NATIONAL INSTITUTE FOR CONGESTION REDUCTION

FINAL REPORT

MARCH 2024

Uses and Limitations of Big Data for Evaluating Transportation Equity

Tia Boyd, Taylor Dinehart, Kristine Williams

National Institute for Congestion Reduction
University of South Florida
Center for Urban Transportation Research | University of South Florida

NICR
NATIONAL INSTITUTE FOR CONGESTION REDUCTION

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Uses and Limitations of Big Data for Evaluating Transportation Equity

Prepared by

Tia Boyd, Taylor Dinehart, Kristine Williams
University of South Florida

Prepared for

National Institute for Congestion Reduction
University of South Florida
Center for Urban Transportation Research

4202 E. Fowler Avenue, ENG030, Tampa, FL 33620-5375
nicr@usf.edu

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Uses and Limitations of Big Data for Evaluating Transportation Equity		5. Report Date March 2024 6. Performing Organization Code
7. Author(s) Tia Boyd, Taylor Dinehart, Kristine Williams		8. Performing Organization Report No.
9. Performing Organization Name and Address Center for Urban Transportation Research University of South Florida 4202 E. Fowler Avenue Tampa, FL 33620-5375		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 69A3551947136, #79075-25
12. Sponsoring Organization Name and Address U.S. Department of Transportation University Transportation Centers 1200 New Jersey Avenue, SE Washington, DC 20590 United States National Institute for Congestion Reduction 4202 E. Fowler Avenue Tampa, FL 33620-5375 United States		13. Type of Report and Period Covered Final Report, [May 2020 – November 2023] 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Social equity has come to the forefront in recent years as a hot topic and essential aspect of transportation planning. Technological advancements are allowing planners to collect more precise data and conduct analysis in real time, as well as improve visualization and forecasting abilities. "Big data" has immense potential to enrich transportation planning and our understanding of equity issues, even as it poses potential limitations regarding social equity. Examples abound of big data applications in transportation, including transportation equity analysis, yet these examples are dispersed in different agency studies, company websites, and publications. This synthesis project scanned and mined the literature, big data company websites, and other sources to identify and document how various entities are using big data to evaluate transportation equity considerations, including how companies that market big data have designed their platforms to support equity analysis. It also documents examples and considerations relative to the limitations of big data in addressing equity. The synthesis culminates in a summary of key applications of big data for equity analysis and ongoing gaps in the use of big data to evaluate social equity in transportation planning.

17. Key Words Big data, transportation equity, underserved communities, transportation disadvantaged		18. Distribution Statement	
19. Security Classification (of this report) Unclassified.	20. Security Classification (of this page) Unclassified.	21. No. of Pages 40	22. Price

Table of Contents

Tables

Executive Summary

Both equity and big data are rapidly advancing as topics of interest in transportation. These topics are identified as (1) priorities for the Biden administration (Executive Order No. 13985, 2021; Executive Order No. 14091, 2023), (2) 2021 planning emphasis areas by the Federal Highway Administration (FHWA) and the Federal Transit Administration (FHWA & FTA, 2021), (3) drivers of change by the American Planning Association (APA) (APA, 2021), and (4) key areas of interest for myriad other agencies and organizations.

Transportation equity analysis relies on data to evaluate the impacts of decisions and their outcomes on affected populations. Through these evaluations, agencies can more clearly understand how people use the transportation system, identify where the system can be improved to meet the needs of all people, and evaluate the distributional equity of transportation investments and impacts. To provide a better understanding of the state of practice for big data in transportation equity analysis, this document synthesizes available research on big data to describe the various uses and limitations of big data applications for evaluating transportation equity at the time the literature review was conducted.

Methodology

This research included a review of several topics related to big data in transportation. The first group of literature addressed the broad uses, applications, and limitations of big data for general transportation analysis purposes. The second group of literature addressed the uses and limitations of big data for transportation equity analysis. After the literature review, the research team interviewed the staff of the Hillsborough Transportation Planning Organization (TPO) on their practices for transportation equity analysis, focusing on their use of big data for this purpose. The research team also included a scan of the websites for three big data platforms that emphasize equity in their products. These platforms were identified and selected based on their inclusion in the existing literature on big data and transportation equity and the research team's existing knowledge of these platforms. The findings from each of these steps were synthesized and documented in the research report.

Findings

A wide variety of big data types and sources are used in the transportation sector for research, planning, and operations. The most widely used sources include smart card and automatic fare collection, GPS and automatic vehicle location (AVL), sensors, smartphone data, and web and social media. The general limitations of big data for transportation analysis include data availability, data quality, bias, agency capacity, and security.

The existing literature and federal legislation and guidance confirm that big data can support transportation equity analysis and equity-related decision-making, while also acknowledging that it has the potential to create inequities. Transportation equity analyses are informed by applications that involve identifying and evaluating the travel behavior of underserved populations and evaluating the equity impacts of policy decisions by using various data types (Desouza & Smith, 2016; Fanibi, 2022; Griffin et al., 2018). Through the literature review, a four-step framework for transportation equity analysis was identified, which includes (1) defining and evaluating the population, (2) assessing the distribution of benefits and burdens, (3) identifying and measuring equity and inequity, and (4) evaluating progress toward equity.

While big data can support each of these steps, various aspects of big data can create several challenges, including the potential for known or unknown inequitable outcomes (Desouza & Smith, 2016; Ruijer et al., 2022). Commonly cited limitations of big data for transportation equity analysis include data availability and gaps in available data, disparate impact and bias, aggregated data, and predictive policing and privacy. In general, concerns may stem from methodological, technological, political, and/or epistemological limitations (Chen et al., 2022). While these limitations are also applicable to big data in general, their effects on the transportation equity analysis process can be detrimental to already overburdened and underserved populations.

Beyond the general limitations of big data for transportation equity analysis, transportation agencies face several challenges in their use. These challenges include funding, data reliability, and political environments. Reassessment of data sources and agency evaluation methods for equity paired with innovative strategies can support the agencies' efforts to advance transportation equity using these data.

Several big data platforms emphasize their products' applicability for transportation equity analysis. These platforms highlight the availability of disaggregated data through their platforms, accenting the potential to address one of the long-standing critiques of equity analysis and limitations of large datasets—aggregation. Transportation practitioners using big data for equity analysis should consider working closely with these companies to ensure that the data and resulting analysis meet the needs of affected communities.

Recommendations

Big data can support efforts to advance equity in transportation. However, this is a process that first requires practitioners to acknowledge big data's limitations and identify strategies to mitigate the potential adverse outcomes of using big data for transportation equity analysis. To help practitioners meet this need, the following list includes suggested areas of future consideration for big data in transportation equity analysis:

- An assessment of current practices on the use of big data in transportation equity analysis.
- The development of dynamic resources, tools, and strategies to mitigate bias from big data in transportation analysis.
- Standardized performance measures, benchmarks, and reporting for transportation equity analysis, specifically focused on the use of big data for this purpose.
- An analysis of the stages in data life cycles (pre-collection planning, data processing, data storage, archiving, analysis, decision-making, visualization, etc.) to identify opportunities for integrated equity considerations.

Chapter 1. Introduction

Both equity and big data are rapidly advancing as topics of interest in transportation. These topics are identified as (1) priorities for the Biden administration (Executive Order No. 13985, 2021; Executive Order No. 14091, 2023), (2) 2021 planning emphasis areas by the Federal Highway Administration (FHWA) and the Federal Transit Administration (FHWA & FTA, 2021), (3) drivers of change by the American Planning Association (APA) (APA, 2021), and (4) key areas of interest for myriad other agencies and organizations. Transportation equity analysis relies on data to evaluate the impacts of decisions and their outcomes on affected populations. Through these evaluations, agencies can better understand how people use the transportation system, identify where the system can be improved to meet the needs of all people, and evaluate the distributional equity of transportation investments and impacts. As a result, the type and quality of data influence the effectiveness of the equity analysis.

Equity is described as either horizontal or vertical. Horizontal equity considers everyone similarly, regardless of their different attributes, whereas vertical equity provides for groups based on their specific attributes and needs (Chen et al., 2022; Litman, 2023). Transportation equity analysis addresses distributional equity assessing the dissemination of benefits and burdens to various populations (Williams & Golub, 2017). Distributional equity can be either proportional or restorative. Proportionality considers if underserved communities benefit from transportation investments in proportion to the broader population. Restorative justice considers if transportation investments are distributed in a manner that reduces inequalities over time. Furthermore, equity is typically defined within the frame of procedural equity (*who is involved and to what degree?*), geographic equity (*how are impacts spatially distributed?*), or social equity (*how are impacts distributed between population groups?*) (Wennink & Krapp, 2020). Therefore, equity in transportation can be broadly defined as justice and fairness in the distribution of benefits and burdens.

Recent assessments of equity data sources, metrics, and analytic methods have revealed that equity analysis is improved through “big data” applications (Chen, 2020; ITE, 2020). Yet, there are gaps in data availability and potential discrepancies in metrics, resulting in several limitations for accurate equity analysis (Pereira & Karner, 2021). For example, location-based data from cell phones is being used as an alternative to traditional surveys and is reported as being more representative of the population (StreetLight Data, n.d.). But, because these platforms use smartphone data and Global Positioning System (GPS) devices, they may not capture persons who do not have access to digital devices, broadband, or a personal vehicle—persons who are oftentimes already classified as underserved (Vogels, 2021a). These gaps have broad consequences for the transportation field and practitioners’ efforts to advance equity. This is just one example of how the limitations of big data affect equity in transportation. Additional examples are explored in the *Limitations of Big Data in Transportation Equity Analysis* section of this report.

In the *2022 Trend Report for Planners*, the American Planning Association (APA) identified seven key insights related to big data, each of which have varying equity implications. These insights include the following (APA & Lincoln Institute of Land Policy, 2022, pp. 34–35):

- **Individual identities:** Data collection on demographics and population will require new approaches that can adequately and accurately reflect diversity.
- **Equity, diversity, and inclusion (EDI) in planning efforts:** There is a rapid growth in mandates from local, state, and federal programs that require communities to explicitly measure and identify equity, diversity, and inclusion efforts in local planning work.

- **Intersectionality:** The profession needs to prepare to do a better job of reaching out to people with lived experiences at the intersection of multiple social identities (e.g., disability and class, in addition to race and gender).
- **Scoring systems:** Planners should prepare to consider the risks of using scoring systems in their work, such as when attempting to measure neighborhood success. [These scoring systems can "further formalize harmful biases…"]
- **Crowdsourcing:** Crowdsourcing can be a formal iteration of civic tech, and one outcome could be large amounts of data provided directly from residents that reflect the preferences of population groups.
- **Wearable technology:** The constant development and refinement of new tools for data collection will continually affect how planners collect data, and what types of data they have access to.
- **Data protection and privacy:** As planners get more access to new kinds of data, they will also need to understand how these data protection regulations affect how they can use data.

This document synthesizes available research on big data to describe the various uses and limitations of big data applications for evaluating transportation equity at the time the literature review was conducted. The first section defines big data and describes its broad uses, applications, and limitations for general transportation analysis purposes. The second section describes the uses and limitations of big data specifically for transportation equity analysis, including a case example from a metropolitan planning organization (MPO) in Florida. The third section provides examples of select big data platforms that emphasize equity in their products. The document concludes with a summary of the report and future research needs.

