



# Rural Mobility and Access: Leveraging Big Data Analytics and Context-Aware Computing

Final Report

Ying Song  
Department of Geography,  
Environment, and Society  
University of Minnesota

CTS 25-07



CENTER FOR  
TRANSPORTATION STUDIES  
UNIVERSITY OF MINNESOTA

## Technical Report Documentation Page

1. Report No. CTS 25-07	2.	3. Recipients Accession No.	
4. Title and Subtitle Rural Mobility and Access: Leveraging Big Data Analytics and Context-Aware Computing		5. Report Date July 2025	
		6.	
7. Author(s) Ying Song, Di Zhu, Xiaohuan Zeng, Meicheng Xiong		8. Performing Organization Report No.	
9. Performing Organization Name and Address Geography, Environment, and Society University of Minnesota 414 Social Science Building, 267 19th Ave S, Minneapolis, MN 55455		10. Project/Task/Work Unit No. # 2025008	
		11. Contract (C) or Grant (G) No.	
12. Sponsoring Organization Name and Address Center for Transportation Studies University of Minnesota 440 University Office Plaza 2221 University Avenue SE Minneapolis, MN 55414		13. Type of Report and Period Covered Final Report	
		14. Sponsoring Agency Code	
15. Supplementary Notes <a href="https://www.cts.umn.edu/publications/researchreports/">https://www.cts.umn.edu/publications/researchreports/</a>			
16. Abstract (Limit: 250 words) <p>This seed project investigates rural mobility and accessibility in Minnesota using mobile phone data and context-aware analytics. Traditional travel surveys, though detailed, often underrepresent rural populations and lack long-term, large-scale coverage. To address these gaps, this study leverages anonymized mobile phone data, in combination with land use and census datasets, to analyze the spatial behavior of rural, suburban, and urban residents.</p> <p>The project develops data-driven methods to classify users into behavior-based subgroups and identify key, group-specific routine activity locations such as home and workplace. Using the inferred home and work locations, the study evaluates the representativeness of the mobile phone sample and finds relatively high coverage in rural areas. Mobility indicators are then mapped and summarized, revealing distinct mobility patterns of rural residents—characterized by more dispersed activity spaces, less structured work schedules, and longer distances to frequent destinations. These findings underscore the value of mobile phone data in complementing travel surveys and offering a representative view of rural mobility and accessibility.</p> <p>In addition to methodological contributions, the project introduces an interactive mapping tool that allow users to visualize mobility flows across geographic scales, from statewide patterns to individual census tracts. These tools provide actionable insights for planners, researchers, and policymakers seeking to develop informed transportation strategies and investments to promote mobility and accessibility in rural communities.</p>			
17. Document Analysis/Descriptors Rural areas, Mobility, Smartphones, Location data, Data analysis, Accessibility		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Alexandria, Virginia 22312	
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 38	22. Price

# **RURAL MOBILITY AND ACCESS: LEVERAGING BIG DATA ANALYTICS AND CONTEXT-AWARE COMPUTING**

## **FINAL REPORT**

*Prepared by:*

Ying Song  
Di Zhu  
Xiaohuan Zeng  
Meicheng Xiong  
Department of Geography, Environment, and Society  
University of Minnesota – Twin Cities

## **JULY 2025**

*Published by:*

Center for Transportation Studies  
University of Minnesota  
440 University Office Plaza  
2221 University Avenue SE  
Minneapolis, MN 55414

This report represents the results of research conducted by the authors and does not necessarily represent the views or policies of the Center for Transportation Studies and/or the University of Minnesota. This report does not contain a standard or specified technique.

The authors, the Center for Transportation Studies, and the University of Minnesota do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to this report.

# TABLE OF CONTENTS

CHAPTER 2: Introduction .....	1
CHAPTER 3: Data Acquisition and Preprocessing .....	3
3.1 Mobile Phone Data .....	3
3.2 Land Use Data .....	3
3.3 Census Data .....	5
CHAPTER 4: Research Design .....	6
4.1 Framework.....	6
4.2 Derive Representative Activity Locations .....	7
4.3 User Segmentation and Activity Classification .....	8
4.4 Representativeness of Mobile Phone Data .....	10
CHAPTER 5: Results and Discussion .....	11
5.1 Representative Activity Locations .....	11
5.2 User Segmentation .....	11
5.3 Activity Classification .....	13
5.4 Representativeness Evaluation.....	23
5.5 Mobility and Accessibility Exploration.....	25
5.5.1 Visual Exploration Using Static Graphs and Maps.....	25
5.5.2 Visual Exploration Using Interactive Maps.....	30
CHAPTER 6: Summary of Key Findings.....	33

## LIST OF FIGURES

Figure 3.1 Research design with three key components .....	6
Figure 4.1 Percentage of devices with at least one cluster and the average number of clusters per device .....	11
Figure 4.2 Percentage of devices with at least one cluster and the average number of clusters per device .....	12
Figure 4.3 Activity clusters, their characteristics, and distances to nearest residential parcels .....	14
Figure 4.4 Activity clusters, their characteristics, and distances to nearest work-related parcels .....	19
Figure 4.5 Comparison of Population and Employment Distributions in PlaceIQ and Census data.....	24
Figure 4.6 Visualization of mobility flows by destination activity type .....	26
Figure 4.7 Comparison of the user group labels and the home location regions .....	28
Figure 4.8 Frequent Activities Within and Outside Residential Neighborhoods .....	29
Figure 4.9 Frequent Activities Within and Outside Residential Neighborhoods .....	29
Figure 4.10 Frequent Activities Within and Outside Residential Neighborhoods .....	30
Figure 4.11 Example view of the scale-free interactive map.....	31
Figure 4.12 Visualization of out-flows from a rural census tract and.....	32

## LIST OF TABLES

Table 4.1 Summary of user groups' key characteristics .....	13
Table 4.2 Summary of activity classification outcomes .....	23

## CHAPTER 1: INTRODUCTION

Access to transportation plays a key role in rural communities and addresses their need to access jobs, education, healthcare, and other activities of daily life. Recent projects such as *Rural Community Transit Strategies* and *Funding Shared Mobility as an Extension of Existing Public Transit Services* have identified the transportation needs and challenges of rural communities in Minnesota and explored strategies for promoting shared mobility services. To ensure the success of these strategies, it would be crucial to gain a comprehensive view of the current mobility and accessibility of rural residents.

Travel survey data have been collected for travel forecasting and accessibility evaluation. Despite the rich information collected in the travel surveys data, it has always been challenging to scale up spatially and temporally. Regarding spatial coverage, survey participants are mostly from urban and suburban areas, leading to a lack of data for rural areas. Regarding temporal coverage, participants are often asked to provide travel diaries for no more than a week, making it challenging to capture variations in mobility patterns in the longer term. For instance, the 2019 Travel Behavior Inventory (TBI) survey was fielded from October 1, 2018, through September 30, 2019, in Minnesota. Participants were asked to provide 7-day travel diaries using a smartphone application and 1-day travel diaries online or via the call center. The 7,837 participating households were mostly from the seven-county metropolitan area<sup>1</sup>.

Recent studies have started to integrate mobile phone data for human mobility analysis. They recognize that 97% of Americans, including kids, now own a cell phone, and 85% of U.S. adults own a smartphone (as of February 2021). Moreover, Americans spend approximately 3 hours on their phones each day on average. So, mobile phone data has great potential to provide timely views of human mobility patterns at large scales for longer periods. However, most studies have focused on the urban area and trip-based metrics such as origin-destination flow by mode. This makes it challenging to address issues such as trip-chaining behaviors. Moreover, the representativeness of the sample offered by the mobile phone data has been understudied since the identifiable information has been removed to protect user privacy. So, it is challenging to determine methods to assign weights to the sample and use it in practice.

Despite the lack of attention to sample representativeness, using mobile phone data to study human mobility has gained increasing attention in recent years. Notably, many studies have identified the need for context-aware methods to derive richer information than what can be obtained from mobile phone data alone, making the insights more applicable to planning practices. For instance, studies have used contextual data such as points of interest (POIs) and weather information to improve the detection of travel modes from GPS data. Therefore, integrating geographic context into the analysis of mobility and

---

<sup>1</sup> (Participating households per county) Hennepin 3,039; Ramsey 1,427; Dakota 752; Anoka 557; Washington 439; Wright 240; St. Croix 194; Scott 193; Sherburne 161; Carver 150; Rice 116; Chisago 100; Goodhue 91; Polk 90; Isanti 83; McLeod 81; Pierce 74; Le Sueur 47; Sibley 3. Source: <https://metrotransitm.nshinyapps.io/travel-survey-explorer/>

accessibility may provide novel insights into different mobility patterns and needs of residents across the geographical regions.

This project aims to address the gaps in investigating rural mobility and accessibility by comprehensively examining using mobile phone data and context-aware analytics. The three specific aims are:

- *Deriving spatial distributions of mobile phone users.* The project will develop context-aware data mining methods to label activities along individuals' daily trajectories (e.g., home and work) and use them to derive distributions of mobile phone users and infer their trip purposes.
- *Evaluating the sample representativeness of mobile phone data.* The project will compare spatial distributions of mobile phone users' home and work locations and spatial distributions of people and jobs recorded in census data, which will be used to evaluate the sample representativeness of mobile phone data.
- *Exploring mobility and accessibility across rural, suburban, and urban areas.* The project will first compare spatial accessibility to jobs, frequent, and occasional activities among rural, suburban, and urban residents at the person level. The project will then create static and interactive maps to allow customized explorations of mobility and accessibility at the zonal level.

This report is organized as follows: Chapter 2 summarizes the mobile phone data and context data used for the study; Chapter 3 presents the research design with procedures and methods to achieve the three specific aims; Chapter 4 presents analysis results and discusses their implications; and finally, Chapter 6 summarizes key findings.

## CHAPTER 2: DATA ACQUISITION AND PREPROCESSING

This chapter introduces the three sets of data used for the study. The first set is the Place IQ data, which contains one month of mobile phone data from July 2021. The second set is the land use data, including land parcel polygons and POI points. The third set is the census data, including the American Community Survey (ACS) and Longitudinal Employer-Household Dynamics (LEHD) data.

### 2.1 MOBILE PHONE DATA

The mobile phone data used in this project was purchased from PlacelQ, a location data and technology company that provides human mobility data and supports business analytics. PlacelQ mobility data is derived from anonymized device-level signals collected through mobile applications that integrate with operating system services. As users interact with their devices, locations and timestamps of their signals are passively recorded. PlacelQ aggregates and processes the raw data by i) clustering spatiotemporal signals into inferred “visits”, each corresponding to a sustained presence at a location, and ii) generating “trips” for movements between distinct visits.

This study uses PlacelQ data collected in Minnesota during July 2021. The original data contains derived trips and visits of each mobile device. Since trips are generated between two consecutive visits, we only use “visit” data in this study. For a sample day, 7,846,562 visits were recorded by 599,676 unique active devices, with an average positioning accuracy of 5.2 meters. For the entire month, 21 million visits were recorded by over 720k active unique devices.

Each recorded visit includes the following attributes:

- **Device Key:** A unique identifier for each mobile device generated by the system. This identifier is anonymized and not linked to any personal information, such as the Social Security Number (SSN).
- **Timestamp:** The time of the visit, recorded as a Unix Timestamp (a.k.a. Epoch time or POSIX time), which represents time as the number of seconds elapsed since January 1, 1970 at 00:00:00 UTC.
- **Duration:** The minimum duration of the visit, calculated based on the range of timestamps from all device signals associated with the visit.
- **Longitude and latitude:** The geographical coordinates of the visit, expressed in degrees under the WGS84 reference system. These coordinates are computed as the centroid of all recorded signal locations during the visit.

All recorded visits are retained for analysis without filtering based on their attributes. However, it is important to note that visits occurring outside of Minnesota are unavailable, even if the user is a Minnesota resident. Conversely, visits made within Minnesota are included regardless of the user's place of residence during the study period.

### 2.2 LAND USE DATA

We collect two datasets to represent land-use patterns to facilitate the context-aware analysis.

- **Point of Interest (POI) Data**

The first dataset is the Point of Interest (POI) data provided by OpenStreetMap (OSM). The OSM POIs represent real-world entities such as restaurants, schools, hospitals, and transportation facilities. Each POI is typically stored as a node (point) or way (line) with associated tags that describe its type and attributes (e.g., amenity=restaurant, name=Punch Pizza). The data is freely available under the Open Database License (ODbL), making it widely used in mapping, research, and planning practices. Unlike commercial services, OSM relies on a decentralized model where contributors manually add and update POI data using tools like JOSM or iD Editor. POI data from OSM can be accessed via exports, APIs, or third-party services, and is highly customizable due to its flexible tagging schema.

For this study, we retrieved OSM POI data in Minnesota, which includes a total of 1,123,691 amenities. The original OSM POI dataset contains approximately 130 officially documented amenity types, along with community-used or regional-specific variations. To facilitate analysis, we group these POIs into 11 generalized categories, including Business (27,077), Retail (18,374), Dining (10,158), Financial (7,090), Travel (6,200), Automotive (6,019), Educational (4,078), Entertainment (2,884), Transportation (1,797), Industrial (1,768), and Real Estate (1,402).

- **Land Use Parcel Data**

The second dataset is the land use parcel data, which contains detailed property records. The dataset includes parcel polygons, tax assessments, and other attributes related to land ownership and land use. In Minnesota, parcel data is managed at the county level, typically by the recorder's, assessor's, or land surveyor's offices. While some counties maintain digital records of land parcels and make them publicly accessible online, others offer web-based mapping tools that allow users to view property information or even rely on paper-based systems. The Minnesota Geospatial Information Office (MnGeo) provides a centralized list of counties with available digital parcel data, along with links to additional resources<sup>2</sup>. We used this list to retrieve land parcel data from individual counties. Out of Minnesota's 87 counties, we identified 53 counties that offer downloadable parcel datasets, and 37 counties have parcel datasets with ownership type and/or land use classification.

Compared to the OSM POI data, the land use parcel data provides broader coverage of land use types, including the distribution of residential properties alongside commercial, industrial, and other land use types. In this study, the parcel data is used to support the evaluation of activity types and trip purposes inferred from mobile phone data, specifically for detecting users' residential and workplace locations. Therefore, while handling variations in land use classification systems across Minnesota counties, we standardized the data by assigning two binary labels to each parcel: one indicating whether the parcel is used for residential purposes (including mixed-use residential); the other indicating whether the parcel can potentially serve as a workplace, excluding categories such as vacant or undeveloped properties.

