic Engineering 58 (Sept. 2015), pp. 240–250. ISSN: 0968-090X. DOI: 10.1016/j.trc.2015.02.018. URL: https://www.sciencedirect.com/science/article/pii/S0968090X1500073X (visited on 08/16/2022).

[4] T. Althoff et al. “Large-scale physical activity data reveal worldwide activity inequality”. In: *Nature* 547.7663 (2017), pp. 336–339. DOI: 10.1038/nature23018. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85025128546&doi=10.1038%2fnature23018&partnerID=40&md5=7c907e5e76f9817f487f4ece5e81ee5d.

[5] A. Amini et al. “The impact of social segregation on human mobility in developing and industrialized regions”. In: *EPJ Data Science* 3.1 (2014), pp. 1–20. DOI: 10.1140/epjds31. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84905501315&doi=10.1140%2fepjds31&partnerID=40&md5=4322fc9198318f7ae5c9bb23a4b046f3.

[6] J.P. Bagrow, D. Wang, and A.-L. Barabási. “Collective response of human populations to large-scale emergencies”. In: *PLoS ONE* 6.3 (2011). DOI: 10.1371/journal.pone.0017680. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-79953317477&doi=10.1371%2fjournal.pone.0017680&partnerID=40&md5=01faf775cb569b180dd12bdbb9aef2fc.

[7] D. Balcan et al. “Multiscale mobility networks and the spatial spreading of infectious diseases”. In: *Proceedings of the National Academy of Sciences of the United States of America* 106.51 (2009), pp. 21484–21489. DOI: 10.1073/pnas.0906910106. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-76049115282&doi=10.1073%2fpnas.0906910106&partnerID=40&md5=08ebd77a8c8110b1403eed5faf20b4f0.

[8] Hugo Barbosa et al. “Human mobility: Models and applications”. In: *Physics Reports* 734 (2018), pp. 1–74.

[9] Edward Barbour et al. “Planning for sustainable cities by estimating building occupancy with mobile phones”. In: *Nature communications* 10.1 (2019), pp. 1–10.

[10] A. Bassolas et al. “Hierarchical organization of urban mobility and its connection with city livability”. English. In: *Nature Communications* 10.1 (2019). ISSN: 2041-1723. DOI: 10.1038/s41467-019-12809-y.

[11] A. Bassolas et al. “Mobile phone records to feed activity-based travel demand models: MATSim for studying a cordon toll policy in Barcelona”. English. In: *Transportation Research Part A: Policy and Practice* 121 (2019), pp. 56–74. ISSN: 0965-8564. DOI: 10.1016/j.tra.2018.12.024.

[12] R.A. Becker et al. “A tale of one city: Using cellular network data for urban planning”. In: *IEEE Pervasive Computing* 10.4 (2011), pp. 18–26. DOI: 10.1109/MPRV.2011.44. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-80054104236&doi=10.1109%2fMPRV.2011.44&partnerID=40&md5=8d21be981995fd07dfd110673b441e1b.

[13] V. Belik, T. Geisel, and D. Brockmann. “Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases”. In: *Physical Review X* 1.1 (2011), pp. 1–5. DOI: 10.1103/PhysRevX.1.011001. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-83655195346&doi=10.1103%2fPhysRevX.1.011001&partnerID=40&md5=c52eb0f7b234a8369543549e0c73f1b2.

[14] Paolo Bellavista, Axel K, and Sumi Helal. “Location-Based Services: Back to the Future”. en. In: *IEEE Pervasive Computing* 7.2 (Apr. 2008), pp. 85–89. ISSN: 1536-1268. DOI: 10.1109/MPRV.2008.34. URL: http://ieeexplore.ieee.org/document/4487093/ (visited on 09/18/2023).

[15] Linus Bengtsson et al. “Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti”. en. In: *PLOS Medicine* 8.8 (Aug. 2011). Publisher: Public Library of Science, e1001083. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001083. URL: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001083 (visited on 08/15/2022).

[16] Nibir Bora, Yu-Han Chang, and Rajiv Maheswaran. “Mobility Patterns and User Dynamics in Racially Segregated Geographies of US Cities”. en. In: *Social Computing, Behavioral-Cultural Modeling and Prediction*. Ed. by William G. Kennedy, Nitin Agarwal, and Shanchieh Jay Yang. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2014, pp. 11–18. ISBN: 978-3-319-05579-4. DOI: 10.1007/978-3-319-05579-4_2.

[17] D. Brockmann, L. Hufnagel, and T. Geisel. “The scaling laws of human travel”. en. In: *Nature* 439.7075 (Jan. 2006). Number: 7075 Publisher: Nature Publishing Group, pp. 462–465. ISSN: 1476-4687. DOI: 10.1038/nature04292. URL: https://www.nature.com/articles/nature04292 (visited on 06/24/2022).

[18] L. Canzian and M. Musolesi. “Trajectories of depression: Unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis”. In: *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2015, pp. 1293–1304. DOI: 10.1145/2750858.2805845. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-

84960941653&doi=10.1145%2f2750858.2805845&partnerID=40&md5=c9f21296e5b4720bb9ddd3c596822ec5.

[19] T.K. Chan et al. “A Comprehensive Review of Driver Behavior Analysis Utilizing Smartphones”. In: *IEEE Transactions on Intelligent Transportation Systems* 21.10 (2020), pp. 4444–4475. DOI: 10.1109/TITS.2019.2940481. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086513620&doi=10.1109%2fTITS.2019.2940481&partnerID=40&md5=a429aeaeff8be37ead5ce8d143e1b05a.

[20] S. Çolak, A. Lima, and M.C. González. “Understanding congested travel in urban areas”. In: *Nature Communications* 7 (2016). DOI: 10.1038/ncomms10793. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84962624972&doi=10.1038%2fncomms10793&partnerID=40&md5=fe75d867304007041fcd0702ef9

[21] Serdar Çolak et al. “Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities”. In: *Transportation Research Record* 2526.1 (2015), pp. 126–135.

[22] V. Colizza et al. “Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions”. In: *PLoS Medicine* 4.1 (2007), pp. 0095–0110. DOI: 10.1371/journal.pmed.0040013. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-33846670320&doi=10.1371%2fjournal.pmed.0040013&partnerID=40&md5=05382f5a4700cd9be56a7e6731966eb8.

[23] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”. en. In: *Machine Learning* 20.3 (Sept. 1995), pp. 273–297. ISSN: 0885-6125, 1573-0565. DOI: 10.1007/BF00994018. URL: http://link.springer.com/10.1007/BF00994018 (visited on 09/18/2023).

[24] Y.-A. De Montjoye et al. “Unique in the Crowd: The privacy bounds of human mobility”. In: *Scientific Reports* 3 (2013). DOI: 10.1038/srep01376. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84875822474&doi=10.1038%2fsrep01376&partnerID=40&md5=e9216afca1ec2126f7b87354851bfd44.

[25] Hengfang Deng et al. “High-resolution human mobility data reveal race and wealth disparities in disaster evacuation patterns”. en. In: *Humanities and Social Sciences Communications* 8.1 (June 2021). Number: 1 Publisher: Palgrave, pp. 1–8. ISSN: 2662-9992. DOI: 10.1057/s41599-021-00824-8. URL: https://www.nature.com/articles/s41599-021-00824-8 (visited on 08/15/2022).

[26] *Department of Transportation, Census Transportation Planning Package (CTPP)*. 2000. URL: https://doi.org/10.21949/1518908 (visited on 12/01/2022).

[27] P. Deville et al. “Dynamic population mapping using mobile phone data”. In: *Proceedings of the National Academy of Sciences of the United States of America* 111.45 (2014), pp. 15888–15893. DOI: 10.1073/pnas.1408439111. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84909619343&doi=10.1073%2fpnas.1408439111&partnerID=40&md5=5dafbfdf35b49fd4a9c81dc80c82da

[28] B. Dewulf et al. "Dynamic assessment of exposure to air pollution using mobile phone data". In: *International Journal of Health Geographics* 15.1 (2016). DOI: 10.1186/s12942-016-0042-z. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84963812530&doi=10.1186%2fs12942-016-0042-z&partnerID=40&md5=37c85db4bed39df463390ae68758a855.

[29] Kirstin Dow and Susan L. Cutter. "Emerging Hurricane Evacuation Issues: Hurricane Floyd and South Carolina". EN. In: *Natural Hazards Review* 3.1 (Feb. 2002). Publisher: American Society of Civil Engineers, pp. 12–18. ISSN: 1527-6988. DOI: 10.1061/(ASCE)1527-6988(2002)3:1(12). URL: https://ascelibrary.org/doi/10.1061/%28ASCE%291527-6988%282002%293%3A1%2812%29 (visited on 08/15/2022).

[30] Sara B. Elagib, Aisha-Hassan A. Hashim, and R. F. Olanrewaju. "CDR analysis using Big Data technology". en. In: *2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)*. Khartoum, Sudan: IEEE, Sept. 2015, pp. 467–471. ISBN: 978-1-4673-7869-7. DOI: 10.1109/ICCNEEE.2015.7381414. URL: http://ieeexplore.ieee.org/document/7381414/ (visited on 09/18/2023).

[31] S. Eubank et al. "Modelling disease outbreaks in realistic urban social networks". In: *Nature* 429.6988 (2004), pp. 180–184. DOI: 10.1038/nature02541. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-2442686815&doi=10.1038%2fnature02541&partnerID=40&md5=a2f694cee3a874c91f11535ac7f747d

[32] Danielle L Ferreira, Bruno Astuto A Nunes, and Katia Obraczka. "Scale-free properties of human mobility and applications to intelligent transportation systems". In: *IEEE Transactions on Intelligent Transportation Systems* 19.11 (2018), pp. 3736–3748.

[33] Marshall Fixman. "Radius of gyration of polymer chains". In: *The Journal of Chemical Physics* 36.2 (1962), pp. 306–310.

[34] Manuel A Florez et al. "Measuring the impact of economic well being in commuting networks–A case study of Bogota, Colombia". In.

[35] Grace Fox et al. "Exploring the competing influences of privacy concerns and positive beliefs on citizen acceptance of contact tracing mobile applications". en. In: *Computers in Human Behavior* 121 (Aug. 2021), p. 106806. ISSN: 0747-5632. DOI: 10.1016/j.chb.2021.106806. URL: https://www.sciencedirect.com/science/article/pii/S0747563221001291 (visited on 08/16/2022).

[36] A. Galeazzi et al. "Human mobility in response to COVID-19 in France, Italy and UK". English. In: *Scientific Reports* 11.1 (2021). ISSN: 2045-2322. DOI: 10.1038/s41598-021-92399-2.

[37] Hemant Gehlot, Arif M. Sadri, and Satish V. Ukkusuri. "Joint modeling of evacuation departure and travel times in hurricanes". en. In: *Transportation* 46.6 (Dec. 2019), pp. 2419–2440. ISSN: 1572-9435. DOI: 10.1007/s11116-018-9958-4. URL: https://doi.org/10.1007/s11116-018-9958-4 (visited on 08/15/2022).

[38] T.C. Germann et al. “Mitigation strategies for pandemic influenza in the United States”. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.15 (2006), pp. 5935–5940. DOI: 10.1073/pnas.0601266103. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-33645802797&doi=10.1073%2fpnas.0601266103&partnerID=40&md5=a19f56c6c10c0cacdc1203a3c3fa5e

[39] J.R. Giles et al. “The duration of travel impacts the spatial dynamics of infectious diseases”. English. In: *Proceedings of the National Academy of Sciences of the United States of America* 117.36 (2020), pp. 22572–22579. ISSN: 0027-8424. DOI: 10.1073/pnas.1922663117.

[40] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. “Understanding individual human mobility patterns”. In: *nature* 453.7196 (2008), pp. 779–782.

[41] Marta C. González, César A. Hidalgo, and Albert-László Barabási. “Understanding individual human mobility patterns”. en. In: *Nature* 453.7196 (June 2008). Number: 7196 Publisher: Nature Publishing Group, pp. 779–782. ISSN: 1476-4687. DOI: 10.1038/nature06958. URL: https://www.nature.com/articles/nature06958 (visited on 08/13/2022).