Chapter 2. Big Data in Transportation

The current and potential applications for big data in the transportation sector are extensive. The ability of big data and machine learning to measure patterns and predict trends has led to their application in a broad range of transportation topics such as the maintenance and operation of transportation assets or assessing public health. The introduction of concepts such as artificial intelligence, machine learning, and big data into the transportation sector, however, has not eliminated bias or equity concerns; rather, it has introduced new avenues for them to become evident in transportation planning activities. Due to the complexity of the topic of big data, there are several big data life cycle models that differ depending on field or organization (Arass et al., 2017; Pouchard, 2016; Vydra et al., 2021). For broad application by the transportation community and with a sharp focus on equity considerations, a truncated life cycle is presented here consisting of data collection and application. The section concludes with identified limitations of big data applications in transportation analysis.

Data Sources and Data Collection

Historically, data for transportation analysis came from what many researchers refer to as "traditional sources" or "traditional techniques." These techniques include manual traffic counts, machine traffic counts using tubes and air switches, traffic cameras, induction loops, manual vehicle occupancy samples, travel surveys (e.g., parking surveys, transit onboard surveys, hotel and visitor surveys, and household travel surveys), transit ridership counts, manual pedestrian counts, interviews, and census data (Klein et al., 2006; Meyer, 2016). The census, in particular, has been an important source for socioeconomic data used for transportation analysis and planning (Lawson et al., 2019).

Today, advancements in information and communication technology have led to new sources of data and new data collection techniques. Given that these data are large in volume, are often generated automatically and dynamically, and are comprised of varying formats (e.g., sensor data, social media posts and metadata, GPS information, and video and image output), they are aligned with the generally accepted definition of big data (Beyer & Laney, 2012; IBM, 2014; Laney, 2001; Scott, 2015). A review of the literature revealed a wide variety of big data types and sources used in the transportation sector for research, planning, and operations. The most widely used sources of data include the following:

- Smart card and automatic fare collection
- GPS and automatic vehicle location (AVL)
- Sensors
- Smartphones
- Web and social media

Each of these data sources are described briefly in the next sections. Many data sources are interrelated or used in conjunction with other sources so routinely that they are lumped together in the literature, therefore certain data types span across multiple categories. This list is not exhaustive and only represents data types and sources referenced frequently and heavily in literature pertaining to big data and transportation.

Smart Card Data and Automatic Fare Collection (AFC)

Smart cards are the size of a credit card and are embedded with an integrated circuit (IC) chip or a radio-frequency identification (RFID) chip. The card can be inserted into a chip reader (in the case of an IC chip) or read by passing it near a terminal (in the case of a contactless RFID chip) (Rankl & Effing, 2010). The purpose of the smart card is to enable quick and easy fare collection, also called automatic fare collection (AFC). Depending on the sophistication of the card, it can also be used to collect various data such as boarding time, pick-up locations, transfers, and travel mode. The Transit Access Pass (TAP) card in Los Angeles is a multimodal smart card that collects data on bus use, rail, bikeshare, and shuttle services throughout the region. Smart cards like the Octopus card in Hong Kong and the Oyster card in London that require passengers to engage the card terminal at both pick-up and drop-off, further fine-tune the transportation data collected from the smart card (Zannat & Choudhury, 2019). Oftentimes smart cards contain some amount of personal identification data that connects it to a unique traveler (Rankl & Effing, 2010). Additionally, smart cards like the EZ-Link card in Singapore are used for functions beyond public transit such as parking and retail, allowing for the potential to evaluate trip purpose (Anwar et al., 2016).

GPS and Automatic Vehicle Location (AVL)

GPS (Global Positioning System) is a system comprised of satellites, ground stations, and receivers. Receivers emit electromagnetic waves to measure their distance to at least four satellites and use that information to triangulate their position (Crato, 2010). GPS embedded devices are ubiquitous today and include smartphones, tablets, devices worn on the body (e.g., Fitbit, Garmin, Apple Watch), fleet tracking systems (e.g., automatic vehicle location and systems for tracking e-scooter, bikeshare equipment, and taxis), various manned and unmanned aircraft, and GPS trackers attached to privately owned vehicles (Desouza & Smith, 2016).

In the case of AVL, a GPS device installed on a fleet vehicle sends latitude, longitude, time, and date to an AVL database at varying time intervals. This information, referred to as GPS traces, can then be combined with smart card data or other data for further analysis and planning research (Neilson et al., 2019; Transportation Research Board, 2014; Welch & Widita, 2019; Zannat & Choudhury, 2019). GPS points and traces can also be obtained from smartphones and any other GPS device that emits GPS information (Lawson et al., 2019; Transportation Research Board, 2014).

Sensors

Sensors have been used to collect transportation data for years. However, the sheer volume of sensors in use, their ability to broadcast data automatically and continuously, and their integration into the Internet of Things (IoT) has increased the use of sensors as a source of big data (Zhang et al., 2016). Sensors are devices that are placed in a fixed location or attached to a moving object (e.g., bus, bicycle, person) to gather information (Welch & Widita, 2019). Some examples of sensors include devices with infrared, microwave, magnetic, pneumatic, piezoelectric, ultrasonic, acoustic, or light detection and ranging (LIDAR) detection capabilities, devices enabled with Bluetooth, Wi-Fi, or radio frequency identification (RFID) technology, and devices that can capture and process video and audio (Zhu et al., 2019). Typical sensors used for transportation planning and research are included on the following (Lu et al., 2021; Neilson et al., 2019; Torre-Bastida, 2018):

- **Public transit vehicles** to count passengers as they board and track weight adjustments. (This type of information is referred to as Automatic Passenger Count (APC) and is often used in conjunction with AVL and AFC data.)
- **Vehicles of all types** to track speed, location, temperatures, and as onboard diagnostic devices that track equipment health.
- **Roadways, sidewalks, waterways, and rail** to monitor multimodal traveler movement, traffic density, vehicle speed, wait times, and pavement or rail condition.

- **Buildings, bridges, parking spots, micromobility storage areas, high-demand curb areas and other infrastructure** to gather information about usage, traffic, the environment, safety, public health, and conditions of the infrastructure itself.
- **Streetlights and traffic lights** to monitor performance.
- **Connected and autonomous vehicles (CAV)** for driving assistance and collision avoidance.
- **People** for a variety of data ranging from traveler behavior to carbon footprint.

Smartphone Data

There are two broad types of data collected from smartphones: (1) sensor data, and (2) cellular network–based data. In terms of sensor data, smartphones have already been mentioned as far as their role in providing GPS traces, but they are also equipped with Wi-Fi and Bluetooth technology, both of which broadcast information to nearby receivers (Zannat & Choudhury, 2019). Cellular network–based data from smartphones can also provide data through a variety of other means that could be used for various analyses. For instance, call detail records (CDRs) containing time-stamped location coordinates are collected by cell phone companies any time a call is made. Though anonymized, this data can be used to supply spatial and temporal information about human mobility patterns ranging from a very localized area to a global scale (Lu et al., 2021; National Academies of Sciences, 2018).

Web and Social Media Data

Data collected from social media networks (e.g., Facebook, Twitter, LinkedIn, YouTube, Yelp, blogs) and from the Internet as a whole (e.g., e-commerce sites, search engines, job boards, tourism websites) can be processed and analyzed for transportation planning purposes. Many of the data traces that are produced from these sources are not only linked to unique users, but are also either geotagged with GPS data, pertain to a specific location, or are linked to trajectory-centric data from apps and websites that host data from wearable technology (e.g., smart watches, fitness trackers) (Kane & Tomer, 2021; Tasse & Hong, 2017). For this reason, web and social media data can provide transportation professionals with data traces that include the time, data, location, demographic information, opinions, values, and feelings of these users as well as their interactions with other users (Tasse & Hong, 2017).

Crowdsourced data is a source of geotagged information that falls within the category of web and social media data. This type of data comes from collaborative applications—large numbers of users voluntarily offer input either passively or actively (Torre-Bastida, 2018). An example of this is the real-time navigation application Waze—when using the app, users agree to provide their passive geolocation information as well as active input when they notice something reportable (Desouza & Smith, 2016).

Application and Uses of Big Data in Transportation

Once data are collected, prepared, and analyzed, there are many potential applications for this information or ways it can impact providers and users in the transportation sector. Similar to the above list of data sources, the way in which the collected data can be utilized is vast—seemingly limited only by imagination and ingenuity. While the following list is not exhaustive, it serves to illustrate the broad scope of big data applications in the transportation sector. This list does not include big data for transportation equity analysis, which is addressed in the next section.

Travel Pattern Analytics

With the use of big data, transportation researchers can analyze travel behavior, make inferences about trip purpose, track activity patterns, and better understand travel demand. The types of data typically used in combination for these purposes include automatic fare collection (AFC) data from smart cards, Automatic Passenger Count (APC) data, automatic vehicle location (AVL) and GPS data, social media data, and smartphone data (Zannat & Choudhury, 2019). This type of analysis can be used to optimize public transit services, improve signal timing, develop travel demand management strategies, reduce traffic congestion, improve and expand multimodal networks throughout a region, improve parking and curb use, or make a case for integrated transportation and land use strategies such as transit-oriented or mixed-use development—in short, big data travel pattern analyses can be used in nearly all facets of transportation (Lu et al., 2021; Torre-Bastida, 2018; National Academies of Sciences, Engineering, and Medicine, 2018).

Real-Time Analytics

The use of big data and machine learning has allowed for the creation of real-time data and predictive models that provide individuals, communities, and organizations new tools for making informed decisions (Desouza & Smith, 2016). For example, a traveler can use apps such as Waze and Google Maps to receive real-time updates on traffic congestion, crashes, road construction, and other information about the traffic environment. These apps can also automatically compute the best route given these real-time traffic conditions and then offer travelers turn-by-turn navigation to avoid delays (National Academies of Sciences, Engineering, and Medicine, 2018). Apps such as Google Maps, CityMapper, and Transit can provide real-time and predictive information for bus arrivals and departures, map out the best option to reach a destination, and give travelers information on the cost of travel (Hurtado et al., 2021). Another example is smart parking, which uses sensors to detect available parking spaces and present this information in real time to travelers. This type of information allows travelers to make more informed decisions about when to travel, what mode to take, and what they can expect along the way.

Transit and traffic managers and operators also benefit from real-time analytics. For instance, Traffic Operations Centers with access to real-time traffic and transit information can monitor current conditions and, in the event of traffic signal failure, vehicle crash, inclement weather, or other unexpected impediments to travel, deploy mitigating measures. Traffic Operations Centers may remotely update dynamic message signs to warn travelers of dangerous conditions or suggest alternate routes, deploy roadside assistance or emergency personnel to stranded vehicles, send maintenance teams to malfunctioning hardware or fleet vehicles, remotely troubleshoot signal issues, or perform any number of other necessary tasks (Transportation Research Board, 2014).