---

<sup>2</sup> MnGeo, Land Ownership: Parcel: [https://www.mngeo.state.mn.us/chouse/land\\_own\\_property.html#](https://www.mngeo.state.mn.us/chouse/land_own_property.html#)

## 2.3 CENSUS DATA

We retrieve two census datasets representing population and job distributions in Minnesota to support the evaluation of the representativeness of mobile phone data samples.

- **American Community Survey (ACS)**

To align with the time frame of the PlacelQ data, we retrieve the 2017-2021 ACS five-year estimates to represent the population distribution in Minnesota. Since ACS data is collected and aggregated based on respondents' residential locations, it can serve as an appropriate benchmark for comparing with the inferred residential locations of mobile users to assess the representativeness of mobile phone data. We obtain the ACS data through the National Historical Geographical Information System (NHGIS)<sup>3</sup> platform. Based on an initial exploration of the PlacelQ data, we selected census tracts, rather than census block groups, as the spatial unit to mitigate the issue of block groups with sparse or no samples. Additionally, considering the typical age at which individuals own mobile devices, we used Table B01001 (Sex by Age) and restricted the analysis to individuals 15 and older. We also used Table B23025 (Employment Status) to compare the total employed population with the number of mobile phone users for whom workplace locations were identified.

- **Longitudinal Employer-Household Dynamics (LEHD)**

Similar to the ACS data, we retrieved the 2021 LEHD Origin-Destination Employment Statistics (LODES) at the census tract level. Specifically, we downloaded the Workplace Area Characteristics (WAC) dataset for Minnesota from the US Census Bureau's official website<sup>4</sup>. The WAC data provides information on the total number of jobs, as well as job distributions across industry sectors, age groups, and income levels. In this study, the WAC data serves as a complementary benchmark to evaluate the representativeness of PlacelQ data by comparing inferred work locations of users from an employment-based perspective.

To support spatial analysis and mapping, we downloaded the 2020 Census TIGER/LINE Shapefiles for census tracts and joined the ACS and WAC datasets to the tract polygons using shared tract identifiers.

---

<sup>3</sup> Steven Manson, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 19.0 [dataset]. Minneapolis, MN: IPUMS. 2024. <http://doi.org/10.18128/D050.V19.0>

<sup>4</sup> US Census Bureau, LEHD Data: <https://lehd.ces.census.gov/data/>

## CHAPTER 3: RESEARCH DESIGN

This chapter presents the research design with procedures and methods to infer activity types and trip purposes and evaluate the representativeness of mobile phone data.

### 3.1 FRAMEWORK

Figure 3.1 illustrates the research design with three major components following the three project aims: 1) labeling activities along individuals' daily trajectories recorded in the PlacelQ data, 2) evaluating the representativeness of the mobile phone data from residential- and employment-based perspectives, and 3) exploring the distinct mobility patterns of rural, suburban, and urban residents.

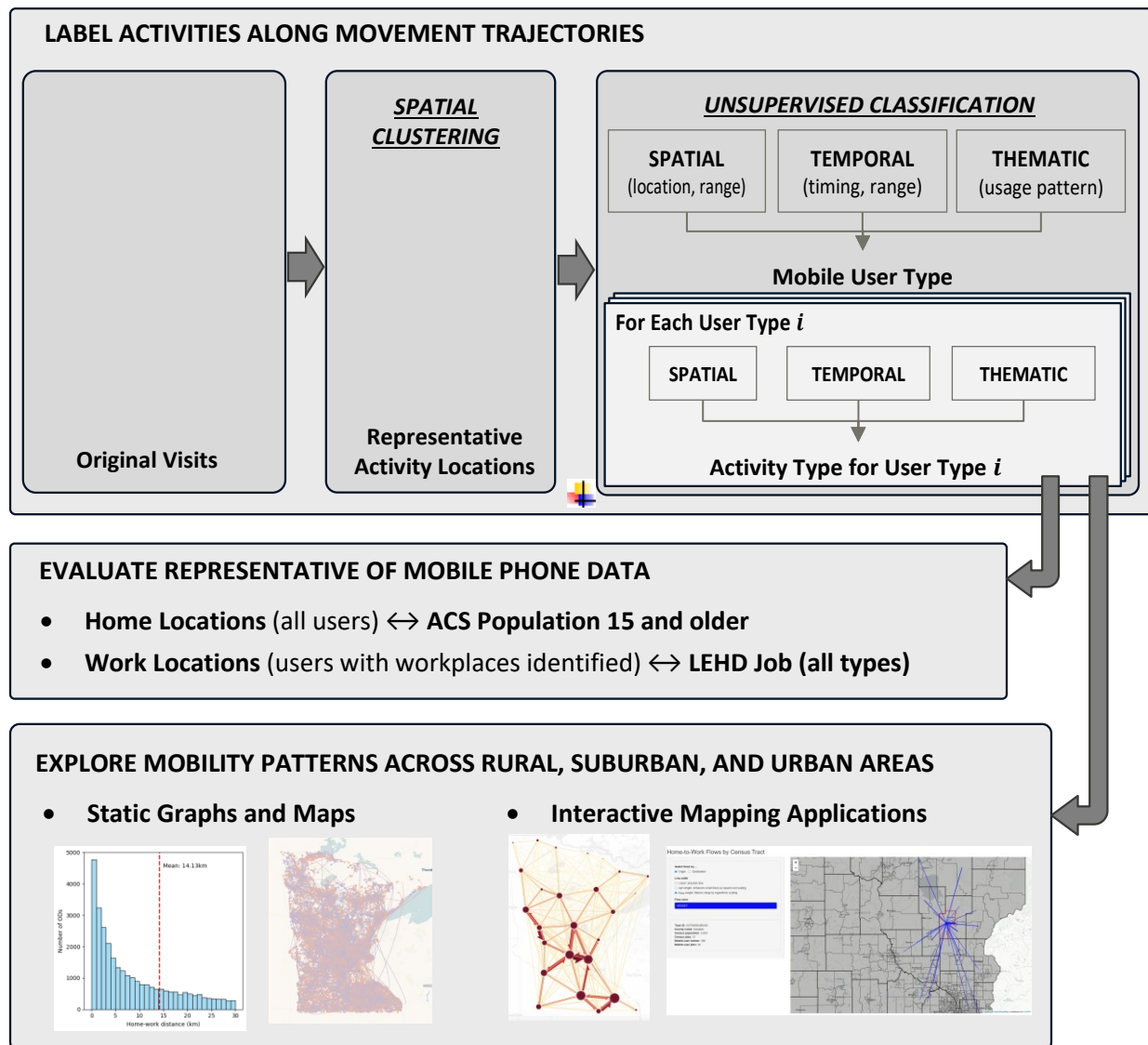


Figure 3.1 Research design with three key components

### 3.2 DERIVE REPRESENTATIVE ACTIVITY LOCATIONS

Each “visit” in the PlaceIQ dataset represents a sustained presence at a location, with the visit location derived from recorded coordinates of all signals during that period. However, visits to the same activity location across different days may exhibit slight variations in their spatial coordinates. Moreover, some activities, such as walking a dog within the neighborhood, are not entirely stationary and thus require a representative location. To deal with these issues and extract meaningful activity locations from the raw visit data, we applied a refined spatial clustering approach based on the Density-Based Spatial Clustering of Application with Noise (DBSCAN) algorithms.

DBSCAN is an unsupervised clustering algorithm that can identify groups of well-connected points, those that lie within a specific distance threshold and exceed a minimum density. Unlike partitioning methods, DBSCAN does not require the number of clusters to be specified in advance, can detect arbitrary shaped clusters, and is robust to noise and outliers. We chose DBSCAN for this study considering the large and unknown number of potential activity locations in the dataset, as well as the substantial variation in the spatial distributions of multi-day visits to the same activity location.

We refined the original DBSCAN algorithm to capture distinct usage patterns of mobile phone users and to accommodate the varying spatial extents of stationary and non-stationary activities. Specifically, we refined the two key parameters that control clustering behavior:

- **minPts (min\_samples).** This model parameter specifies the minimum number of points to form a cluster. Instead of using a constant number for all devices, we defined the minPts to be device-specific, based on the number of active days for each mobile device:

$$\text{minPts}_i = \max(\text{ndate}_i \times \text{freq}, \text{mincnt}) \quad (3.1)$$

where  $\text{ndate}_i$  is the number of days with visit data for device  $i$  and  $\text{freq}$  is the required percentage of days that a location must be visited to qualify as a representative activity location. A higher  $\text{freq}$  value increases the threshold for cluster formation, which helps reduce the noise but may exclude meaningful locations with lower visit frequency. We set  $\text{freq}$  to 60% to focus on routinely visited activity locations. To ensure robustness for devices with limited data, we introduced a lower bound,  $\text{mincnt}$ , representing the minimum number of visits required to define a valid activity location. We set  $\text{mincnt}$  to 2 in this study.

- **eps (epsilon).** This model parameter defines the maximum distance between two points for them to be considered as part of the same cluster. To account for different land use patterns across urban, suburban, and rural areas, we implemented a two-stage hierarchical clustering approach. In the first stage, we focused on fine-scale activities at the block level, using a distance threshold of 50 meters to detect tightly clustered visits within adjacent blocks. The second stage targeted activity locations with broader spatial extent, such as those associated with dispersed or non-stationary activities. For this stage, we selected the distance threshold based on a sensitivity analysis that assesses its effect on the number of detected clusters.

### 3.3 USER SEGMENTATION AND ACTIVITY CLASSIFICATION

We implemented a two-stage hierarchical approach to label activities along individuals' daily trajectories as home, work, routine, and occasional activities. In the first stage, we performed user segmentation to group mobile phone users based on similarities in their device usage and activity-travel behavior. In the second stage, we first identified home locations for users in each user group and then classified the non-home activity locations into work, frequent, and other activities. This two-stage approach enables more accurate activity classification by accounting for variations in lifestyle and mobility patterns across rural, suburban, and urban residents.

We adopted the K-means clustering method, which is an unsupervised machine learning algorithm that partitions an input dataset into K distinct, non-overlapping clusters. Each data point is represented in an n-dimensional attribute space, and the algorithm iteratively minimizes the within-cluster distance, that is, the sum of squared distances between data points and the centroid of their assigned cluster. Cluster assignments and centroids are updated repeatedly until the algorithm converges. The algorithm works well with large datasets and is easy to implement and interpret.

We generated data points by computing a set of user-level and activity-level attributes to capture the spatial, temporal, and thematic characteristics of users and their activity locations. We describe each attribute and its implications below.

- **SPATIAL ATTRIBUTES**

- **Percentages of visits in urban, suburban, and rural areas.** Urban areas are defined by the city boundaries of Minneapolis, Saint Paul, and Duluth. Suburban areas include regions in the seven-county Twin Cities Metropolitan Area outside Minneapolis and Saint Paul and regions in St. Louis County outside Duluth. All remaining areas in Minnesota are viewed as rural areas. At the user level, the relative proportions of visits to urban, suburban, and rural areas can infer the probable residential context of mobile phone users. At the activity level, this relative proportions can help us determine the most appropriate activity classes.
- **Activity space.** We used the area of the convex hull encompassing all visits associated with each mobile phone device to represent the user's activity space. Additionally, we included the lengths of the major and minor axes of the minimum bounding ellipse that contains all visits, in order to capture the shape and directional extent of the activity space. The user-level activity space can capture the spatial mobility and accessibility of users. At the activity level, this measure can help us determine the most appropriate activity classes.

- **TEMPORAL ATTRIBUTES**

- **Unique dates.** We included the total number of days with visit data for each mobile phone user as a control variable, as it may influence other derived attributes such as activity space and visit frequency. At the user level, this measure can also serve as a proxy for the frequency of mobile phone usage. At the activity level, this measure can help us distinguish occasional activities from routine activities such as home and work.

- **Percentages of visits at different periods of the day.** To capture the temporal variation in visit patterns, we computed the proportion of visits occurring during five time periods throughout the day: i) early morning (midnight to 6:00 a.m.), ii) morning rush hours (6:00 a.m. to 9:00 a.m.), iii) mid-day (9:00 a.m. to 3:00 p.m.), iv) afternoon rush hours (3:00 p.m. to 6:30 p.m.), and v) nighttime (6:30 p.m. to midnight). At the user level, the relative distribution of visits across these time intervals can provide insights into users' employment status, commuting behaviors, and other temporal patterns of activities. At the activity level, this measure can help us to infer most likely home and work locations.

- **THEMATIC ATTRIBUTES**

- **Visit Frequency.** We included both the total number of visits and the average number of visits per day to capture the overall activity level of each mobile phone user. At the user level, these measures indicate how active a user is in terms of daily mobility and mobile device usage. At the activity level, they can reflect the relative popularity or importance of specific activity locations for each user.
- **Visit Duration.** At the user level, we used the minimum visit durations to assess how fragmented a user's daily schedule is. At the activity level, we used the median visit duration to estimate the typical length of time spent at that location.
- **Total observations.** Each visit in the PlacelQ dataset consists of multiple recorded mobile phone signals. At the user level, the total number of observations serves as direct indicator of mobile phone usage frequency. At the activity level, this measure can complement visit frequency and visit duration to capture frequency and duration at a specific activity location.

To account for the different units and scales of the input attributes, we applied min-max normalization to rescale the values of each attribute to a range between 0 and 1. This approach ensures comparability across attributes while preserving the original distribution and relative differences within each attribute.

We determined the optimal number of clusters (K) for K-mean clustering algorithm using a combination of the Elbow method and domain knowledge to balance data fit and model interpretability. In the Elbow method, we plotted the within-cluster sum of squares (WCSS) for different values of K. WCSS generally decreases with increasing K, and the "elbow point" indicates where the rate of improvement begins to level off. However, if the elbow occurs at a large K, it would result in an excessive number of user and activity classes, reducing the practical utility of the segmentation. To address this, we also examined the characteristics of the resulting clusters and selected a K value that yields user groups and activity classes with plausible and interpretable features. It is worth noting that, for different user groups, the optimal number of activity classes may be different.

The final products of the user segmentation and activity classes include: i) a group label for each mobile device, along with a summary of group features, ii) a class label for each representative activity location, along with descriptions of each class, and iii) the identified home and workplace coordinates, if any, for each mobile device.

### **3.4 REPRESENTATIVENESS OF MOBILE PHONE DATA**

We systematically evaluated the representativeness of mobile users as a sample population within the study area by comparing the spatial distributions of inferred home and work locations of users with the corresponding census-based population and employment distributions. To facilitate this comparison, we aggregated the inferred home and workplace locations to the census tract level and summarized the three key metrics for each tract: the total number of users (unique devices), the number of users with identified work locations, and total number of inferred work locations. These were then compared to the corresponding census metrics: the total population 15 and older, the employed population, and the total number of jobs per census tract. We visualized the results to assess spatial variations in the sample rates and evaluate the extent to which the distributions of mobile users and their workplaces align with the actual population and job distributions.

