FREE MOVEMENT: ENHANCING OPEN DATA TO FACILITATE INDEPENDENT TRAVEL FOR PERSONS WITH DISABILITIES

by

SHILOH L. DEITZ

A DISSERTATION

Presented to the Department of Geography
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

March 2021

DISSERTATION APPROVAL PAGE

Student: Shiloh L. Deitz

Title: Free Movement: Enhancing Open Data to Facilitate Independent Travel For Persons With Disabilities

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Geography by:

Amy Lobben	Chairperson
Henry Luan	Core Member
Colin Koopman	Core Member
Stephen Fickas	Institutional Representative

and

Kate Mondloch	Interim Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2021

DISSERTATION ABSTRACT

Shiloh L. Deitz

Doctor of Philosophy

Department of Geography

March 2021

Title: Free Movement: Enhancing Open Data to Facilitate Independent Travel for Persons with Disabilities

Nearly 40 million Americans report a disability, and of this population, 70 percent travel less because of the challenges they face. When they do travel, those with limited mobility are more likely to be pedestrians or public transit users. Today, free commercial routing applications such as Google Maps offer a robust suite of tools for the able-bodied public to walk, ride bikes, take public transportation, or hail a taxi. Yet, such tools for persons with limited mobility to determine a safe and perhaps even pleasant urban route are experimental, limited, and only available in select cities (e.g. accessmap.io, chisafepath.com). This project intervenes by tackling the challenge of missing environmental data. First, I assess a regionally stratified sample of municipalities across the United States on their collection and maintenance of open data on environmental features that impact accessible travel for persons with disabilities. Based on this assessment, I evaluate options for filling in missing curb ramp data using machine learning and supplemental open data such as open street map, LiDAR, and aerial imagery. Finally, I look at the relevance and replicability of these GeoAI methods for filling in missing curb ramp data. Centering the needs of community members with disabilities, this research creates tools for improving mobility, increasing community strength and inclusivity while also critiquing the data driven scientific paradigm.

CURRICULUM VITAE

NAME OF AUTHOR: Shiloh L. Deitz

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Southern Illinois University, Carbondale
Whitworth University, Spokane

DEGREES AWARDED:

Doctor of Philosophy, Geography, 2021, University of Oregon
Master of Science, Geography, 2016, Southern Illinois University
Master of Arts, Sociology, 2014, Southern Illinois University
Bachelor of Arts, Peace Studies, 2008, Whitworth University

AREAS OF SPECIAL INTEREST:

data ethics, spatial analysis, GeoAI, cartography, infrastructure, social inequality

PROFESSIONAL EXPERIENCE:

Community Data Coordinator, Spokane Public Library, 2020-present

Visiting Scholar, Paul Simon Public Policy Institute, Southern Illinois University, 2016-2021

Graduate Assistant, Department of Geography, University of Oregon, 2018-2019

Water Resources Technician, CDM Smith, 2016-2018

Coordinator and Researcher, Paul Simon Public Policy Institute, Southern Illinois University, 2015-2016

Celia Howard Fellow, Paul Simon Public Policy Institute, Southern Illinois University, 2014-2015

Graduate Assistant, Department of Sociology, Southern Illinois University, 2014-2015

GRANTS, AWARDS, AND HONORS:

Doctoral Fellowship, National Institute of Transportation and Communities, 2020

Rippey Dissertation Writing Grant, Department of Geography, University of Oregon, 2020

M. Gregg Smith Fellowship, College of Arts and Sciences, University of Oregon, 2019-2020

Bill Loy Award for Cartographic Design, Department of Geography, University of Oregon, 2019

First Year Fellowship, Graduate School, University of Oregon, 2017-2018

David Arey Memorial Award for Creativity in Research, Department of Geography, Southern Illinois University, 2016

Ben Dziegielewki Graduate Award, Department of Geography, Southern Illinois University, 2015

Master's Graduate Fellowship, Graduate School, Southern Illinois University, 2012-2013

Presidential Scholarship, Whitworth University, 2006-2009

PUBLICATIONS:

Deitz, Shiloh, and Katie Meehan. 2019. "Plumbing poverty: mapping hot spots of racial and geographic inequality in U.S. household water insecurity." *Annals of the American Association of Geographers.* DOI: 10.1080/24694452.2018.1530587

Jackson, John, Charles Leonard, and Shiloh Deitz. 2019. "The Climate of Opinion in Illinois 2008-2018: Gridlock Broken?" *The Simon Review.* Carbondale, IL: Paul Simon Public Policy Institute.

Jackson, John, Charles Leonard, and Shiloh Deitz. 2016. "The Climate of Opinion in Illinois 2008-2016: Roots of Gridlock." *The Simon Review.* Carbondale, IL: Paul Simon Public Policy Institute.

Deitz, Shiloh. 2016. "Healthy Communities in Southern Illinois." *The Simon Review.* Carbondale, IL: Paul Simon Public Policy Institute.