[42] M.E. Halloran et al. “Modeling targeted layered containment of an influenza pandemic in the United States”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.12 (2008), pp. 4639–4644. DOI: 10.1073/pnas.0706849105. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-42449139659&doi=10.1073%2fpnas.0706849105&partnerID=40&md5=88a61a33bff11f07f5930058e9ee3ab2.

[43] Haosheng Huang, Yi Cheng, and Robert Weibel. “Transport mode detection based on mobile phone network data: A systematic review”. en. In: *Transportation Research Part C: Emerging Technologies* 101 (Apr. 2019), pp. 297–312. ISSN: 0968090X. DOI: 10.1016/j.trc.2019.02.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S0968090X1831369X (visited on 07/20/2023).

[44] W. Huang et al. “An exploration of the interaction between urban human activities and daily traffic conditions: A case study of Toronto, Canada”. English. In: *Cities* 84 (2019), pp. 8–22. ISSN: 0264-2751. DOI: 10.1016/j.cities.2018.07.001.

[45] Sibren Isaacman et al. “A tale of two cities”. In: *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications*. HotMobile ’10. New York, NY, USA: Association for Computing Machinery, Feb. 2010, pp. 19–24. ISBN: 978-1-4503-0005-6. DOI: 10.1145/1734583.1734589. URL: https://doi.org/10.1145/1734583.1734589 (visited on 08/15/2022).

[46] Shan Jiang, Joseph Ferreira, and Marta C Gonzalez. “Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore”. In: *IEEE Transactions on Big Data* 3.2 (2017), pp. 208–219.

[47] Shan Jiang et al. “The TimeGeo modeling framework for urban mobility without travel surveys”. In: *Proceedings of the National Academy of Sciences* 113.37 (2016). _eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1524261113, E5370–E5378. DOI: 10.1073/pnas.1524261113. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1524261113.

[48] F. Johari et al. “Urban building energy modeling: State of the art and future prospects”. English. In: *Renewable and Sustainable Energy Reviews* 128 (2020). ISSN: 1364-0321. DOI: 10.1016/j.rser.2020.109902.

[49] S. Jothilakshmi and V.N. Gudivada. “Large Scale Data Enabled Evolution of Spoken Language Research and Applications”. en. In: *Handbook of Statistics*. Vol. 35. Elsevier, 2016, pp. 301–340. ISBN: 978-0-444-63744-4. DOI: 10.1016/bs.host.2016.07.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169716116300463 (visited on 09/04/2023).

[50] A. Kaltenbrunner et al. “Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system”. In: *Pervasive and Mobile Computing* 6.4 (2010), pp. 455–466. DOI: 10.1016/j.pmcj.2010.07.002. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-77956402716&doi=10.1016%2fj.pmcj.2010.07.002&partnerID=40&md5=30982dda3da0d66c8b56501a568

[51] Y. Kang et al. “Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic”. English. In: *Scientific Data* 7.1 (2020). ISSN: 2052-4463. DOI: 10.1038/s41597-020-00734-5.

[52] Moritz U. G. Kraemer et al. “Mapping global variation in human mobility”. en. In: *Nature Human Behaviour* 4.8 (Aug. 2020). Number: 8 Publisher: Nature Publishing Group, pp. 800–810. ISSN: 2397-3374. DOI: 10.1038/s41562-020-0875-0. URL: https://www.nature.com/articles/s41562-020-0875-0 (visited on 08/13/2022).

[53] Moritz U. G. Kraemer et al. “The effect of human mobility and control measures on the COVID-19 epidemic in China”. In: *Science* 368.6490 (May 2020). Publisher: American Association for the Advancement of Science, pp. 493–497. DOI: 10.1126/science.abb4218. URL: https://www.science.org/doi/full/10.1126/science.abb4218 (visited on 08/16/2022).

[54] Shengjie Lai et al. “Global holiday datasets for understanding seasonal human mobility and population dynamics”. en. In: *Scientific Data* 9.1 (Jan. 2022). Number: 1 Publisher: Nature Publishing Group, p. 17. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01120-z. URL: https://www.nature.com/articles/s41597-022-01120-z (visited on 08/13/2022).

[55] Shirley Laska and Betty Hearn Morrow. “Social Vulnerabilities and Hurricane Katrina: An Unnatural Disaster in New Orleans”. In: *Marine Technology Society Journal* 40.4 (Dec. 2006), pp. 16–26. DOI: 10.4031/002533206787353123.

[56] W.D. Lee, M. Qian, and T. Schwanen. “The association between socioeconomic status and mobility reductions in the early stage of England’s COVID-19 epidemic”. In: *Health and Place* 69 (2021). DOI: 10.1016/j.healthplace.2021.102563. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103419728&doi=10.1016%2fj.healthplace.2021.102563&partnerID=40&md5=e35d43cad9de3ef4842c7e6071be7965.

[57] Y. Liu et al. “Urban land uses and traffic ’source-sink areas’: Evidence from GPS-enabled taxi data in Shanghai”. In: *Landscape and Urban Planning* 106.1 (2012), pp. 73–87. DOI: 10.1016/j.landurbplan.2012.02.012. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84862792974&doi=10.1016%2fj.landurbplan.2012.02.012&partnerID=40&md5=e6946ed2c6a7301a06056c488

[58] X. Lu, L. Bengtsson, and P. Holme. “Predictability of population displacement after the 2010 Haiti earthquake”. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.29 (2012), pp. 11576–11581. DOI: 10.1073/pnas.1203882109. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84863893880&doi=10.1073%2fpnas.1203882109&partnerID=40&md5=c1c6f03e2b5fbc1311a388bd80a2a198.

[59] Massimiliano Luca et al. “Modeling international mobility using roaming cell phone traces during COVID-19 pandemic”. en. In: *EPJ Data Science* 11.1 (Dec. 2022). Number: 1 Publisher: SpringerOpen, pp. 1–17. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-022-00335-9. URL: https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-022-00335-9 (visited on 08/16/2022).

[60] Feixiong Luo et al. “Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago”. In: *Applied Geography* 70 (2016), pp. 11–25. ISSN: 0143-6228. DOI: https://doi.org/10.1016/j.apgeog.2016.03.001. URL: https://www.sciencedirect.com/science/article/pii/S0143622816300194.

[61] J Macqueen. “SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS”. en. In: *MULTIVARIATE OBSERVATIONS* ().

[62] E. Massaro, D. Kondor, and C. Ratti. “Assessing the interplay between human mobility and mosquito borne diseases in urban environments”. In: *Scientific Reports* 9.1 (2019). DOI: 10.1038/s41598-019-53127-z. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85075193792&doi=10.1038%2fs41598-019-53127-z&partnerID=40&md5=6f7701efa6af15f4e2fdac91442edde2

[63] N. Mohammadi and J.E. Taylor. “Urban energy flux: Spatiotemporal fluctuations of building energy consumption and human mobility-driven prediction”. In: *Applied Energy* 195 (2017), pp. 810–818. DOI: 10.1016/j.apenergy.2017.03.044. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85016426259&doi=10.1016%2fj.apenergy.2017.03.044&partnerID=40&md5=d1736d3c9b1ac5cebf888b166af216a6.

[64] D. Monsivais et al. “Seasonal and geographical impact on human resting periods”. In: *Scientific Reports* 7.1 (2017). DOI: 10.1038/s41598-017-11125-z. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85028915844&doi=10.1038%2fs41598-017-11125-z&partnerID=40&md5=01cdf0f9f9c90e326083

[65] M. E. J. Newman. “Modularity and community structure in networks”. en. In: *Proceedings of the National Academy of Sciences* 103.23 (June 2006), pp. 8577–8582. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0601602103. URL: https://pnas.org/doi/full/10.1073/pnas.0601602103 (visited on 07/11/2023).

[66] M. E. J. Newman. “Power laws, Pareto distributions and Zipf’s law”. In: *Contemporary Physics* 46.5 (Sept. 2005). _eprint: cond-mat/0412004, pp. 323–351. DOI: 10.1080/00107510500052444.

[67] Nuria Oliver et al. “Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle”. In: *Science Advances* 6.23 (June 2020). Publisher: American Association for the Advancement of Science, eabc0764. DOI: 10.1126/sciadv.abc0764. URL: https://www.science.org/doi/10.1126/sciadv.abc0764 (visited on 08/15/2022).

[68] Luis E Olmos et al. “A data science framework for planning the growth of bicycle infrastructures”. In: *Transportation research part C: emerging technologies* 115 (2020), p. 102640.

[69] John R. B. Palmer et al. “New Approaches to Human Mobility: Using Mobile Phones for Demographic Research”. In: *Demography* 50.3 (Nov. 2012), pp. 1105–1128. ISSN: 0070-3370. DOI: 10.1007/s13524-012-0175-z. URL: https://doi.org/10.1007/s13524-012-0175-z (visited on 08/15/2022).

[70] C. Panigutti et al. “Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models”. In: *Royal Society Open Science* 4.5 (2017). DOI: 10.1098/rsos.160950. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85019731265&doi=10.1098%2frsos.160950&partnerID=40&md5=3905276290869f68edb7fff74fff6242.

[71] L. Pappalardo et al. “Understanding the patterns of car travel”. In: *European Physical Journal: Special Topics* 215.1 (2013), pp. 61–73. DOI: 10.1140/epjst/e2013-01715-5. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84873146100&doi=10.1140%2fepjst%2fe2013-01715-5&partnerID=40&md5=8080167b23f122123e75b6f43aa1d1e0.

[72] Luca Pappalardo et al. “Evaluation of home detection algorithms on mobile phone data using individual-level ground truth”. en. In: *EPJ Data Science* 10.1 (Dec. 2021), p. 29. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-021-00284-9. URL: https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-021-00284-9 (visited on 04/18/2023).

[73] Luca Pappalardo et al. “Evaluation of home detection algorithms on mobile phone data using individual-level ground truth”. In: *EPJ data science* 10.1 (2021), p. 29.

[74] Luca Pappalardo et al. “Returners and explorers dichotomy in human mobility”. en. In: *Nature Communications* 6.1 (Sept. 2015). Number: 1 Publisher: Nature Publishing Group, p. 8166. ISSN: 2041-1723. DOI: 10.1038/ncomms9166. URL: https://www.nature.com/articles/ncomms9166 (visited on 08/13/2022).

[75] Luca Pappalardo et al. “scikit-mobility: A Python library for the analysis, generation and risk assessment of mobility data”. In: *arXiv preprint arXiv:1907.07062* (2019).

[76] Luca Pappalardo et al. “Using big data to study the link between human mobility and socio-economic development”. In: *2015 IEEE International Conference on Big Data (Big Data)*. IEEE. 2015, pp. 871–878.

[77] T. Pei et al. “A new insight into land use classification based on aggregated mobile phone data”. In: *International Journal of Geographical Information Science* 28.9 (2014), pp. 1988–2007. DOI: 10.1080/13658816.2014.913794. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84907591712&doi=10.1080%2f13658816.2014.913794&partnerID=40&md5=da3ab234e81701d79141fd322129d5

[78] Daniela Perrotta et al. “Comparing sources of mobility for modelling the epidemic spread of Zika virus in Colombia”. en. In: *PLOS Neglected Tropical Diseases* 16.7 (July 2022). Publisher: Public Library of Science, e0010565. ISSN: 1935-2735. DOI: 10.1371/journal.pntd.0010565. URL: https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0010565 (visited on 08/16/2022).

[79] S. Phithakkitnukoon et al. “Activity-aware map: Identifying human daily activity pattern using mobile phone data”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6219 LNCS (2010), pp. 14–25. DOI: 10.1007/978-3-642-14715-9_3. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-77958483645&doi=10.1007%2f978-3-642-14715-9_3&partnerID=40&md5=9e23b53c5f16fad398afb641b7

[80] N. Pourebrahim et al. “Trip distribution modeling with Twitter data”. English. In: *Computers, Environment and Urban Systems* 77 (2019). ISSN: 0198-9715. DOI: 10.1016/j.compenvurbsys.2019.101354.