Service and Performance Evaluation

Big data and machine learning allow agencies to optimize operations through the evaluation of past and real-time performance and predictive modeling. Big data analytics for service and performance evaluation can uncover unseen patterns in transportation issues and offer insight for possible solutions (Dickens & Hughes-Cromwick, 2019). For instance, an analysis of near misses and crashes on roadways allows transportation agencies to reassess safety measures that are in place or propose infrastructure changes (Neilson et al., 2019); an analysis of traffic flow or transit ridership may lead decision-makers to add bus routes, modify traffic signal timing, or adjust toll or bus fares (Welch & Widita, 2019); and an analysis of railway failures could lead to adjusted schedule maintenance timelines for tracks and trains (Zhu et al., 2019). Service and performance evaluations can also be used to predict day-to-day staff shortages and to assist in talent acquisition (Dickens & Hughes-Cromwick, 2019).

Predictive Maintenance

Predictive maintenance uses historic data and machine learning to predict when faults will occur in vehicles, hardware, and infrastructure, and then acts based on those predictions to prevent possible failure (Zhu et al., 2019). Commonly, maintenance activities are reactive (conducted after a fault has occurred), or preventative (conducted before a fault has occurred). Preventative maintenance is further categorized as either time-directed (scheduled maintenance) or condition-based (monitoring the condition of an asset and conducting maintenance only when the asset needs it) (Faiz & Singh, 2009; Ghofrani et al., 2018). Predictive maintenance using machine learning is a sophisticated type of condition-based maintenance that uses data from sensors on equipment, weather conditions, historic data, video, and so on to observe the condition of an asset and perform maintenance based on those observations (Faiz & Singh, 2009;; Zhu et al., 2019). Predictive maintenance has the potential to reduce costs and damages caused by equipment failure and save time and money by more accurately scheduling maintenance when compared to time-directed maintenance (Dickens & Hughes-Cromwick, 2019; Ghofrani et al., 2018; Killeen et al., 2019).

Safety, Security, and Public Health

Big data is currently being used in the transportation sector for a wide range of topics related to safety, security, and public health. Some topics already mentioned include the prediction and prevention of crowding, analyzation of near misses and collisions, identification of hazardous road conditions, and incident management operations (Welch & Widita, 2019). Regarding security, big data technologies are also being used to help identify and neutralize cyber threats and to forecast the likelihood of crime in certain areas—also known as “predictive policing” (Welch & Widita, 2019; Desouza & Smith, 2016; Nguyen & Boundy, 2017).

Big data is also being used to predict, detect, and respond to environmental abnormalities and public health threats. Data collected from sensors and social media platforms can alert decision-makers to the existence of degraded air quality, dangerous chemicals, and infectious substances in and around transportation infrastructure or assets (Desouza & Smith, 2016; Hurtado et al., 2021). Predictive modeling can be used to predict flu trends, seismic activity, and dangerous weather phenomenon and to assess infrastructure resiliency (Alam & Sadri, 2022; Grinberger et al., 2017; Lu et al., 2021). Research using big data can also be used to map and model various trends such as access to food and health care for people who use public transit, the effects of transportation infrastructure on physical and mental health, and transportation-related environmental justice issues such as the disproportionate impact of harmful vehicle emissions on certain neighborhoods (Hurtado et al., 2022; Lu et al, 2021; Welch & Widita, 2019).

Public Sentiment and Public Participation

Sentiment analysis is the study of people’s opinions, emotions, and evaluations toward any entity—such as a product, service, place, person, or idea—using text mining and machine learning (Liu, 2012). Using sentiment analysis, transportation professionals can gain insight into the transportation experiences of the traveling public. The added dimension of space (i.e., geotagged text) allows analysts, planners, and decision-makers to visualize how the public feels about certain transportation-related issues in specific places (Desouza & Smith, 2016; Lock & Pettit, 2020; Verma, 2022).

Big data and gamification have opened new pathways for public participation through interactive and highly visual representations of urban infrastructure and systems. Gamification takes the presentation of big data to the public one step further. Gaming allows the public and transportation professionals to experience the implementation of transportation plans in a virtual, prototypical setting where they can test outcomes, experiment with travel behavior, and discover deficiencies in a low-cost, safe, and engaging environment (Desouza & Smith, 2016; Hurtado et al., 2022). In many communities, the public now has access to "civic tech"—interactive visualizations of data related to their city or community—which allows a clearer understanding of urban issues that affect them (Desouza & Smith, 2016; Hurtado et al., 2021).

Connected Vehicles and Smart Cities

Connected vehicles use wireless technology to communicate and coordinate with sensors on vehicles, infrastructure, and people. This exchange of information allows for in-vehicle applications that range across all levels of vehicle automation, from driver assistance systems (i.e., vehicles driven by a human but assisted to some degree by automation) to fully autonomous systems (i.e., driverless vehicles) (Crute et al., 2018; Lian et al., 2020). Autonomous vehicles use a complex combination of technologies to navigate city space. Real-time data allow them to create a 3D model of the world around them and machine learning allows them to analyze past "experiences" shared among other connected and autonomous vehicles (CAVs) to make decisions about routing, maneuvering, and safety (Crute et al., 2018).

Smart cities generate big data via all data sources listed earlier in this report, then use that data to carry out a wide range of functions and services (Nguyen & Boundy, 2017). Many of those functions and services directly impact transportation such as the operation of streetlights, traffic lights, transit, tolling, dynamic message signs, and variable speed limits, as well as apps and dashboards available to the public. Smart cities use sensors on infrastructure to feed information to connected vehicles for assisted driving (e.g., lane departure warning and smart speed adaptation) and autonomous vehicle systems (Zheng, et al., 2016). Smart cities can also use information collected pertaining to weather, noise levels, pollution, nature, resiliency, and the community to make better planning decisions in the future (Hurtado et al., 2021).

Limitations of Big Data

While the benefits of big data are far-reaching, there are several limitations that should be considered when using big data in transportation. Some of these limitations include the availability of data; the quality and credibility of the data collected; bias in the selection, collection, analysis, and implementation of data; agency capacity to collect, analyze, and implement the data; and security concerns. Each of these limitations are briefly described in the following sections.

Data Availability

In general, the application and implementation of big data requires a continuously updated flow of data from a variety of sources. The most useful, most beneficial, or most complete data are not always readily available or easily accessible for transportation professionals. The four most common reasons behind lack of data availability are: (1) limited funds for either primary data collection or the purchasing of data from a third party, (2) difficulties in navigating data ownership matters such as licensing agreements or confidentiality roadblocks with a third party, (3) regulatory or legal restrictions concerning privacy protection for citizens, and (4) a lack of data sharing policies or infrastructure between entities with open access data (Desouza & Smith, 2016; Hurtado et al., 2021; Grinberger et al., 2017; Transportation Research Board, 2014).

Data Quality and Credibility

Difficulties in securing high quality, reliable data can be a limitation to using big data in the transportation sector. Given that big data comes to transportation from a diversity of data collection methods, in disparate formats, and by various data collection entities (e.g., public, private, and nonprofit agencies), ensuring that the data used are complete, homogeneous, and accurate can be a considerable challenge (Torre-Bastida, 2018).

Bias

The vastness of big data does not preclude it from bias. The existence of bias without identification and mitigation measures limits the usefulness and intended impact of big data when applying it in the transportation sector. All bias normally encountered in the selection, collection, analysis, and implementation of data can potentially appear in big data. However, literature on the subject highlights three key types of bias usually encountered when working with big data and machine learning in particular—selection bias, information bias, and algorithmic bias. Each of these biases can lead to a systemic error in an association or outcome.

Selection bias happens when individuals or groups in a study differ from the population of interest (Nunan et al., 2017). As Richard Shearmur (2015) writes, "big data are not about society, but about users and markets… they are therefore inherently biased in that they do not track people who fall outside the particular markets or activities being tracked" (p. 967). As a result, common types of selection bias inherent in research using big data include sampling bias, coverage bias, social desirability, and self-selection and mode bias (Griffin et al., 2018; Griffin et al., 2020; Transportation Research Board, 2014).

Information bias, also called measurement bias, happens when key study variables are inaccurately measured or classified (Alexander et al., 2015). Examples of possible information bias in transportation studies include low-resolution sensors that provide inaccurate raw data (Bai et al., 2020; Prozzi et al., 2008) and datasets, such as from mobile phone data, which do not correctly or precisely represent their target users (Griffin et al., 2018).

Algorithmic bias is oftentimes observed when artificial intelligence and machine learning are used to find associations or establish outcomes. Algorithms are a set of instructions that computers use to perform a task. For machine learning, a computer model is given training material (datasets) and a series of algorithms instructions to find patterns and make predictions based on the data given (Hooker, 2021). Therefore, there are two basic paths for machine learning models to be biased: (1) through the datasets they are trained on, and (2) through the algorithms written to direct the computer model—each of which is subject to various biases such as implicit bias, selection bias, information bias, historical bias, and institutional bias (Hurtado et al., 2021).

Agency Capacity

Traditional data storage infrastructure cannot meet the capacity or processing needs that big data requires. Storing and managing large datasets require funds, personnel, and a robust IT architecture that most transportation agencies are not equipped with (Zhang et al., 2016). Often agencies rely on storage infrastructure and processing power from private data storage companies such as Google, AWS, and Microsoft (Zhu et al., 2019).

Security

The accumulation of transportation data and the need to use those data for day-to-day transportation functions create an opportunity for serious security concerns (Neilson et al., 2019). Cybersecurity threats to transportation agencies and systems are not new; for instance, transportation agencies in San Francisco, Sacramento, Philadelphia, and New York, as well as state transportation departments in Colorado and Texas, have all received ransomware attacks in the last 10 years (Pargman, 2020). In each of these cases, IT systems were disrupted, partial functionality of the agencies was lost, and public data risked exposure. Often, transportation agencies do not have the funds, manpower, or systems necessary to properly combat these security risks (NCHRP, 2015).

Chapter 3. Big Data for Transportation Equity Analysis

The previous section of this report describes common sources for and uses of big data in transportation, addressing some of the frequently noted limitations identified in the literature. Although many of these sources address the influence of big data on transportation generally, the implications on transportation equity analysis are of key importance. This focus on equity and the resulting outcomes of the analysis affect how transportation investments are distributed and whether they create or exacerbate burdens for already disadvantaged and underserved populations.

The 2022–2026 U.S. Department of Transportation's (U.S. DOT) Strategic Plan and the U.S. DOT Equity Action Plan (developed as an initiative from Executive Order 13985) call for "data-driven equity assessments" (U.S. Department of Transportation, 2022). According to Stobierski (2019), data-driven analysis must use data to inform decision-making processes and potential future actions. Section 9 of Executive Order 13985 established an equitable data working group, referring to the lack of aggregated data and the resulting consequences for measuring and advancing equity.

Additionally, a 2023 Executive Order, "Further Advancing Racial Equity and Support for Underserved Communities through the Federal Government," places emphasis on "equitable data practices." In this Executive Order, equitable data are defined as "data that allow for rigorous assessment of the extent to which Government programs and policies yield consistently fair, just, and impartial treatment of all individuals" (Executive Order No. 14091, 2023). Despite the acknowledgement of the need for data-driven equity analysis, limited data are available for this purpose (Chen et al, 2022).

This section synthesizes the available, relevant literature on the uses and limitations of big data for transportation equity analysis. The first section outlines a framework for equity analysis as identified in the literature. The second section describes the limitations of big data in the transportation equity analysis process. The third section offers a case example of the Hillsborough Transportation Planning Organization (TPO) covering Tampa, Florida, briefly describing their experience with big data and transportation equity analysis.