Deitz, Shiloh, and Kristen Barber. 2015. "Geographies of Inequality: Urban Renewal and the Race, Gender, and Class of Post-Katrina New Orleans." *Race, Gender & Class* (January): New Orleans for Whom and What*?*

Barber, Kristen, and Shiloh Deitz. 2015. "Missing in the Storm: The Gender Gap in Hurricane Katrina Research and Disaster Management Efforts." In *Rethinking Disaster Recovery: A Hurricane Katrina Retrospective,* edited by Jeannie Haubert. Lanham, MD: Lexington Books.

Deitz, Shiloh. 2015. "Envisioning Carbondale's Creative Community: Current Challenges and Opportunities." *Paul Simon Public Policy Institute,* November.

Deitz, Shiloh. 2015. "Internet Access in Southern Illinois: Results from the Jackson/Williamson County Survey." *Paul Simon Public Policy Institute,* July.

Deitz, Shiloh. 2015. "Gender Sensitive Disaster Policy in Southern Illinois." *The Simon Review.* Carbondale, IL: Paul Simon Public Policy Institute.

Deitz, Shiloh. 2014. "Religiosity and Attitudes toward Immigrants and Immigration." *SIU Online Research Papers*, January 1.

For the renegade synthesizers

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Most of us living in the 21st century produce a steady stream of data throughout the day while conducting a variety of tasks that are also facilitated by data-driven services. Data are ubiquitous. Our data are used by an ever-growing number of government and commercial entities/services to make decisions that affect our lives - whether we may receive a loan, cross a border, or are stopped by the police.

However, there remain large gaps in data collection and use. For example, over 1 in 10 Americans report some disability.[1] Yet the digital tools available – including routing applications – continue to mostly cater to fully abled persons. An application like Google Maps runs on massive quantities of *proprietary* data and the skills of some of the best and brightest programmers and data scientists in the world. The fact that the app offers routes optimized for cyclists but not wheelchair users, for example, is not a matter of the impossibility of the task, instead it reflects corporate (and perhaps social) priorities and awareness.

Despite the examples above, data science has been incredibly useful for advancing scientific understanding and ameliorating social ills. Underpinning my research interest is a desire to understand power and inequality in data science and to better understand how data science can be mobilized for social equity. I approach my research with a focus on both scientific rigor (replicability, validity, and uncertainty

[1] ACS 2017, 5-year estimates

measurement) and social relevance (usefulness of outcomes or findings for specific populations or in application to particular social problems).

In this dissertation, I explore equity in data science through the case of accessible mobility in the United States. Preliminary analysis of a regionally stratified sample of the largest cities in the United States reveals that only 76 percent provide open government data. Of those that have made some datasets available to the public, only 14 percent have at least half of the datasets on environmental features that might facilitate accessible routing for persons with disabilities. Rather than letting available data determine the analysis, this dissertation takes a transformative detour by interrogating missing data sets through an empirical case study. I aim to both identify *what* is missing and *how* to fill in missing datasets or points - asking how we can produce more data to *better* serve marginalized populations.

a. To what extent do municipalities collect and maintain open-source data on environmental features that impact accessible travel for persons with disabilities?
b. How can predictive spatial modelling be used to impute missing data measuring environmental accessibility features?
c. What can we learn from this case about the relationship between social and scientific relevance in data science?

This dissertation is in journal article format. Chapter 2 addresses research question *a*. This paper evaluates the open data on environmental accessibility across a regionally stratified sample of 178 municipalities in the United States. I argue that while there is a robust discussion in the literature on data privacy, exclusion from big data represents a substantial injustice as well. Specifically, the exclusion of ADA and other accessible features from transportation databases is a systematic and systemic exclusion of people with disabilities. In Chapter 3, I address research question *b* on using predictive spatial modeling to impute missing data on environmental accessibility features. I focus on using machine learning algorithms to classify curb ramp locations in 9 urban areas across the United States. In this paper, I outline a methodology which achieves high classification accuracy and go on to explore how the key strength of machine learning algorithms – powerful classification and prediction on big datasets – is also a harmful weakness. I examine how error varies across context in ways that are not systematic, and this *outlier bias* cannot just be coded in. I propose a kind of rich or thick description of data error, which is slow and tedious, but direly needed if we truly intend to develop equitable AI. In Chapter 4, I address questions of relevance and replicability in the data-driven scientific paradigm (research question *c*). I answer these larger questions by looking at the impact of different data inputs and the subsequent machine learning classifications of curb ramp locations in Seattle, WA. On the way to thinking about replicability and relevance, I consider key aspects of the scientific method and the impact of new data and algorithms on these processes. Specifically, I look at the implications for scientific knowledge of abandoning careful and reflexive conceptualization, operationalization, and measurement. I argue that because of complexity, big data and AI do in fact “speak” and careful consideration of what they have to say is important.

