Collaborative Research for USDOT Tier 1 UTC: Urban Mobility & Equity Center

**Equity in Accessibility to Opportunities: Insights, Measures, and Solutions based on Mobile Device Location Data**

Project PI: Dr. Lei Zhang

Herbert Rabin Distinguished Professor of Civil Engineering

Director, Maryland Transportation Institute, University of Maryland (UMD)

Phone: (301) 405-2881; Email: [lei@umd.edu](mailto:lei@umd.edu)

Project Co-PI: Dr. Hyeon-Shic Shin

Assistant Professor

School of Architecture and Planning, Morgan State University (Morgan)

Phone: (443) 885-3208; Email: hyeonshic.shin@morgan.edu

Project Manager: Dr. Sepehr Ghader

Assistant Director, Maryland Transportation Institute

University of Maryland, College Park, Maryland, 20742, United States

Phone: (703) 638-4046; Email: [sghader@umd.edu](mailto:sghader@umd.edu)

Project Team: Aref Darzi, Guangchen Zhao, Aliakbar Kabiri

# ABSTRACT

This report summarizes the study of accessibility to opportunities among different population groups and neighborhoods in Baltimore City. The study is the first of its kind in utilizing observed multimodal mobile device location data from individual devices to systematically study accessibility to opportunities. Passively collected mobile device location data used in this study reveal day-to-day travel patterns of more than 25% of the U.S. population for an entire year across the nation. To showcase the application of this data, we selected the Baltimore city as our testbed. This new data source with very high sampling rates, combined with point of interest data and census data, allows us to analyze how residents in each neighborhood travel to work or seek their essential needs such as food and healthcare. The study introduces a data-driven accessibility measure based on the observed location data, which can also be calculated using individual-level outputs of a typical activity-based model. Research findings directly identify accessibility gaps among neighborhoods. In addition to the above, accessibility and equity measures from mobile device location data are compared with traditional measures, and the comparison results are discussed. Furthermore, this study draws on information from the data-driven method to capture the differences in accessibility among different income groups.

# TABLE OF CONTENTS

[ABSTRACT 1](#_Toc68868905)

[TABLE OF CONTENTS 2](#_Toc68868906)

[LIST OF FIGURES 3](#_Toc68868907)

[LIST OF TABLES 4](#_Toc68868908)

[1. INTRODUCTION 5](#_Toc68868909)

[2. LITERATURE REVIEW 7](#_Toc68868910)

[3. TRADITIONAL APPROACH 9](#_Toc68868911)

[3.1. Primal Accessibility 9](#_Toc68868912)

[3.2. Dual Accessibility 10](#_Toc68868913)

[4. DATA-DRIVEN APPROACH 12](#_Toc68868914)

[4.1. Data 12](#_Toc68868915)

[4.1.1. Mobile device location data 12](#_Toc68868916)

[4.1.2. Point of Interest (POI) data 12](#_Toc68868917)

[4.1.3. American Community Survey 12](#_Toc68868918)

[4.2. Methodology 13](#_Toc68868919)

[4.2.1. Home location and work location imputation 13](#_Toc68868920)

[4.2.2. Trip identification 13](#_Toc68868921)

[4.2.3. Trip purpose identification 15](#_Toc68868922)

[4.2.4. Accessibility measure 15](#_Toc68868923)

[4.3. Results 16](#_Toc68868924)

[4.3.1. Job accessibility 16](#_Toc68868925)

[4.3.2. Food and healthcare accessibility 17](#_Toc68868926)

[5. STATISTICAL TESTING OF ACCESSIBILITY DIFFERENCE AMONG INCOME GROUPS 18](#_Toc68868927)

[5.1. Job Accessibility 18](#_Toc68868928)

[5.2. Food and Healthcare Accessibility 19](#_Toc68868929)

[6. CONCLUSION 21](#_Toc68868930)

[REFERENCES 23](#_Toc68868931)

# LIST OF FIGURES

[Figure 1. Multimodal mobile device location data for studying accessibility 7](#_Toc68869786)

[Figure 2.The percentile distribution of number of job opportunities available in each time threshold 10](#_Toc68869787)

[Figure 3. The percentile distribution of average time to job opportunities within a certain time threshold 11](#_Toc68869788)

[Figure 4. Heat map of observed home location density (a) and work location density (b) in the Baltimore and Washington, D.C., metropolitan area (device/km2) 13](#_Toc68869789)

[Figure 5. Recursive algorithm for trip identification 15](#_Toc68869790)

[Figure 6. (a) Density of the data-driven job accessibility measure, (b) regional data-driven job accessibility in the Baltimore and Washington, D.C., metropolitan area 16](#_Toc68869791)

[Figure 7. Regional accessibility for food (a) and healthcare (b) in the Baltimore and Washington, D.C., metropolitan area 17](#_Toc68869792)

[Figure 8. Polynomial regression between median household income in 2017 and the data-driven job accessibility measure 18](#_Toc68869793)

[Figure 9. Polynomial regression between median household income in 2017 and the data-driven food (a) and healthcare (b) accessibility 20](#_Toc68869794)

# LIST OF TABLES

[Table 1. Summary statistics of mobile device location data in Baltimore in September 2017 12](#_Toc47974383)

[Table 2. Summary statistics of the data-driven job accessibility measure 16](#_Toc47974384)

[Table 3. Summary statistics of the data-driven accessibility measures for food and healthcare 17](#_Toc47974385)

[Table 4. One-way ANOVA results for the data-driven job accessibility 19](#_Toc47974386)

[Table 5. t-test results for the data-driven job accessibility 19](#_Toc47974387)

[Table 6. One-way ANOVA results for the data-driven food and healthcare accessibility 20](#_Toc47974388)

[Table 7. t-test results for the data-driven food and healthcare accessibility 21](#_Toc47974389)

# INTRODUCTION

One of the earliest definitions of accessibility is “the potential of opportunities for interaction” (Hansen, 1959). Many other definitions were introduced in the literature afterward, mainly defining accessibility in terms of opportunities and travel impedance. Higher accessibility means more options to fulfill activity needs and lower generalized cost – travel time, monetary cost, etc. – to reach these options (Van Wee, 2016). Accessibility can be seen as the ease with which any land-use activity can be reached from an origin given a specific transportation system (Burns and Golob, 1976). Considering these definitions, accessibility mainly depends on two factors: opportunities and travel impedance or resistance. Accessibility depends on four main components: land-use, which represents the spatial distribution of opportunities; transport, which represents the impedance of travel for an individual to reach the opportunities; time, which represents the time constraints related to the availability of opportunities and the schedule of an individual; and person, which represents the needs and abilities of an individual to participate in activities (Geurs and Van Wee, 2004). Focusing on different components of accessibility has led to various indicators and methodologies for measuring accessibility. Models and measures of accessibility fall into four general categories: infrastructure-based, location-based, person-based, and utility-based measures (Guzman, et al., 2017, Handy and Niemeier, 1997, Kwan, 1998, Lei and Church, 2010, Liu and Zhu, 2004, Van Wee, et al., 2013). The calculation of accessibility mainly involves the generalized cost of travel between the origin zone to the opportunities in other zones (Guzman, et al., 2017). Traditional accessibility studies usually use land-use data in terms of population or employment as a proxy for measuring the attractiveness of destinations (Niemeier, 1997). They also use estimates or measures of travel-time (or distance) as an indicator of impedance.