Big Data for Transportation Equity Analysis

Big data supports equity-related decision-making, informed by applications that involve identifying and evaluating the travel behavior of underserved populations and evaluating the equity impacts of policy decisions (Chen et al., 2022; Desouza & Smith, 2016; Fanibi, 2022; Griffin et al., 2018; Gooden et al., 2017). The equity implications of data relate to how data are selected, collected, analyzed, and used (Ruijer et al., 2022). This section focuses on data use in the transportation equity analysis process and addresses other functions of data where necessary to expand on the related findings in the literature.

Although there are several ways to conduct transportation equity analysis, Litman (2023) divides equity analysis into the following five steps (p. 61):

1. Define the type of equity to be considered (horizontal, vertical, social justice).
2. Define the impacts (benefits and costs) to be considered (funding, facility supply, cost burdens, etc.).
3. Define what distribution of impacts is considered fair and appropriate.
4. Define the population groups considered (demographics, income, geography, mode users), and which are disadvantaged.
5. Evaluate the degree to which the distribution of impacts is considered fair and appropriate.

As shown in step 1 of the transportation equity analysis process, the analysis is driven by the dimension of equity selected for the analysis (Bills & Walker, 2017; Carleton & Porter, 2018; Guo et al., 2020; Litman, 2023). As stated in the introduction of this report, the dimensions of equity are horizontal or vertical—horizontal equity treats everyone similarly and does not consider discrete differences, whereas vertical equity provides for groups based on their needs and differences between individuals and/or different population groups (Chen et al., 2022; Litman, 2023). Horizontal equity can be categorized as fair share and external costs, and vertical equity can be categorized as inclusivity, affordability, and social justice (Litman, 2023). Bills & Walker (2017) further explained these terms as follows, "income, age, and gender are variables that represent vertical equity, while location, travel mode, and time-of-day represent the horizontal equity dimension" (p. 63).

Another framework for transportation equity analysis is presented by Chen et al. (2022) and Guo et al. (2020). This framework uses a three-step approach that is applicable to both horizontal and vertical equity. These steps include (1) defining and evaluating the population, (2) quantifying the distribution of benefits and burdens, and (3) measuring inequality (also identified by Gooden et al., 2017). These processes are interrelated and interdependent. Chen et al. and Guo et al. explain that inequality is measured by comparing the outcomes of the population evaluation and the benefits and burdens analysis. Gooden et al. (2017) stress that analysis should also evaluate progress toward a more equitable transportation system through benchmarking and other strategies. Therefore, the remainder of this section uses the three-step data framework for equity analysis described by Chen et al. and Guo et al. to outline the uses of big data for equity analysis and includes the evaluation of progress toward equity by Gooden et al. as a fourth step in the transportation equity analysis framework.

Define and Evaluate the Population

The first step in the equity analysis process is determining who is being studied (Bills & Walker, 2017; Guo et al., 2020). The population evaluation includes groups or individuals that are analyzed based on spatial location or population characteristics (Chen et al., 2022; Guo et al, 2020). This evaluation typically involves dividing the population into segments or groups based on established thresholds using a variety of relevant variables (Bills & Walker, 2017; Litman, 2023). In the definition of equity, E.O. 13985 (2021) lists the following populations as being a part of underserved communities:

> Black, Latino, and Indigenous and Native American persons, Asian Americans and Pacific Islanders and other persons of color; members of religious minorities; lesbian, gay, bisexual, transgender, and queer (LGBTQ+) persons; persons with disabilities; persons who live in rural areas; and persons otherwise adversely affected by persistent poverty or inequality (Sec. 2).

Other socioeconomic factors used when defining transportation disadvantaged populations include the following (Carleton & Porter 2018, p. 65):

- Employment status
- Language fluency
- Immigration status
- Single parent status

In addition to these broad socioeconomic factors, other factors that specifically define disadvantaged populations may be included in the analysis (Gou et al., 2020). These factors commonly include (1) unemployment rate, (2) mobility need and ability, and (3) cumulative impacts. These factors intersect and should be considered both holistically and separately (Carleton & Porter, 2018).

The level of analysis, variables, and other analysis components are dependent on the dimension of equity selected for the analysis (horizontal or vertical) and affect the outcome of the analysis (Bills & Walker; Guo et al.). For example, Carleton & Porter (2018) describe the challenges of measuring transit equity and explain that, when analyzing vertical equity, the factors selected for population analysis, including how underserved populations are defined and analyzed, influence the accuracy of the results.

Assess the Distribution of Benefits and Burdens

Distributional comparisons are most commonly used for transportation equity analysis (Bills & Walker, 2017; Carleton & Porter, 2018; Litman, 2023; Shi, 2021). These processes are also referred to as cost/benefit measurement, which quantifies the outcomes of transportation investment (positive or negative) on the study population as compared to other populations (Guo et al., 2020; Twaddell et al., 2019). The process for the distributional equity analysis proposed by Bills & Walker (2017) includes (1) identifying equity indicators and segmenting the population, (2) calculating the indicators for each unit of segmentation, (3) comparing the distribution of the indicators between the population segments, and (4) ranking the scenarios from the activity-based model based on selected criteria.

Distributional equity criteria can be divided in a variety of ways. For example, Wennink & Krapp (2020) use five categories with varying degrees of potential equity impacts. The first four criteria relate to spatial components of equity and include location burdens–based, location benefits–based, impact benefits–based, and access to destinations–based. The final criteria, which is identified as having the greatest potential equity impact, is user-based criteria. These criteria are defined as follows (Wennink & Krapp, p. 7):

- Location burdens–based: Considers the location of the proposed project in relation to predefined areas with high concentrations of marginalized populations and awards points if the project is not located within them.
- Location benefits–based: Considers the location of the proposed project in relation to predefined areas with high concentrations of marginalized populations and awards points if the project is located within them.
- Impact benefits–based: Considers the potential positive impacts the proposed project will have on predefined areas with high concentrations of marginalized populations, which may include—but goes beyond—an assessment of only spatial proximity.
- Access to destinations–based: Considers accessibility improvements that projects will provide to areas with high concentrations of marginalized populations. This is called out separately due to the higher specificity of this analysis and the value in focusing on transportation's essential function of providing access to basic needs and economic opportunity.
- User-based: Considers the number of users of the proposed project that will belong to the population defined as marginalized and awards more points to projects with more marginalized users.

To conduct the assessment, appropriate factors and indicators specific to equity must be selected as they form the basis for measuring outcomes (Bills & Walker, 2017; Guo et al., 2020). The type of data, source of the data, and data measurement level influence the results of the equity analysis. The four major categories of equity data defined by Chen et al. (2022) include (1) population data, (2) transportation infrastructure data, (3) mobility data (aggregate and disaggregate), and (4) other data. These data sources are described by Chen et al. as follows (pp. 6–10):

- Population data – sociodemographic information such as ancestry, age, gender, education attainment, income, language proficiency, disability, employment, housing characteristics, and so on.

- Transportation infrastructure data – information on the multimodal transportation network consisting of roads (automobiles privately owned and shared), sidewalks (walking), bike paths (cycling, both privately owned and shared), transit routes (public transit modes), shared mobility stations (bike-sharing, e-scooter sharing, car-sharing; note that ride-hailing does not need stations) and charging stations (electric vehicles or EVs).
- Mobility data – traffic counts, speed, and travel times of each link in a transportation network. At the disaggregated level, mobility data refer to individual travel itineraries, such as the sequence of trips that a person makes over a day.
- Other data – air pollution, facilities, traffic incidents, etc.

Bills & Walker (2017) define equity indicators as "a set of records or observations that measure the costs or benefits associated with implementing a transportation plan" (p. 64). The indicators selected and how they are calculated are critical to the outcomes of the equity analysis (Bills & Walker, 2017; Litman, 2023). Litman (2023) describes the equity implications of different measurement units as shown in Table 1.

Table 1. Equity Implications of Different Measurement Units

Unit	Description	Equity Implications
Congestion impacts	Transportation funds are allocated based on their expected congestion reductions.	Favors people who frequently drive on congested roads.
Vehicle Miles Traveled (VMT)	Transportation funds are allocated based on vehicle-miles driven in an area.	Favors people who drive their automobile more mileage than average.
Passenger Miles Traveled (PMT)	Transportation funds are allocated based on passenger-miles travelled in an area.	Favors people who travel by any mode, with more funding for longer trips.
Passenger Trips	Transport investments are evaluated according to where trips occur.	Provides more support for shorter trips, including active modes and local travel.
Access	Transport investments can support many types of transport improvements.	Can benefit the largest range of users, particularly non-drivers.
Mobility Need	Transport investments maximize benefits to people with mobility impairments.	Favors people with disabilities and other special needs.
Affordability	Transport user fees are evaluated with respect to users' ability to pay.	Favors more affordable modes and lower-income people.
Cost Recovery	Transport expenditures are evaluated according to whether users pay their costs.	Favors wealthier travelers because they tend to spend the most.

How travel is measured can have equity impacts. Some units favor people who drive more than average.
Source: Litman, 2023

The equity indicators in the study conducted by Bills & Walker (2017) were developed using an activity-based travel model and include population data, travel behavior data, transportation network data, and land-use data. These data types and levels of measurement are described as follows (p. 65):

- Population data at
    - The individual level, which includes ethnicity, age, gender, employment status, employment sector.
    - The household level, which includes size, income, residential location, # workers, # children, # vehicles.

- Travel behavior data at
    - The trip level, which includes location, purpose, mode, time-of-day, and
    - The tour level, which includes tour class (home-based mandatory, home-based non-mandatory work-based, etc.), stop frequency, primary mode, primary origin and destination.
- Travel network data for
    - Day-Pattern, which includes tour frequency
    - Travel Time Skims (by mode), which include in-vehicle times, wait times, access times
    - Travel Cost Skims (by mode), which include vehicle operating costs, tolls, parking costs, transit fares
    - Travel Distance Skims (by mode), and
    - Volumes, which include vehicle-miles-traveled.
- Land-use data for
    - Locations (i.e., Zones, neighborhoods, etc.),
    - Population,
    - # households,
    - Employment by sector, and
    - Amenities (shopping, hospitals, banks, etc.).

Identify and Measure Equity and Inequity

Data for equity analysis support equity outcomes as well as community empowerment (Ruijer et al, 2022; Nguyen & Boundy, 2017). More specifically, big data is used to promote equitable outcomes by analyzing disparities between population groups, with the goal of improving the quality of life of underserved populations (Nguyen & Boundy, 2017). Big data is also used to evaluate inequities and demonstrate evidence of inequity (Ruijer et al., 2022).

The inequity measurement compares the outcomes of the population analysis and the cost/benefit analysis (Chen et al., 2022). This measurement is either aggregate or disaggregate and could be for all individuals or spatially distributed population groups (horizontal), or it could be by population subgroups defined by demographic characteristics (vertical) (Bills & Walker, 2017; Guo et al., 2020). For example, in an analysis of the distributional impacts of transportation improvements, Bills &Walker (2017) compare disaggregate indicators between population segments at an aggregate level and at an individual level.