CHAPTER II

MISSING DATA[2]

40.7 million Americans report some disability[3] and the majority (70 percent) reduce their travel because of their disability.[4] When they do travel, they are more likely than the general population to be pedestrians or public transit users. Wheelchair riders are significantly more likely than other pedestrians to be killed in traffic accidents (Kraemer and Benton 2015; Poon 2015). Thirty years after the landmark Americans with Disabilities Act legislation, these data suggest that pedestrians with disabilities still face several barriers to safely accessing their communities.

There is an abundance of transportation data made available through Google Maps and other routing applications, yet these applications overlook the unique needs of people with disabilities. If data on the accessibility of transportation infrastructure do exist, they are the intellectual property of private companies and unavailable to the public. While Google Maps has recently piloted an accessibility option for public transit, common conceptions of what constitutes a mode of transportation can contribute to the exclusion of people with disabilities. For example, why does Google's routing application define a bicycle as a mode of transportation but not a wheelchair?

Missing data on the accessibility of US municipalities is a significant barrier to safe mobility for persons with disabilities. Of the 178 largest municipalities across the United States that we sampled, only 60 percent provided both open data and at least one

[2] Authors: Shiloh Deitz, Amy Lobben, Arielle Alfarez
[3] ACS 2018, 5-year estimates
[4] Bureau of Transportation Statistics 2018

piece of information vital to the safe routing of persons with ambulatory disabilities. This information barrier transforms an impairment into a disability – excluding those who depend on features such as curb cuts from safe route planning. Moreover, the dearth of information on accessible features reduces the effectiveness of the Americans with Disabilities Act. As an unfunded mandate the ADA is toothless, but information (particularly quantitative information) is politically powerful (Jasanoff 1998; Koopman 2019b). Open and accessible data on public infrastructure would give advocates the tools to demand ADA compliance while also fulfilling the Title II administrative requirement for self-evaluation and transparency (ADA National Network 2020)[5].

The absence of useful data aggravates disability in complex and often unintentional ways, shining light on the mundane contours of life-threatening bias. This social reality starkly contrasts the excitement and bold claims that have bolstered and shaped a new research paradigm of data-driven science (Hey 2009). This paradigm marvels at big data or new volumes, velocities, and varieties of data (Laney 2001) and claims that available data can now provide an exhaustive picture of the world (see C. Anderson 2008; Hey 2009; Prensky 2009). But exhaustive for whom?

In this paper we focus on the lack of data on environmental features that might promote independent living and safe travel for persons with disabilities. Specifically, we ask to what extent municipalities collect and maintain open data on environmental features that impact accessible travel for people with disabilities, and what are the

[5] The Department of Justice administrative regulation for Title II include performing a self-evaluation, notifying the public about compliance, designating an employee to coordinate ADA compliance, developing a procedure for resolving ADA complaints, and developing a transition plan.

consequences of specific data regimes? We argue that looking at missing data reveals the social nature of data, which, like infrastructure itself, is stubborn to change and often biased in insidious ways[6].

Data-Driven Bias

While all research is driven by data, the buzzword 'data-driven' refers to quantitative information within the realm of data science. Hey (2009) uses the term data-intensive to describe a fourth paradigm of scientific discovery. According to that work, the preceding paradigms were empirical evidence, scientific theory, and computational science. While this framework might be debatable, the rise in reliance on data science and data-driven discovery is undeniable. The term "data-driven" was found in 15 publications in 1980, 177 in 1990, 502 in 2000, 2,408 in 2010, and 11,992 in 2019.[7] The popularity of the term in both public and private sectors has risen alongside advances in the internet and computing. Oft-cited, Anderson's (2008) proclamation of the end of theory went on to claim: "Petabytes allow us to say: 'Correlation is enough'….We can analyze the data without hypotheses about what they might show." These sentiments are not unique. A year later Prensky claimed that researchers could "mine the complete set of data for patterns that reveal effects, producing scientific conclusions *without* further experimentation" (2009, 5). At the same time, new analytics software was developed that claimed to have totally removed the human element and resultant bias (cf. Kitchin 2014). Various academic disciplines have been reshaped by these ideas, for example, prominent

[6] This kind of harm has been elsewhere characterized as "slow violence" or harm that happens so slowly and out of view that it is hardly measurable (Laurie and Shaw 2018).
[7] According to Semantic Scholar, an AI-powered aggregator of scientific literature.

geographers recently published an article outlining data-driven geography (Miller and Goodchild 2015).

The juxtaposition between human bias and objective (non-human) data has a long history (Haraway 1996; D'Ignazio and Klein 2020; Perez 2019). However, as data have gotten bigger and algorithms more complex, biases have taken on greater heft. These biases are often unintentional and confusing to researchers who trust the objectivity of data, but who lose comprehension of what is going on within the ever more opaque black box of big data and complex algorithms ((Burrell 2016; Suresh and Guttag 2020). Data are also infrastructures like roads, plumbing, and sidewalks—as such, they are embedded in history, stubborn to change, and only made visible when they fail (Star 2016; Selin and Sadowski 2015).