The mainstream studies of accessibility suffer from some of the following limitations: (1) difficulties in measuring multimodal accessibility, (2) low spatial resolution of the network and activity data, (3) simplifying or neglecting access/egress travels, (4) dependence on assumptions about activity space, and (5) lack of detailed individual-level data (Chen, et al., 2017, Páez, et al., 2012, Van Wee, 2016). Challenges in accessibility studies stem from the requirement of detailed knowledge of activity locations, transportation network, and individual travel behaviors (García-Albertos, et al., 2018). Despite the challenges, accessibility is a critical subject to study, as the availability of meaningful, affordable, and accessible opportunities such as healthcare, employment, food, and education is essential for any family and individual. Social inequalities and social exclusion to a large extent result from accessibility inequalities (Guzman, et al., 2017, Ohnmacht, et al., 2009).

Passively collected location data can offer a new perspective in measuring accessibility. This report presents the first study that uses mobile device big data to systematically study accessibility with a data-driven approach. While traditional studies used land-use data as a proxy, previous research also relied on model estimates of travel-time (or distance) as an indicator of impedance. As a new data source, mobile device location data offers a novel paradigm for measuring and analyzing accessibility. First, observed location trajectories in big data reveal where travelers choose to go and how they get there every single day. Therefore, no land-use proxy, revealed-preference survey, or destination choice model is even needed. Second, travel time, distance, and other travel impedances can be directly observed from mobile device location data with timestamps. Therefore, no travel time model, traffic assignment model, or traffic simulation is needed. Last but not least, the high spatial and temporal resolution, much larger sample size, and continuous observations day after day in mobile device location data to a large extent remove previous data constraints in accessibility research. With traditional accessibility measures such as the number of jobs reachable within a certain travel time budget (e.g., 30 minutes), a low-income community next to a downtown could be considered a place with high accessibility to jobs, even if none of its residents work in the downtown. With a data-driven approach based on observed destination and travel choices, accessibility can be measured with a higher accuracy, especially for low-income and underserved communities.

This study uses observed multimodal travel big data from mobile devices to systematically study accessibility at the metropolitan scale. One month of passively collected mobile device location data from location-based services, covering more than 25% of the population, is used to measure how accessible employment, food, and healthcare are in different neighborhoods in the City of Baltimore, USA. We also compare the accessibility measure derived from mobile device location data with traditional accessibility measures. **Figure 1** shows the multimodal mobile device location data used for the study. The research findings help identify accessibility gaps among various population groups and neighborhoods.

The rest of the report is organized as follows. In section two, we review the accessibility literature to explain why we want to take a data-driven approach. In section three, we describe a traditional measure of accessibility and calculate accessibility to jobs for the Baltimore area. The traditional approach is presented as an introduction to the state of the practice and its results can serve a comparison purpose. In section four, we describe our data-driven approach and present the results for the accessibility to jobs, food, and healthcare. In section five, we use the data-driven approach results to compare the accessibility among income groups to evaluate if the more vulnerable population groups are suffering from lower accessibility to their essential needs. Section six offers concluding remarks.