## CHAPTER 4: RESULTS AND DISCUSSION

This chapter presents the results for representative activity location derivation, user segmentation and activity classification, sample representative evaluation, and mobility pattern exploration.

### 4.1 REPRESENTATIVE ACTIVITY LOCATIONS

We applied the refined DBSCAN algorithm, as described in section 3.1, to derive representative activity locations from the recorded visits. Using a distance threshold (eps) of 50 meters, approximately 17% of mobile devices did not yield any detected clusters. For these devices, we increased the eps value to 500 meters, based on sensitive analysis results (see Figure 4.1), to capture representative locations for users with more dispersed or infrequent activity patterns.

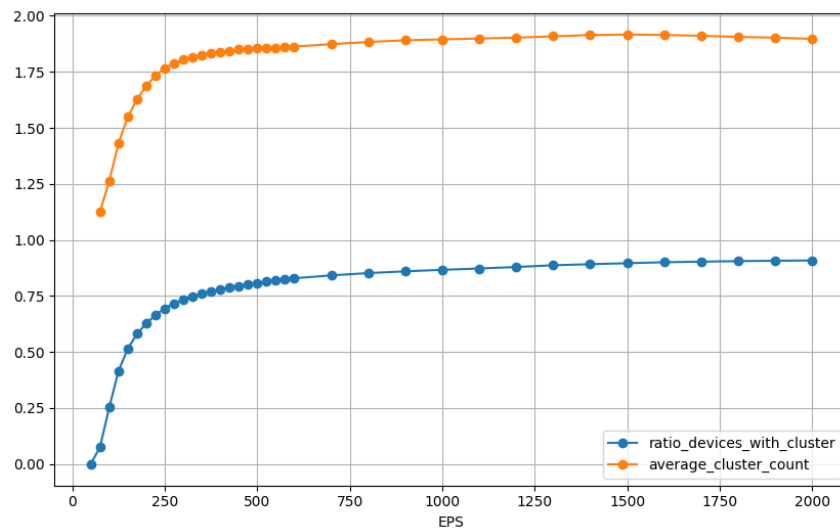
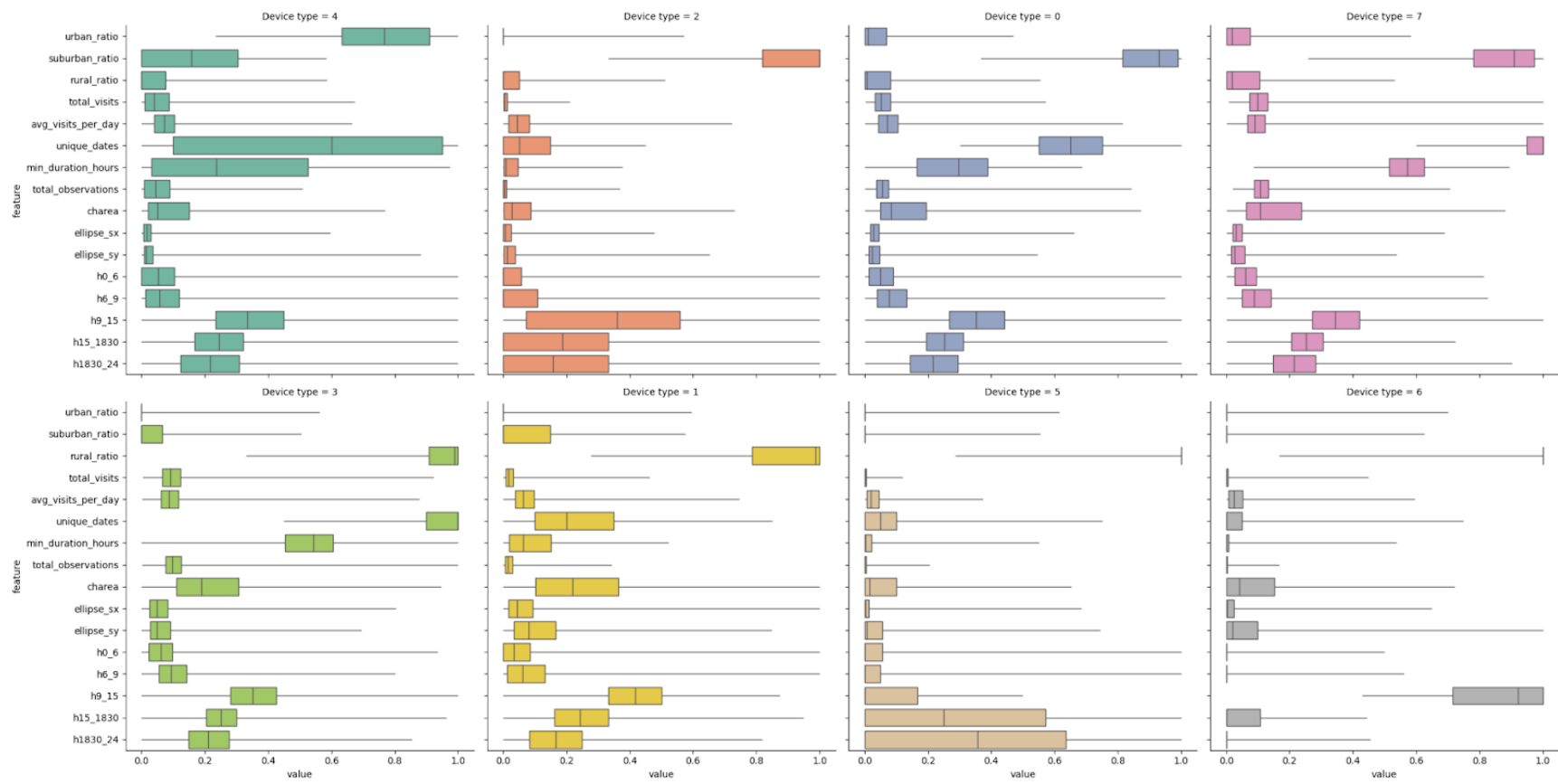


Figure 4.1 Percentage of devices with at least one cluster and the average number of clusters per device

### 4.2 USER SEGMENTATION

We applied the K-mean clustering algorithm, as described in section 3.2, to group mobile phone devices (users) based on their attribute similarity. The optimal number of clusters was determined to be eight, guided by the Elbow method and visual exploration of group characteristics. Figure 4.2 illustrates the attribute profiles of each user group, while Table 4.1 provides a summary of their key characteristics. Each subgraph in Figure 4.2 presents the characteristics of one user group. Each box plot corresponds to one attribute; the middle line shows the median value; the left and right bounds show the first and third quartiles; and the horizontal line shows the value range. The results indicate that i) the predominant geographic context of activities – whether rural, suburban, or urban – is a primary factor distinguishing the user groups, ii) the activity space are strongly associated with the usage frequency, and iii) most groups have visits across the entire day except for two groups in rural areas. We also notice that the numbers of devices belonging to rural, suburban, and urban groups align well with the proportions of population across rural, suburban, and urban areas in Minnesota.



**Figure 4.2 Percentage of devices with at least one cluster and the average number of clusters per device**

**Table 4.1 Summary of user groups’ key characteristics**

Device Type	# of Mobile Devices	Most Visited Region	Usage Frequency	Activity Space	Temporal Range
4	22,643	Urban	Medium	Medium	09:00 – 24:00
2	57,027	Suburban	Low	Low	09:00 – 24:00
0	22,043	Suburban	Median	Median	09:00 – 24:00
7	64,288	Suburban	High	High	09:00 – 24:00
3	86,401	Rural	High	High	09:00 – 24:00
1	43,303	Rural	Median	Median	09:00 – 24:00
5	25,727	Rural	Low	Low	15:00 – 24:00
6	18,013	Rural	Low	Low	15:00 – 18:30

### 4.3 ACTIVITY CLASSIFICATION

We further applied the K-means clustering algorithm to label activity locations for each user group as home, work, frequent, and infrequent activities. To improve the accuracy of activity classification, we first focused on identifying home locations. For each user group, we selected the optimal number of clusters as the one that yielded the most plausible home clusters, characterized by intuitive spatial and temporal attributes. Once home locations were identified, we calculated the distance from each non-home activity location to the corresponding user’s home location to distinguish between activities occurring within residential neighborhoods and those farther away. We then proceeded to classify non-home activities, with a particular focus on identifying work locations, and selected the optimal number of clusters accordingly<sup>5</sup>.

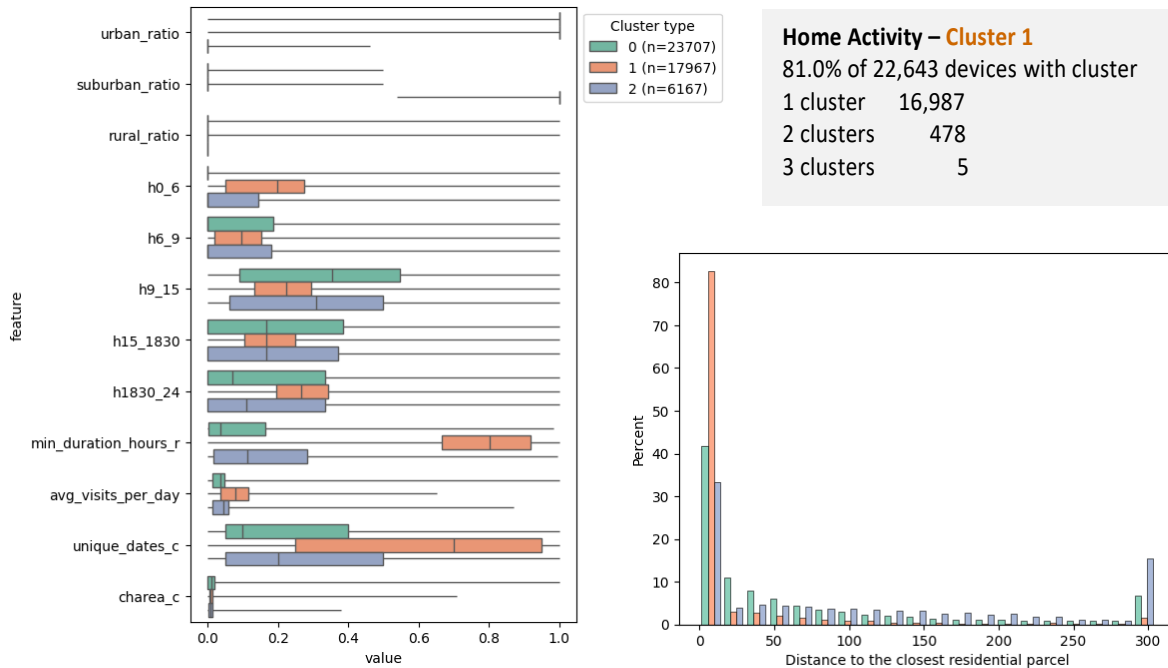
- **Home Activity Locations**

Figure 4.3 presents the classification results for home location identification. Each subgraph corresponds to a specific user group and includes: i) box plots illustrating the characteristics of each identified cluster, ii) a histogram showing the distances from activity locations to their nearest residential parcels, and iii) a summary of the identified home activity cluster for that user group.

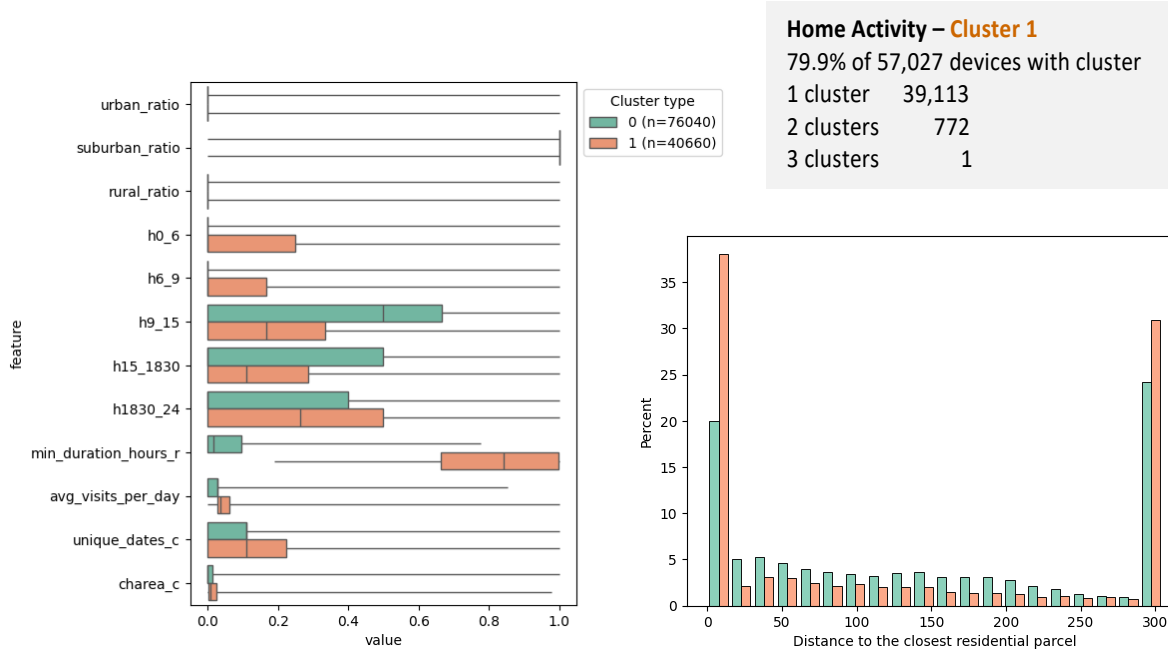
The overall performance of home location identification is satisfactory. For six of the eight user groups, over 80% of devices have at least one identified home activity cluster, and the majority of these home locations are located within 10 meters of a residential land parcel. In contrast, only about 30% of devices in rural subgroups with low usage frequency have a detected home activity cluster, which exhibit longer activity durations compared to other clusters. The lower detection rate in these subgroups is primarily attributed to the limited availability of visit data during nighttime and early morning hours (midnight to 9:00 a.m.), when home activities are most likely to occur.

---

<sup>5</sup> This study focuses on in-person, full-time employment conducted during regular working hours. Other types of employment or activities, such as part-time or remote work, are categorized as frequent routine activities.