Approaches for the inequality analysis identify gaps in needs and investments (supply and demand) or demonstrate degree of inequality (Carleton & Porter, 2018). These approaches are described as the base for a wide range of possible equity analysis methods including mismatch analysis (e.g., using GIS mapping), calculation of inequality metrics, and regression modeling (Guo et al., 2020).

Evaluate Progress Toward Equity

In addition to measuring inequities, it is also necessary to create measures to monitor and report progress toward accomplishing equity goals (Gooden et al., 2017). An approach that uses standardized performance measures, benchmarks, and reporting that support evidence-based comparative analysis is recommended for this purpose. Without these checks and balances, as well as careful consideration for underserved populations and awareness of the potential adverse impacts in the use of big data, disadvantaged groups may continue to experience unfair burdens from decision-making processes and the transportation system (Mahendra et al., 2021).

Emerging technology and big data may present a challenge for transportation agencies due to the volume, velocity, and variety of big data (Hurtado et al., 2021). While there is evidence of the ways that big data supports equity analysis, as with any type of data, there are also a number of limitations. The next section identifies some of the limitations identified in the literature.

Limitations of Big Data in Transportation Equity Analysis

The newness of big data contributes to conflicts in thought on its use for transportation equity analysis (Griffin et al., 2020). Some research highlights the potential positive outcomes of big data for transportation analysis such as speed and cost savings (Griffin et al., 2020), precision (Desouza & Smith, 2016), increased data availability (Desouza & Smith, 2016; Griffin et al., 2020), accuracy (Ruijer et al., 2022), and innovation (Chen et al., 2016; Daepp et al., 2022; Griffin et al., 2020). On the other hand, some researchers have explained that the potential for inequity is a rational outcome of big data (Griffin et al., 2018). Much of this research identifies potential negative outcomes such as bias, disparate impacts (Griffin et al., 2018), policing and privacy concerns, data misuse (Ruijer et al., 2022), and other limitations influenced by data selection, collection, analysis, and implementation (Ruijer et al., 2022).

In addition to the aforementioned limitations, there are also ethical considerations when using passively collected data (big data) and algorithms (APA, 2022; Crawford, 2013; Wigan & Clarke, 2013). These considerations are based on the size of the dataset (Crawford, 2013), the exclusion of observed populations in data collection (Ruijer et al., 2022), over-surveillance and privacy concerns (Crawford, 2013; Griffin et al., 2018; Mahendra et al., 2021; Ruijer et al, 2022; Wigan & Clarke, 2013), and the interpretation of correlation vs. causation (Crawford et al., 2013). Many of these limitations lead to a phenomenon Ruijer et al. (2022) identified as “datafied marginalization where the risks of datafication are borne by data subjects and the benefits enjoyed by controllers” (p. 326). In other words, equity data can become dehumanized, and people become a variable or data subject with little to no personal input into the information that is collected from them (Crawford, 2013; Ruijer et al., 2022; Wigan and Clarke, 2013).

The volume, velocity, and variety of big data create several challenges, including the potential for known or unknown inequitable outcomes (Desouza & Smith, 2016; Ruijer et al., 2022). Big data may, for example, lead to discriminatory practices when vague and limited information is used to make broad assumptions (inferences) about individuals in certain population groups (Chen et al., 2016; Wigan & Clarke, 2013). Chen et al. (2022) set out to develop a transportation equity big data library. In this process, the authors divided the limitations of big data in transportation equity analysis into four broad categories (Chen et al., 2022):

- Methodological – challenges exist in developing efficient, explainable, and fair analytics to analyze transportation equity with big data.
- Technological – difficulty in data collection, standardization, integration, storage, processing, transmission, and dissemination.
- Political – the political, institutional, and ethical concerns in the use of individual-level data, constrained by the agendas and actions of various institutions, stakeholders, and processes involved with the data.
- Epistemological – the knowledge generation process.

A lack of disaggregate data, gaps in available data, the potential for disparate impact and bias, including bias stemming from the analysis methods used to evaluate the data, and limited agency resources to collect, manage, share, and use data are cited as supporting evidence for the limitations of big data for equity analysis (Chen et al., 2022; Desouza & Smith, 2016; Pereira & Karner, 2021; University of Utah, 2015). While this list is not exhaustive, it reflects the commonly cited limitations described in the literature. This section describes these limitations.

Gaps in Available Data

Analysis, for any purpose, cannot rely on factors for which there are no data (Carleton & Porter, 2018). There is the potential for available data to miss certain populations or misrepresent the population, obscure or miss data relevant to underserved populations, and/or have significant gaps and limitations relevant to the data's ability to fully capture users in the transportation system (Chen et al., 2016; Crawford et al., 2014; Griffin et al., 2020; Hurtado et al., 2021; Ruijer et al., 2022).

Gaps in data are of particular consequence for "vulnerable groups that lack access to the technology that collects the data" (Ruijer et al 2022, p. 323). For example, big data from cell phones, smart watches, apps, and so on only represents a portion of the population (those with access to the technology, those who have adopted the technology, and those who have access to broadband) and oftentimes does not capture the entire trip chain or all of the trips made by individuals (Chen et al., 2016; Crawford et al., 2014; Griffin et al., 2020; Mahendra et al., 2021). In short, data are only available for the people who use the technology, and even that information may be limited (Shaheen et al., 2017).

These limitations are evident in the difference between data collected passively and those collected through traditional sources where, for example, passive data are not as easily validated as data from traditional sources (Chen et al., 2016). Although passively collected data (e.g., using GIS) have greater potential for accuracy and can improve efficiency (Mahendra et al., 2021; Nguyen & Boundy, 2017), these methods introduce a variety of challenges. They create issues related to the populations that data are collected from (selection bias), the methods used to measure or classify data (information bias), as well as the resulting outcomes (algorithmic bias) and conclusions drawn using that data (Chen et al., 2016; Griffin et al., 2020). These biases are described in more detail in the next section on *Disparate Impact and Bias*. Desouza and Smith (2016) encourage agencies to use traditional sources if they cannot use big data with precision and in an equitable manner.

Only a few organizations, many of which are private, own passive data, creating difficulties for public agencies responsible for transportation decision-making (Ehrlich et al., 2020; Hurtado et al., 2021; Shaheen et al., 2017). As a result, data on users of shared mobility services are not widely available in national datasets (Shaheen et al., 2017). Regulations for open data and data sharing policies could support public agencies and decision-makers in addressing this challenge (Shaheen et al, 2017; Nguyen & Boundy, 2017). Where data are available, the resources needed to analyze large volumes of big data create additional challenges for public agencies (Mahendra et al., 2021; Nguyen & Boundy, 2017). This results in the modifiable area unit problem (MAUP), "[d]ifferent scales can lead to inconsistent results in equity evaluation" (Dark and Bram, 2007 as cited by Guo et al., 2020 , p. 4).

Other challenges relate to data resolution, disparate data sources with differing information, and difficulties with data availability and data sharing between agencies (Chen et al., 2022). These challenges are described by Chen et al., (2022) as follows:

> Cross-sectional studies comparing transportation equity outcomes and dynamics across different modes of transportation and across cities would allow the extraction of systematic, institutional, and structural factors behind transportation outcomes inequity. This knowledge would enable the

explanation of heterogeneity between and within cities, using sociodemographic and other factors, such as how transportation system structures affect transportation equity outcomes. (p. 3)

Disparate Impact and Bias

Existing methods for collecting and analyzing data have been described as systematically inequitable. These inequities occur because the existing systems and procedures used to collect, assess, and interpret the data are fundamentally biased and inform policy or enforcement that result in disparate impacts (Hurtado et al., 2021; Pereira & Karner, 2021; Ruijer et al, 2022; Wilson, 2022). Disparate impact is described as an ostensibly unbiased policy or practice that results in negative outcomes for specific population groups based on race, religion, sexual orientation, nationality, or other protected status, where more equitable alternatives would serve the same purpose (Desouza & Smith, 2016; Shaheen et al., 2017; University of Utah, 2015).

Bias is difficult to identify, and while a variety of factors contribute to bias, inferences from big data can lead to bias and disparate impact, as well as other discriminatory practices (Griffin et al., 2018; Griffin et al., 2020; Wigan and Clarke, 2013). Including context during the analysis can help to mitigate potential bias, but this information may be difficult to get from passive data (Gooden et al., 2017).

Algorithms are becoming increasingly difficult to regulate (APA, 2022; Ruijer et al., 2022). Algorithms and automated systems can create and reinforce bias and disparate impact, even when demographic information is anonymized (APA, 2022; Desouza & Smith, 2016; Executive Order No. 14091, 2023; Pereira & Karner, 2021; Ruijer et al., 2022). The potential for algorithmic bias is described as follows: “screening processes are designed and interpreted to produce disparate treatment on the backend… pernicious feedback loops in data analysis that drive an overabundance of policy or enforcement and result in disparate treatment” (Ruijer et al., 2022, pp. 324–325).

In an example of algorithmic bias presented by APA (2022) and Hurtado et al. (2021), autonomous vehicles (AVs), which use artificial intelligence to learn how to detect people and avoid crashes, were not programmed to detect people of color and, as a result, were more likely to crash into people of color than white people. To this end, the literature presents three precursors for algorithmic bias:

- Algorithms reflect any bias of the algorithm’s creator (Hurtado et al., 2021; Ruijer et al., 2022).
- If groups are missing from the data, they will not be represented in the algorithm (Hurtado et al., 2021).
- Algorithms programmed without an equity lens cannot consider equity impacts (Hurtado et al., 2021).

Zweig (2019), as cited by Hurtado et al. (2021), lists the following considerations for avoiding algorithmic bias (p. 91):

- The quality and quantity of the data used,
- The nature of the question or problem that needs to be resolved, how it is defined, and transparency around it (i.e., what do we ask the algorithm to solve and do we ask the question in the right way?), and
- The definition of “common good” and the identification of the ethically correct or morally acceptable outcome.

The digital divide exacerbates potential bias in analysis relying on technology and using big data (Hurtado et al., 2021). Desouza and Smith (2016) describe the digital divide as follows:

> The term “digital divide” refers to the gap between groups of individuals with access to modern technology, such as digital devices and the Internet, and groups such as lower-income individuals and the elderly that do not have such access to or cannot afford digital services… it is difficult to collect data on individuals who are not generating data in more common ways, such as through Internet usage, social media usage, or credit card transactions. (p. 11)

A well-documented example of the digital divide exacerbating bias is the use of cell phone data resulting in coverage bias, sampling bias, and/or non-response bias because it cannot capture persons who do not own or have access to a cell phone (Suich et al., 2021; Griffin et al., 2020). Cell phone data are more likely to exclude persons in low-income households, those living in rural areas, or older adults (Faverio, 2022; Vogels, 2021a; Vogels, 2021b)—groups that are identified in the definition of underserved communities (E.O. 13985) and are therefore included in the equity analysis.

Aggregated Data

Aggregated data create bias in transportation analysis by obscuring individual data and limiting the effectiveness of the analysis applied (Carleton & Porter, 2018; Chen et al., 2022; Griffin et al., 2020). Data resolution levels such as travel analysis zones (TAZ), census tracts, and census blocks, as well as the analysis approach (e.g., using the mean) result in aggregated data (Bills & Walker et al., 2017; Guo et al., 2020). To mitigate aggregation bias and ensure that disparities are effectively measured between population groups, disaggregation based on the dimension of equity being analyzed (horizontal or vertical) and at the individual level, as well as the use of comprehensive equity indicators, are recommended (Bills & Walker, 2017; Chen et al., 2022; Crawford, 2013; Litman, 2023).