Real-world biases – systemic and repeatable errors that create inaccurate and often unfair outcomes - are built into our physical and research infrastructures with embodied consequences. These consequences are felt not only by women and minorities, but more so for those with physical impairments. While the data-driven research paradigm has been bolstered by big data and the accompanying belief that available data can now provide an exhaustive picture of the world, the world that is presented still defaults to the experience of white, able-bodied men (D'Ignazio and Klein 2020; Haraway 1996; Perez 2019). For example, because data about men were predominantly used in past heart disease research, women have long been endangered by overlooked and untreated heart disease symptoms that are either unique to or more prevalent in women (Garcia et al. 2016). Predictive policing over-targets black men because the training data (based on racial profiling practices) is biased in that direction (Richardson, Schultz, and

Crawford 2019; Selbst 2017). Facial recognition software does not easily recognize black women's faces because it is trained on white men and image recognition algorithms over-identify subjects in the kitchen as women (Simonite 2017; 2018). New software for automating hiring decisions overwhelmingly discriminates against disabled people by weeding out anyone with characteristics not found in the training data (Engler 2019; Whittaker et al. 2019). And autonomous vehicles have been found to not "see" wheelchair riders or pedestrians using other assistive technologies (Whittaker et al. 2019). All these biases contribute to real-world inequalities, are likely accidental, and stem from a lack of complete information or missing data. These examples show that no matter the size of data, it defaults towards further embedding whatever is believed to be socially "normal" at the dangerous exclusion of all kinds of constituency groups. Furthermore, biases towards constituency groups organized by race, age or gender are easier to identify than those affecting people with disabilities due to the size and diversity of the group. Examining data-driven biases that crop up for the disabled community is particularly instructive for understanding inequality. While we can organize people based on race, gender, sexuality, or any other demographic characteristics, social scientists have established that any such identity intersects with other facets of a person to produce unique experiences of the world (Collins 2008).

People are grouped based on socio/economic/political/physical attributes. Those grouping choices are social constructs (D'Ignazio and Klein 2020). For example, there is a feedback between our cultural conception of gender and its formatting in demographic surveys (binary, male/female).[8] Demographic classification has gotten more complex as more of our lives move online but no less restrictive in most cases. Our "biography" on various platforms takes on the shape of the categories given. As Koopman points out, "within a few short years we have all been retrained to present ourselves again through exactly the terms specified by the conventions of the newest social networks" (2019b, 7). Or as the disabilities movement called attention to, the definition of abnormal is socially and medically defined with consequences for both data formats and human lives. For example, homosexuality was a mental illness in the Diagnostic and Statistical Manual of Mental Disorders (DSM) until 1973 (Drescher 2015), underscoring the interplay between how we defined homosexuality and how gay people were treated in society. Women have long been thought of as mentally and physically deficient men (Garland-Thomson 2005) and the same is true of race. These ideas manifest in persistent race and gender-based pay gaps. In 2015, African Americans earned 75 percent as much as whites and women earned 83 percent as much as men in the United States (Patten 2016).

Data, classification, equations – these are socio-technical systems. That is, they are socially created and trusted, and are products of their time, making biases hard to weed out (Crawford 2017; Ananny 2016; Beer 2017). These systems can be seen as constituting an *infopower* that, as Koopman defines it, is "an exercise of power through the work of its varied and flexible formats" (2019b, 12). The process of data collection, analysis, and dissemination allows for information to amplify itself. In this way it both

[8] When we refer to "formats" or "formatting" we are referring to how information is conceptualized for quantitative measurement. For example, on demographic surveys gender could be male/female or a wider range of options (Koopman 2019b).

“pins us down” to the constraints of formats and speeds up by making things easier to find, analyze, and classify (ibid). This process is the subtle work of conceptualization. Formatting is simply the mundane choices about what and how to measure. Data size does not erase biases but amplifies them by overwhelming us with information. As data size increases, so does the opacity of the science conducted with it (Kwan 2016). In the very mundane work of formatting, certain possibilities are put in place which “preclude[d] what could have been other options” (Koopman 2019b, 180). Past injustices, like redlining or racial profiling in police work, cannot be solved by more data; rather, when coded into more data they are amplified.

Big data has generated many of the wildest claims about what data can do. However, the data-driven paradigm does not discriminate based on data size or quality and the sheer size of big data makes it much harder to track biases. We argue that to understand bias in quantitative analyses, there needs to be more serious review of the data themselves - the underlying processes and the mechanisms for biased results. While progress has been made in understanding the dangers of too much data and quantification, we lend another perspective suggesting that there actually are not enough data in areas that would benefit historically marginalized groups. For example, there wasn’t any comprehensive data on people killed by the police until activists demanded it (Onuoha [2016] 2019).