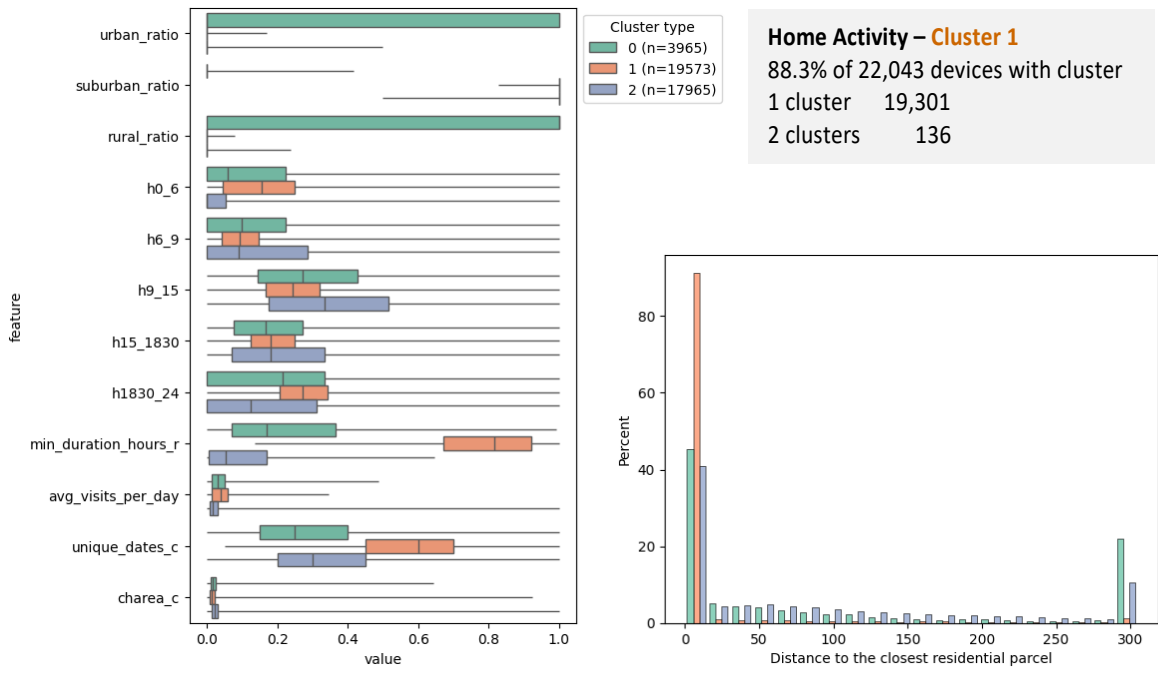


(a) User Group 4: Urban with Median Usage Frequency

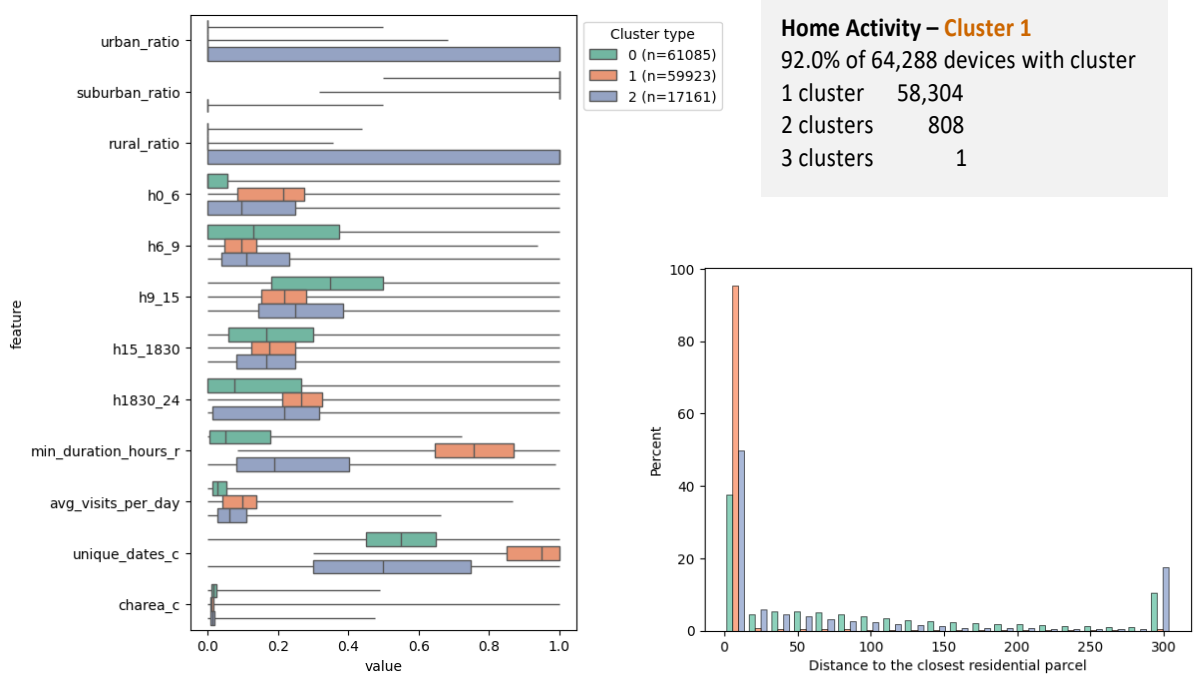


(b) User Group 2: Suburban with Low Usage Frequency

Figure 4.3 Activity clusters, their characteristics, and distances to nearest residential parcels

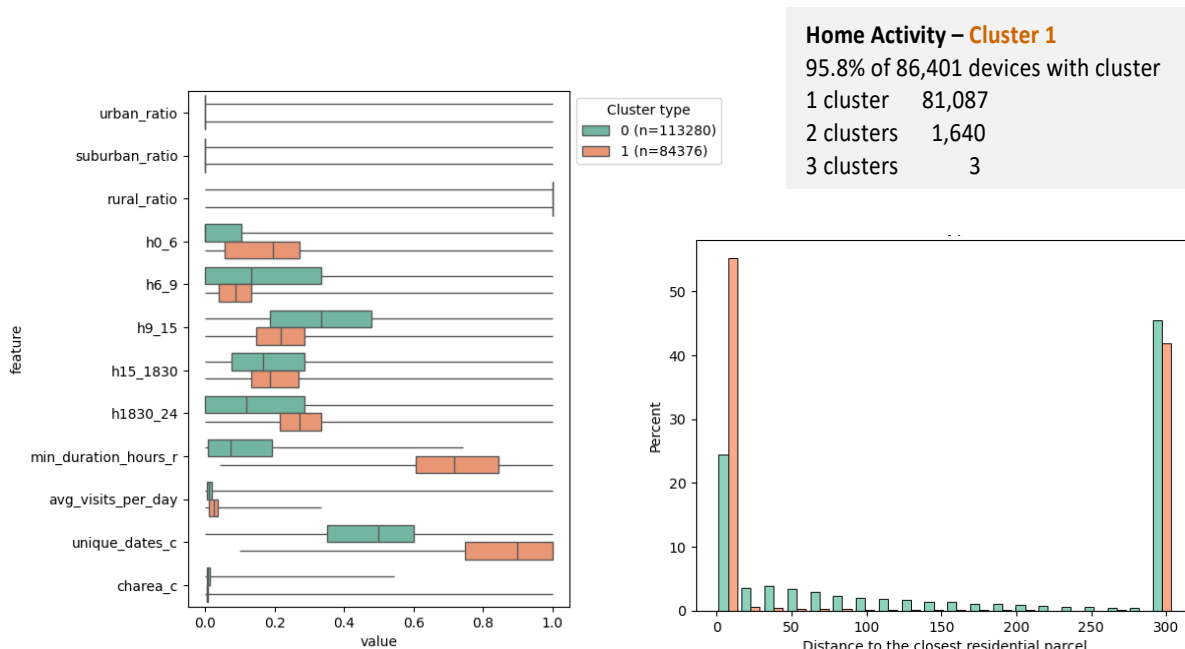


(c) User Group 0: Suburban with Median Usage Frequency

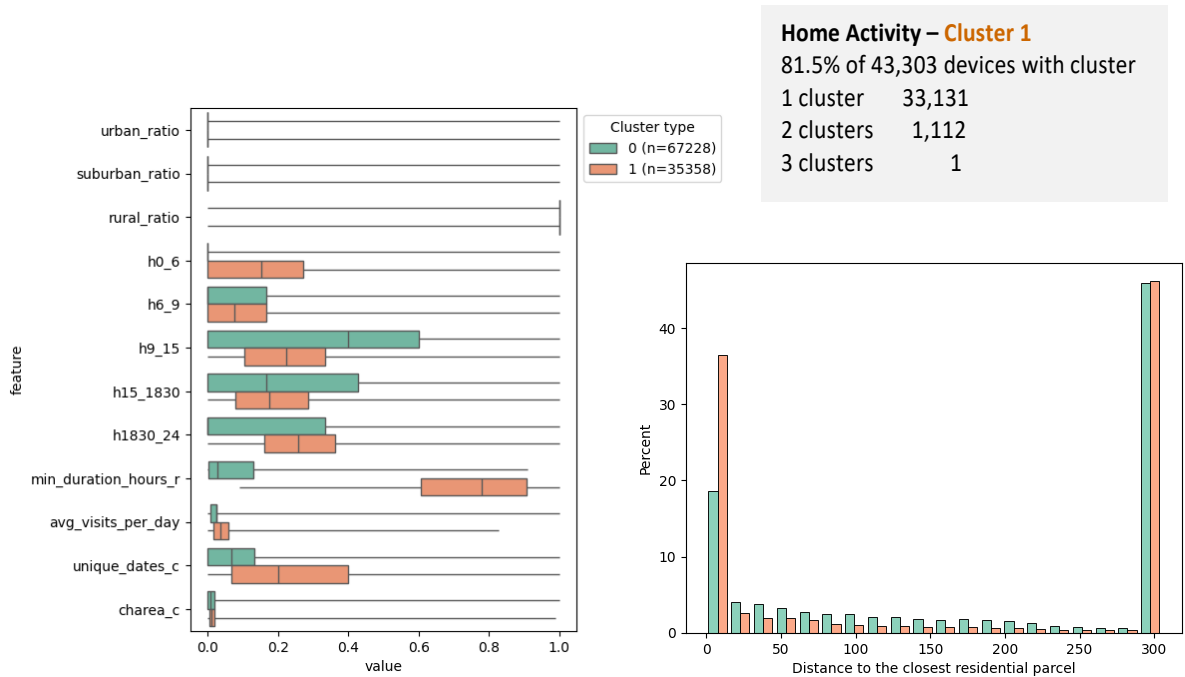


(d) User Group 7: Suburban with High Usage Frequency

Figure 4.3 Activity clusters, their characteristics, and distances to nearest residential parcels (continued)

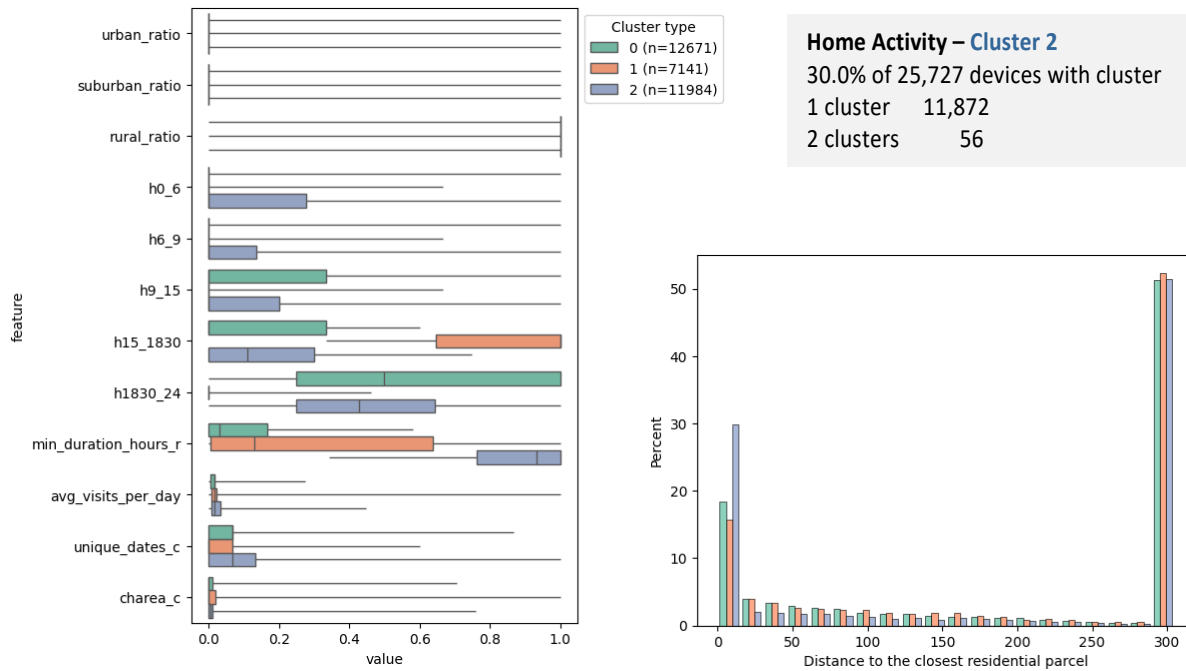


(e) User Group 3: Rural with High Usage Frequency

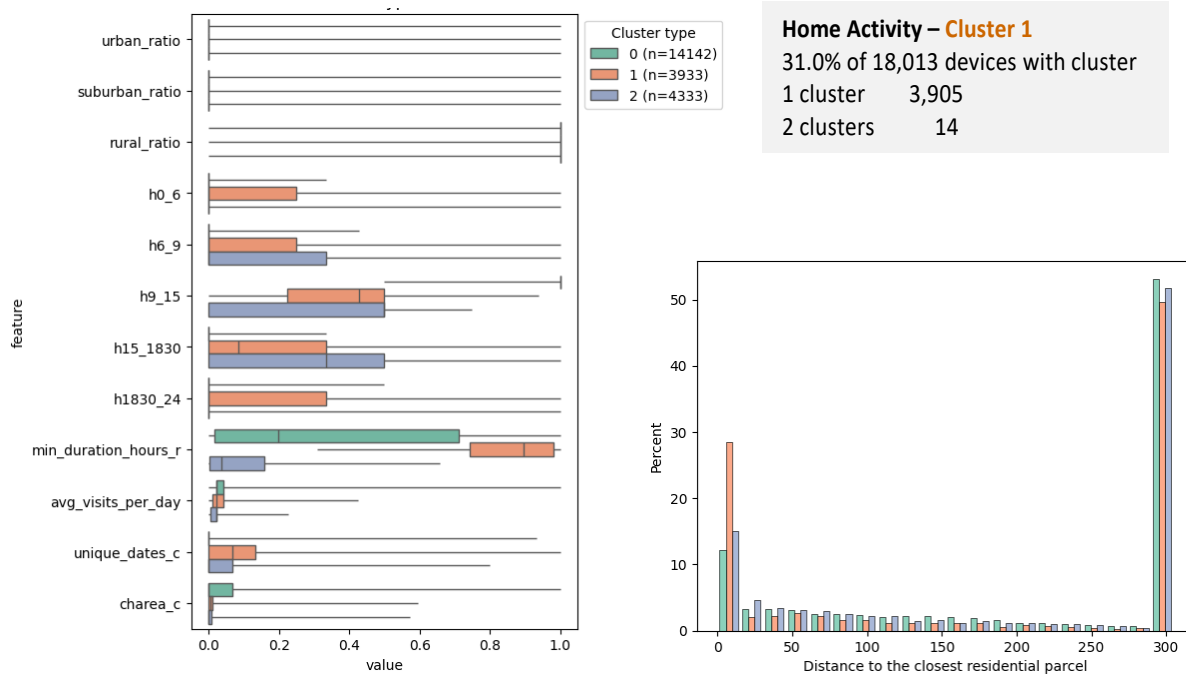


(f) User Group 1: Rural with Median Usage Frequency

Figure 4.3 Activity clusters, their characteristics, and distances to nearest residential parcels (continued)