Although this need is well noted, disaggregated data (by population group and neighborhood) that supports comprehensive and accurate equity analysis are often not readily available (Chen et al., 2022; Mahendra et al., 2021). For example, in horizontal equity analysis, the difficulty in disaggregating data has been cited as resulting from “the number of people, the complexity of human travel behavior in a multimodal transportation system, and the lack of individual-level demographic data due to privacy issues” (Guo et al 2020, p. 3). Activity-based travel demand models and technology that can disaggregate data are suggested to address limitations related to aggregated data (Bills & Walker, 2017; Guo et al., 2020; Mahendra et al., 2021).

Predictive Policing and Privacy Concerns

Policing and privacy are frequently cited issues relating to passively collected data. For example, big data is used for predictive policing, a preventative measure that anticipates behavior (Nguyen & Boundy, 2017). The impacts of this practice related to privacy and justice have been debated in the literature and are described in this section.

Passively collected data present significant personal privacy concerns related to how the data is collected, used, and stored—who has permission to access the data, how and where data are sold, and what happens in the event of data breach (Crawford et al., 2014; Griffin et al., 2018; Mahendra et al., 2021; Wigan and Clarke, 2013). Even when big data is anonymized, oftentimes individuals are able to be reidentified using the personal information collected through these passive data sources (Taylor, 2015; Wigan and Clarke, 2013).

Data collected without direct awareness of the subjects are described as contributing to the “erosion of civil liberties and privacy” (Crawford et al., 2014, p. 4). These practices raise a number of ethical considerations for researchers, particularly those who are studying historically marginalized groups (Crawford et al., 2014; Ruijer et al., 2022; Taylor, 2015). To mitigate these privacy concerns, some government agencies may restrict or limit access to certain technology or set conditions for data applications and/or individuals may choose not to adopt certain technology (Griffin et al., 2020; Griffin et al., 2018; Ruijer et al., 2022; Mahendra et al., 2021). While these measures have the potential to address privacy concerns, they also introduce gaps in data as well as sampling bias (described in more detail in the *Disparate Impact and Bias* section of this report) (Griffin et al., 2018).

Big data can be used to predict safety and security concerns based on population characteristics, past behavior, social media activity, social events, weather, land uses/development, and transportation (Desouza & Smith, 2016; Dickens & Hughes-Cromwick, 2019; Neilson et al., 2019; Ruijer et al., 2022; Nguyen & Boundy, 2017). This practice has been coined *dataveillance*—"a more economical method for monitoring individuals than physical and electronic surveillance" (Wigan and Clarke 2013, p. 46). The use of big data for surveillance can save agencies resources and allow for more efficient data sharing between agencies (Crawford et al., 2014; Dickens & Hughes-Cromwick, 2019).

The American Public Transportation Association (APTA) surveyed transit agencies on their use of big data and reported that more than half of the survey respondents used big data for safety and security purposes including cybersecurity as well as surveillance of public transportation users and environments to spot abnormalities (Dickens & Hughes-Cromwick, 2019). The APA PAS Report 585, *Big Data and Planning*, includes examples of big data's use for predictive policing in Glasgow, Japan, Boston, and Chicago. These examples demonstrate how big data is used to predict behavior to prevent crime, but they also address the potential for profiling and other equity concerns (Desouza & Smith, 2016).

While improved safety and security is supported by big data, there are also several drawbacks, particularly for marginalized populations. Underserved populations are already over-surveilled and passively collected data have the potential to exacerbate inequitable practices (Crawford et al., 2014; Ruijer et al., 2022, Taylor, 2015). Predictive policing is one such example of how technology can adversely impact already burdened groups due to bias and disparate impacts in data practices and algorithms (this topic is discussed in more detail in the *Disparate Impact and Bias* section of this report) (Hurtado et al., 2021; Ruijer et al., 2022). In these instances, inferences from data are used to disproportionately increase monitoring or targeting of certain groups or individuals with certain characteristics (Ruijer et al., 2022; Wigan and Clarke, 2013).

Case Example: Hillsborough Transportation Planning Organization

In March 2023, the research team interviewed staff from the Hillsborough Transportation Planning Organization (TPO) in Tampa, Florida, to identify their practices for transportation equity analysis, focusing on their use of big data for this purpose. The Hillsborough TPO is an officially designated metropolitan planning organization (MPO). A summary of this interview is provided in this section and describes the TPOs practices at the time of the interview.

MPOs are required to comply with federal laws, executive orders, and policies, including those related to equity. To meet federal requirements, the Hillsborough TPO has been working to more effectively incorporate equity into the TPO's decision-making processes. One such approach was an assessment of proposed projects and their impacts on traditionally underserved communities. This analysis used census data to identify communities as either underserved or not underserved. Once populations were defined, projects were prioritized in the Transportation Improvement Program (TIP) based on the equity impacts on Environmental

Justice (EJ) communities. Using this approach, points were awarded based on the percentage of the project areas that fall within an EJ community. This analysis was used to determine if everyone was getting their fair share and if there was underinvestment in underserved communities. At the time of this interview, TPO staff indicated that they have not gotten as far with this level of analysis for persons with disabilities and several other protected classes.

TPO staff have been working to assess how they evaluate the impacts of major projects to consider how projects are perceived and accepted by affected populations (e.g., highway expansion, fixed guideway transit projects, etc.). The Hillsborough TPO is also interested in evaluating the distribution of funds to local governments. This evaluation would determine if money was being distributed in an equitable fashion and to track Disadvantaged Business Enterprise (DBE) participation on contracts.

The Hillsborough TPO conducted a health impact assessment of complete streets projects in the long-range transportation plan (LRTP). The goal of this assessment was to identify disparities in transportation infrastructure using a combination of Centers for Disease Control and Prevention (CDC) places data and data from a variety of other sources such as the Environmental Protection Agency's (EPA's) EJ index, land-use data, and transportation performance measures. The assessment produced a set of neighborhood profiles comparing health and transportation outcomes in EJ areas and the entire county. Factors used in the assessment included the following:

- Walkability
- Street intersection density
- Transportation cost burden
- Smart location index
- Access for vehicles and active commuters
- Active transportation infrastructure (sidewalks, bike lanes)
- Crashes
- Lead paint
- Wastewater discharge

When developing the 2018 non-discrimination plan, the TPO used communities of concern (COCs), which was a shapefile of aggregated indicators. For the 2021 non-discrimination plan, these indicators were disaggregated and EJ was made the focus of the analysis, now including minority status and income as discreet variables.

Due to the volume and resolution of data, decisions are often based on assumptions from the demographic data for large census geographies (TAZ, block groups, etc.). Tools and strategies that provide more granular level data were identified as important to advancing the TPO's transportation equity analysis processes. As a result, the TPO has started using data at the parcel-level to ensure accuracy in the data and analysis.

As is the case for many agencies, metropolitan planning organizations (MPOs) work with constrained budgets, and the Hillsborough TPO is no exception. These limited budgets can make it difficult to invest in all the needed resources and tools that support more advanced transportation equity analysis. Over the years, the Hillsborough TPO has used a variety of big data platforms through subscriptions or on an ad hoc basis. Cost and concerns regarding accuracy for granular analysis were cited as reasons the subscriptions were ended. For example, the Hillsborough TPO previously used software to create walk and bike sheds, allowing them to identify the number of critical services within that transportation shed. These transportation sheds were used to evaluate access to grocery stores, health care facilities, schools, and other essential destinations. Although it is described as an effective tool, the subscription was ended due to its cost.

The most significant challenge for the Hillsborough TPO when conducting transportation equity analysis is ensuring that they have enough accurate granular level data. This challenge is compounded by disparate data sources with differing data availability. Many of the TPO committees raise questions about the accuracy of the data, data sourcing, data collection methods, and so on.

Chapter 4. Big Data Platforms

This section summarizes information about a small selection of platforms that use big data and describes their use for equity analysis. These platforms include StreetLight Data, Replica, and Urban SDK. These summaries offer a description of the platforms and serve as a snapshot of the information available on each company's website. For more information, links to each platform's website are provided at the end of the summaries.

StreetLight Data

StreetLight Data uses smartphone data to provide mobility metrics including average annual daily traffic (AADT) counts, average travel distances, origins and destinations, turning movement counts for intersections, vehicle miles traveled (VMT), vehicle hours of delay (VHD), demographic data, link analysis, top routes, trip purpose, trip speed, travel time, and length. The StreetLight Data website offers several examples of how the platform can be used including anticipating and planning for special events, identifying underserved areas, optimizing bike and pedestrian infrastructure, optimizing freight travel, and prioritizing spending. Modes measured include bicycles and pedestrians, bus and rail, electric vehicles, personal vehicles, ride hailing and delivery, and trucks.

Transportation equity is listed as one of the featured solutions on the platform website. The website includes several webinars, guidebooks, case studies, blogs, and other resources that describe how StreetLight Data has been used to evaluate and address transportation equity. Key capabilities for these analyses include aggregating demographics by selecting sociodemographic characteristics of the study population, analyzing origin and destination by modes and time of day to compare how and why different populations travel, and performing before and after studies to understand the outcomes of transportation projects in various population groups.

StreetLight Data can be accessed here: https://www.streetlightdata.com/

Replica

The Replica website states that it uses disparate datasets to provide "near real-time" data about the built environment, including mobility, land use, people, and economic activity. The platform can be used for origin-destination (O-D) analysis, corridor studies, transit studies, freight studies, workforce profiles, pandemic recovery, residential profiles, active transportation, and tax forecasting. Replica links to a variety of reports and studies that used Replica data, including several that address equity.

Transportation equity is listed as a key initiative of the platform. Replica offers insights on a variety of equity related topics such as transportation access, access to opportunity, environmental health and safety, and intersectional impact. The website explains that Replica provides disaggregate data for individual trips and trip makers, with information on household surveys, traffic counts, ground-truth calibration, central source of truth, socioeconomic and demographic information, O-D flows, trip purpose, trip attributes, comprehensive land use, and consumer spending.

Replica can be accessed here: https://www.replicahq.com/

Urban SDK

Urban SDK provides location and mobility data and tools for visualization and GIS analysis. The platform uses daily probe and location data sources. Probe data is defined as “data that is generated by monitoring the position of individual vehicles (i.e., probes) over space and time rather than measuring characteristics of vehicles or groups of vehicles at a specific place and time” (FHWA, 2017). The platform allows users to explore trends, upload data onto maps and templates and share reports, and download data and performance measures to support analysis. Urban SDK data have been used for a variety of analyses such as O-D studies, speed and reliability studies, traffic counts, crash reports, and incident management. Urban SDK includes GIS datasets for speed, travel time, congestion, traffic counts, trips, foot traffic, crashes, crash rates, crash risk, emissions, sustainability, equity, places, points of interest, land use, demographics, roads, transit, and infrastructure.