With respect to our focus here on accessible transportation, a dearth of data on key safety features such as curb cuts, cross controls and cross walks has material consequences for the safe routing of persons with disabilities while also reducing the power of the ADA. This deficiency in transportation data persist despite rapid advances in similar areas - high quality routing applications (for those without physical impairments) having become free and ubiquitous, and many of us worry that our phones are starting to collect *too much* information about where we've been and where we're going. Within this transportation data deluge – which both helps us navigate and threatens our privacy – what are the barriers to collecting and creating tools around data that would expand the independence and safety of persons with disabilities in their urban travel?

In this work we fill that gap by looking at the relationship between legislation, municipal data practices, and the consequences for persons with disabilities. Specifically, we trace the history of the ADA and open data. We look at the open data practices of a regionally stratified sample of municipalities to reveal the interplay between "common sense" data priorities, political power, and routine injustices. We confirm that there is a dearth of data on accessible transportation across a wide range of municipalities in the United States and suggest alongside artist Mimi Onuoha that these data are missing because those who have the resources to collect data lack the incentive to and are actually more incentivized to hide or obscure it due to complex enforcement of the ADA ([2016] 2019).

Opening the Americans with Disabilities Act

The Americans with Disabilities Act of 1990 was landmark legislation and a great achievement for the disability rights movement. The movement rhetorically shifted understandings of disability from purely medical to questions of social justice and design. The social model of disability situates disability as the product of disabling environments and attitudes. These barriers, which include physical environments, lack of assistive technologies, negative attitudes, and public policies transform impairments into disabilities. This model has been adopted by the World Health Organization, Centers for Disease Control, and United Nations (World Health Organization 2001; Siebers 2008). ADA legislation aims to reduce these barriers in the areas of employment (Title I), public accommodations and services (Titles II & III), and telecommunications (Title IV) in the United States.

Title II of the ADA bans disability-based discrimination in state and local governments. This legislation comes with many guidelines for the accessibility of pedestrian right of ways including curb ramps, removal of obstructions, and safe road crossings. New accessible infrastructure standards were released in 2010. The regulations also require that local governments: (1) perform a self-evaluation; (2) notify the public about ADA compliance; (3) designate an employee to carry out ADA responsibilities; (4) develop a procedure for resolving complaints; and (5) develop a transition plan for achieving ADA compliance.

Directly after the ADA's passage, municipalities struggled to comply in order to avoid a $50,000-$100,000 fine (Mills 1995). However, the Unfunded Mandate Reform Act of 1995 restricted the federal government's enforcement ability and ADA compliance has since then largely been enforced by citizens through complaints and lawsuits. A recent study of New England municipalities found that less than 1 in 10 municipalities were compliant with Title II requirements (Brault et al. 2019). When asked why, many participants responded that they had not had issues or did not think it was necessary. In fact, lack of knowledge or personnel were the most common reasons for ADA non-compliance, followed by money and time (ibid).

The administrative requirements of Title II to perform a self-evaluation, inform the public, and develop a transition plan all require data and data sharing. While open data is only mandated at the federal level through the OPEN Government Data Act (passed January 14, 2019), many municipalities have followed suit with their own policies or initiatives. The term "open data" describes data that can be freely accessed, used, modified, and shared for any purpose. The open data movement, active since the mid-2000s, attempts to bring data to more people in the interest of transparency, accountability, and innovation. Advancements in transparency and inclusion in data analysis have facilitated the identification of built-in bias in data sets and systems. Open data are a rich source not only for innovation, but transparency and accountability, too. That is, their value lies both in what they measure, and in what they ignore, giving citizens a metric for understanding their city's priorities *and* its blind spots.

There are no established metadata standards for open data and few guidelines about what to include, making them a wild west of localized quantification. Limited public funding for technology and open data infrastructure has often resulted in data openness without context. These dumps of "zombie data" are accessible and open but without provenance, meaning, or purpose (D'Ignazio and Klein 2020). Out of context, data cannot "speak for themselves" as many have claimed. The biases mentioned in the preceding section are only amplified when analyzed by persons unfamiliar with data collection practices and context. For example, if a person unfamiliar with police profiling aimed to garner insights from data provided by a police department with a history of profiling, the data might confirm their own racial bias.

These issues reveal the messy relationships between information, infrastructure, legislation, social justice, and American ideals. There is a scarcity of empirical work on data quality and bias. This includes a lack of systematic analysis of what data are missing and what the social consequences might be of that missingness. In the case of accessibility, the lack of information about transportation infrastructure not only impacts an individual's safety and ability to navigate the world freely, but also has implications for political power, legislative enforcement, and governmental transparency. For example, a municipal transportation database that collects information about street conditions such as potholes but not about the presence of curb cuts on sidewalks is a technical infrastructure that has potentially positive impacts for some (drivers who can avoid pothole-ridden routes or lobby for local street repair) and potentially negative impacts for others (wheelchair users who cannot plan safe curb access routes to new destinations or lobby for such access). Such data collection practices also result in a prioritization of one traveler over another, whether intended or not.