(g) User Group 5: Rural with Low Usage Frequency during Late-Day



(h) User Group 6: Rural with Low Usage Frequency during Mid-Day

Figure 4.3 Activity clusters, their characteristics, and distances to nearest residential parcels (continued)

- **Work and Other Activities**

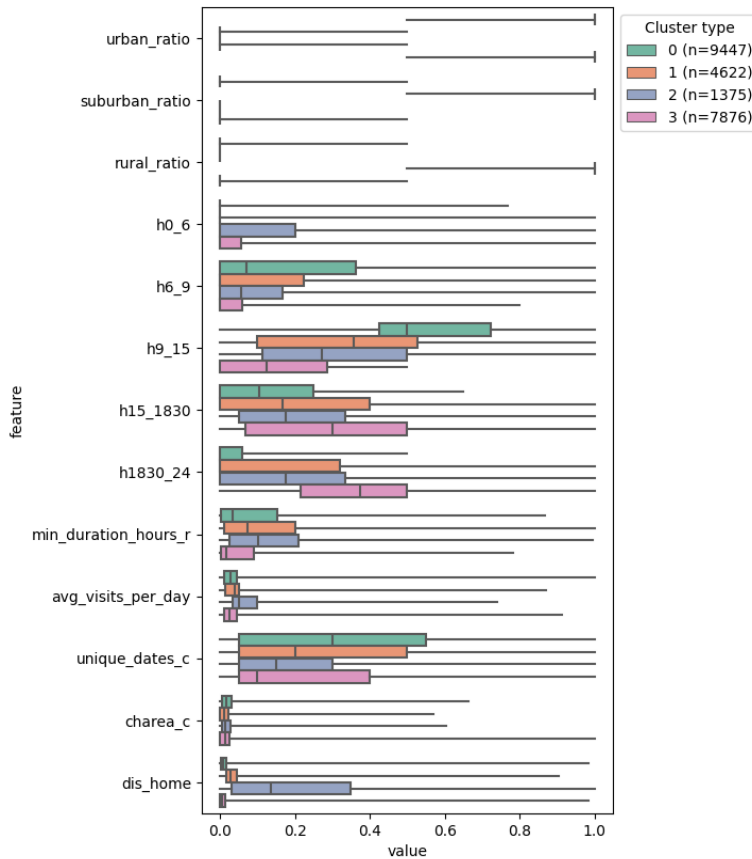
We continued to classify non-home activity locations, incorporating an additional attribute representing the distance from each activity location to the corresponding home location. Figure 4.4 presents the classification results. Each subgraph corresponds to a specific user subgroup and includes: i) box plots illustrating the characteristics of each identified activity cluster, and ii) a summary of identified activity clusters, including the number of unique devices, the cluster descriptions, and the median distance to the nearest work-related land use parcels.

Across all eight user subgroups, 35% to 55% of devices have at least one identified work cluster. These clusters are characterized by visits that predominantly occur during regular working and commuting hours (6:00 a.m. to 6:30 p.m.) and have relatively long durations and high visit frequencies. For urban and suburban subgroups, the distances from identified work locations to the nearest work-related land parcels are typically less than 200 meters. In contrast, for rural subgroups, these distances often exceed 1,500 meters, suggesting that workplaces in rural areas are less likely to be concentrated on designated work-related land uses, such as office buildings or warehouses. Additionally, clusters in rural subgroups exhibit relatively larger median work-home distances and greater variability compared to those in urban and suburban subgroups.

We created two generalized labels for the remaining types of clusters: frequent activities and infrequent activities. In addition to their lower visit frequency (fewer unique dates), infrequent activity clusters are typically located much farther from home compared to work and frequent activity clusters. Moreover, these infrequent clusters exhibit greater variation in spatial distribution and temporal patterns.

23,320 identified clusters for 11,604 devices

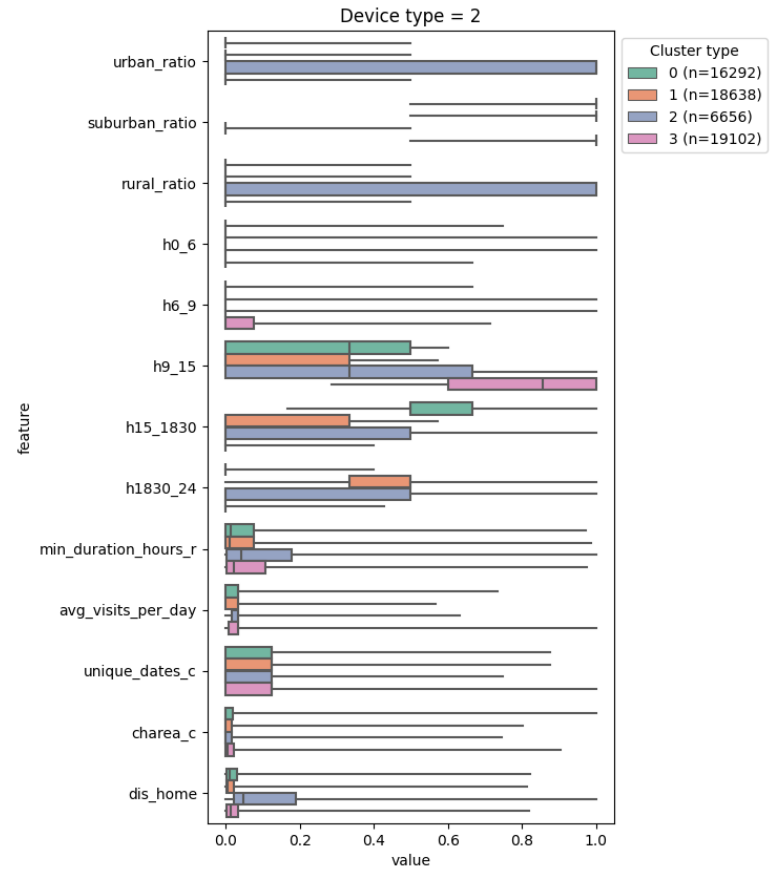
- Cluster 0 (6,240 devices) urban work, 3.7 m
- Cluster 1 (3,533 devices) suburban work, 172.2 m
- Cluster 2 (6,240 devices) infrequent activities far from home, 0.0 m
- Cluster 3 (6,240 devices) activities near home, 11.9 m



(a) User Group 4: Urban with Median Usage Frequency

60,688 identified clusters for 23,288 devices

- Cluster 0 (10,367 devices) suburban late work, 0.0 m
- Cluster 1 (11,590 devices) activity near home, 0.0 m
- Cluster 2 ( 4,845 devices) infrequent activities, 23.8 m
- Cluster 3 (11,594 devices) suburban early work, 0.0 m



(b) User Group 2: Suburban with Low Usage Frequency

Figure 4.4 Activity clusters, their characteristics, and distances to nearest work-related parcels

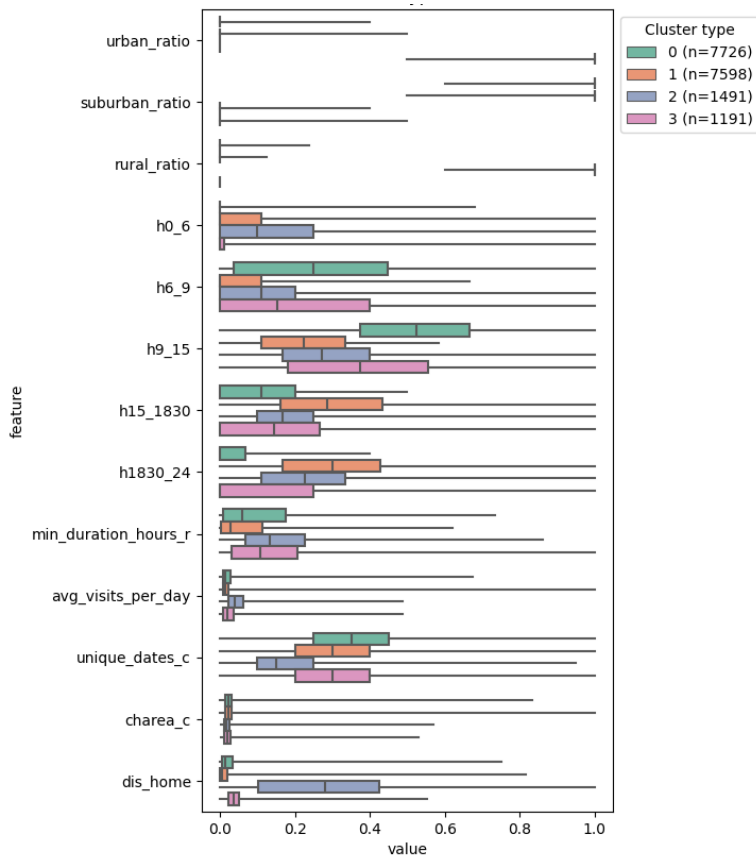
**18,006 identified clusters for 10,711 devices**

**Cluster 0** (5,726 devices) **work**, 0.0 m

**Cluster 1** (5,485 devices) nighttime activities near home, 48.9 m

**Cluster 2** (1,325 devices) rural infrequent activities, 569.0 m

**Cluster 3** (1,022 devices) daytime activities near home, 0.0 m



**(c) User Group 0: Suburban with Median Usage Frequency**

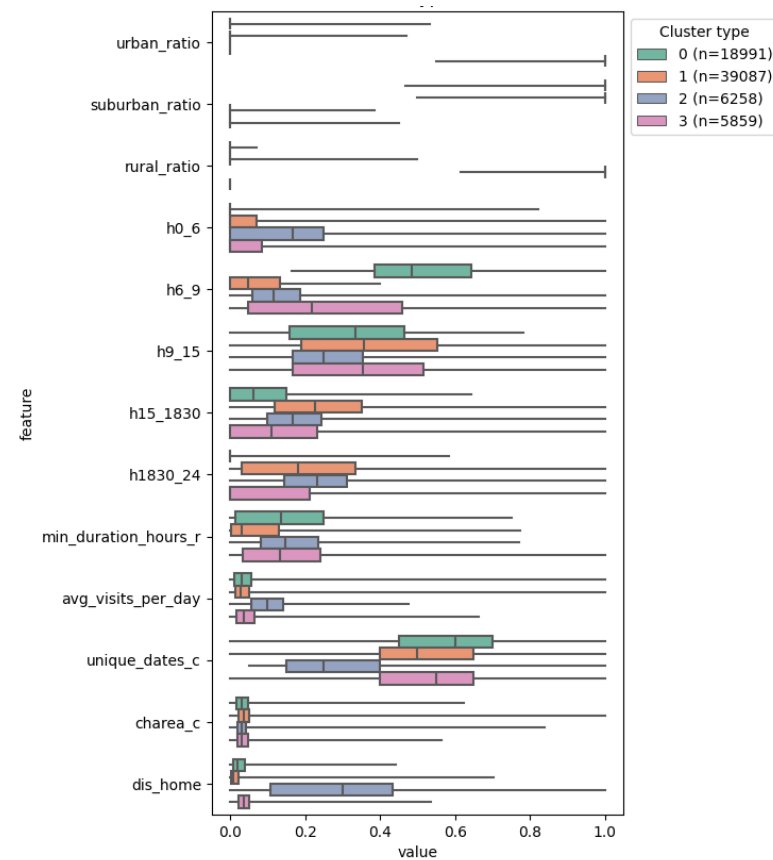
**70,195 identified clusters for 41,465 devices**

**Cluster 0** (15,428 devices) **work**, 0.0 m

**Cluster 1** (25,655 devices) activity near home, 21.7 m

**Cluster 2** ( 5,602 devices) rural infrequent activities, 648.2 m

**Cluster 3** ( 4,782 devices) urban infrequent activities, 0.0 m



**(d) User Group 7: Suburban with High Usage Frequency**

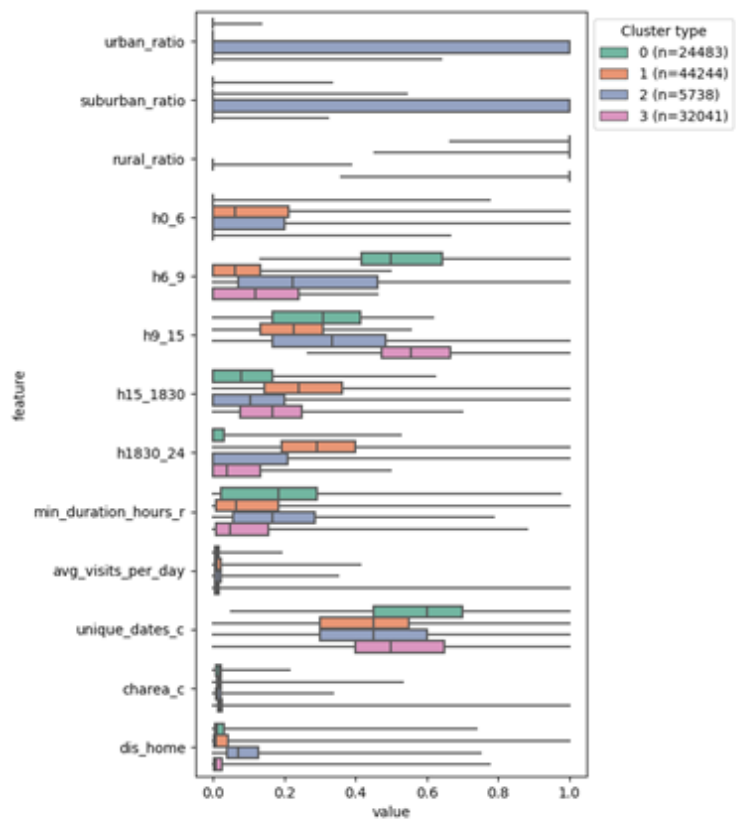
**Figure 4.4 Activity clusters, their characteristics, and distances to nearest work-related parcels (continued)**