[81] R. Prieto Curiel et al. “Mobility between Colombian cities is predominantly repeat and return migration”. English. In: *Computers, Environment and Urban Systems* 94 (2022). ISSN: 0198-9715. DOI: 10.1016/j.compenvurbsys.2022.101774.

[82] C. Ratti et al. “Mobile landscapes: Using location data from cell phones for urban analysis”. In: *Environment and Planning B: Planning and Design* 33.5 (2006), pp. 727–748. DOI: 10.1068/b32047. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-33749679764&doi=10.1068%2fb32047&partnerID=40&md5=752f814d944b0708df23bc3a5255c784.

[83] Samuel Ribeiro-Navarrete, Jose Ramon Saura, and Daniel Palacios-Marqués. “Towards a new era of mass data collection: Assessing pandemic surveillance technologies to preserve user privacy”. In: *Technological Forecasting and Social Change* 167 (June 2021), p. 120681. ISSN: 0040-1625. DOI: 10.1016/j.techfore.2021.120681. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8019834/ (visited on 08/16/2022).

[84] Alex Ruiz-Euler et al. *Mobility Patterns and Income Distribution in Times of Crisis: U.S. Urban Centers During the COVID-19 Pandemic*. en. SSRN Scholarly Paper. Rochester, NY, Apr. 2020. DOI: 10.2139/ssrn.3572324. URL: https://papers.ssrn.com/abstract=3572324 (visited on 08/15/2022).

[85] Nick Warren Ruktanonchai et al. "Using Google Location History data to quantify fine-scale human mobility". In: *International Journal of Health Geographics* 17.1 (July 2018), p. 28. ISSN: 1476-072X. DOI: 10.1186/s12942-018-0150-z. URL: https://doi.org/10.1186/s12942-018-0150-z (visited on 08/13/2022).

[86] Paria Sadeghian, Johan Håkansson, and Xiaoyun Zhao. "Review and evaluation of methods in transport mode detection based on GPS tracking data". en. In: *Journal of Traffic and Transportation Engineering (English Edition)* 8.4 (Aug. 2021), pp. 467–482. ISSN: 20957564. DOI: 10.1016/j.jtte.2021.04.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S2095756421000623 (visited on 07/11/2023).

[87] T. Sakaki, M. Okazaki, and Y. Matsuo. "Earthquake shakes Twitter users: Real-time event detection by social sensors". In: *Proceedings of the 19th International Conference on World Wide Web, WWW '10*. 2010, pp. 851–860. DOI: 10.1145/1772690.1772777. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-77954571408&doi=10.1145%2f1772690.1772777&partnerID=40&md5=504c2bc057e15d6a0497a4b9b248476a.

[88] Christian M. Schneider et al. "Unravelling daily human mobility motifs". In: *Journal of The Royal Society Interface* 10.84 (July 2013). Publisher: Royal Society, p. 20130246. DOI: 10.1098/rsif.2013.0246. URL: https://royalsocietypublishing.org/doi/10.1098/rsif.2013.0246 (visited on 08/16/2022).

[89] M. Šimon et al. "Activity spaces of homeless men and women measured by GPS tracking data: A comparative analysis of Prague and Pilsen". English. In: *Cities* 86 (2019), pp. 145–153. ISSN: 0264-2751. DOI: 10.1016/j.cities.2018.09.011.

[90] K. Smolak et al. "Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models". English. In: *Urban Water Journal* 17.1 (2020), pp. 32–42. ISSN: 1573-062X. DOI: 10.1080/1573062X.2020.1734947.

[91] Chaoming Song et al. "Limits of Predictability in Human Mobility". In: *Science* 327.5968 (Feb. 2010). Publisher: American Association for the Advancement of Science, pp. 1018–1021. DOI: 10.1126/science.1177170. URL: https://www.science.org/doi/full/10.1126/science.1177170 (visited on 06/24/2022).

[92] Chaoming Song et al. "Modelling the scaling properties of human mobility". In: *Nature physics* 6.10 (2010). Publisher: Nature Publishing Group, pp. 818–823.

[93] Peter R. Stopher and Stephen P. Greaves. "Household travel surveys: Where are we going?" en. In: *Transportation Research Part A: Policy and Practice* 41.5 (June 2007), pp. 367–381. ISSN: 09658564. DOI: 10.1016/j.tra.2006.09.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0965856406001182 (visited on 10/04/2023).

[94] Emanuele Strano et al. “Mapping road network communities for guiding disease surveillance and control strategies”. en. In: *Scientific Reports* 8.1 (Mar. 2018). Number: 1 Publisher: Nature Publishing Group, p. 4744. ISSN: 2045-2322. DOI: 10.1038/s41598-018-22969-4. URL: https://www.nature.com/articles/s41598-018-22969-4 (visited on 08/13/2022).

[95] Zongshun Tian et al. “Characterizing the activity patterns of outdoor jogging using massive multi-aspect trajectory data”. en. In: *Computers, Environment and Urban Systems* 95 (July 2022), p. 101804. ISSN: 0198-9715. DOI: 10.1016/j.compenvurbsys.2022.101804. URL: https://www.sciencedirect.com/science/article/pii/S0198971522000485 (visited on 08/16/2022).

[96] J.L. Toole et al. “Inferring land use from mobile phone activity”. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2012, pp. 1–8. DOI: 10.1145/2346496.2346498. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84866008013&doi=10.1145%2f2346496.2346498&partnerID=40&md5=6b4784c464ccf4168e5fa167fdd968a1.

[97] J.L. Toole et al. “The path most traveled: Travel demand estimation using big data resources”. In: *Transportation Research Part C: Emerging Technologies* 58 (2015), pp. 162–177. DOI: 10.1016/j.trc.2015.04.022. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84940451094&doi=10.1016%2fj.trc.2015.04.022&partnerID=40&md5=c6523779c5361df522579fd49a7b21db.

[98] Uber. *H3: Uber’s Hexagonal Hierarchical Spatial Index*. Jan. 2022. URL: https://eng.uber.com/h3/.

[99] P. Valgañón et al. “Contagion-diffusion processes with recurrent mobility patterns of distinguishable agents”. English. In: *Chaos* 32.4 (2022). ISSN: 1054-1500. DOI: 10.1063/5.0085532.

[100] Maarten Vanhoof et al. “Detecting home locations from CDR data: introducing spatial uncertainty to the state-of-the-art”. In: *arXiv preprint arXiv:1808.06398* (2018).

[101] M.M. Vazifeh et al. “Optimizing the deployment of electric vehicle charging stations using pervasive mobility data”. In: *Transportation Research Part A: Policy and Practice* 121 (2019), pp. 75–91. DOI: 10.1016/j.tra.2019.01.002. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85059803171&doi=10.1016%2fj.tra.2019.01.002&partnerID=40&md5=6564879d6e27ca3206dd413da3f8

[102] N.-N. Wang et al. “Epidemic spreading with migration in networked metapopulation”. English. In: *Communications in Nonlinear Science and Numerical Simulation* 109 (2022). ISSN: 1007-5704. DOI: 10.1016/j.cnsns.2022.106260.

[103] P. Wang et al. “Understanding road usage patterns in urban areas”. In: *Scientific Reports* 2 (2012). DOI: 10.1038/srep01001. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84871806481&doi=10.1038%2fsrep01001&partnerID=40&md5=5184b3a9fbadffe8710ba34b1d6af605.

[104] Qi Wang et al. “Urban mobility and neighborhood isolation in America’s 50 largest cities”. In: *Proceedings of the National Academy of Sciences* 115.30 (July 2018). Publisher: Proceedings of the National Academy of Sciences, pp. 7735–7740. DOI: 10.1073/pnas.1802537115. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1802537115 (visited on 08/15/2022).

[105] B. Wellman and B. Leighton. “Networks, Neighborhoods, and Communities: Approaches to the Study of the Community Question”. In: *Urban Affairs Review* 14.3 (1979), pp. 363–390. DOI: 10.1177/107808747901400305. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0018697642&doi=10.1177%2f107808747901400305&partnerID=40&md5=58072c97030387a2c65ecd8c74992b64

[106] Amy Wesolowski et al. “The Use of Census Migration Data to Approximate Human Movement Patterns across Temporal Scales”. en. In: *PLoS ONE* 8.1 (Jan. 2013). Ed. by John P. Hart, e52971. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0052971. URL: https://dx.plos.org/10.1371/journal.pone.0052971 (visited on 10/04/2023).

[107] Jean Wolf, Marcelo Oliveira, and Miriam Thompson. “Impact of Underreporting on Mileage and Travel Time Estimates: Results from Global Positioning System-Enhanced Household Travel Survey”. en. In: *Transportation Research Record: Journal of the Transportation Research Board* 1854.1 (Jan. 2003), pp. 189–198. ISSN: 0361-1981, 2169-4052. DOI: 10.3141/1854-21. URL: http://journals.sagepub.com/doi/10.3141/1854-21 (visited on 10/04/2023).

[108] Longgang Xiang, Meng Gao, and Tao Wu. “Extracting Stops from Noisy Trajectories: A Sequence Oriented Clustering Approach”. en. In: *ISPRS International Journal of Geo-Information* 5.3 (Mar. 2016), p. 29. ISSN: 2220-9964. DOI: 10.3390/ijgi5030029. URL: http://www.mdpi.com/2220-9964/5/3/29 (visited on 06/28/2023).

[109] Rui Xin et al. “Impact of the COVID-19 pandemic on urban human mobility - A multiscale geospatial network analysis using New York bike-sharing data”. en. In: *Cities* 126 (July 2022), p. 103677. ISSN: 0264-2751. DOI: 10.1016/j.cities.2022.103677. URL: https://www.sciencedirect.com/science/article/pii/S0264275122001160 (visited on 08/16/2022).

[110] Y. Xu and M.C. González. “Collective benefits in traffic during mega events via the use of information technologies”. In: *Journal of the Royal Society Interface* 14.129 (2017). DOI: 10.1098/rsif.2016.1041. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85019208029&doi=10.1098%2frsif.2016.1041&partnerID=40&md5=7656286a0087fe71e5b35e04ab66a955.

[111] Y. Xu et al. “Human mobility and socioeconomic status: Analysis of Singapore and Boston”. In: *Computers, Environment and Urban Systems* 72 (2018), pp. 51–67. DOI: 10.1016/j.compenvurbsys.2018.04.001. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85046130493&doi=10.1016%2fj.compenvurbsys.2018.04.001&partnerID=40&md5=c5a9ec65db0d434b9ead72058f3

[112] Yanyan Xu, Riccardo Di Clemente, and Marta C. González. “Understanding vehicular routing behavior with location-based service data”. en. In: *EPJ Data Science* 10.1 (Feb. 2021), p. 12. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-021-00267-w. URL: https://doi.org/10.1140/epjds/s13688-021-00267-w (visited on 08/13/2022).

[113] Yanyan Xu et al. “Planning for electric vehicle needs by coupling charging profiles with urban mobility”. In: *Nature Energy* 3.6 (2018), pp. 484–493.

[114] Yanyan Xu et al. “Unraveling environmental justice in ambient PM2. 5 exposure in Beijing: A big data approach”. In: *Computers, Environment and Urban Systems* 75 (2019), pp. 12–21.

[115] Takahiro Yabe and Satish V. Ukkusuri. “Effects of income inequality on evacuation, reentry and segregation after disasters”. en. In: *Transportation Research Part D: Transport and Environment* 82 (May 2020), p. 102260. ISSN: 1361-9209. DOI: 10.1016/j.trd.2020.102260. URL: https://www.sciencedirect.com/science/article/pii/S1361920919311101 (visited on 08/16/2022).