The website offers case studies, articles/blogs, news stories, and white papers describing how the platform has been used for various types of analyses. A blog post titled *Using Data to Increase Transportation Equity* describes how Urban SDK data can be used to advance equity. The blog post states, “Sensors and real time data collection can help cities identify neighborhoods being underserved, where infrastructure is insufficient, and the demands placed upon it” (Robare, 2023, para. 26). Several other blog posts describe the importance of data for equity analysis, also giving examples of how their data has been used in equity analysis.

Urban SDK can be accessed here: https://www.urbansdk.com/

Chapter 5. Summary

A wide variety of big data types and sources are used in the transportation sector for research, planning, and operations. The most widely used sources were described in this report and include smart card and automatic fare collection, GPS, and automatic vehicle location (AVL), sensors, smartphone data, and web and social media. While the proliferation of big data into the transportation sector supports transportation equity analysis and equity-related decision-making, it has the potential to create additional equity concerns. Federal legislation and guidance have also recognized the need for data-driven equity analysis, while also acknowledging these restrictions.

This report synthesized existing literature on the uses and limitations of big data for transportation in general as well as for transportation equity analysis. Transportation equity analyses are informed by applications that involve identifying and evaluating the travel behavior of underserved populations and evaluating the equity impacts of policy decisions (Desouza & Smith, 2016; Fanibi, 2022; Griffin et al., 2018). To this end, this report described a four-step framework for transportation equity analysis that was derived from existing literature. The framework includes (1) defining and evaluating the population, (2) assessing the distribution of benefits and burdens, (3) identifying and measuring equity and inequity, and (4) evaluating progress toward equity. The volume, velocity, and variety of big data create several challenges, including the potential for known or unknown inequitable outcomes (Desouza & Smith, 2016; Ruijer et al., 2022). Commonly cited limitations of big data for transportation equity analysis include gaps in available data, disparate impact and bias, aggregated data, and predictive policing and privacy. In general, concerns may stem from methodological, technological, political, and/or epistemological limitations (Chen et al., 2022). To support equity, it is suggested that adjustments are made in the way big data is defined to consider two additional “V’s”—value and validity (Hurtado et al., 2021).

A case example of the Hillsborough TPO in Tampa, Florida, was used to illustrate how transportation agencies are using big data for transportation equity analysis, while contending with a variety of limitations and constraints including funding, data reliability, and political environments. Reassessment of data sources and agency evaluation methods for equity paired with innovative strategies support the agency’s efforts in advancing equity.

The report concluded with descriptions of three big data platforms that emphasize their products’ use for transportation equity analysis. Among other functions, these companies highlight the availability of disaggregated data through their platforms. This emphasis on disaggregated data has the potential to address one of the long-standing critiques of equity analysis and limitations of large datasets—aggregation.

Big data can be used as a resource to advance equity by reducing inequities and improving quality of life (Thakuriah et al., 2017). Although big data has several limitations that practitioners must contend with, addressing these limitations can reveal pathways to more effective equity analysis. The following list describes suggested areas of need related to big data for transportation equity analysis:

- Assess current practices on the use of big data in transportation equity analysis. Although this research scanned and mined existing literature, research is needed to document the current state of practice and identify notable practices from an agency perspective, expanding the findings of this synthesis and other related works. The outcomes of this research can inform decision-making and support analysis that advances equity.

- Develop dynamic resources, tools, and strategies to mitigate bias from big data in transportation analysis. The type and quality of data available contribute to the effectiveness of the analysis. As technology continues to advance rapidly, the practitioner's toolkit needs to evolve at a comparable pace. Research is needed to ensure that the available resources support data applications and analysis methods that advance equity and do not introduce inequities.
- Create standardized performance measures, benchmarks, and reporting for transportation equity analysis, specifically focused on the use of big data for this purpose. This report summarized literature that emphasized the need for a more holistic approach to social equity analysis that included steps for iterative performance measurement and monitoring using innovative data-driven approaches. Big data can support these additional approaches, but research is needed to ensure that these data are applied in a way that advances equity and minimizes potential adverse impacts and bias.
- Analyze the stages in data life cycles (pre-collection planning, data processing, data storage, archiving, analysis, decision-making, visualization, etc.) to identify opportunities for integrated equity considerations. This report focused on data collection and applications, and there is a need to identify other areas in the data life cycle where bias and equity issues are likely to emerge. Identifying areas for which planners and implementers can have the most meaningful impact is of key importance to this analysis and can inform the roles and responsibilities of those involved in transportation equity analysis.

References

Alam, M. R., & Sadri, A. M. (2022). Examining the communication pattern of transportation and transit agencies on Twitter: A longitudinal study in the emergence of COVID-19 on Twitter. *Transportation Research Record*. https://doi.org/10.1177/03611981221082564

Alexander, L. K., Lopes, B., Ricchetti-Masterson, K., & Yeatts, K. B. (2015). *Sources of systematic error or bias: Information bias*. Chaple Hill, NC: UNC Gillings School of Global Public Health. https://sph.unc.edu/wp-content/uploads/sites/112/2015/07/nciph_ERIC14.pdf

American Planning Association (APA) and Lincoln Institute of Land Policy. (2022). *2022 APA foresight: Trend report for planners*. https://www.planning.org/publications/document/9228382/

American Planning Association (APA). (2021). 2021 year in review: Six topics that had planners clicking. https://www.planning.org/blog/9221486/2021-year-inreview-six-topics-that-had-planners-clicking/

Anwar, A., Odoni, A., & Toh, N. (2016). BusViz big data for bus fleets. *Journal of the Transportation Research Board, 2544*, 102–109. doi:10.3141/2544-12

Arass, M. E., Tikito, I., & Souissi, A. N. (2017). Data lifecycles analysis: Towards intelligent cycle, 1–8. doi:10.1109/ISACV.2017.8054938

Bai, L., Liu, R., & Li, Q. (2020). Data-driven bias correction and defect diagnosis model for in-service vehicle acceleration measurements. *Sensors, 6*(20), 872. doi:10.3390/s20030872.

Beyer, M., & Laney, D. (2012). The Importance of "big data": A definition. Gartner. https://www.gartner.com/doc/2057415

Bills, T. S., & Walker, J. L. (2017). Looking beyond the mean for equity analysis: Examining distributional impacts of transportation improvements. Transport Policy, 54, 61-69.

Carleton, P. R., & Porter, J. D. (2018). A comparative analysis of the challenges in measuring transit equity: definitions, interpretations, and limitations. Journal of transport geography, 72, 64-75.

Chen, Z., Long, K., Stuart, A., Mannering, F., & Li, Xiaopeng. (2022). The Promise of big data for transportation equity. Preprint.

Crato, N. (2010). How GPS Works. Figuring it out. https://doi.org/10.1007/978-3-642-04833-3_12

Crawford, K. (2013). The hidden biases in big data. *Harvard business review*, *1*(4).

Crawford, K., Gray, M. L., & Miltner, K. (2014). Big Data| critiquing Big Data: Politics, ethics, epistemology| special section introduction. International Journal of Communication, 8, 10.

Crute, J., Riggs, W., Chapin, T. S., & Stevens, L. (2018). Pas Report 592: *Planning for autonomous vehicles*. Chicago: American Planning Association.

Daepp, M. I., Binet, A., Gavin, V., Arcaya, M. C., & Healthy Neighborhoods Research Consortium. (2022). The moving mapper: Participatory action research with big data. *Journal of the American Planning Association*, *88*(2), 179-191.

Desouza, K. C., & Smith, K. L. (2016). PAS Report 585: *Big data and planning*. Chicago: American Planning Association. https://www.planning.org/publications/report/9116397/

Dickens, M., & Hughes-Cromwick, M. (2019). *Leveraging big data in the public transportation industry*. New York, NY: American Public Transportation Association (APTA). https://www.apta.com/research-technical-resources/research-reports/leveraging-big-data/

Ehrlich, J., Hard, E., Komanduri, A., & Anderson, R. S. (2019). A Century of Travel Surveys Informing Transportation Investments.

Executive Order No. 13985, 86 Fed. Reg. 7009 (Jan. 20, 2021)

Executive Order No. 14091, 88 Fed. Reg. 10825 (Feb. 16, 2023)

Faiz, R. B., & Singh, S. (2009). Information analysis of rail track for predictive maintenance. *WSEAS Transactions on Computers, 8*(7), 1123–1133. https://dl.acm.org/doi/abs/10.5555/1639477.1639486

Fanibi, V. (2022). Working on purpose: Using big data to make progress on transportation equity. https://www.streetlightdata.com/transportation-equity-big-data/?type=blog/

Faverio, M. (2022). Share of those 65 and older who are tech users has grown in the past decade. https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/

Federal Highway Administration (FHWA) & Federal Transit Administration (FTA). (2021). 2021 planning emphasis areas. https://www.transit.dot.gov/sites/fta.dot.gov/files/2022-01/Planning-Emphasis-Areas-12-30-2021.pdf

Federal Highway Administration (FHWA). (2017). Work zone performance measurement using probe data. https://ops.fhwa.dot.gov/wz/resources/publications/fhwahop13043/ch2.htm

Ghofrani, F., He, Q., Goverde, R. M., & Liu, X. (2018). Recent applications of big data analytics in railway transportation systems: A survey. *Transportation Research Part C*, 226–246. https://doi.org/10.1016/j.trc.2018.03.010

Gooden, S. T. (2017). Social equity and evidence: Insights from local government. https://onlinelibrary.wiley.com/doi/epdf/10.1111/puar.12851

Griffin, G. P., Mulhall, M., & Simek, C., & Riggs, W. W. (2020). Mitigating bias in big data for transportation. *Journal of Big Data Analytics in Transportation, 2,* 49–59. https://doi.org/10.1007/s42421-020-00013-0

Griffin, G., Mulhall, M., & Simek, C. (2018). *Sources and mitigation of bias in big data for transportation safety*. College Station, TX: Safe-D National UTC. http://hdl.handle.net/10919/88893

Grinberger, A. Y., Lichter, M., & Felsenstein, D. (2017). Seeing cities through big data dynamic agent based simulation of an urban disaster using synthetic big data. In P. Thakuriah, N. Tilahun, & M. Zellner, *Seeing Cities Through Big Data* (pp. 349–382). Switzerland: Springer. doi:10.1007/978-3-319-40902-3

Guo, Y., Chen, Z., Stuart, A., Li, X., Zhang, Y. (2020). A systematic overview of transportation equity in terms of accessibility, traffic emissions, and safety outcomes: From conventional to emerging technologies. *Transportation Research Interdisciplinary Perspectives, 4*. https://doi.org/10.1016/j.trip.2020.100091

Hooker, S. (2021). Moving beyond "algorithmic bias is a data problem." *Patterns.* https://doi.org/10.1016/j.patter.2021.100241

Hurtado, P., Hitchings, B. G., & Rousem, D. C. (2021). PAS report 599: *Smart cities integrating technology, community, and nature*. https://www.planning.org/publications/report/9226594/

Hurtado, P., Shah, S., DeAngelis, J., & Gomez, A. (2022). *APA 2022 trend report for planners*. Chicago: American Planning Association. https://www.planning.org/publications/document/9228382/

IBM. (2014). The four V's of big data. IBM Big Data and Analytics Hub. http://www.ibmbigdatahub.com/infographic/four-vs-big-data