**106,506 identified clusters for 58,764 devices**

**Cluster 0** (20,440 devices) morning activities near home, 1202.4 m

**Cluster 1** (31,618 devices) nighttime activities near home, 1958.8 m

**Cluster 2** (1,325 devices) infrequent activities, 0.0 m



**55,193 identified clusters for 23,936 devices**

**Cluster 0** (13,023 devices) **work**, 1482.0 m

**Cluster 1** ( 3,352 devices) suburban infrequent activities, 0.0 m

**Cluster 2** ( 6,361 devices) daytime activities near home, 1031.2 m

**Cluster 3** (12,453 devices) nighttime activities near home, 1878.2 m

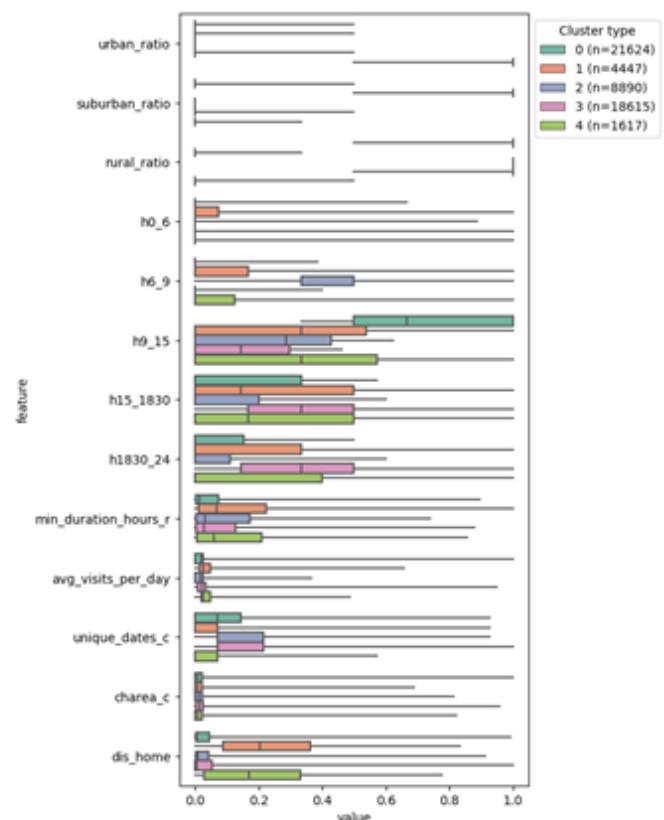
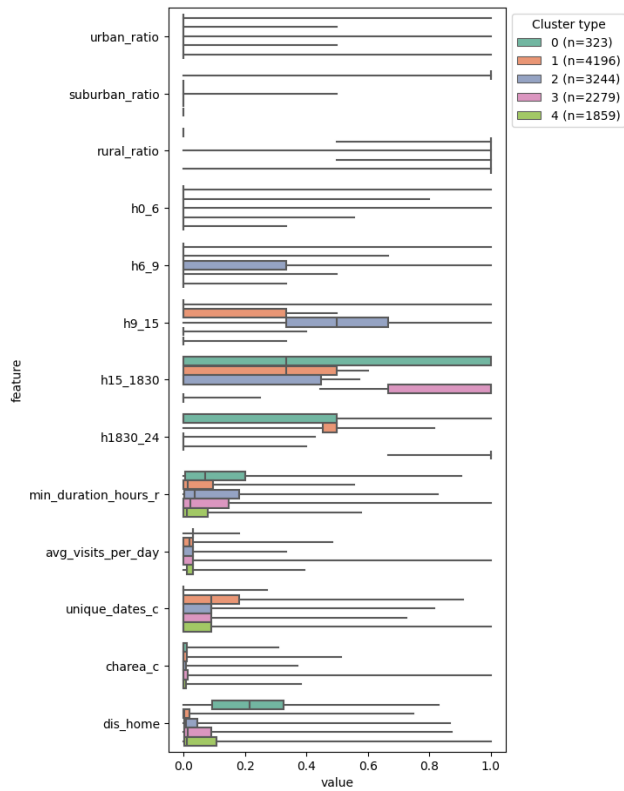


Figure 4.4 Activity clusters, their characteristics, and distances to nearest work-related parcels (continued)

**11,901 identified clusters for 5,996 devices**

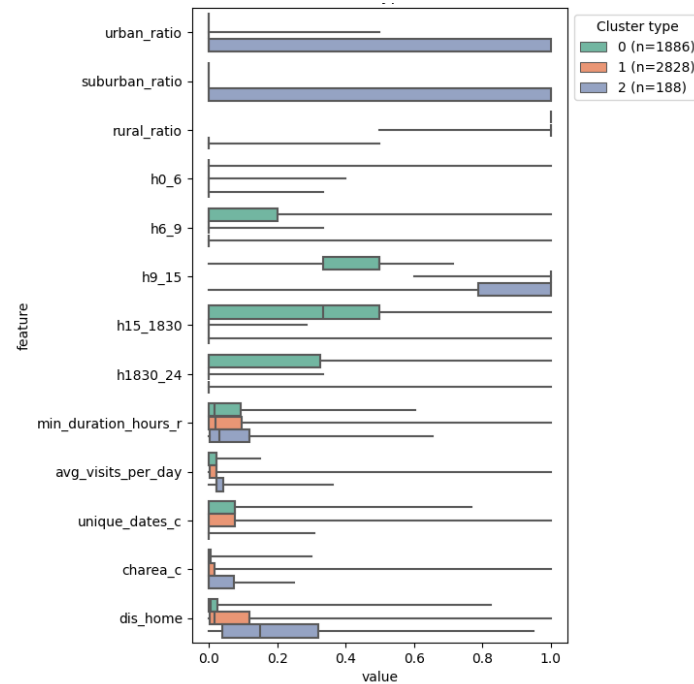
- Cluster 0** ( 268 devices) non-rural infrequent activities, 0.0 m
- Cluster 1** (2,733 devices) activities near home, 2970.9 m
- Cluster 2** (2,333 devices) work, 2331.3 m
- Cluster 3** (1,776 devices) afternoon activities, 2352.3 m
- Cluster 4** (1,468 devices) nighttime activities, 3809.6 m



**(e) User Group 5: Rural with Low Usage Frequency, Late-Day**

**4,902 identified clusters for 2,456 devices**

- Cluster 0** (1,278 devices) activities near home, 3168.4 m
- Cluster 1** (1,732 devices) work, 4027.8 m
- Cluster 2** ( 147 devices) non-rural infrequent activities, 0.0 m



**(f) User Group 6: Rural with Low Usage Frequency, Mid-Day**

**Figure 4.4 Activity clusters, their characteristics, and distances to nearest work-related parcels (continued)**

#### 4.4 REPRESENTATIVENESS EVALUATION

Considering that the census population data only accounts for the primary residential location, we further identify the primary residential location for each device that has more unique dates of visit data, a higher percentage of visits occurring in early morning and late night, and longer durations. Similarly, we further refine the work locations for each device and include only those with a higher percentage of visits occurring from 9:00 a.m. to 3:00 p.m. than other periods and more than one unique date of visit data. Excluded work locations are reclassified as frequent activities instead.

Table 4.2 summarizes the final processing outcomes. Overall, our method performs well for most user subgroups but faces greater challenges while dealing with the two rural user subgroups who have lower mobile phone usage frequency that limits the identification of both home and workplace locations. This highlights that the representativeness of mobile phone data is strongly influenced by users' mobile phone usage patterns, especially in rural areas.

Notably, only about 34% of users have an identified workplace, significantly lower than Minnesota's 68.2% labor force participation rate<sup>6</sup>. This discrepancy may be attributed to three primary factors: i) work-from-home cases are not captured, as all visits recorded at home locations are classified as home activities, ii) work with non-standard working hours may be classified as frequent activities rather than work, and iii) discrepancies between actual workplace and registered employment locations, such as for mobile or field-based workers (e.g., maintenance staff), are not accounted for.

**Table 4.2 Summary of activity classification outcomes**

Device Type	Total # of Devices	Most Visited Region	Usage Freq	% of Device with Home	% of Device with Work	% of Device with Frequent Activity	% of Device with Infrequent Activity
4	22,643	Urban	Medium	90.10%	37.17%	70.39%	6.74%
2	57,027	Suburban	Low	82.87%	21.12%	93.21%	14.08%
0	22,043	Suburban	Median	96.97%	30.17%	47.10%	6.98%
7	64,288	Suburban	High	97.46%	40.11%	61.94%	9.99%
3	86,401	Rural	High	97.64%	40.51%	78.94%	6.80%
1	43,303	Rural	Median	92.04%	43.76%	79.50%	15.21%
5	25,727	Rural	Low	65.75%	12.54%	44.43%	13.38%
6	18,013	Rural	Low	54.32%	7.00%	41.18%	1.92%
<b>Total</b>	<b>339,445</b>			<b>89.15%</b>	<b>34.22%</b>	<b>71.75%</b>	<b>9.93%</b>

<sup>6</sup> State and National Employment and Unemployment: <https://mn.gov/deed/data/current-econ-highlights/state-national-employment.jsp>

We further compared the spatial distributions of inferred home and work locations from the PlacelQ dataset with census-based population and employment distributions. Figure 4.5 presents the tract-level comparison results, which reveal relatively higher sampling rates in rural areas compared to urban and suburban areas, for both population and employment. While the PlacelQ data closely mirrors the spatial distributions of population and jobs in urban and suburban areas, it captures a disproportionately larger share of rural areas. Notably, the ACS five-year estimates achieve an approximate cumulative sampling rate of 12.5% (equivalent to 1 in 8 households over five years), whereas the PlacelQ dataset captures an average sampling rate of 5.5% in just a single month. These results demonstrate the strong potential of using PlacelQ data to study mobility and accessibility, particularly in rural contexts where traditional travel survey data sources are often limited.

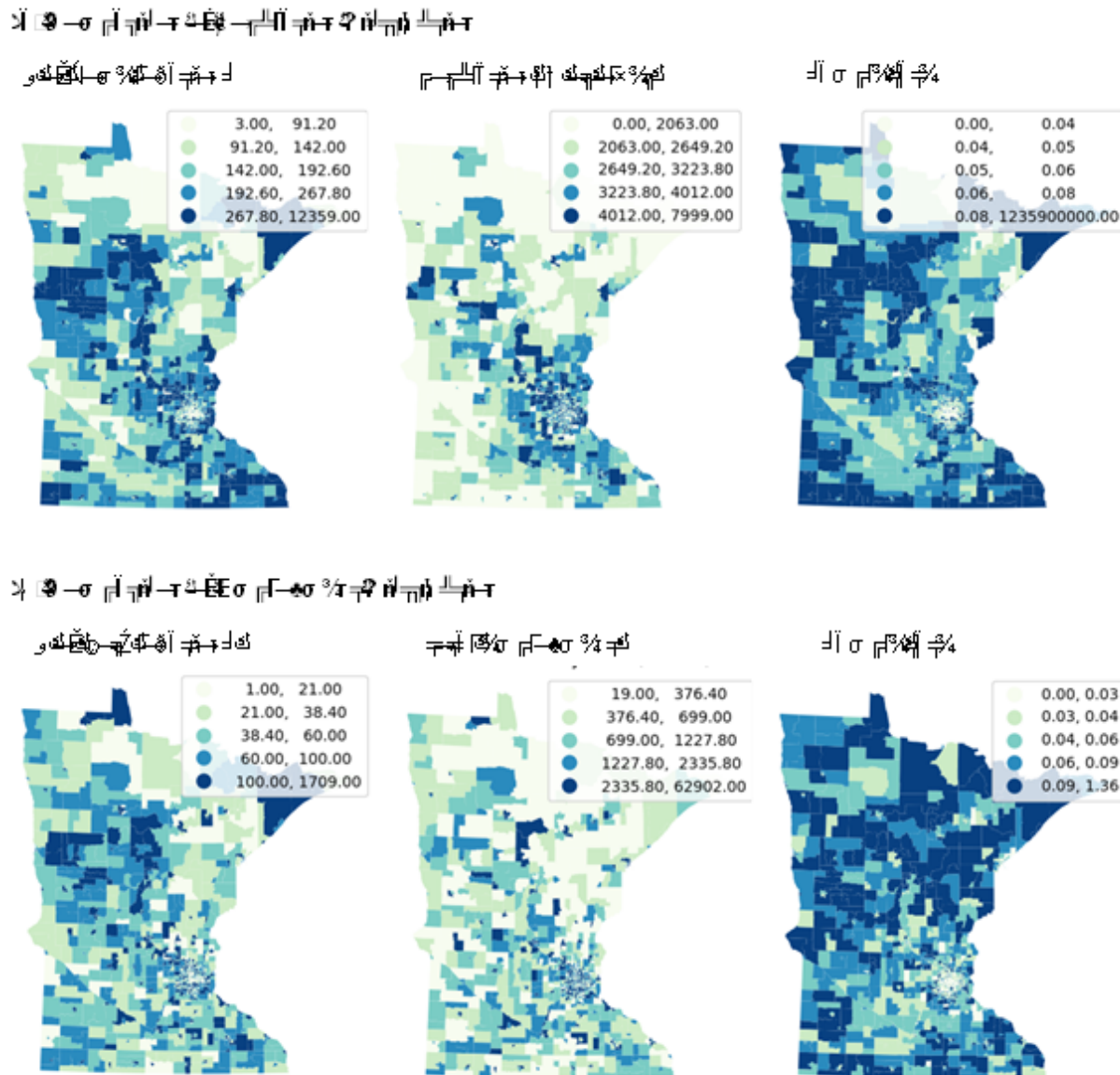


Figure 4.5 Comparison of Population and Employment Distributions in PlacelQ and Census data

## 4.5 MOBILITY AND ACCESSIBILITY EXPLORATION

### 4.5.1 Visual Exploration Using Static Graphs and Maps

To start with, we directly visualized the mobility flows and comparing patterns across eight user groups<sup>7</sup>. Each subgraph in Figure 4.6 corresponds to one user group, which includes maps for the home-to-work, home-to-frequent, and home-to-infrequent activity flow and histograms of their distances.