[116] Takahiro Yabe, Satish V. Ukkusuri, and P. Suresh C. Rao. “Mobile phone data reveals the importance of pre-disaster inter-city social ties for recovery after Hurricane Maria”. en. In: *Applied Network Science* 4.1 (Dec. 2019). Number: 1 Publisher: SpringerOpen, pp. 1–18. ISSN: 2364-8228. DOI: 10.1007/s41109-019-0221-5. URL: https://appliednetsci.springeropen.com/articles/10.1007/s41109-019-0221-5 (visited on 08/15/2022).

[117] Xiaodong Yang et al. “Does the development of the internet contribute to air pollution control in China? Mechanism discussion and empirical test”. In: *Structural Change and Economic Dynamics* 56 (2021), pp. 207–224.

[118] Y. Yang et al. “Potential of low-frequency automated vehicle location data for monitoring and control of bus performance”. In: *Transportation Research Record* 2351 (2013), pp. 54–64. DOI: 10.3141/2351-07. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84919723948&doi=10.3141%2f2351-07&partnerID=40&md5=75921fb5c78ef9eca13d18c8f9ee8b78.

[119] Junjun Yin et al. “Depicting urban boundaries from a mobility network of spatial interactions: a case study of Great Britain with geo-located Twitter data”. In: *International Journal of Geographical Information Science* 31.7 (2017). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/13658816.2017.1282615, pp. 1293–1313. DOI: 10.1080/13658816.2017.1282615. URL: https://doi.org/10.1080/13658816.2017.1282615.

[120] H. Zang and J. Bolot. “Anonymization of location data does not work: A large-scale measurement study”. In: *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*. 2011, pp. 145–156. DOI: 10.1145/2030613.2030630. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-80053584816&doi=10.1145%2f2030613.2030630&partnerID=40&md5=2efac161a232b382cb14152f3bb2881e.

[121] F. Zhang et al. “Assessment of urban human mobility perturbation under extreme weather events: A case study in Nanjing, China”. English. In: *Sustainable Cities and Society* 50 (2019). ISSN: 2210-6707. DOI: 10.1016/j.scs.2019.101671.

14 Glossary of terms, abbreviations, and symbols

Application Programming Interface (API) A set of protocols that allows access to software or databases with a programming language. 12

Call Detail Records (CDR) Electronic records that contain information about cellphone calls, including the location of the cell phone tower that the call is conducted through. 10

Coronavirus Disease 2019 (COVID-19) A contagious disease caused by the Severe Acute Respiratory Syndrome Coronavirus 2. 10

Gaussian Mixture Model (GMM) A clustering method that represents complex data distributions with a set of Gaussian distributions. 12

Location-Based Services (LBS) Electronic records that contain location-related information, usually generated through mobile phone applications. 10

Radius of Gyration (r_g) A metric that measures the spread of a user's travel patterns. 10, 18

Shelter-in-Place (SIP) Shelter-in-place order was announced on 3/19/2020 by Governor Gavin Newsom, which requires all residents of the state of California to shelter in place for all but essential activities. 10

Support Vector Machine (SVM) A supervised machine learning algorithm used for regression and classification. 20

Vehicle Miles Traveled (VMT) An indicator of vehicle usage. 10

15 Appendix I The Human Mobility White Paper: Human Mobility Data in the 21st Century

Understanding human mobility means understanding how and when people move from location to location in their everyday lives. In the past decade or so, advances in technology have opened up new opportunities for understanding human mobility through novel data sets. These data sets, such as call detail records (CDRs) [30] from cell phone providers and location histories collected by smartphone apps, referred as Location Based Services (LBS) [14], allow researchers to observe individuals on a new scale. The advent of these large data sets requires us to empirically derive new metrics, algorithms, and even models to understand and capture how human behavior relates to human mobility. Improved models for human mobility have far reaching impacts, from models of disease spread [94] to transportation demand modeling [112], or to understanding behavior in natural disasters [55].

Here, we review several major discoveries in human mobility research from the last 15 years, and the novel data sets that enabled these breakthroughs. We discuss the universality of these discoveries, showing that they can be replicated with two novel data sources using different resolutions and sampling methods. We briefly discuss some of the numerous applications of human mobility research, and we draw from these recent discoveries and insights to identify future directions and challenges in the field.

In order to discuss the major discoveries of the past decade, it is helpful to first understand the empirical data sources used in human mobility research. The breakthroughs discussed in this paper were made possible by the availability of novel data sets, and validated using traditional sets used in previous research of the field. Before the advent of the novel data sets discussed in this paper, researchers only relied on actively collected data sets, like census data and travel surveys [93, 106, 107]. Censuses often include questions regarding home and work locations for respondents, which can be used to estimate commuting flows, and questions regarding previous and current residence locations, which can be used to estimate migration. In the United States, information about aggregated commuter flows is available through the Census Transportation Planning Package (CTPP) [26].

Other conventional data sources to study human mobility are local travel surveys or travel diaries. Travel surveys contain individual records of consecutive locations visited. These diaries often include self reported information on the time of the trip, the purpose of the trip, and mode of transportation at the individual level. Travel surveys offer very granular information, but are expensive to collect and often include records for a relatively small number of participants and time spans. Census records and travel surveys remain important in the field of human mobility, as they can be used as ground truth data to validate results from more novel data sources that are passively collected as the result of providing an information service as well as simulation models [88, 3, 47].

In recent years, the human mobility research community has demonstrated great success in using non-traditional data sets to estimate human mobility patterns. One of the first novel data sets to be used in human mobility research was from a currency tracking website. Because people carry and disperse banknotes when they make transactions, data sets of banknote movements encode information about human mobility [17].

Perhaps the biggest advance in human mobility data comes from the usage of call detail records. CDRs are collected and maintained by cellphone service providers. CDRs are charac-

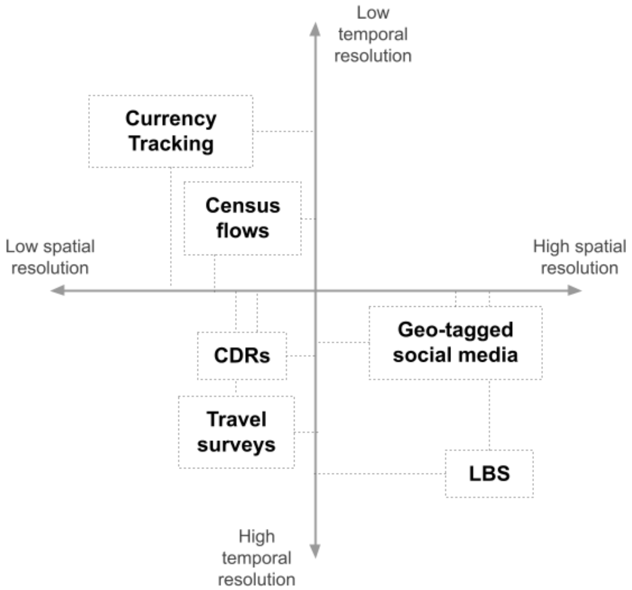


Fig 14: Relevant data sources and their spatial-temporal characteristics

terized by a unique user ID, a timestamp indicating the time of the activity, and a geo-locatable cell tower ID for every call, SMS message, or data session that a person conducts with their cell phone. These records can be used to estimate individual mobility patterns comparable to travel surveys and aggregate flows comparable to Census estimates. The advent of using CDRs to estimate individual mobility is responsible for many of the major breakthroughs described in this paper [2, 41, 88, 91, 92, 74].

Other novel data sets that have been shown to be useful for estimating human mobility include geo-tagged social media posts and location-based services (LBS) data. Both of these data sets are made possible by the widespread use of GPS enabled smartphones. Some social media websites, like Flickr, Foursquare, and Twitter, allow users to geo-tag their posts. These geo-tags, along with their corresponding timestamps, can be used to construct mobility patterns for social media users using their publicly available posts [1, 104, 80]. Mobility data from social media sites can be coupled with demographic data (e.g., census), using neighborhood demographics or via novel methods such as name analysis [60]. LBS data refers to a category of data collected from applications on smartphones that utilize a user's location to provide a specific service to the user. These data streams are collected by the app, associated with a unique user ID and timestamp. LBS data can be bought or licensed from individual apps or from data aggregators, who collect LBS data for individual users across multiple applications. One of the main advantages of LBS data is that they can provide high/accurate spatial and temporal resolutions, however, it can be unwieldy to use without a clear objective because of its large size (and thus, computational cost) and the presence of users with different/non-comparable temporal resolution data points.

The emergence of these large-scale data sources has presented huge opportunities to progress our understanding of human mobility, but using them presents multiple challenges. Rarely are mobility insights straightforward to extract, but rather, mobility information is embedded within data sets collected for other purposes. Additionally, the size of these data sets can easily become a challenge by itself, making basic operations such as data cleaning, statistical analyses, and data storage distinct challenges for researchers. Therefore, we often face a set of trade-offs between the ease with which researchers can work with the data, its manipulation costs, the temporal and spatial resolutions, its accessibility (in terms of cost and general availability), privacy concerns, accuracy of the samples, and the absolute amount of information about individual mobility encoded in the data. We illustrate how some of these data sets compare when evaluated for spatial and temporal resolution in Fig.14, providing a visual aid of the inherent trade-offs involved for researchers interested in human mobility.

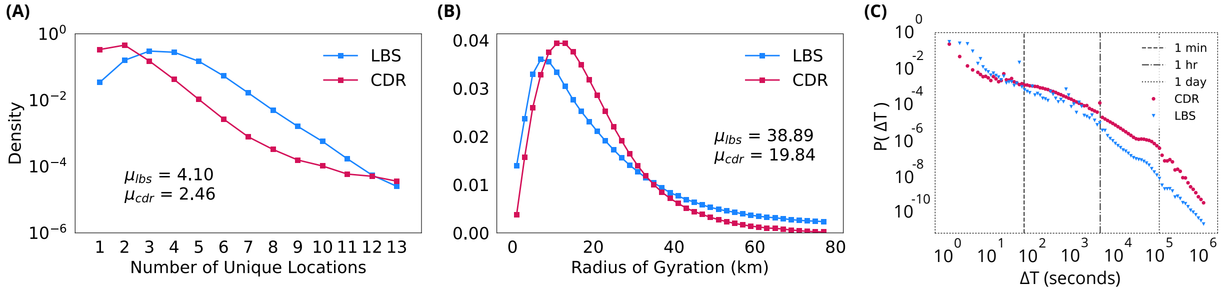
15.1 Universality of Human Mobility

As data from LBS and CDR are passively collected, it is not a direct task to create travel demand models from them. In both cases, it is necessary to identify and extract non-trivial patterns from the data, that would in turn allow us to reconstruct complete mobility profiles from the sparse traces of multiple individuals records. In this section, we compare a set of patterns of individual mobility, contrasting the results obtained by both CDR and LBS records.

15.1.1 Regularity of Human Mobility.

One of the first metrics of interest when using CDRs data sets was to capture the high degree of regularity in human mobility. To measure each individual's hinterland, or area of influence, González et al. [41] borrowed a quantity from polymer physics known as the radius of gyration (r_g) [33]. For each individual, it measures the average distance to their mean location (e.g., their center of mass). In essence, the radius of gyration is a measurement of the characteristic distance an individual travels during an observation period in the order of several months to a year. Calculating r_g and the number of unique locations visited by individuals (figure 15A and figure 15B), we observe differences between CDR and LBS data sets. We can attribute these differences to both data resolution and demographics. First, LBS data captures closer locations than its CDR counterpart (i.e., higher resolution), as antennas processing the signals operate on a specific region, frequently labeling multiple (close) locations under the same position. Second, CDR and LBS populations may differ in their demographics characteristics, as LBS data collects samples from multiple smartphone applications, normally associated with a younger population that could be more active in their daily mobility. Focusing on the time between calls/samples (figure 15C), we observe how both CDR and LBS data sets show a “bursty” pattern, with CDR data showing, on average, longer waiting times between events. This could be attributed to the data collection process. While CDR requires an active process (i.e., the individual calls), LBS data collection is passive and thus, tends to be collected at regular/most frequent intervals.