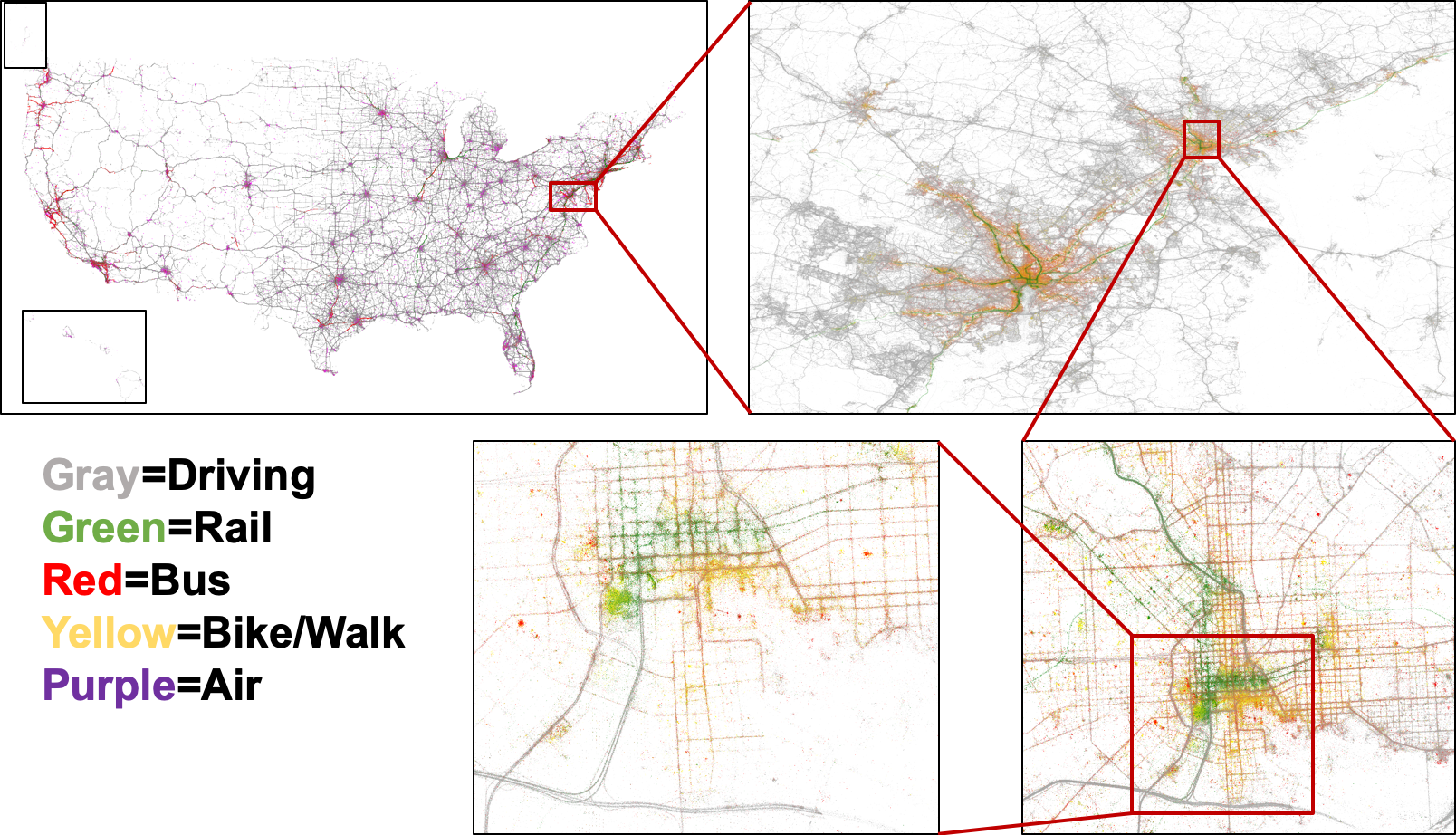


Figure . Multimodal mobile device location data for studying accessibility

# LITERATURE REVIEW

Accessibility studies are challenging, as they require detailed knowledge of activity locations (land-use component), transportation network (transportation component), and individual travel behaviors (time and person component) (García-Albertos, et al., 2018). However, these studies can reveal valuable information about the availability of social and economic opportunities for individuals in a specific neighborhood or a particular population segment (Geurs and Van Wee, 2004). Availability of employment, healthcare, food, and education are essential for healthy living; therefore, the availability of these opportunities for all groups of individuals should be studied and considered in decision-making. Improving accessibility and spatial equity of opportunities should be considered when evaluating any land-use or transportation policy and prioritizing relevant projects. Studies focusing on the relationship between accessibility and equity have highlighted the importance of equity in development and evaluation of public policies, suggesting that the approaches such as benefit-cost analysis and multi-criteria analysis fail to fully address equity concerns (Lucas, et al., 2016). The contribution of accessibility inequalities in producing social inequalities and social exclusion is highlighted in the literature (Guzman, et al., 2017, Ohnmacht, et al., 2009). It has been shown that the more vulnerable segments of the population suffer more from insufficiencies of transportation services such as longer travel time, higher exposure to pollution, and risk of accident (Titheridge, et al., 2014). Measuring accessibility can be used to analyze social inequalities and decide about transportation investments. Accessibility can help analyze the inadequacies of transportation service and provide evidence of disparities among neighborhoods. The importance of accessibility in issues of social and spatial inequalities is significantly highlighted in the literature (Guzman, et al., 2017). As a result, many planning agencies around the world have included social exclusion and equity in their planning processes (Bocarejo S and Oviedo H, 2012). However, there is no adequate tool for identifying accessibility needs; moreover, there is no consensus about the methods for causal analysis of accessibility issues and prediction of accessibility implication of projects within the literature.

Early accessibility studies utilized graphical techniques to study accessibility (Black and Conroy, 1977, Cohen and Basner, 1972). In this approach a cumulative distribution curve is created representing the distribution of the proportion of opportunities accessible by time. These curves can be created and compared for different study areas, different modes, or different socio-demographic groups, but comparison of these curves is difficult in practice. This difficulty led to single value measures of accessibility. Early single value accessibility studies measured accessibility as the cumulative number of opportunities accessible within a predefined time (Wachs and Kumagai, 1973, Wickstrom, 1971). One drawback of this measure is the arbitrary selection of time threshold (Niemeier, 1997). The other drawback is that all opportunities are treated similarly, assuming they are equally desirable for all individuals, regardless of their distance or type (Vickerman, 1974). To address the dependency of opportunity desirability on distance or travel time (impedance in general), researchers specified functional forms that allowed accessibility to decrease by increasing impedance. One such functional form widely used in the literature is the denominator of the gravity model (Handy, 1992, Niemeier, 1997). The main drawback of such measures is that they neglect the person component of accessibility. All individuals would have similar accessibility by such measures, regardless of their travel behavior and preferences (Ben-Akiva and Lerman, 1985). Also, there are issues related to spatial aggregation, defining measure of attractiveness, and the construction of the friction factor (Hanson and Schwab, 1987). These issues motivated researchers to introduce accessibility measures that depend on individual choice based on Random Utility models. In such utility-based measures, the accessibility is the realization of a set of opportunities, which can be identified by the maximum expected utility, or logsum (Niemeier, 1997).

Accessibility measures can be divided into four categories (Geurs and Van Wee, 2004); the first is the infrastructure-based measures, such as “average travel speed” or “level of congestion,” which are mainly measures of mobility or describe the network performance or level-of-service; the second is the location-based measures, such as “number of jobs accessible within 30 minutes,” that quantify the accessibility of locations on a macro-level scale, describing the accessibility of a location to the spatially distributed opportunities; the third is the person-based measures that describe individual level accessibility based on an individual’s spatial and temporal constraints, such as “ number of activities an individual can participate in a given time”; and the fourth is the utility-based measures such as logsum, measuring the economic benefits or utility of accessing spatially distributed opportunities. The benefit of infra-structure-based measures is in their simplicity and interpretability, but such measures only capture the transportation component of accessibility and they are not suitable for social analysis or economic evaluation of land-use changes. The location-based measures such as “number of opportunities” or gravity-based measures capture both transportation and land-use components of accessibility, but they are not sensitive to the person component. The person-based and utility-based measures better capture all accessibility components, but they get more complex and require more individual-level observations.

Passively collected location data can offer a new perspective in measuring accessibility. Passively collected location data can provide observed information about the chosen destinations and their corresponding impedance and reduce the dependency on models and assumptions. The measure of accessibility we introduce in this study relies on passively collected mobile device location data to provide timely and empirically observed information about accessibility.

# TRADITIONAL APPROACH

In this section, we use a typical approach to measure accessibility to job locations for the Baltimore metropolitan area. The main purpose is to compare the results with the data-driven accessibility measure we introduce in the next section. As we discussed in the previous section, in order to measure the accessibility of a region to jobs, we can look either into the travel distance or travel time of the regions as a proxy of impedance. We can also use land-use data as a proxy of opportunities. Here, we show the measure using free-flow travel time as the impedance. The travel distance approach showed similar results. The number of workers by block group from Smart Location Database, 2010, is used as the proxy for the number of job opportunities (Ramsey and Bell, 2014). To measure the impedance, we have used off-peak skim matrices from the Baltimore Metropolitan Council’s activity-based model (InSITE) (Zhang, et al., 2018).