The early years of the big data revolution ushered in legal, political, and social debates centered around the risks of inclusion, i.e. privacy and civil liberties concerns. While we do not discount the merits of those concerns, we argue that exclusion from big data represents a substantial injustice as well. Specifically, the exclusion of ADA and other accessible features from transportation databases is a systematic and systemic exclusion of people with disabilities from these databases. In the following pages we present our empirical exploration of this missingness.

Methods

Our overall goal was to identify the extent to which municipalities make publicly available databases that include features that represent facilitators to environmental access for people with disabilities. Our process involved the collection as well as scoring/evaluation of municipal databases throughout the United States. We then looked for patterns in their collection practices to better understand the conditions for inclusion.

Study Area and Data Collection

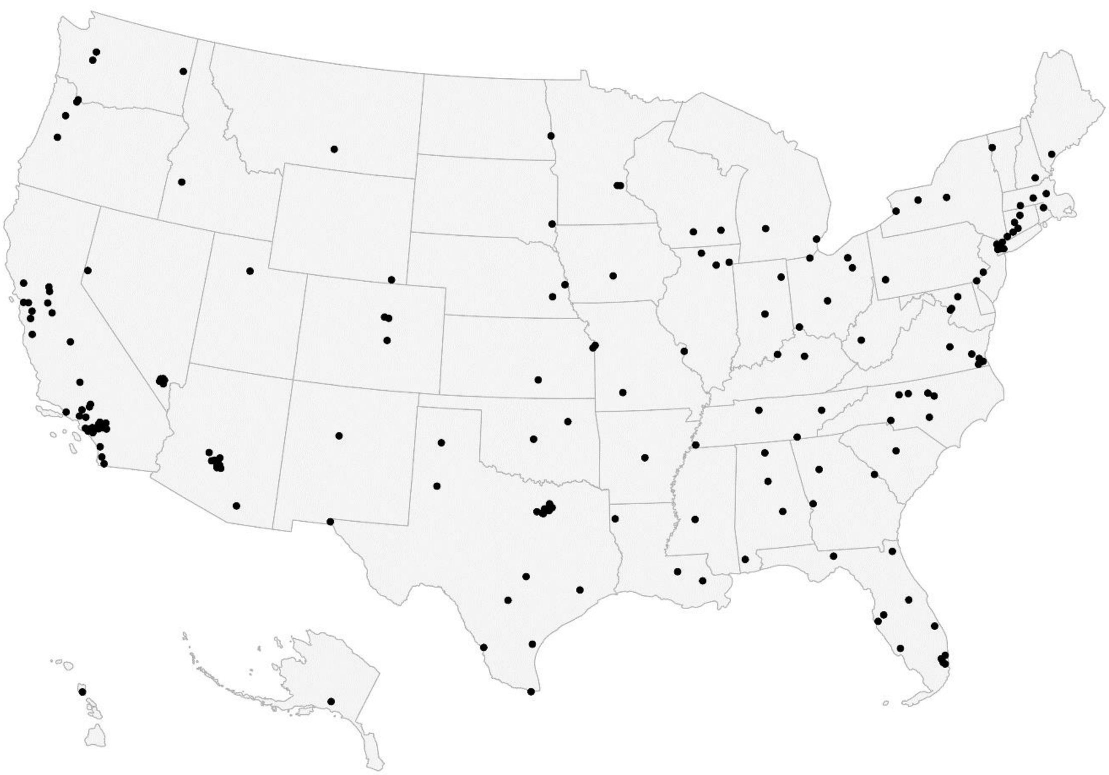

Figure 1. Municipality Sample

We used a regionally stratified sample of open data portals across 178 United States municipalities to answer the question: to what extent do municipalities collect and maintain open-source data on environmental features that impact accessible travel for persons with disabilities? The sample of 178 municipalities is based on 2010 U.S. Census population counts and includes: all municipalities with populations over 150,000, the 10 most populous municipalities for each census subregion, and the most populous municipality of each state (see figure 1).

Open data are freely available to access, share, and use. While it is conceivable that physical versions of open data might be freely available within a library, for example, the standard is increasingly that they are downloadable over the internet (Open Knowledge Foundation 2020). To ascertain the presence of an open data portal, we conducted internet searches[9] from June 2019 to March 2020. Search terms included the municipality name and “open data”, “data”, or “gis.” If nothing promising resulted from those searches, we would investigate the municipalities’ website for any information on data or transparency. We gave municipalities with no open data portal a score of 0 (the lowest score in our evaluation scheme; see below). We checked all municipalities with scores of 0 again at the end of the review period (April 2020) as some municipalities were in the process of developing a portal.

Database Scoring

Following the data collection, we then reviewed the existing open data portals for their inclusion of data on 14 environmental features that are barriers to or facilitators of safe travel for disabled pedestrians. The list of features is based on a two-year empirical study by Lobben and Perdue (in progress) in which interviews, focus groups, and national

[9] Using Google and DuckDuckGo search engines.

surveys identified environmental features that represented either barriers or facilitators to environmental access for people with disabilities. These features were identified as most impactful for safe routing across four impairment types: vision, hearing, electric wheelchair, and non-electric wheelchair. Table 1 combines Lobben and Perdue's 14 environmental features (left column) with ADA standards (McNally 2011; Department of Justice 2010).