Using the radius of gyration, we can measure the likelihood of a user traveling a distance based on their radius of gyration (known as conditional jump lengths), and find that the distribution of jump lengths is actually truncated by r_g. This indicates that, once jump length probabilities are re-scaled by radius of gyration, the distributions collapse into a single curve, unveiling that there is a fixed relationship between characteristic distance traveled by individuals and their jump length distributions [40]. Figure 16A) recreates these findings using CDR and LBS data from the San Francisco Bay Area and the state of California, respectively. While the distribution of jump lengths reflects the distance of the chosen locations, we measure the probability of finding an individual in a randomly chosen location to add the temporal component. From this analysis, it is shown that it follows a regular pattern (periodicity) every 12 hours [41]. This pattern can be reproduced using



A) Distribution of the number of unique locations visited by individuals from CDR and LBS data sets. We observe a dominant exploration of LBS users. B) Distribution of radius of gyration from individuals' trajectories. We observe individuals from LBS data presenting a higher mean r_g. C) The time between consecutive calls (CDR) and passive data collection (LBS). We observe that most calls/data points are placed/collected soon after the previous one, with dominance of LBS. Only occasionally, we observe long periods without any activity (higher probability on CDR). Mean values (μ) are reported for both CDR and LBS data sets.

Fig 15: Radius of gyration, unique locations, and burstiness

both CDR and LBS data as shown in figure 16B), contrasting it with the diffusive movement of random walks for reference.

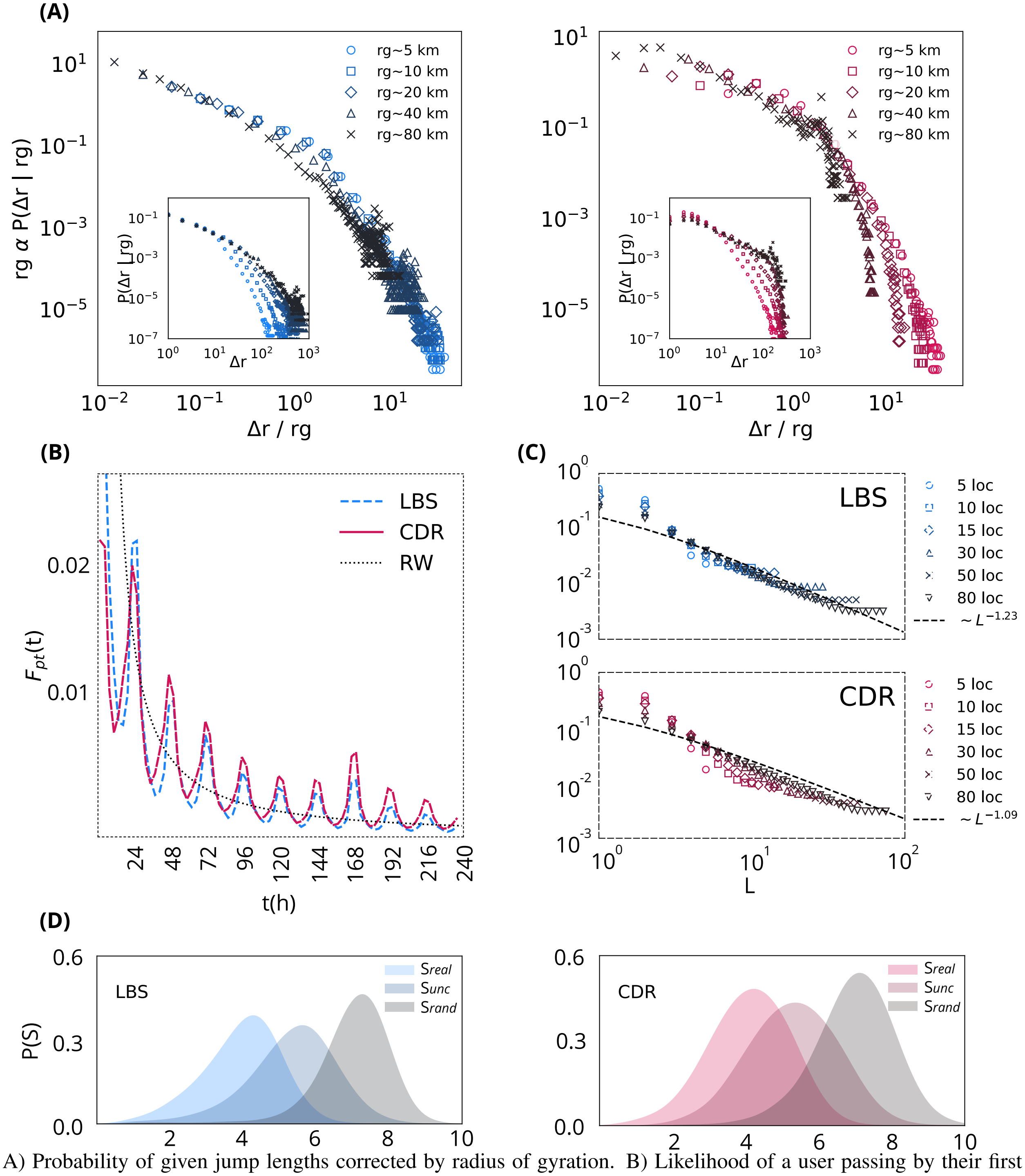
This is a direct result of people most likely visiting locations that they have already visited many times before. The probability of visiting a location based on its visit frequency rank (commonly denoted as L or K) closely follows a Zipf law [41, 92]. We recreated the probability distribution for visiting a location based on its visitation frequency rank, finding the expected Zipf Law relationship (figure 16C).

15.1.2 Predictability of Human Mobility.

One of the next major discoveries in the field was that human mobility is highly predictable, in other words, individuals tend to move in highly predictable trajectories. Drawing from the field of information science, Song et al. showed that human mobility is more predictable than a random variable [91]. Using CDR data, the authors calculated different measures of informational entropy for each individual. In this context, informational entropy is a measure of how "regular" a user's behavior is. By calculating random, uncorrelated (defined as the temporal uncorrelated entropy of a set of individuals), and "true" entropy (depending on both the visits frequency, order, and time spent at each location), Song et al. found that individual predictability is much higher than would be expected if individuals moved randomly. In this context, we have shown and confirmed this behavior using empirical observations from both CDR and LBS data, highlighting how individuals mobility patterns are in fact more predictable than random variables (figure 16D).

15.1.3 The Scaling Laws Human Mobility.

One of the first breakthroughs of the field in using novel data sources to estimate human mobility came from tracking banknotes in the US [17]. This study was able to model the mobility inferred from bank notes using a continuous time random walk model (CTRW) with two parameters: jump lengths and wait times. Jump length refers to the distance between two instances of the bank note's location being recorded. Wait times represent the time between instances of the bank note's location being recorded. From the study, the authors found that jump lengths for dollar bills



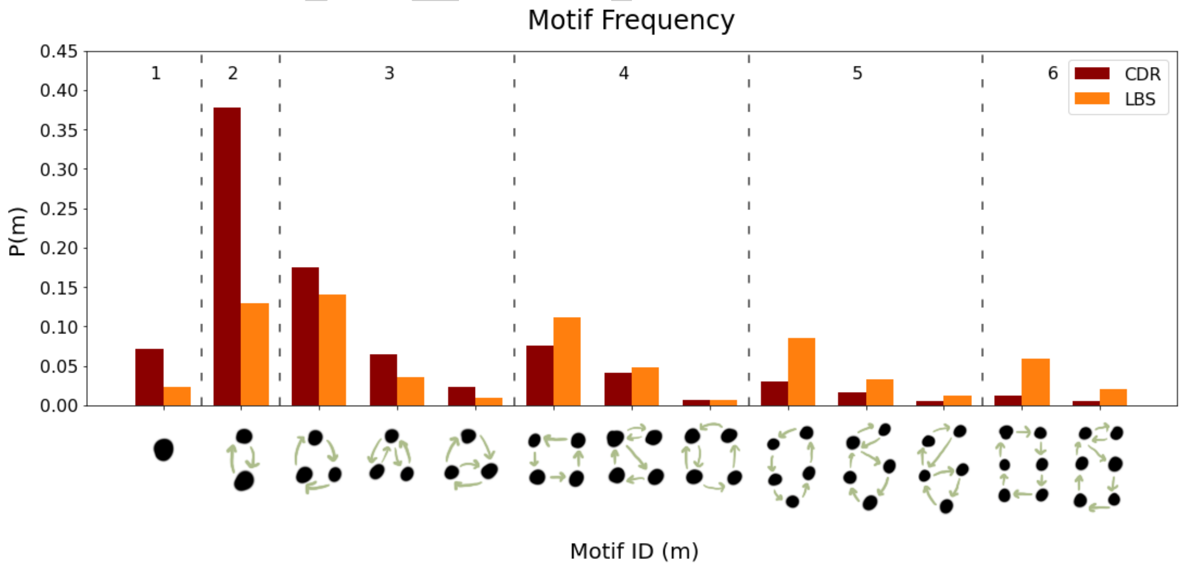
A) Probability of given jump lengths corrected by radius of gyration. B) Likelihood of a user passing by their first visited location at a given time t relative to their first observed location. C) Location rank (by frequency of visits) versus likelihood of finding a user at that location. D) Distributions of random (right curve), uncorrelated (middle curve), and “true” (left curve) entropy across all users in both data sets.

Fig 16: Comparison of universal patterns in human mobility using CDR and LBS data sets

follow a power law distribution. This same power law relationship was later reproduced for individuals using CDR data. We note that power law distributions are considered “scale-free” because they are the only distribution that does not have a typical scale [66].

While a CTRW model, using wait times and jump lengths as parameters, was shown to provide a reasonably good estimate for human mobility, it did not account for several empirically observed phenomena. These phenomena include: ultra-slow diffusion (people tend to return home, rather than following Brownian motion), and the Zipf law relationship between number of unique locations visited over time [8]. Song et al. [92] showed that the CTRW models were missing two important factors driving human mobility: exploration and preferential returns. Exploration is the observed behavior that people tend to visit fewer new locations over time [92]. Preferential returns, as discussed in Section 15.1.1, refer to the observed behavior that people tend to return to locations they have visited before. By including these two factors in their model, Song et al. were able to recreate empirical scaling laws and analytically predict scaling exponents [92].

In a more recent breakthrough, it has been discovered that the scale-free nature of human mobility patterns may arise in part from the nested spatial scales in which humans operate [2]. Prior empirical findings had found that jump lengths follow a power law distribution, implying that human mobility is geographically scale-free [17]. This was a counter-intuitive finding because people do tend to operate at meaningful spatial scales (e.g., at a neighborhood, city, province, or country scales) and also conceptualize space at relevant scales. The authors of Alessendreti et al. were able to unify these contradictory patterns/observations by developing a model that uses spatial “containers” to restrict mobility behavior [2]. In their research, they found that mobility within spatial containers tended to follow normal or log-normal distributions, but that by mixing these distributions at different scales, they could recreate the scale-free distribution found in numerous previous studies [32]. They used two GPS data sets to validate their model.