Two types of measurements are widely used to study accessibility. The first is the number of opportunities reachable within a certain amount of impedance, typically known as the primal accessibility. The second is the average impedance to a certain number of opportunities, typically known as the dual accessibility (Cui and Levinson, 2019). Here, we present the two measures for job accessibility in the Baltimore Metropolitan Area.

## Primal Accessibility

One way of quantifying the accessibility of different regions to job opportunities is to count the number of job opportunities available to a region within a certain amount of time, like 5 minutes, 10 minutes, etc. **Figure 2** shows the results for the primal method by presenting the percentile distribution of number of job opportunities available to each block group within six different time thresholds.

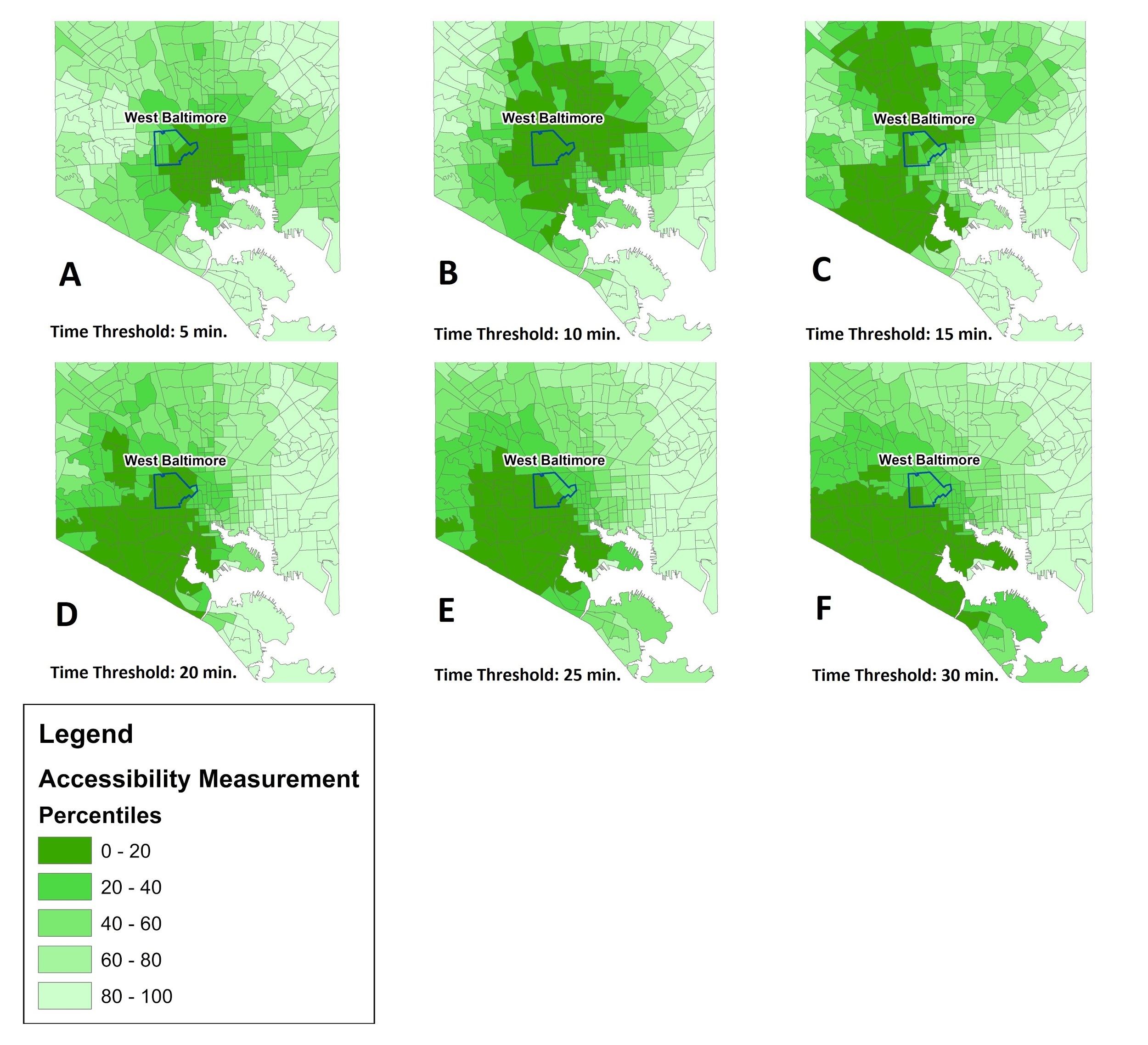
****

Figure .The percentile distribution of number of job opportunities available in each time threshold

The figure shows that when we consider a short time threshold, the center of the city and downtown area show the highest number of potential job locations. As we increase the time threshold, the locations farther from downtown, especially in the southwestern part of Baltimore City, show a higher accessibility. The reason is that block groups in the southwestern region are closer to major opportunities in other regions such as Columbia and Prince George’s County and will have a higher number of jobs accessible when we increase the time threshold.

## Dual Accessibility

We consider the average travel time to all the job opportunities that are accessible to each block group within a time threshold as the measurement of accessibility for each block group. **Figure 3** shows the percentile distribution of travel time to all the potential job opportunities reachable within six different time thresholds.