Table 1. Environmental data types for accessibility & ADA standards

Feature	ADA Requirement
Cross control	Pedestrian push button requirements
Audible cross control	
Flashing cross control	
Vertical obstructions	Pedestrian ROW must be unobstructed
Ground obstructions	
Curb ramps	Connecting every path of travel; Width: >36"; Cross slope: <2%; Running slope: <8.33%
Slope	Recommendations for sidewalks
Crosswalks	
Sidewalks	Pedestrian ROW must be unobstructed
Sidewalk material	
Sidewalk condition	
Streets	If new construction pedestrians right of ways must comply with ADA
Street number of lanes	
Street speed limit	

Our scoring process was based on simple presence or absence of the features of interest. Therefore, a "perfect" score of 14 was achieved if the database included fields and data for all 14 features. Again, an absence of publicly available data resulted in a score of 0. Most municipalities included some combination of data on the 14 possible features (see supplementary data).

We then took a closer look at the metadata and fields for sidewalks, cross controls, curb ramps, and crosswalks as these features fall under ADA accessible design standards (Department of Justice 2010, see table 1). This was based on an emerging observation of a relationship between ADA compliance efforts and open data completeness. Possible fields included feature type (point, line, polygon), slope, ADA compliance, and signal type.

Demographic data for correlation analysis came from the American Community Survey 2018 5-year estimates for census designated places.

Analysis

Using these data, we compiled descriptive statistics of overall scores and each of the 14 accessibility features. We focused only on municipalities with an open data portal and at least one feature of interest (n=107). To better understand the score, we broke it into quartiles. We then looked at the rate of inclusion for each accessibility feature. To understand the relationship between data completeness and inclusion of features, we also calculated the rate of inclusion of each feature based on score. For example, 92 percent of municipalities with a score of 2 included spatial information about streets in their open data portal.

Next, we looked at the Pearson correlations between the presence of feature information, score, and demographics across the entire sample (n=178). We did this to look for a relationship between demographic or spatial characteristics of cities and their open data practices. In exploratory analysis, we looked at a variety of municipality characteristics including median age, income, poverty rate, disability rate, housing insecurity, median income, population size, latitude, and longitude. We also explored spatial correlations.

Based on the Pearson correlations, we ran a basic linear regression on each municipality's data inclusion score. The independent variables were municipality median income, population size, latitude, and longitude[10]. We conducted analysis of the outliers using regression residuals and standard deviation of the residuals. The residual is the difference between the actual and predicted value. We looked at all municipalities with predicted value residuals at least 1 standard deviation above or below the actual value.

Results: Accessibility and Access

The results reported here include both the entire municipality sample (n=178) and the municipalities with an open data portal and at least one accessibility feature (n=107). Tables 2 to 4 and figure 2 cover the more limited sample because we are looking at trends in those municipalities with data online. Tables 5 to 7 and figure 3 cover the entire sample to capture the relationships between municipality characteristics and score.

Data Inclusion

Just under 78 percent of the 178 municipalities sampled had an open data portal (138 total). Among these, 107 municipalities had information on at least one of the features we were looking for (77.5 percent). Half of those municipalities had data on 3 or fewer

[10] Latitude and longitude were used to capture geographic variability in open data practices. We observed that northern cities were more likely to have more inclusive open data portals.

features and only the top quartile had more than half of the 14 features (23 municipalities had 7 or more, see table 2).

Table 2. Quartiles of Municipality Scores (n=107)

Quartile	Score Range
1st	6-13
2nd	3-6
3rd	2-3
4th	1-2

Table 3. Feature Inclusion Rate (n=107)

Feature	%
Streets	90
St: Speed	53
Slope	44
Vertical Obstructions	37
St: Lanes	36
Sidewalks	34
Ground Obstructions	30
Crosswalk	19
Swk: Condition	18
Curb Ramp	17
Cross controls	17
Swk: Material	14
CC: flashing beacon	14
CC: audible	7

Ninety percent of municipalities with information on at least one accessibility feature had spatial information about streets (see table 3). Just more than half had information about the speed limit on streets (53%), followed by slope (44%), vertical obstructions (37%), number of lanes in streets (36%), sidewalks (34%), ground obstructions (30%), crosswalks (19%), sidewalk condition (18%), curb ramps (17%), cross controls (17%), sidewalk material (14%), flashing beacon cross controls (14%) and audible cross controls (7%)[11].