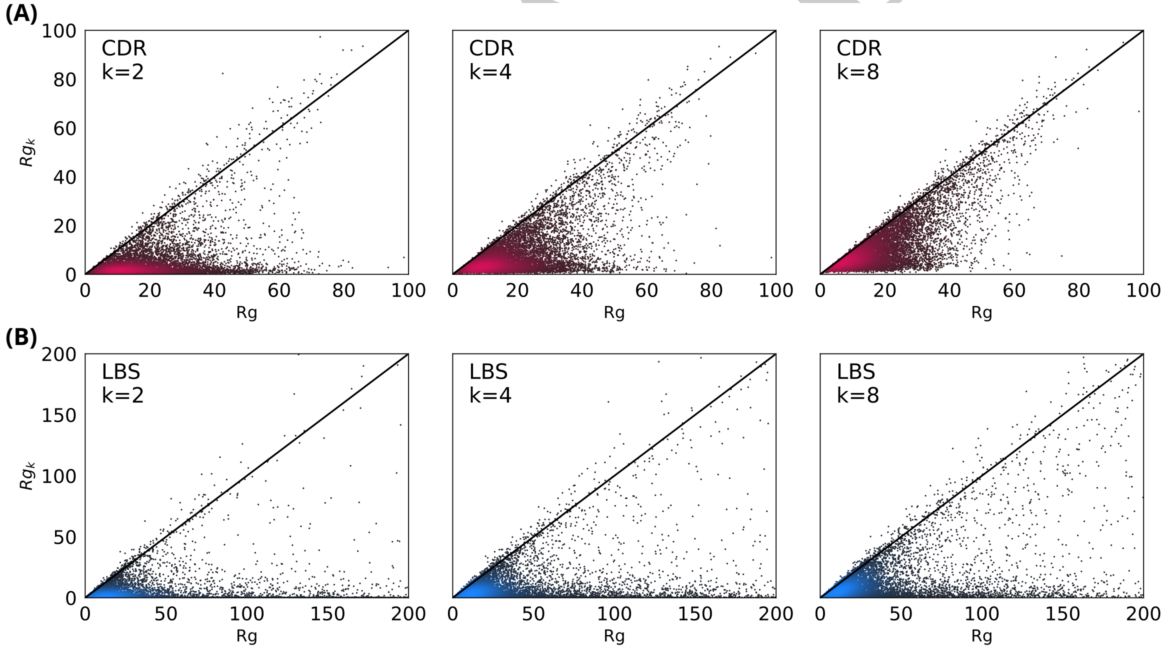
Daily mobility patterns are limited, with only 13 different motifs. The probability P(m) to find one of these motifs in both CDR and LBS is presented. The motifs are grouped according to their size separated by dashed lines. Most mobility motifs can be classified by complete tours and back and forward trips.

Fig 17: Mobility motifs and their probability

15.1.4 Individual Mobility Patterns.

Previously discussed findings concern aggregated patterns of several individuals spanning months of observations. An interesting body of work has emerged based on characterizing individual mobility on daily timescales. For example, Schneider et al., found that people exhibit a statistically small number of daily mobility patterns, termed "motifs" [88]. They found that just 17 unique networks could describe 90% of daily mobility networks found in empirical data, validating these results with both CDRs and travel surveys. Individuals tend to exhibit a characteristic mobility motif that remains stable over months of observation. In figure. 17, we depict a comparison of daily mobility motifs using both CDR and LBS data.

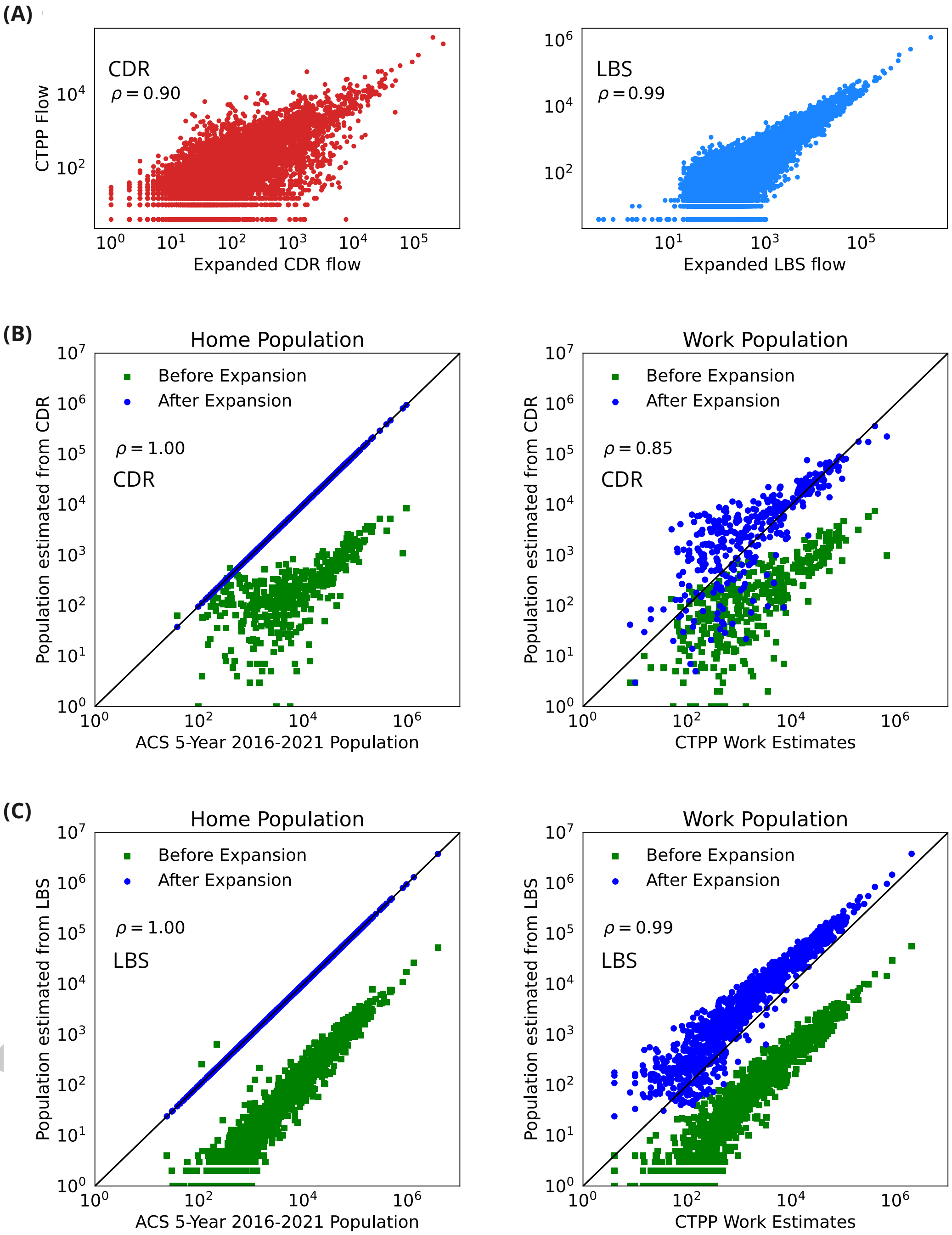
Another notable work concerns the classification of individuals based on their mobility behaviors. Using both CDR and GPS data sets, Pappalardo et al. were able to identify a dichotomy – clear distinction – between people that can be characterized as "returners" or "explorers" [74]. A returner is a person who's mobility can be summarized using a few of their locations, meaning their mobility can be characterized by a limited subset of their most frequently visited locations (e.g., recurrent home-to-work patterns). On the other hand, an explorer is someone whose mobility cannot be summarized by using only their most frequently visited locations; instead, their mobility tends to be more evenly distributed across many locations. Simply put, returners visit a few fixed locations, while explorers go to many locations, being less predictable.



We show the true radius of gyration (Rg) vs the k-radius of gyration (Rg_k) for CDR (A) and LBS (B) data sets for $k \in \{2, 4, 8\}$.

Fig 18: Returner vs. Explorer dichotomy

This dichotomy was captured by developing a metric called k-radius of gyration [74]. The k-radius of gyration expands on the prior discovery of radius of gyration as a relevant descriptor of human mobility (Section 15.1.1), but examines the characteristic distance traveled when considering an individual's top k locations. At a given value for k, individuals can be described as either "explorers" or "returners" based on the relationship between their k-radius of gyration and their



Comparison of CDR and LBS expanded flows, residence, and worker populations versus traditional data sources estimations: the 2016-2021 American Community Survey and CTTP. All estimates are at the census designated place scale. (A) Overall expanded flows versus reported CTPP flows comparison for CDR (left) and LBS (right) data sets. We observe correlations greater than 0.9 in both cases. (B) Residence population estimates versus population estimated from CDR data (left). Number of workers estimated by CTTP versus workers estimated from CDR data (right). (C) Residence population estimates versus population estimated from LBS data (left). Number of workers estimated versus workers estimated from LBS data (right).

Fig 19: Flows validation

true radius of gyration, which includes all their visited locations within the timespan of interest. We were able to recreate this dichotomy between individuals mobility patterns using both LBS and CDR data (Fig. 19). As expected, more explorations are observed in the LBS data due to the higher resolution of the observations and the potential demographic characteristics of the main users of the smartphone apps.

15.2 Validating Flows

Considerable work has been done in the past years to ensure that these novel data sets contain mobility information that is comparable to traditional data sets used in those studies, like census data and travel surveys. In this context, Alexander et al. showed that CDRs can be used to estimate commuter flows with a high degree of accuracy in Massachusetts, validating their results with data from the US Census and the US Census Transportation Planning Package [3]. Similar validations were done in diverse cities from the U.S., Portugal, Brazil, and Colombia [97, 34, 21]. We have recreated these methodologies and shown that similar results can be achieved in California using both CDR and LBS expanded data (Fig. 19). On an individual level, Jiang et al. showed that CDRs can be used to approximate results from travel surveys using the TimeGeo modeling framework [47].

15.3 Applications

The applications of improved data sources like CDRs and LBS for an understanding of human mobility are far reaching. In this review, we will focus our discussion on three general areas of applications with high impact outcomes: i) general public health, disease spread modeling, and COVID-19 pandemic analysis; ii) evacuation and disaster relief; and iii) urban planning and transportation modeling.

15.3.1 Public Health, Disease Spread Modeling, and COVID-19.

It is generally accepted that novel and better understanding of empirical human mobility leads to improvements of disease spread models [13, 38]. Mobility data can be used to model physical contacts for understanding and controlling outbreaks [31]. Recurrent mobility patterns can be used to improve disease spread models [99], and mobility-based metrics, such as trip duration, have been shown to inform disease spread modeling [39]. While there is a complex relationship between migration and disease spread [102], uncovering and understanding the dynamics of air travel has proven to be crucial in characterizing disease spread, obtaining valuable insights [22]. Novel data sources allow for mobility to be analyzed at many scales, from individual to country-to-country flows. This multi-scale characteristic provides the kind of mobility estimates and understanding required to further enrich and improve disease spread models [7].

Understanding the impact of social distancing measures has shown to be important for understanding disease spread and estimating disease containment [42]. In recent years, the COVID-19 pandemic has sparked particular interest in understanding the relationships between mobility restrictions and contagion spread, leading to multiple studies. Researchers have been able to analyze the effects of travel restrictions on COVID-19 spread using novel mobility data sources [53]. As an example, LBS data have been used to create granular mobility flow observations during the COVID-19 pandemic [51], allowing for analysis on multiple spatial scales. Similarly, CDRs have been used to model international travel and regional flows during the COVID-19 pandemic [59]. Moreover, mobility patterns inferred from Facebook interaction data sets have been used to quantify the impact of imposed COVID-19 travel restrictions [36].

In the context of the COVID-19 pandemic, it was observed that mobility reductions during lock downs differed based on socioeconomic groups in the U.S. [84]. Similarly in England, socioeconomic status was found to influence individual mobility reduction in the early stages of the COVID-19 pandemic, with the magnitude of the relationship varying depending on the region [56]. In the same line, we find another example using bike-share data in the city of New York, showing that the COVID-19 pandemic not only impacted macro-mobility patterns but had a significant impact on micro-mobility as well, as bike share usage decreased significantly in 2019-2020 [109].

Focusing on the general disease spreading phenomenon, mobile phone data showed to be especially useful for modeling disease spread in populous, urban areas [70]. Human mobility has been shown to be a critical factor in inter-urban vector-borne disease spread, like dengue, which is spread by mosquitoes [62]. In this context, the application of CDRs data have been shown to be useful for predicting Zika outbreaks in Colombia [78]. Overall, multiple studies have demonstrated that mobile phone data helps to create actionable data sets in pandemic management, if properly leveraged [67].