- **Urban and suburban user groups exhibit similar mobility patterns.** Most of their work and frequent activities are near where they live and concentrated within the Twin Cities metropolitan area, which occur at an average distance of about seven miles from home. In contrast, infrequent or occasional activities are dispersed throughout the entire state of Minnesota. For suburban groups with medium to high mobile device usage frequency, these infrequent activities often occur far from home, with an average distance of approximately 90 miles.
- **Rural user groups display diverse mobility patterns.** Their home locations are widely dispersed across the state, making it difficult to extract clear insights from static maps alone. However, some common spatial trends emerge. For example, frequent activity locations are generally more widely distributed than work locations, while infrequent activities tend to cluster around metro areas.

Compared to their urban and suburban counterparts, rural subgroups typically travel farther to reach work and frequent activity locations, indicating a lower level of spatial access in rural areas. Interestingly, rural users have travel distances to infrequent activity locations that are comparable to – or even shorter than – those of urban and suburban users. This may be attributed to the nature of infrequent activities, such as camping, which often occur in rural areas and are thus closer to rural residents. We will further examine such differences using several mobility metrics in this subsection.

---

<sup>7</sup> This section only includes mobile devices with at least one home location identified.

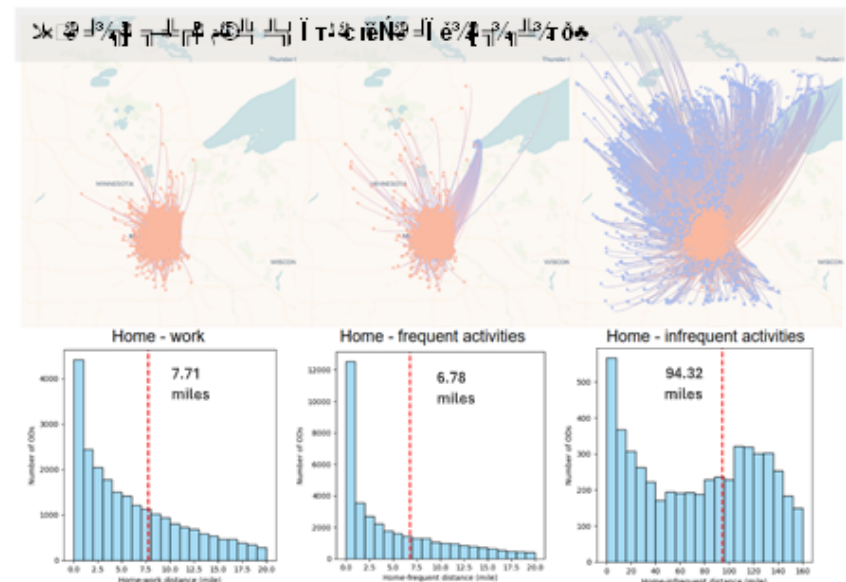
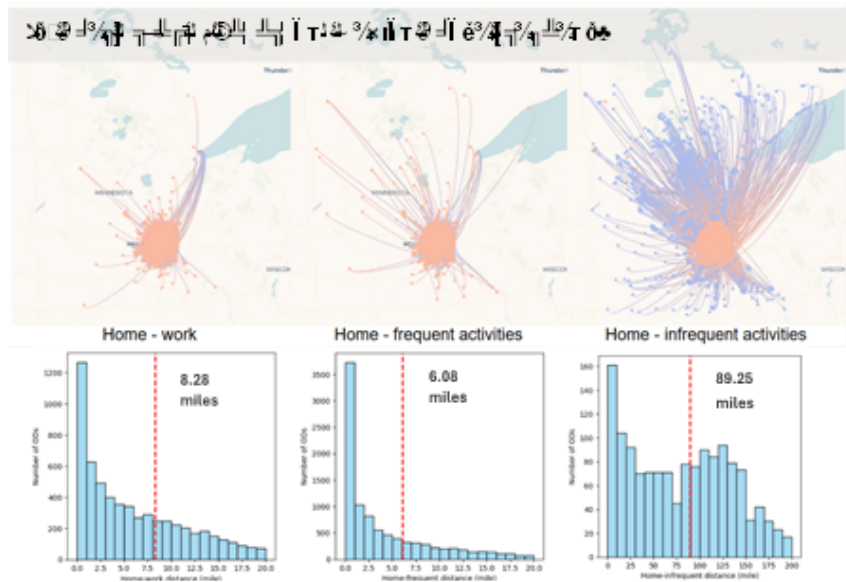
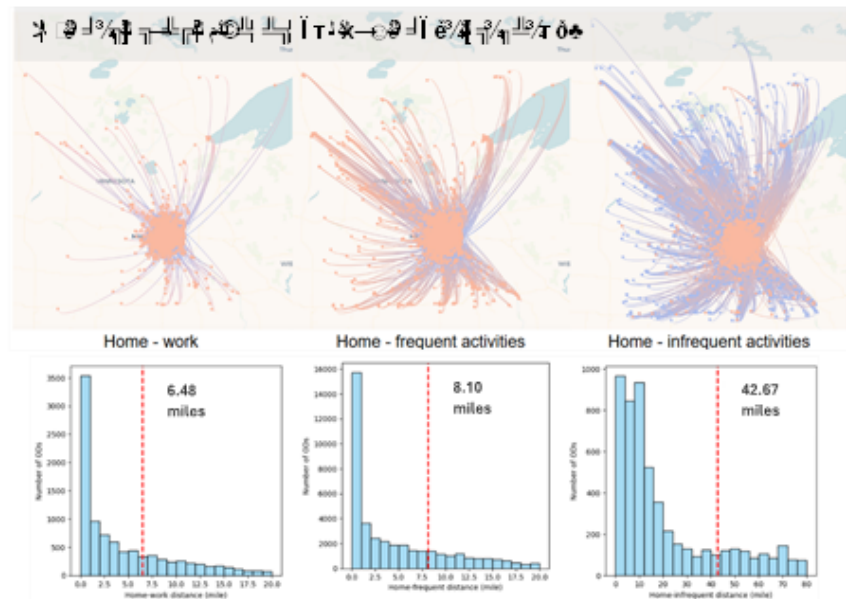
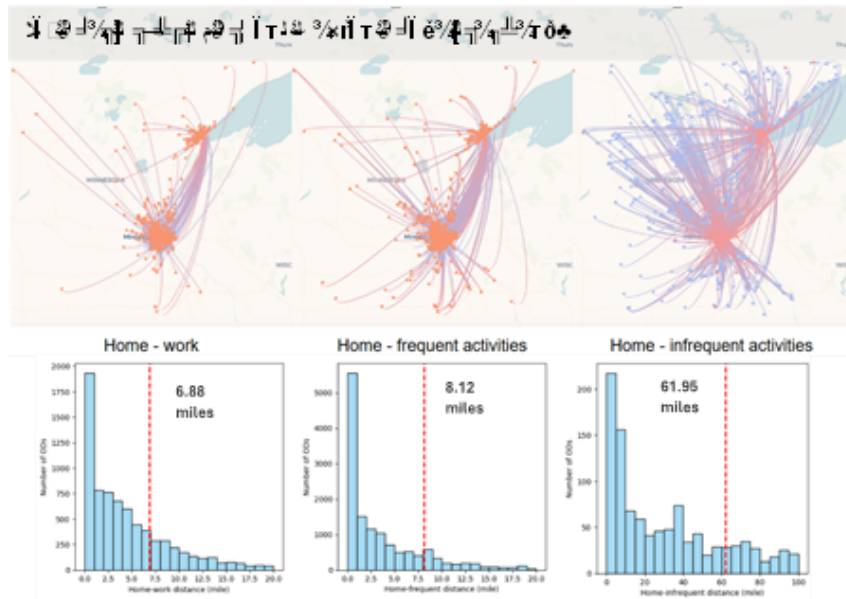
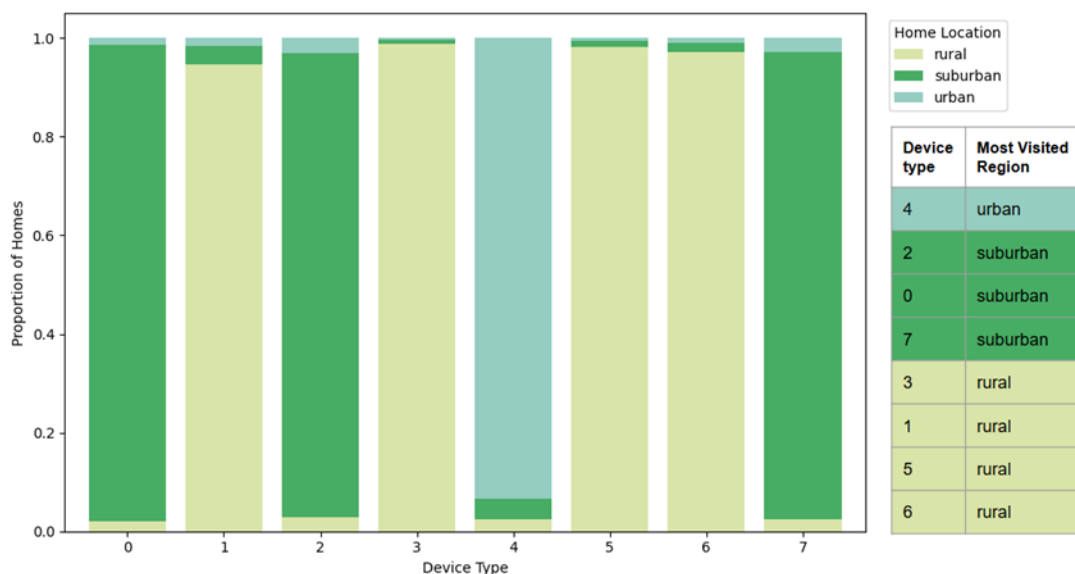


Figure 4.6 Visualization of mobility flows by destination activity type



Before deriving and comparing mobility metrics across rural, suburban, and urban areas, we reclassified mobile users into rural, suburban, and urban groups based on their identified home locations. Figure 4.7 compares this home-based classification with the initial user group labels for the most frequently visited region. The results show a strong alignment between the two classifications, indicating that the most frequently visited region effectively reflects users' home locations. This finding also highlights a broader pattern: individuals tend to conduct most of their activities within the region where they reside, with relatively infrequent travel across different regions.



**Figure 4.7 Comparison of the user group labels and the home location regions**

Using the reclassified user groups, we calculated three metrics for each mobile user to compare mobility and accessibility patterns across rural, suburban, and urban areas.

### 1. Distance to Frequent Activity Locations

Distances to frequent, routine activity locations can serve as a proxy for users' spatial access to essential resources such as schools, grocery stores, and other daily needs. As shown in Figure 4.6, users across all subgroups tend to have a concentration of frequent activities within one mile of their home location. To better understand this spatial access, we calculated the percentage of frequent activities located within a quarter mile, half mile, and one mile from home, and visualized the distance distribution for those beyond the one-mile radius. Results in Figure 4.8 indicate that while rural residents have a similar level of neighborhood access (within one mile) as urban and suburban residents, they tend to experience a lower level of spatial access beyond their immediate neighborhood.

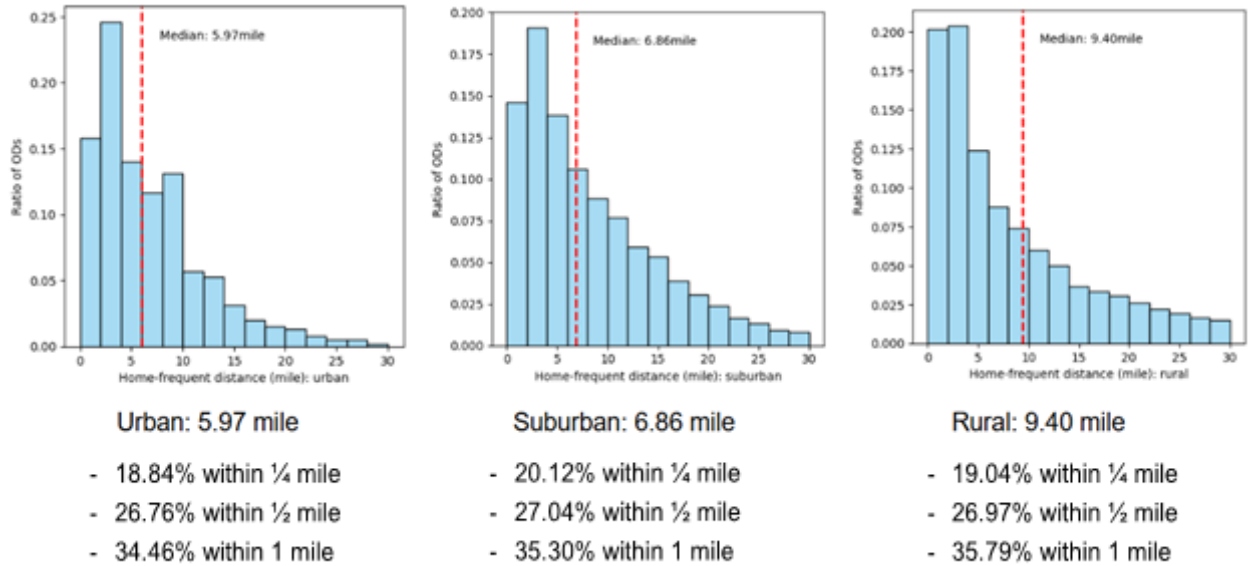


Figure 4.8 Frequent Activities Within and Outside Residential Neighborhoods

## 2. Average Daily Travel Distance

We computed the average daily travel distance for each mobile user based on all trips to work, frequent, and infrequent activity locations, and compared distributions of these distances across rural, suburban, and urban areas (see Figure 4.9). Interestingly, rural residents tend to have shorter daily travel distances than suburban residents. This pattern may reflect a combination of factors, including the reduced travel frequency of rural residents due to the lower level of access, as well as their generally shorter distances to infrequent activity locations discussed earlier.