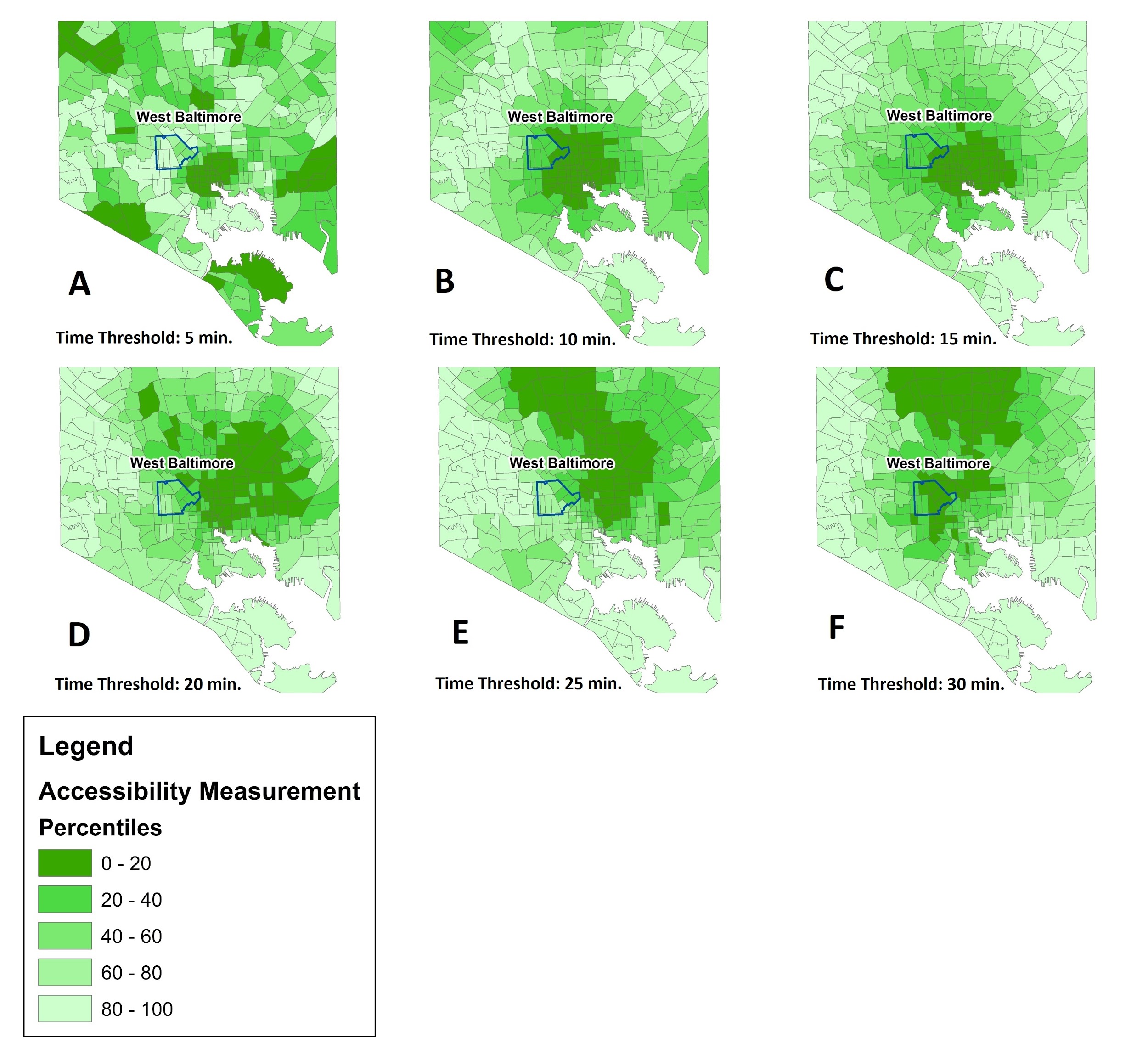
****

Figure . The percentile distribution of average time to job opportunities within a certain time threshold

The figure shows that with a short time threshold (5 minutes, **Figure 3.A**) the high-accessibility regions are widespread, clustered near local downtowns and town centers. By increasing the time threshold, the high-accessibility region moves toward downtown Baltimore. By further increasing the time threshold, places in the northern part of Baltimore City will have a better accessibility to job opportunities.

It is worth noting that in the dual method, the northern block groups were showing higher accessibility by increasing the time threshold, in contrast with the southwestern regions in the primal method. This illustrates that even though there are many job opportunities in the southwestern part of the Baltimore region, the distance to these opportunities is high. Therefore, these opportunities are not as accessible as the opportunities in the northern regions in the dual method. The comparison between the primal and the dual results highlights the importance of the accessibility measure in interpreting the accessibility results.

# DATA-DRIVEN APPROACH

In the data-driven approach, we use observed trips from mobile device location data to measure how accessible opportunities related to jobs, food, and healthcare are to each census block group. This measure depends on the residents choosing the opportunities that are under observation, not all opportunities. In this sense, the measure can lead to different results, as some opportunities that are reachable in the distance or travel time sense might not be suitable opportunities for residents of a neighborhood, as they may not be financially reachable. In the data-driven approach, we consider the observed decisions to incorporate the person component of accessibility. We first introduce the data used for the data-driven approach, then we describe the methodology used to calculate the data-driven measure for accessibility to jobs, food, and healthcare and present the results.

## Data

### Mobile device location data

We used one month of mobile device location data for September 2017 covering the Baltimore Metropolitan Area. **Table 1** summarizes several aggregate-level statistics of the data. The raw mobile device location data includes an anonymized device ID, latitude, longitude, and timestamp for each location sighting at a high level of location accuracy (less than 30 feet). Since the anonymized device ID is consistent, the mobility pattern and activity space of each device and each individual in Baltimore can be observed continuously with 117 location points per device per day on average.

Table . Summary statistics of mobile device location data in Baltimore in September 2017

|  |  |  |  |
| --- | --- | --- | --- |
| **No. of raw location points/month** | **No. of devices in the sample** | **Median no. of points per device per day** | **Average no. of points per device per day** |
| 1.2 billion | 1.1 million | 33 | 117 |

### Point of Interest (POI) data

In order to know the location of food and healthcare opportunities, we have used Point of Interest (POI) data. The POI data collected has extensive coverage and high accuracy. It has more than 39 million POIs around the United States. Each POI has a place ID, category (restaurant, fast food, etc.), name (Walmart, McDonald's, etc.), latitude, longitude, zip code, city, county and state, etc.

### American Community Survey

In order to study the accessibility of different income groups, we have used the 2017 American Community Survey 5-year estimates conducted by the U.S. Census Bureau. It contains the estimated median household income information at the census block group level. The number of block groups in our study area is 4,376, each of which is labeled with a unique geographic ID.

## Methodology

### Home location and work location imputation

We first cluster the data to obtain the main activity locations. Density-based spatial clustering of applications with noise (DBSCAN), which is one of the most common clustering algorithms for spatial data, is used to impute the location of residence and destinations for employment (if exists) for all individuals included in the mobile device data sample (Ester, et al., 1996). For home imputation based on the DBSCAN algorithm we use every day of the study period, from 7 PM to 7 AM the next day. For work location imputation, we use every working day in the study period, from 11 AM to 5 PM. For home location imputation, among all the clusters obtained from the DBSCAN algorithm for each device ID, the center of the cluster with the most observation points is identified as its home location. If a device does not have any cluster, the device is entirely removed from the sample. For the work location imputation, the cluster with the most observation points during the assumed work hours may still be the home location, as some people might work with a different schedule or just work from home. Therefore, in order to impute a fixed location, we first identify if the device has one or more clusters during the daytime which are at least 500 meters away from its imputed home location. Among them, the center of the cluster with the most observation points is identified as the device’s work location (Alexander, et al., 2015).