In addition to applications in disease spread modeling, human mobility data has relevant applications in the broader field of public health. Here, high resolution mobility trajectories can be leveraged to estimate, for example, individual exposure to environmental hazards, like air pollution [28, 117, 114]. Another effective application of mobility patterns inferred from mobile phone data consists of thesuccessful monitoring of mood, one of the most important indicators for mental health [18]. Mobility data can be also used to understand health-promoting behaviors, like jogging [95], and the widespread availability of these data make it possible to develop large-scale and accurate comparative studies. One of these studies uses minute-by-minute data from a step counter app to measure physical activity. Then, the authors performed an analysis to reveal which factors impact physical activity level across 46 different countries [4].

15.3.2 Emergency Evacuations and Natural Disasters.

Social ties, travel duration, and departure times are all important factors in understanding disaster evacuation behaviors [37]. In addition, understanding routing choices is important for evacuation planning [29]. Moreover, neighborhood demographics can influence natural disaster outcomes and even future mitigation policies [55].

In this context, we find a series of relevant studies exploiting mobility data to draw significant insights and useful support tools to deal with these challenging situations. For example, both CDRs and Twitter data sets can be used to identify disaster events in real-time [87], or near real-time [15], making it possible to generate a constant stream of up-to-date reports to guide people during such events. CDRs make it possible to compare mobility in emergency scenarios to non-emergency crowd scenarios, such as large social events [6]. Novel mobility data allows researchers to compare the impacts of extreme weather events, such as snowstorms and rainstorms, on demand for different modes of transportation [121]. On a longer planning horizon view, mobility data inferred from mobile phones can be used to model displacement post-disaster, such as after the 2010 Haiti earthquake [58] and after Hurricane Harvey in 2017 in Houston, Texas [25].

The capacity for mobility data to be coupled with demographic characteristics has allowed for critical post-disaster analysis on a very large scale. High resolution mobility data has revealed disparities in disaster evacuation patterns along racial and wealth lines, with more white, wealthier people being more likely to evacuate and have more cohesion in their evacuation destinations than their poorer counterparts [25]. In Florida, after Hurricane Irma in 2017, researchers found that evacuees with higher income were more likely to evacuate, reach safer locations, and suffer less damage on their housing and infrastructure by studying mobile phone records [115]. Similarly, higher social connectivity (i.e., a denser support network), or higher numbers of people moving in and out of counties in Puerto Rico, were shown to correlate to a faster recovery from Hurricane Maria devastation in 2017 [116]. From these studies, we note how important insights for informing municipalities in their pre-disaster planning, and post disaster recovery plans can be obtained by understanding and exploiting mobility data.

15.3.3 Urban Planning and Transportation Modeling.

Some of the most studied applications for human mobility research come from the desire to understand how people interact with cities [32]. Therefore, it is not a surprise that cell phone derived mobility data, such as CDRs and LBS, have been deeply used in urban planning research [82].

The applications of mobility research in urban studies are far reaching. For example, mobility data have been used to empirically identify neighborhood boundaries [105, 119]. CDRs and other mobility data sets can be used to infer different land use types [96, 57, 77]. CDRs can be used to estimate laborshed, the areas in which city's workers live, and partyshed, the areas in which those who visit a city for social reasons [12]. Human mobility can be coupled with energy-related information like water consumption data sets to model short-term water demand [90], used in electric vehicle charging station planning [101, 113], and even applied to modeling energy demand [63, 48, 9]. Human migration mobility data analysis has been used in Colombia to study return migration, or migration to a city previously inhabited by an individual [81]. More recently, a study uses Google Location History data (i.e., LBS records) to uncover the relationship between hierarchical mobility and different urban metrics, like emissions, walkability, and health indicators [10].

Data derived from mobile phones, including mobility traces, have been shown to be useful in demographic research [69]. This is particularly useful when dealing with shorter timescales than traditional data sets (e.g., census) due to their constant recording nature and availability, allowing researchers to estimate and map population levels over shorter time intervals [27]. Novel mobility data collection methods can be also used to study traditionally hard to reach demographics, like the unhoused [89]. While it is tempting to draw universal conclusions about the relationship between mobility and socio-economic factors, these relationships have been shown to vary, as it is influenced by segregation, employment opportunities that are unique to specific geographic regions [76]. Still, some mobility metrics, such as number of activity locations, activity entropy, and travel diversity have been shown to be similar across socioeconomic classes across different countries, leading to obtain general insights across multiple population profiles and regions [111].

Moreover, different kinds of mobility data sets can be coupled to study more complex phenomena and patterns accurately, characterizing complex interacting systems. For example, CDRs can be used to estimate travel demand [97] and GPS trajectories of vehicles can be used to model transportation demand, by estimating car travel patterns [71]. LBS data can be used to recreate vehicle paths, and these paths can be associated with individuals [112]. CDRs can be used to generate activity-based models [11] and daily activity patterns [79, 46]. LBS data from bicycle share apps have been leveraged to understand and model micro-mobility demand [50] and similarly, CDR data have been researched to inform bike path planning [68].

Mobile phone data have been studied to understand traffic congestion [20]. Using mobility data, researchers have been able to identify the few driver sources contributing to the major congestion [103]. Novel mobility data sources, like social media derived mobility traces or CDRs, make it possible to analyze the impact of human activities (like social gatherings) on traffic congestion [44] and can be used to reduce congestion due to mega events, such as the 2016 Olympic Games in Rio de Janeiro [110]. Related, improved methods in using non-traditional data sets to estimate mobility allow for better public transportation management [118]. In addition to understanding congestion, mobility data have been used in previous studies focusing on the use of smartphones to sense driver behavior in near real-time, with the aim of monitoring for behaviors that increase the likelihood of traffic accidents [19].

15.4 Future Directions and Challenges

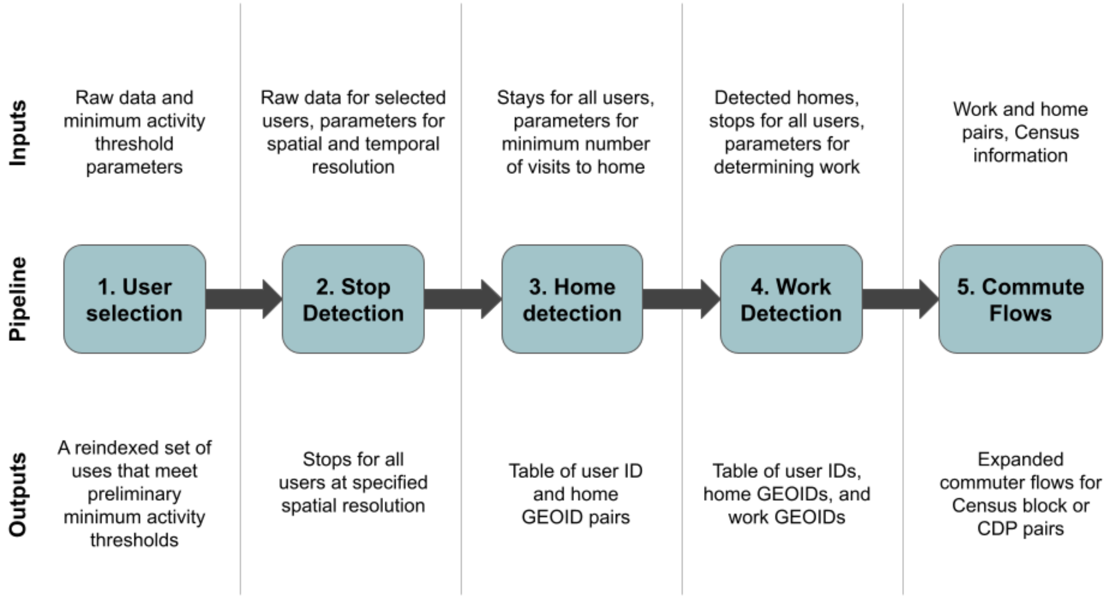
The widespread availability of these novel data sources make comparative studies between different neighborhoods, cities, and even countries more viable [45, 64]. Global data sets to aid in mobility research, such as global school and work holiday data sets, have been compiled [54], and LBS data, taken from Google Location History, has been used to generate high-quality global mobility data sets [85, 52]. Many studies have compared mobility traces in different regions, showing that while the empirical methods can be transferred, not all results are general/universal between different study areas. This has been shown, for example, by comparing CDRs from Portugal to those from Cote d'Ivoire [5], as well as by comparing commuters in Boston and Singapore [111], among other studies.

The advent of passively collected mobility data opens up new avenues for research, but it also presents new threats and issues. Human mobility traces have been shown to be highly unique, meaning that it is challenging to truly anonymize mobility data [24, 120]. And, while these mobility data sets can be a lens through which to study social issues, like racial segregation [16], it can also be used to monitor minority groups. The tension between the benefits of applications of mobile phone tracking and concerns about privacy has been particularly evident during the COVID-19 pandemic, with efforts such as contact tracing [35, 83].

15.5 Methods for the White Paper

15.5.1 Raw Data

The call detail records dataset (CDR) ranges temporally from December 2018 to July 2019, and spatially covers a major US metropolitan region. Our raw data comes from three sources: SMS logs, call logs, and data usage logs. Each record has an associated timestamp, cell phone tower ID closest to the location of the cell phone, and user ID (∼22.6M valid ids). Across these datasets, user IDs remain consistent, allowing for the three datasets to be combined to form a broader picture of an individual's mobility habits across different activities/channels. The location based services (LBS) ranges temporally from October 2018 to February 2019, and spatially covers a US state (California) that contains the metropolitan area covered by the CDR dataset. Each record contains a timestamp, user ID (∼6.3M unique ids), and a (latitude, longitude) coordinate pair.


Inputs
Raw data and minimum activity threshold parameters
Raw data for selected users, parameters for spatial and temporal resolution
Stays for all users, parameters for minimum number of visits to home
Detected homes, stops for all users, parameters for determining work
Work and home pairs, Census information
Pipeline
1. User selection
2. Stop Detection
3. Home detection
4. Work Detection
5. Commute Flows
Outputs
A reindexed set of uses that meet preliminary minimum activity thresholds
Stops for all users at specified spatial resolution
Table of user ID and home GEOID pairs
Table of user IDs, home GEOIDs, and work GEOIDs
Expanded commuter flows for Census block or CDP pairs

Generalized pipeline to produce commuter flows from raw data, showing the inputs and outputs at every step of the pipeline

Fig 20: Mobility data processing pipeline

15.5.2 User Selection

Because not all users in either dataset are present long enough or active enough to provide meaningful mobility insights, we begin by filtering users with a set of minimum activity thresholds. This is often an iterative empirical process to avoid introducing significant bias into the study, as too narrow of filters results in only the most active users, giving a limited view of the mobility patterns of the population, and too wide of filters results in a noisy data set with no clear patterns to analyze and learn from. Because of the large scale of these data sets, user selection is also an important step to avoid unnecessary processing steps later in the pipeline for inactive users. As seen in figure 20, this is the first step in the data processing pipeline.

Based on our initial data exploration, we observe that for CDR users, we require users to have at least one record every other day on average to be considered active enough for further (and meaningful) analysis. For LBS users, we required that users had an average of at least two records per day and be present in the data set for at least a third of all dates to be included in further analysis. Ultimately, we retained approximately 30% of the original users (∼6.8M and ∼1.9M for CDR and LBS, respectively), whose data accounted for approximately 85% of the raw records.

15.5.3 Radius of gyration

In the context of human mobility, the radius of gyration refers to a measure of the spatial extent of an individual's movements within a given area or region. Mathematically, it is a scalar value defined as the average distance of all the locations that a person visits from the centroid of an area in a specific time window of interest. To estimate it, researchers typically use data about the individual's location over a period of time t. These data could be, e.g., GPS, CDR, or LBS data. The calculation involves the following steps: i) Identify the centroid of the area of interest, such as a city or neighborhood; ii) For each time interval of interest (e.g., hours, days, weeks), calculate the distance between the individual's location and the centroid of the area; iii) Finally, we calculate the average distance of all the locations visited by the individual for a time period t. We note that we can repeat the calculation for multiple time periods t to obtain a more representative estimate of the individual's typical radius of gyration.