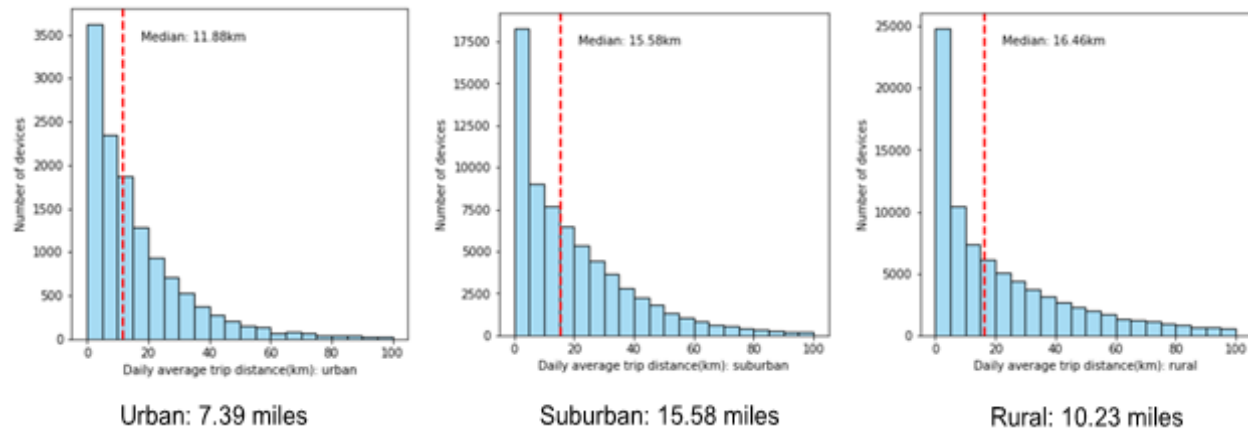
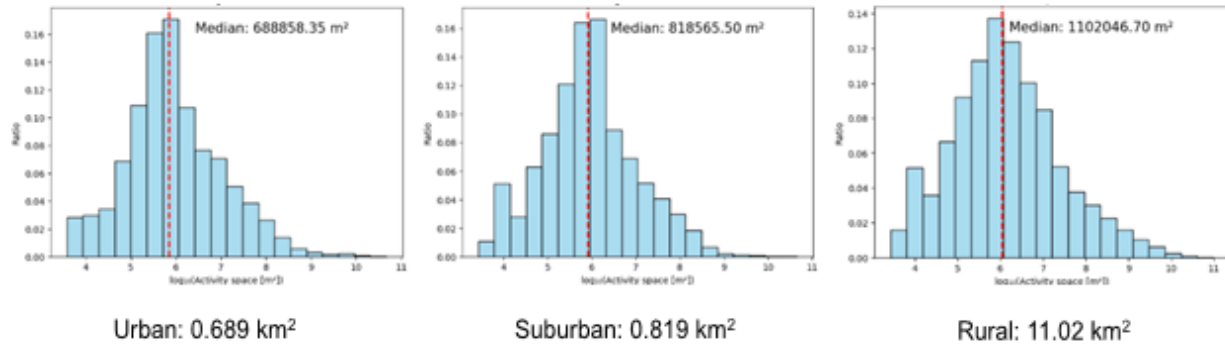


Figure 4.9 Frequent Activities Within and Outside Residential Neighborhoods

## 3. Typical Daily Activity Space

We used home and frequent activity locations to generate each user's typical daily activity space. To measure the size of this activity space, we generated concave hulls from the identified locations and

calculated their areas. For users with only one or two activity locations, we applied a 50-meter buffer around the point or line to approximate their activity space. Similarly, a 50-meter buffer was applied to the concave hulls before calculating their areas to ensure consistency. Figure 4.10 presents results that exclude users with only one activity location. Although rural residents generally travel shorter distances compared to suburban residents, they tend to have larger activity spaces. This is likely due to the sparser distribution of activity opportunities in rural areas, which necessitates broader spatial movement and results in activity spaces that are larger and more isotropic in shape, rather than narrow or elongated.



**Figure 4.10 Frequent Activities Within and Outside Residential Neighborhoods**

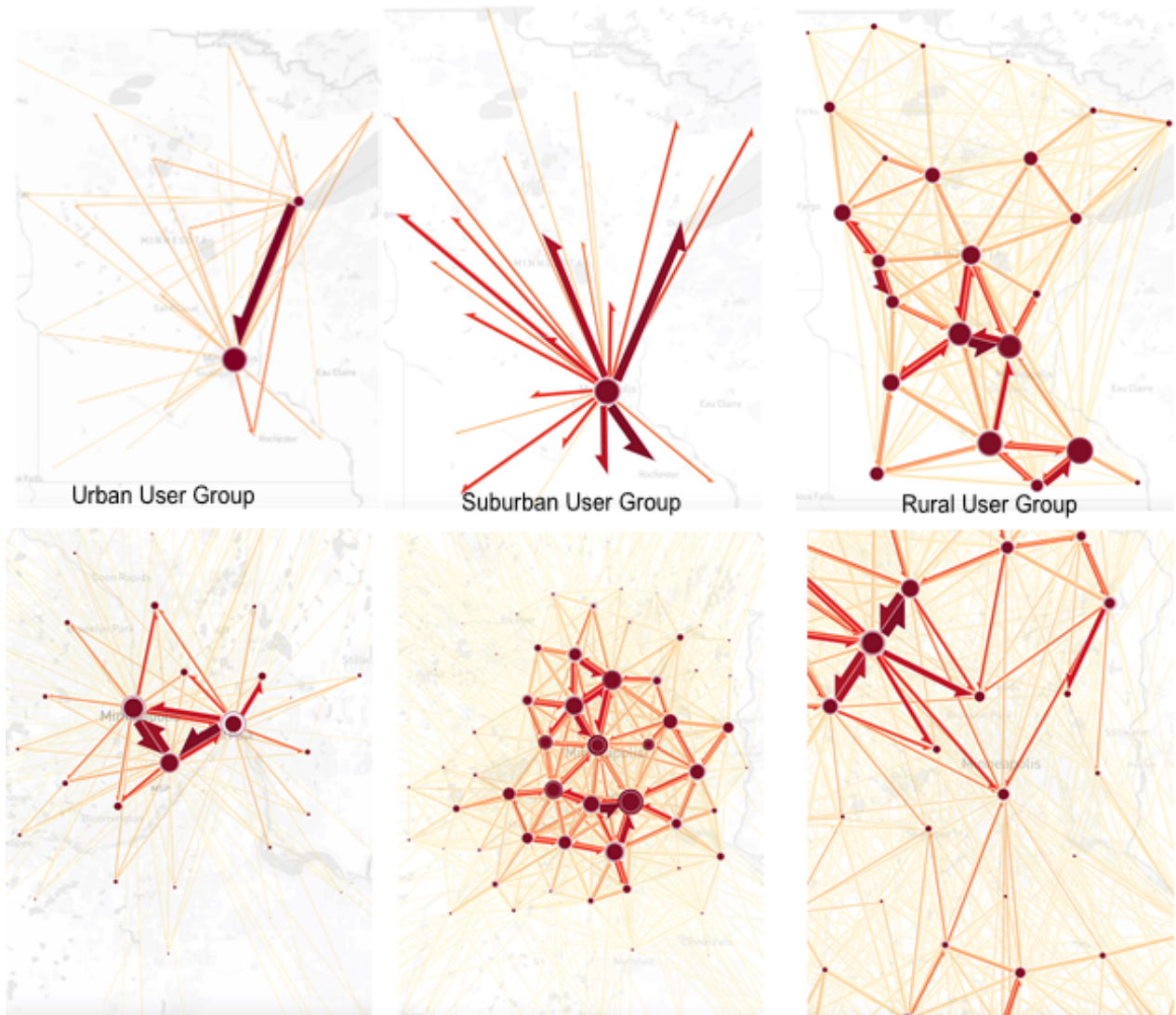
#### 4.5.2 Visual Exploration Using Interactive Maps

To facilitate more in-depth analysis of mobility patterns across the region, we created one interactive map for scale-free exploration of mobility flows and one online mapping application for exploration of flows at the census tract level.

##### 1. Scale-free Interactive Map

We developed an interactive map that enables exploration of flow patterns at multiple spatial scales. Figure 4.11 illustrates home-to-work flows across the entire state and within a targeted area as an example of how to use the interactive map. Each dot represents a cluster of locations within a region; the size of the dot reflects the volume of flows within that region. The width of the line connecting two dots indicates the magnitude of flows between regions, while the arrow at the end of the line denotes the direction of movement.

In contrast to the static maps in Figure 4.6, the interactive map offers a more intuitive and transparent representation of regional connectivity in terms of mobility flows, particularly for rural residents. For example, the global view highlights a strong flow from Duluth to the Twin Cities Metro Area for urban residents, and the local scale clearly shows strong connections between Minneapolis and Saint Paul. For rural residents, the region northwest of Minneapolis shows notable connections with several other regions to the west, north, and south in the global view. When zooming into this region, the local view reveals two subregions – one to the northeast and the other to the southwest – with relatively large volumes of flows between them.



**Figure 4.11 Example view of the scale-free interactive map**

## **2. Interactive Online Mapping**

We developed two online mapping applications that enable the exploration of flow patterns related to work and frequent activities at the census tract level<sup>8</sup>. Figure 4.12 displays the application interface along with an example spatial query result. The top-left panel allows users to customize data layers and map symbology. The bottom-left panel presents basic information about the currently selected tract, while the map on the right visualizes the flows originating from that tract. Based on the information, we can conclude that the selected census tract in the rural area has a relatively high sample rate of 7.83%

---

<sup>8</sup> Home to Work by Census Tract: <https://cslodyit1998.shinyapps.io/HometoWork/>. Home to Frequent Activities by Census Tract: <https://cslodyit1998.shinyapps.io/HometoFrequent/>

and has connections to census tracts mostly toward southwest. The map is publicly accessible with limited number of concurrent users.

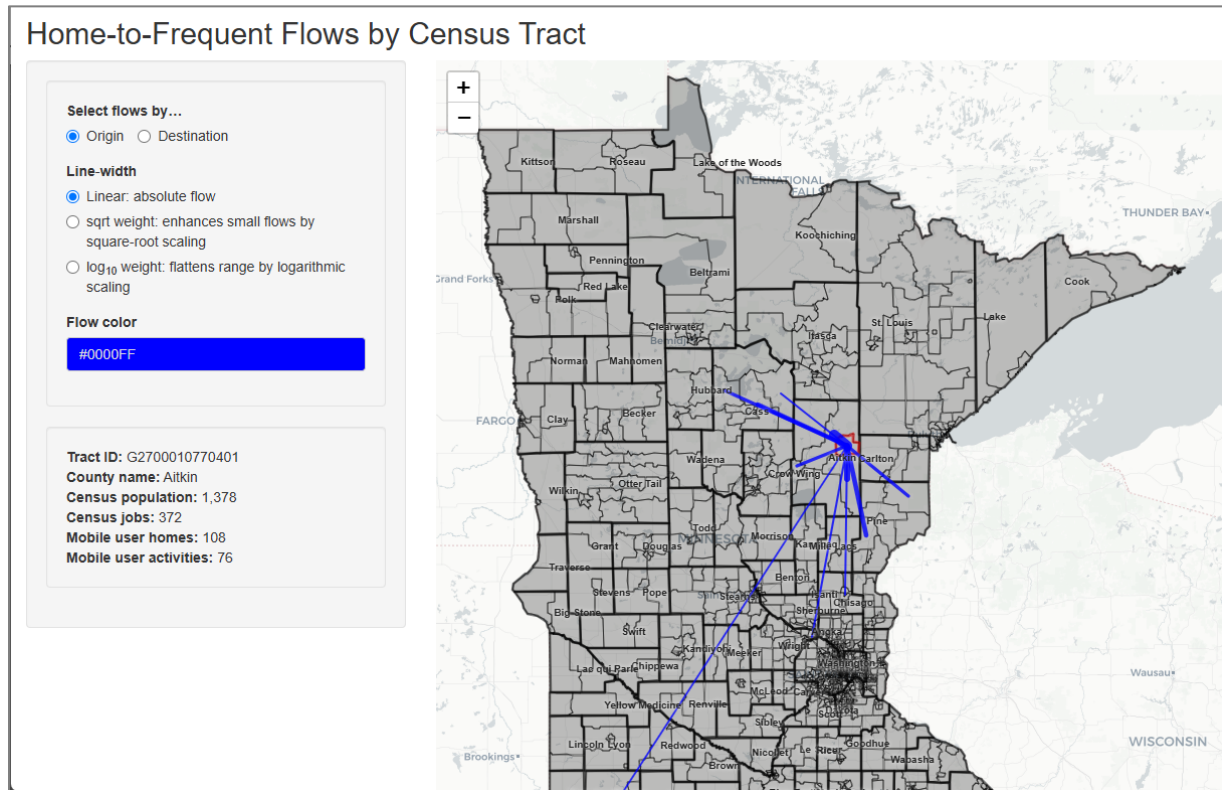


Figure 4.12 Visualization of out-flows from a rural census tract and

## CHAPTER 5: SUMMARY OF KEY FINDINGS

This chapter summarizes the key findings of this research project.

- **Mobile phone data has a higher sampling rate in rural areas.** Analysis of detected home locations from mobile phone data reveals that rural areas typically have a higher sampling rate – 6% or higher – compared to urban and suburban areas. This suggests that mobile data coverage is relatively robust in rural regions and provides a reliable basis for analyzing user mobility patterns, despite concerns that such areas may be underrepresented in large-scale digital datasets.
- **Rural users exhibit distinct mobility behaviors.** Four rural user groups have mobility patterns that are noticeably different from one another. Specifically, each rural user group shows unique visitation frequencies, spatial distributions, and temporal behaviors. These patterns reflect the diverse and often less-structured daily routines typical in rural settings.
- **Work-related activities differ in rural areas.** Fewer rural users appear to have fixed work locations during standard working hours. Instead, they engage in more frequent activities that are likely work-related but do not conform to conventional job patterns with regular working hours. This indicates the presence of more flexible or nontraditional employment arrangements in rural areas, such as agricultural and seasonal works.
- **Rural users travel farther for work and frequent activities.** Rural residents generally travel longer distances from home to reach their workplaces and frequent activity locations compared to users in urban and suburban environments. This is likely due to the more dispersed nature of services, jobs, and amenities in rural areas, which increases the spatial footprint of daily mobility.
- **Rural users have larger and more isotropic activity spaces.** Activity space analysis shows that rural users tend to have larger and more evenly shaped spatial footprints, rather than elongated or corridor-like patterns often seen in urban areas. This reflects the need to travel in multiple directions to access various dispersed opportunities in rural settings, resulting in activity spaces that are broader and more radially balanced.
- **Daily trip distances are shorter, but overall activity space is larger.** Interestingly, although rural users tend to travel farther for specific activities such as work, their average daily travel distance is generally shorter than that of suburban users. Despite this, their activity spaces are larger, which suggests less frequent but more spatially distributed travel patterns that encompass a wider area with fewer concentrated nodes.
- **Scale-free and tract-level interactive maps reveal multi-scale mobility patterns.** The development of a scale-free interactive map allows for seamless exploration of mobility flows across spatial scales from statewide to local patterns. This tool enhances the ability to detect meaningful flow structures, such as strong inter-urban connections and nuanced intra-regional interactions. The flexibility to zoom in and out without loss of detail and select interested census tract enables a comprehensive understanding of how mobility behaviors vary spatially, particularly within and/or across tracts.