**Figure 4** shows the density of imputed home and work locations in the Baltimore and Washington areas.

|  |  |
| --- | --- |
| **A close up of a map  Description automatically generated**  **(a)** | **A close up of a map  Description automatically generated**  **(b)** |

Figure . Heat map of observed home location density (a) and work location density (b) in the Baltimore and Washington, D.C., metropolitan area (device/km2)

### Trip identification

Trips are the unit of analysis in our study. Mobile device location data does not include trip information. Location observations are continuously being generated while the device moves, stops, stays static, or starts a new trip. As a result, we developed a trip identification algorithm, which can detect which location observations together form a trip. We take the following steps to identify trips. The algorithm runs on the observations of each device separately.

1. Pre-Processing: We first sort device observations by time. The algorithm assigns a random ID to each trip it identifies. Many location points in the dataset may belong to no trips. The algorithm assigns “0” to the trip ID of these locations to tag them as static points. For every location point, we calculate distance, time, and speed between the point and its immediate previous and next points, if they exist. Three hyperparameters need to be set for the algorithm: distance threshold (300 meters), time threshold (5 minutes), and speed threshold (1.4 meters per second). The speed threshold is used to identify if a location point is recorded on the move. The distance and time thresholds are used to identify stay locations and trip ends. At this step, the algorithm identifies the device’s first observation with . This identified location point is recorded on the move, so a hashed trip ID is generated and assigned to this point. All points recorded before this point, if existing, are set to have “0” as their trip ID. Next, a recursive algorithm identifies if the next points are on the same trip and should have the same trip ID.
2. Iterative Algorithm. This algorithm checks every point to identify if they belong to the same trip as their previous point (**Figure 5**). If they do, they are assigned the same trip ID. If they do not, they are either assigned a new hashed trip id (when their ) or their trip ID is set to “0” (when their . Identifying if a point belongs to the same trip as its previous point is based on the point’s “speed to,” “distance to” and “time to” attributes. If a device is seen in a point with but is not observed to move there (), the point does not belong to the same trip as its previous point. When the device is on the move at a point (), the point belongs to the same trip as its previous point; but when the device stops, the algorithm checks the radius and dwell time to identify if the previous trip has ended. If the device stays at the stop (points should be closer than the distance threshold) for a period of time shorter than the time threshold, the points still belong to the previous trip. When the dwell time reaches above the time threshold, the trip ends, and the next points no longer belong to the same trip. The algorithm does this by updating “time from” to be measured from the first observation in the stop, not the point’s previous point. The algorithm may identify a local movement as a trip if the device moves within a stay location. To filter out such trips, all trips that are shorter than 300 meters are removed.



Figure . Recursive algorithm for trip identification

### Trip purpose identification

Our purpose is to study accessibility to job, food, and healthcare opportunities. As a result, we need to identify if a trip is made to the person’s job location, a food-related POI, or a healthcare-related POI. We identify trips from home to any of these opportunities by taking the following steps:

1. Filter out all trips in September 2017 taken by devices with identified home location;
2. For each of the trips, if the distance between either of the trip ends and the imputed home location of that device is within 200 meters, this trip is labeled as a “home-based trip”;
3. In each “home-based trip,” for the trip end that is not the home location, consider a 200-meter-radius circle around the trip end. If it includes the imputed work location, the trip is labeled as “home-based work.” If not, find the closest POI location to the trip end located inside the circle. If the category of the POI is related to food, this trip is labeled as “home-based trip for food”; similarly, if the category of the POI is related to healthcare, this trip is labeled as “home-based trip for healthcare.”

### Accessibility measure

For each block group, we calculate the average observed travel time for all “home-based work” trips of its residents as our data-driven measure of accessibility to jobs. Similarly, we calculate the average observed travel time for all “home-based food” trips and “home-based healthcare” trips of the block group residents as the data-driven measures of accessibility to food and healthcare. Therefore, we take the following steps:

1. For each device, find the average trip duration of home-based trips for job/food/healthcare, as its representative travel time to job/food/healthcare.
2. For each census block group, find the average travel time to job/food/healthcare of all the individuals who are imputed to reside there as the block group’s accessibility to job/food/healthcare.

## Results

### Job accessibility

**Table 2** shows the descriptive statistics for the data-driven job accessibility calculated based on the methodology described. We can see that the average job accessibility in the area is around 28 minutes and the distribution is skewed, as the median is about 10 minutes shorter than the mean. It shows that there are high observations that increase the mean value. We can also see the distribution in **Figure 6a**.

Table . Summary statistics of the data-driven job accessibility measure

|  |  |
| --- | --- |
| **Statistic** | **Minutes** |
| **Minimum** | 0.01 |
| **25th percentile** | 7.60 |
| **Median** | 18.55 |
| **Mean** | 27.76 |
| **75th percentile** | 36.12 |
| **Maximum** | 180.00 |
| **ST.D.** | 30.17 |

|  |  |
| --- | --- |
| **(a)** | **(b)** |

Figure . (a) Density of the data-driven job accessibility measure, (b) regional data-driven job accessibility in the Baltimore and Washington, D.C., metropolitan area

**Figure 6b** shows the job accessibility results at different census block groups. We can see a significant difference between the results of the data-driven approach with the results from the traditional approach, whether dual or primal (**Figure 2** and **Figure 3**). The key difference is that the clear patterns observed in the traditional approach results are not observed in the data-driven results. The patterns in the traditional approach results are direct outcomes of the models and assumptions, and the simple models lead to simple accessibility patterns as observed in **Figure 2** and **Figure 3**. However, the data-driven results show that the accessibility can be different between close neighborhoods, as the residents of these neighborhoods may have different socio-demographic characteristics or may show different travel patterns.

### Food and healthcare accessibility

**Table 3** shows the data-driven accessibility to food and healthcare based on the described methodology. We can see that the accessibility to healthcare is generally lower (higher number) than the accessibility to food. An average resident of the study area needs 25 minutes to reach food opportunities and 27 minutes to reach healthcare opportunities. **Figure 7** shows the accessibility to food and healthcare in different regions. We can again see that close neighborhoods may show different access to food and healthcare based on the data-driven approach. For instance, we can see that the denser zones in Baltimore City and Washington, D.C., show different colors, as they vary in their resident’s income distribution.