ITE. (2020). 2020 ITE developing trends. https://www.ite.org/pub/?id=E0E0B1BC-B252-06AA-3321-784D43999A2E

Kane, J., & Tomer, A. (2021). Geolocation data for smart cities. In P. Hurtado, B. Hitchings, & D. Rouse, PAS Report 599: *Smart Cities: Integrating Technology, Community, and Nature*. Chicago: APA. https://www.planning.org/publications/report/9226594/

Killeen, P., Ding, B., Kiring, I., & Yeap, T. (2019). IoT-based predictive maintenance for fleet management. *Procedia Computer Science*, 607–613. doi:10.1016/j.procs.2019.04.184

Klein, L. A., Mills, M. K., & Gibson, D. R. (2006). *Traffic Detector Handbook: Third Edition—Volume I*. Publication No. FHWA-HRT-06-108. McLean, VA: Federal Highway Administration. https://www.fhwa.dot.gov/publications/research/operations/its/06108/index.cfm

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. Application delivery strategies. META Group Inc.: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lawson, C., Muro, A., & Krans, E. (2019). Forecasting bus ridership using a "blended approach." *Transportation, 48*, 617–641. https://doi.org/10.1007/s11116-019-10073-z

Lian, Y., Zhang, G., Lee, J., & Huang, H. (2020). Review on big data applications in safety research of intelligent transportation systems and connected/automated vehicles. *Accident Analysis and Prevention*. https://doi.org/10.1016/j.aap.2020.105711

Litman, T. (2023). Evaluating transportation equity: Guidance for incorporating distributional impacts in transport planning. Evaluating Transportation Equity (vtpi.org)

Liu, B. (2012). *Sentiment analysis and opinion mining*. Switzerland: Springer. doi:10.1007/978-3-031-02145-9

Lock, O., & Pettit, C. (2020). Social media as passive geo-participation in transportation planning – how effective are topic modeling & sentiment analysis in comparison with citizen surveys? *Geo-spatial Information Science, 24*(4), 275–292. https://doi.org/10.1080/10095020.2020.1815596

Lu, K., Liu, J., Zhou, X., & Han, B. (2021). A review of big data applications in urban transit systems. *IEEE Transactions on Intelligent Transportation Systems, 22*(5), 2535–2552. doi: 10.1109/TITS.2020.2973365

Mahendra, A., King, R., Du, J., Dasgupta, A., Beard, V., Kallergis, A., & Schalch, K. (2021). Towards a more equal city: Seven transformations for more equitable and sustainable cities. World Resources Institute. 10.46830/wrirpt.19.00124.

Meyer, M. D. (2016). *Transportation planning handbook* (Fourth edition ed.). Hoboken, NJ: John Wiley & Sons, Inc.

National Academies of Sciences, Engineering, and Medicine. (2018). *Cell phone location data for travel behavior analysis.* The National Academies Press. http://nap.nationalacademies.org/25189

NCHRP. (2015). *Protection of transportation infrastructure from cyber attacks: A primer*. Washington DC: NCHRP. https://nap.nationalacademies.org/read/23520/chapter/2

Neilson, A., Indratmo, Daniel, B., & Tjandra, S. (2019). Systematic review of the literature on big data in the transportation domain: Concepts and applications. *Big Data Research*, *17,* 35–44. https://doi.org/10.1016/j.bdr.2019.03.001

Nguyen, M. T., & Boundy, E. (2017). Big data and smart (equitable) cities. In P. Thakuriah, N. Tilahun, & M. Zellner, *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics* (pp. 517–542). Switzerland: Springer. doi:10.1007/978-3-319-40902-3

Nguyen, M. T., & Boundy, E. (2017). Big data and smart (equitable) cities. *Seeing cities through big data: research, methods and applications in urban informatics*, 517-542.

Nunan, D., Bankhead, C., & Aronson, J. (2017). *Catalogue of bias collaboration: Selection bias*. Oxford, UK: University of Oxford. http://www.catalogofbias.org/biases/selection-bias/

Pargman, R. (2020). New ransomware attacks. Mass Transit. https://www.proquest.com/trade-journals/new-ransomware-attacks/docview/2460544846/se-2

Pereira, R. H. M., & Karner, A. (2021). Transportation equity. In R. Vickerman (Eds*) International Encyclopedia of Transportation*, Vol. 1, p. 271–277. Elsevier. DOI: https://doi.org/10.1016/B978-0-08-102671-7.10053-3

Pouchard, L. (2016). Revisiting the data lifecycle with big data curation. *International Journal of Digital Curation, 10*(2), 176–192. http://dx.doi.org/10.2218/ijdc.v10i2.342

Prozzi, J. A., Hong, F., & Leung, A. (2008). Effect of traffic load measurement bias on pavement life prediction: A mechanistic-empirical perspective. *Transportation Research Record: Journal of the Transportation Research Board*, 91–98. doi:10.3141/2087-10

Rankl, W., & Effing, W. (2010). *Smart Card Handbook, Fourth Edition*. Wassenaar. Netherlands: John Wiley & Sons, Inc. https://ebookcentral.proquest.com/lib/usf/detail.action?docID=533998

Robare, M. (2023). Using data to increase transportation equity, https://www.urbansdk.com/blog/transportation-equity-data

Ruijer, E., Porumbescu, G., Porter, R., & Piotrowski, S. (2022). Social equity in the data era: A systematic literature review of data-driven public service research. *Public Administration Review,* 1–17. https://doi.org/10.1111/puar.13585

Scott, B. (2015). Big data with volume, velocity, variety, veracity and value. Insights. http://www.sagese.com/resources/volume-velocity-value

Shaheen, S., Bell, C., Cohen, A., Yelchuru, B., & Hamilton, B. A. (2017). *Travel behavior: Shared mobility and transportation equity* (No. PL-18-007). United States. Federal Highway Administration. Office of Policy & Governmental Affairs.

Shearmur, R. (2015). Dazzled by data: Big data, the census and urban geography. *Urban Geography, 36*(7), 965–968. http://dx.doi.org/10.1080/02723638.2015.1050922

Shi, F. (2021). Research on accessibility and equity of urban transport based on multisource big data. *Journal of Advanced Transportation*, *2021*, 1-18.

Stobierski, T. (2019). The advantages of data-driven decision-making. https://online.hbs.edu/blog/post/data-driven-decision-making

StreetLight Data. (n.d.). LBS data: Benefits for environmental justice studies. https://learn.streetlightdata.com/using-lbs-data-for-environmental-justice-studies-thank-you?submissionGuid=64e490f5-4aab-4cdd-b821-253d00781daf

Suich, H., Yap, M., & Pham, T. (2022). Coverage bias: the impact of eligibility constraints on mobile phone-based sampling and data collection. *International Journal of Social Research Methodology*, *25*(6), 849-860.

Tasse, D., & Hong, J. (2017). Using user-generated content to understand cities. In N. T. Piyushumita Thakuriah & M. Zellner, *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics* (pp. 49–64). Cham, Switzerland: Springer. https://link.springer.com/book/10.1007/978-3-319-40902-3

Taylor, L. (2016). No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *Environment and Planning D: Society and Space*, *34*(2), 319-336.

Torre-Bastida, A. D.-C. (2018). Big Data for transportation and mobility: recent advances, trends and challenges. IET Intelligent Transport Systems. *IET Intelligent Transport Systems*, 742–755. https://doi.org/10.1049/iet-its.2018.5188

Transportation Research Board. (2014). *Applying GPS data to understand travel behavior, Volume I: Background, Methods, and Tests*. Washington, DC: National Academies Press. https://doi.org/10.17226/22370

Twaddell, H., Zgoda, B., Aimen, D., & Morris, A. (2019). *Environmental justice analysis in transportation planning and programming: State of the practice* (No. FHWA-HEP-19-022). United States. Federal Highway Administration.

U.S. Department of Transportation. (2022). Equity action plan. https://www.transportation.gov/sites/dot.gov/files/2022-04/Equity_Action_Plan.pdf

University of Utah. (2015). Programming and prejudice. http://unews.utah.edu/programming-and-prejudice/

Verma, S. (2022). Sentiment analysis of public services for smart society: Literature review and future research directions. *Government Information Quarterly, 39*(3). https://doi.org/10.1016/j.giq.2022.101708

Vogels, E. (2021a). Digital divide persists even as Americans with lower incomes make gains in tech adoption. https://www.pewresearch.org/short-reads/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/

Vogels, E. (2021b). Some digital divides persist between rural, urban and suburban America. https://www.pewresearch.org/short-reads/2021/08/19/some-digital-divides-persist-between-rural-urban-and-suburban-america/

Vydra, S., Poama, A., Giest, S. N., Ingrams, A. R., & Klievink, A. J. (2021). Big data ethics: A life cycle perspective. *Erasmus Law Review, 14*(1), 24–44. doi:10.5553/ELR.000190

Welch, T. F., & Widita, A. (2019). Big data in public transportation: a review of sources and methods. *Transport Reviews*, *6*, 795–818. doi:10.1080/01441647.2019.1616849

Wennink, A. & Krapp, A., (2020). Equity-Oriented Performance Measures in Transportation Planning. *American Planning Association Planning Advisory Service (PAS) Memo.* https://www.planning.org/publications/document/9196998/

Wennink, A., & Krapp, A. (2020). Equity-oriented performance measures in transportation planning. *memo, American Planning Association, Planning Advisory Service*.

Wigan, M. R., & Clarke, R. (2013). Big data's big unintended consequences. *Computer*, *46*(6), 46-53.

Williams, K., & Golub, A., (2017). *Evaluating the distributional effects of regional transportation plans and projects.* NITC-RR-862. Portland, OR: Transportation Research and Education Center (TREC). https://doi.org/10.15760/trec.175

Wilson, K. (2022). What's on the US DOT 'Equity Action Plan' — And what's missing. https://usa.streetsblog.org/2022/04/21/whats-in-the-us-dot-equity-action-plan-and-whats-missing/

Zannat, K. E., & Choudhury, C. F. (2019). Emerging big data sources for public transport planning: A Systematic review on current state of art and future research directions. *Journal of the Indian Institute of Science, 99*(4), 601–619. https://doi.org/10.1007/s41745-019-00125-9

Zhang, Y., Li, W., Zhou, P., Yang, J., & Shi, X. (2016). Big sensor data: A survey. In *Internet and Distributed Computing Systems: IDCS 2016*. Lecture Notes in Computer Science (Vol. 9864). Springer. https://doi.org/10.1007/978-3-319-45940-0_14

Zheng, X., Chen, W., Wang, P., Shen, D., Chen, S., Wang, X., & Yang, L. (2016). Big data for social transportation. *IEEE Transactions on Intelligent Transportation Systems, 17*(3). doi:10.1109/TITS.2015.2480157

Zhu, L., Yu, F. R., Wang, Y., Ning, B., & Tang, T. (2019). Big data analytics in intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems, 20*(1), 383–398. doi:10.1109/TITS.2018.2815678

The National Institute for Congestion Reduction (NICR) will emerge as a national leader in providing multimodal congestion reduction strategies through real-world deployments that leverage advances in technology, big data science and innovative transportation options to optimize the efficiency and reliability of the transportation system for all users. Our efficient and effective delivery of an integrated research, education, workforce development and technology transfer program will be a model for the nation.

www.nicr.usf.edu