In this work, we calculate the radius of gyration $r_g^a(t)$ of each individual's trajectory up to time t following the approach of previous seminal works [41] (Eq. 1). We define $n_a(t)$:= the total number of positions recorded for an individual a up to time t; $r_i^{\vec{a}}$ the vector of dimension $n_a(t)$ of position $i \in \{1, ..., n_a(t)\}$ recorded for the individual up to time t; and $r_{cm}^{\vec{a}}$ the center of mass of the trajectory estimated by $r_{cm}^{\vec{a}} := 1/n_a(t) \sum_{i=1}^{n_a(t)} r_i^{\vec{a}}$.

$$r_g^a(t) = \sqrt{\frac{1}{n_a(t)} \sum_{i=1}^{n_a(t)} \left(r_i^{\vec{a}} - r_{cm}^{\vec{a}} \right)^2} \tag{1}$$

Similarly, the $k-$radius of gyration r_g^k can be defined as a extension of the original radius of gyration [74]. Here, k refers to the number of top k different places that a person visits during the time period t. For example, if $k = 1$, $r_g^{k=1}$ represents the average distance that a person travels from their most frequently visited place (e.g., their work location), considering all the places they visit during the time period t. Similarly, if $k = k^*$, then $r_g^{k^*}$ will consider only the k^* most frequently visited places. We calculate the $k-$radius of gyration $r_g^{a,k}(t)$ for an individual a up to time t by defining (Eq. 2): $n_a^k(t)$:= the total number of positions recorded for an individual a up to time t for their k most frequent locations; $r_i^{\vec{a,k}}$ the vector of dimension $n_a^k(t)$ of position $i \in \{1, ..., n_a^k(t)\}$ recorded for the individual up to time t; and $r_{cm}^{\vec{a,k}}$ the center of mass of the trajectory, only considering the visits to the k most frequent locations. Thus, replacing these expression in Eq. 1 we have:

$$r_g^{a,k}(t) = \sqrt{\frac{1}{n_a^k(t)} \sum_{i=1}^{n_a^k(t)} \left(r_i^{\vec{a,k}} - r_{cm}^{\vec{a,k}} \right)^2} \tag{2}$$

We note that r_g^k is a relative measure. It depends on the specific period of time t and the locations k considered. Therefore, it is important to choose a consistent time frame and definition of visited locations when comparing r_g^k values across different individuals/populations.

15.5.4 Entropy

We use the concept of entropy to quantify the predictability of a person's movements. It measures how evenly (or not) a person's visits are distributed among the locations they visit. In other words, how likely it is to predict the next location that a person will visit, based on their previous locations. Known also as "real entropy", is it calculated by the following expression [91]:

$$E_a^{real} = -\sum_{ST_a} \mathbf{P}(ST_a) \log_2 \left(\mathbf{P}(T_a^i) \right) \tag{3}$$

where T_a := the trajectory of individual a; ST_a := a particular time-ordered sub-trajectory of the original trajectory T_a; and $\mathbf{P}()$ the probability operator. From Eq. 3, we observe how the expression depends on both the frequency and the order in which the nodes are visited, as well as the time spent by the individual at each location.

Uncorrelated entropy [91] refers to the estimation of a temporal-uncorrelated entropy. Also known as Shannon entropy, it is a measure of the diversity of a person's visited locations without considering their temporal order. Therefore, a higher value of uncorrelated entropy indicates a more even distribution of visits across locations, and thus a higher degree of unpredictability. It is calculated as the negative sum of the products of the probability of an individual a visiting each location j, $\mathbf{P}(a, j)$, and the logarithm of that probability. $\mathbf{P}(a, j)$ is estimated from historical data, as the frequency of visits to each location divided by the total number of visits. Thus, we calculate E_a^{unc} using Eq. 4:

$$E_a^{unc} = -\sum_{j=1}^{n_a} \mathbf{P}(a, j) \log_2 \left(\mathbf{P}(a, j)\right) \tag{4}$$

with n_a the total number of distinct locations visited by individual a.

Finally, we calculate the random entropy for each individual trajectory [91]. In simple words, it measures the degree of randomness or unpredictability of the order in which a person visits locations. We calculate it as the logarithm (base two) of the total number of distinct visited locations by an individual a, n_a, capturing the degree of predictability if each location is visited with equal probability. Therefore, a higher value of random entropy indicates a more random and unpredictable sequence of visited locations. In Eq. 5 we can see the formula for E_a^{rand}:

$$E_a^{rand} = \log_2(n_a) \tag{5}$$

In summary, entropy measures the predictability of a person's movements, while uncorrelated entropy measures the diversity of the visited locations, and random entropy measures the randomness of the temporal order of the visits.

15.5.5 Stop Detection

Once the users that meet the activity thresholds have been identified, their raw location data must be converted into stops, or meaningful locations where (multiple) users spent a significant amount of time. There are many methods available in human mobility literature for performing stop detection [73, 100]. For this analysis, we chose to use a tessellation based approach, such as the one used in the known scikit-mobility Python package [75].

The tessellation approach requires all location data to be aggregated into a continuous tessellation of the Earth's surface over the coverage area. For the LBS data, we aggregated the raw (latitude, longitude) coordinates to a global hexagonal tessellation using the known H3 package [98]. In our experiments, we tested two resolutions levels characterized by an average hexagon area of 0.74 and 0.1 km^2; and an average hexagon edge length of 0.46 and 0.17 km, respectively. These correspond to 539,133 and 3,773,919 hexagons covering the state of California. Based on the results (less than $\sim$ 1% of difference) we keep the tessellation with the lower resolution, decreasing the computational burden. For the CDR data, we first geo-located the cell tower IDs for each record, combining nearby cell towers into a unique coordinate pair (e.g., multiple cells available in the same antenna). We then created a Voronoi tessellation over the coverage area using the cell tower coordinates that we used as our geographic scale.

Once we converted the geographic scale of our data to our selected tessellations and resolution, we began performing stop detection (see Alg. 1). Based on the data distribution, if a user remains in the same tessellation polygon for at least t minutes (20 minutes in our case), those

Algorithm 2 Stop detection and initial user selection

```
**Step 1: User selection**
**for** $user \in Users$ **do**                              ▷ Loops through each user in the raw data
    Count the timespan and the number of records
    **if** $timespan < days_{min}$ or $\# \; records < rec_{min}$, **then**
        Remove the user's records
    **end if**
**end for**
**return** selected users' data
**Step 2: Tessellation**
**for** $user \in Users$ **do**                         ▷ Loops through each user in the selected data
    **for** $item \in records$ **do**                          ▷ Loop through each item in user's records
        Replace $item$ locations with tessellation polygon IDs, $geoid$s
    **end for**
**end for**
**return** tessellated data
**Step 3: Stop Detection**
**for** $user \in Users$ **do**                         ▷ Loops through each user in the selected data
    **for** $item \in records$ **do**                          ▷ Loop through each item in user's records
        Calculate the time difference between the current and next item, $\Delta t$
        **if** $\Delta t \geq 20 \; min$ **then**
            Check to see if the user's location has changed
            **if** $geoid_i \neq geoid_{i+1}$ **then**
                Save trace as beginning of new stop
            **end if**
        **end if**
    **end for**
**end for**
**User's stops are ready.**
**return** stops for all users
```

records are aggregated and recorded as a stop location. All other records not satisfying this condition are discarded.

15.5.6 Home and Work Detection

Once the raw data have been converted to meaningful stops for the selected users, we detect the locations of those users' homes (see figure 20, step 3). Homes are defined as the most frequently visited location between a pre-defined time interval, commonly associated with sleeping time. Again, multiple methods have been proposed to estimate homes locations [75]. In this study, we focus on the observations between 8 p.m. and 7 a.m.. Users who did not visit their identified home at least once a week on average, or did not have any stops during night hours, were discarded.

Next, we infer the work locations of those individuals. For any given user, their work location is defined as the location that maximizes $n \times d$, where i) n is the number of visits to that location during weekdays; ii) d is the distance between that location and the identified home location; and iii) given that the location is at least 0.5 miles from the identified home, following the methodology presented in Alexander et al. [3]. Note that, under this methodology, not all users will have identified home and work locations. After applying this logic, we have i) a total of 1,908,961 and 1,818,116 users with identified home locations for CDR and LBS data sets, respectively; and ii) a total of 1,641,401 and 1,815,876 users with identified work locations for CDR and LBS data sets, respectively.

15.5.7 Estimating Commuter Flows and Expansion Factor.

Once we have established a (home, work) pair for most users in both data sets (i.e., CDR and LBS records), we can expand our data to population level using census data. Again, following the methodology presented in Alexander et al. [3], we count the number of users with homes inside each of our respective tessellation polygons (i.e., hexagons). We then join these counts to match the geography of our Census data (tract and Census-designated or CDP, for this analysis). Then, we calculate an expansion factor for each Census geographic unit for both the CDR and the LBS data sets, defined as the ratio of the Census population estimate to the number of residents estimated with our data set. Thanks to this factor, we can upscale our residence and worker estimates to be representative at total population levels, resulting in expanded flows. This is observed in figure 19, where we show the estimates for home and work locations both before and after expansion for both data sets as well as for overall flows. Finally, these expanded flows are used as inputs for travel demand models for the region of study.

Algorithm 3 Estimating Individual Commute Pattern

```
Step 1: Home Detection
for $user \in Users$ do                                ▷ Loops through each user in the stops
    Take only records between 8 p.m. and 7 a.m.
    Count number of visits to each location, determining most frequently visited nightly location, $geoid_{home}$, and the number of visits to that location, $n_{home}$
    if $n_{home} <$ once per week then
        Remove the user's records
    end if
end for
return selected data
Step 2: Work Detection
for $user \in Users$ do                                ▷ Loops through each user with identified home
    Take only records between 8 a.m. and 7 p.m. on weekdays
    for $geoid \in Geoids$ do                          ▷ Loops over unique locations visited by user
        Calculate distance $d_{geoid}$ between $geoid$ and user's home
        Calculate number of visits to this geoid, $n_{geoid}$
    end for
    Find $geoid_{work}$, that maximizes $n_{geoid} \times d_{geoid}$
    if $d_{work} <$ 0.5 miles OR $n_{work} <$ once a week then
        Discard $geoid_{work}$ for this user
    end if
end for
return $(geoid_{home}.geoid_{work})$ pairs for all selected users
```

16 Appendix II

16.1 Synthetic User Generation

We create synthetic users to validate the home change detection algorithm. We first define stationary users as users who keep the same home location for every month in 2020, meaning that their most frequently-visited location did not change between 7pm and 7am. We then sample 50,000 stationary users and select the ones that have at least 180 days in their time span, which is the difference, in number of days, between their last record and the first record available. We sort the users by the date of first appearance in their records and calculate the median date of records for each pair of users available. Let d_i be a set of the number of days since 1/1/2020 that user i initiates a trajectory. If the user has n trajectories in a single day, then the number of days since 1/1/2020 of this date is repeated n times. We define the median of the trajectories of user i as the median over the set d_i. For each pair of users (x, y), we define $min(median(d_x), median(d_y))$ as the *cutoff date*, by which we use to recombine the records. We delete all records of the user with the smaller $median(d_i)$ after the cutoff date and append all the records of the user with the larger $median(d_i)$ after this cutoff date. Then, the *cutoff date* is defined as the date of relocation for this newly synthesized user. The two home locations of the initial users are the home before and after relocation of the synthesized user. With this process, we created a sample of 15,413 synthetic users.

16.2 Coverage Area of Mobility-Related Initiatives

Below is a complete list of the census tracts considered in Sacramento county. 06 is the California state code and 067 is the Sacramento county code.

- 06.067.000400
- 06.067.000501
- 06.067.000502
- 06.067.000600
- 06.067.000700
- 06.067.000800
- 06.067.001201
- 06.067.001202
- 06.067.001300
- 06.067.001400
- 06.067.001900
- 06.067.002000
- 06.067.002100