Table . Summary statistics of the data-driven accessibility measures for food and healthcare

|  |  |  |
| --- | --- | --- |
| **Statistic** | **Food (Minutes)** | **Healthcare (Minutes)** |
| **Minimum** | 1.00 | 1.00 |
| **25th percentile** | 16.60 | 15.61 |
| **Median** | 25.76 | 27.40 |
| **Mean** | 30.15 | 35.16 |
| **75th percentile** | 39.18 | 46.21 |
| **Maximum** | 119.98 | 179.98 |
| **ST.D.** | 18.96 | 27.84 |

|  |  |
| --- | --- |
| **(a)** | **(b)** |

Figure . Regional accessibility for food (a) and healthcare (b) in the Baltimore and Washington, D.C., metropolitan area

# STATISTICAL TESTING OF ACCESSIBILITY DIFFERENCE AMONG INCOME GROUPS

In the two previous sections, we presented two different methods of quantifying accessibility. In this section, we will apply the data-driven approach to a bigger study area, the State of Maryland, and see whether there is any difference between the accessibility of the low-income, medium-income, and high-income residents to their observed job locations, food stores, and healthcare centers. We defined low-income, medium-income, and high-income groups based on the first and the third quartiles of income in the area. As a result, block groups with the median household income less than $55,000, between $55,000 and $105,000, and higher than $105,000 per year were categorized as low-income, medium-income, and high-income, respectively.

## Job Accessibility

**Figure 8** presents an interesting result: Both the low-income and high-income communities show relatively high job accessibility (low values), while the mid-income communities show relatively lower accessibility (higher values). The possible explanation could be that people in low-income communities mostly work in sales and service. These types of job opportunities tend to be widely spread, so there would be no need for their workers to travel far from home, considering the economic and time cost for commuting. Therefore, the characteristics of the job lead to higher accessibility to work for low-income communities. As for people who live in mid-income and high-income communities, their occupations are most likely to be related to business, finance, education, government, medical services, and information technology, which tend to be concentrated in downtown or other sub-center areas. In these areas, the house prices and rents are usually high for mid-income workers, while affordable for high-income workers. Therefore, although the work locations of people in mid-income communities and high-income communities tend to be close, different affordability for cost of living leads to different accessibility to work.

**A screenshot of a cell phone

Description automatically generated**

Figure . Polynomial regression between median household income in 2017 and the data-driven job accessibility measure

The one-way ANOVA test (**Table 4**) and pairwise t-test (**Table 5**) are both in line with the observation that there is a statistically significant difference among the data-driven job accessibility of different income groups, with the medium-income group having lower job accessibility than the other two groups. The one-way ANOVA shows that the difference in the means is significant. The pairwise t-test shows what pairwise differences are statistically significant.

Table . One-way ANOVA results for the data-driven job accessibility

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Mean of low-income** | **Mean of medium-income** | **Mean of high-income** | **F value** | **Significant?**  **(P-value <=0.05)** |
| **ONE-WAY ANOVA** | 31.28 | 32.76 | 31.96 | 6.339 | YES (~0.0018) |

Table . t-test results for the data-driven job accessibility

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Mean of the first group** | **Mean of the second group** | **t-value** | **Significant?**  **(P-value <0.05)** |
| Low- and Medium-income | 31.28 | 32.76 | -3.017 | YES, (~ 0.0025) |
| Low- and High-income | 31.28 | 31.96 | -1.139 | NO, (~ 0.1873) |
| Medium- and High-income | 32.76 | 31.96 | 2.188 | YES, (~ 0.0288) |

## Food and Healthcare Accessibility

The result for the food and healthcare accessibility shows a different trend with income. **Figure 9** shows that both accessibilities increase by increasing income (lower value means higher accessibility). As a result, higher-income neighborhoods are shown to be more accessible to essential needs such as food and healthcare.

The statistical tests confirm the significance of the results. The one-way ANOVA for both food accessibility and healthcare accessibility shows significant differences (**Table 6**). Pairwise tests show that the difference between the low-income and medium-income zones is not significant, but high-income zones have significantly higher accessibility for both food and healthcare.

|  |  |
| --- | --- |
| A close up of a map  Description automatically generated  **(a)** | A close up of a map  Description automatically generated  **(b)** |

Figure . Polynomial regression between median household income in 2017 and the data-driven food (a) and healthcare (b) accessibility

Table . One-way ANOVA results for the data-driven food and healthcare accessibility

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Mean of low-income** | **Mean of medium-income** | **Mean of high-income** | **F value** | **Significant?**  **(P-value <=0.05)** |
| **One-Way ANOVA for food** | 32.30 | 32.03 | 30.92 | 16.14 | YES (~0) |
| **One-Way ANOVA for healthcare** | 38.38 | 38.02 | 34.76 | 45.53 | YES (~0) |

Table . t-test results for the data-driven food and healthcare accessibility

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Mean of the first group** | **Mean of the second group** | **t-value** | **Significant?**  **(P-value <0.05)** |
| **Food** | | | | |
| **Low- and Medium-income** | 32.30 | 32.03 | 1.046 | NO, (~ 0.2958) |
| **Low- and High-income** | 32.30 | 30.92 | 4.958 | YES, (~ 0) |
| **Medium- and High-income** | 32.03 | 30.92 | 5.468 | YES, (~ 0) |
|  |  | **Healthcare** |  |  |
| **Low- and Medium-income** | 38.38 | 38.02 | 0.836 | NO, (~ 0.4032) |
| **Low- and High- income** | 38.38 | 34.76 | 7.885 | YES, (~ 0) |
| **Medium- and High-income** | 38.02 | 34.76 | 9.698 | YES, (~ 0) |

# CONCLUSION

Equity has righteously been one of the main discussions of the year 2020. The topic gained more attention with the rise of COVID19 and its disproportional effect on communities of color and dominated the social and political arena after the black lives matter protests. Societies are now more aware of the disparities in opportunities. Now is the time for us, researchers, to offer our best to overcome challenges in eliminating disparities in order to have more just and equitable societies. It is important to have a clear understanding of equity issues, based on real-world observations, to be able to design policies to fight them. Having an equal, or preferably equitable, accessibility to opportunities among different population segments can alleviate parts of the equity problems.

Fortunately, the enormous amount of data that is continuously being collected can help us monitor the accessibility situation and plan for a more equitable future. In this study, we have utilized mobile device location data to measure accessibility, based on observed chosen opportunities. We have introduced a data-driven measure and shown how it can give a different perception about the areas with accessibility limitations when compared to the traditional methods. An example of West Baltimore, a low-income neighborhood right near Baltimore’s inner harbor, a hub for high-income jobs, highlights why many traditional accessibility models may fail to correctly identify which opportunities are accessible to people, not just physically, but also financially.

It is important to note that our method takes the accessibility modeling to an extreme of only depending on observed chosen opportunities. The other extreme of accessibility modeling considers all opportunities as relevant, whether being chosen or not. A future report may focus on finding a middle ground, in which all opportunities are considered, with possible weights in their level of importance in accessibility based on observed choices. Another future direction is related to forensic analysis to evaluate transportation causes for observed accessibility issues. Identifying the cause may lead to determining possible solutions for mitigating the problems of accessibility for vulnerable neighborhoods.

# REFERENCES

Alexander, L., S. Jiang, M. Murga and M. C. González. (2015), Origin–Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data, *Transportation Research Part C: Emerging Technologies* Vol. 58, pp. 240-250.

Ben-Akiva, M. E. and S. R. Lerman. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press.

Black, J. and M. Conroy. (1977), Accessibility Measures and the Social Evaluation of Urban Structure, *Environment and Planning A* Vol. 9, No. 9, pp. 1013-1031.

Bocarejo S, J. P. and D. R. Oviedo H. (2012), Transport Accessibility and Social Inequities: A Tool for Identification of Mobility Needs and Evaluation of Transport Investments, *Journal of Transport Geography* Vol. 24, pp. 142-154.

Burns, L. D. and T. F. Golob. (1976), The Role of Accessibility in Basic Transportation Choice Behavior, *Transportation* Vol. 5, No. 2, pp. 175-198.

Chen, J., J. Ni, C. Xi, S. Li and J. Wang. (2017), Determining Intra-Urban Spatial Accessibility Disparities in Multimodal Public Transport Networks, *Journal of Transport Geography* Vol. 65, pp. 123-133.

Cohen, D. S. and C. Basner. (1972), *Accessibility – Its Use as an Evaluation Criterion in Testing and Evaluating Alternative Transportation Systems*, U.S. Department of Transportation, Federal Highway Administration.

Cui, M. and D. Levinson. (2019), Primal and Dual Access, *Geographical Analysis*.

Ester, M., H.-P. Kriegel, J. Sander and X. Xu. (1996), A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, (Ed.)^(Eds.), *Kdd*.

García-Albertos, P., M. Picornell, M. H. Salas-Olmedo and J. Gutiérrez. (2018), Exploring the Potential of Mobile Phone Records and Online Route Planners for Dynamic Accessibility Analysis, *Transportation Research Part A: Policy and Practice*.

Geurs, K. T. and B. Van Wee. (2004), Accessibility Evaluation of Land-Use and Transport Strategies: Review and Research Directions, *Journal of Transport Geography* Vol. 12, No. 2, pp. 127-140.

Guzman, L. A., D. Oviedo and C. Rivera. (2017), Assessing Equity in Transport Accessibility to Work and Study: The Bogotá Region, *Journal of Transport Geography* Vol. 58, pp. 236-246.

Handy, S. L. (1992), Regional Versus Local Accessibility: Neo-Traditional Development and Its Implications for Non-Work Travel, *Built Environment (1978-)*, pp. 253-267.

Handy, S. L. and D. A. Niemeier. (1997), Measuring Accessibility: An Exploration of Issues and Alternatives, *Environment and Planning A* Vol. 29, No. 7, pp. 1175-1194.

Hansen, W. G. (1959), How Accessibility Shapes Land Use, *Journal of the American Institute of Planners* Vol. 25, No. 2, pp. 73-76.

Hanson, S. and M. Schwab. (1987), Accessibility and Intraurban Travel, *Environment and Planning A* Vol. 19, No. 6, pp. 735-748.

Kwan, M. P. (1998), Space‐Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point‐Based Framework, *Geographical Analysis* Vol. 30, No. 3, pp. 191-216.

Lei, T. L. and R. L. Church. (2010), Mapping Transit‐Based Access: Integrating GIS, Routes and Schedules, *International Journal of Geographical Information Science* Vol. 24, No. 2, pp. 283-304.

Liu, S. and X. Zhu. (2004), Accessibility Analyst: An Integrated GIS Tool for Accessibility Analysis in Urban Transportation Planning, *Environment and Planning B: Planning and Design* Vol. 31, No. 1, pp. 105-124.

Lucas, K., B. Van Wee and K. Maat. (2016), A Method to Evaluate Equitable Accessibility: Combining Ethical Theories and Accessibility-Based Approaches, *Transportation* Vol. 43, No. 3, pp. 473-490.

Niemeier, D. A. (1997), Accessibility: An Evaluation Using Consumer Welfare, *Transportation* Vol. 24, No. 4, pp. 377-396.

Ohnmacht, T., H. Maksim and M. M. Bergman. (2009), *Mobilities and Inequality*, Ashgate Publishing, Ltd.

Páez, A., D. M. Scott and C. Morency. (2012), Measuring Accessibility: Positive and Normative Implementations of Various Accessibility Indicators, *Journal of Transport Geography* Vol. 25, pp. 141-153.

Ramsey, K. and A. Bell. (2014), Smart Location Database, *Washington, D.C*.

Titheridge, H., R. Mackett, N. Christie, D. Oviedo Hernández and R. Ye. (2014), Transport and Poverty: A Review of the Evidence.

Van Wee, B. (2016), Accessible Accessibility Research Challenges, *Journal of Transport Geography* Vol. 51, pp. 9-16.

Van Wee, B., K. Geurs and C. Chorus. (2013), Information, Communication, Travel Behavior and Accessibility, *Journal of Transport and Land Use* Vol. 6, No. 3, pp. 1-16.

Vickerman, R. W. (1974), Accessibility, Attraction, and Potential: A Review of Some Concepts and Their Use in Determining Mobility, *Environment and Planning A* Vol. 6, No. 6, pp. 675-691.

Wachs, M. and T. G. Kumagai. (1973), Physical Accessibility as a Social Indicator, *Socio-Economic Planning Sciences* Vol. 7, No. 5, pp. 437-456.

Wickstrom, G. V. (1971), Defining Balanced Transportation – A Question of Opportunity, *Traffic Quarterly* Vol. 25, No. 3.

Zhang, L., Yang, D., Ghader, S., Carrion, C., Xiong, C., Rossi, T. F., ... & Barber, C. (2018). An integrated, validated, and applied activity-based dynamic traffic assignment model for the Baltimore-Washington region. *Transportation Research Record*, *2672*(51), 45-55.