Table 4. Feature Inclusion Rate (%) by Score

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
N	71	15	26	15	13	7	6	8	6	2	2	3	3	1
Streets	0	67	92	100	85	86	100	100	100	50	100	100	100	100
St: Speed	0	0	38	53	69	71	67	63	100	50	100	100	100	100
St: Lanes	0	0	8	27	38	57	83	75	67	50	100	67	100	100
Sidewalks	0	0	12	7	23	57	33	88	83	100	100	100	100	100
Swk: Material	0	0	0	0	8	29	17	13	33	100	50	67	67	100
Swk: Condition	0	0	0	0	8	29	17	63	0	100	100	100	67	100
Curb Ramps	0	0	0	0	8	0	0	50	67	50	50	100	100	100
Cross Controls	0	0	0	0	8	14	50	38	50	50	50	33	100	100
CC: Audible	0	0	0	0	8	14	33	25	33	50	50	33	100	100
CC: Flasher	0	0	0	0	8	0	50	13	0	0	0	33	0	100
Crosswalks	0	0	0	7	0	29	17	50	33	50	100	100	100	100
Slope	0	13	15	33	54	71	50	63	100	100	50	100	100	100
Vertical Obs.	0	7	15	33	54	14	33	50	83	100	100	100	100	100
Ground Obs.	0	13	19	40	31	29	50	13	50	50	50	67	67	0

Looking at municipality score by inclusion of data, we found street information in most municipalities with scores under 5 (see table 4 and figure 2). Sixty-seven percent of the municipalities with a score of 1 have street information, 92 percent of those with 2, all of those with 3, 85 percent of those with 4 and 86 percent of those with a score of 5 (see table 4). Only among those with 8 or more features, do features other than streets appear in equal proportions. For those municipalities, there are an even number with information on streets, speed limit, and slope (6 municipalities, see figure 2). For scores higher than 8,

[11] From this point forward we will mostly refer to the feature (rather than 'data about said feature') for conciseness. For example, we will refer to 'curb ramps' and in doing so we do not mean the ramps themselves but rather the presence of information about them.

there are almost equal numbers of municipalities with each feature and few municipalities had those scores (see N in table 4). Information on curb ramps, an important part of ADA design requirements, are only found in one municipality with a score of 4 (Los Angeles) and this information is incomplete and only includes ramps installed in 2014. Otherwise, curb ramp information was only included in municipalities with scores of 7 or higher. Notably, the most common feature with any application to ADA compliance is crosswalks or sidewalk condition, both found in less than 20 percent of municipalities with information on any feature of interest.

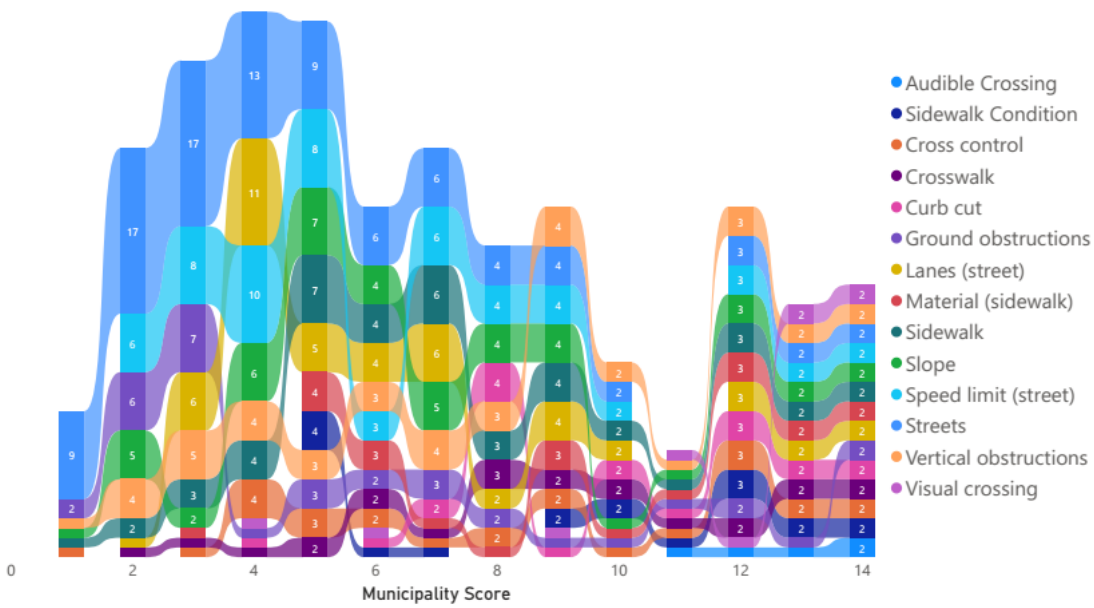

Figure 2. Municipality Score by Feature Inclusion

Patterns and Relationships

There are significant Pearson correlations between inclusion of feature information and score (see table 5). For example, audible cross controls are a descriptive field within a cross control dataset so there is a strong correlation between the two (0.904). There are also correlations between less-related features. For example, between data on curb ramps and cross walks (0.648) and curb ramps and sidewalks (0.527). The strongest feature correlation with score is between score and sidewalks (0.750), meaning municipalities with more accessibility features (higher score) were more likely to include sidewalks.

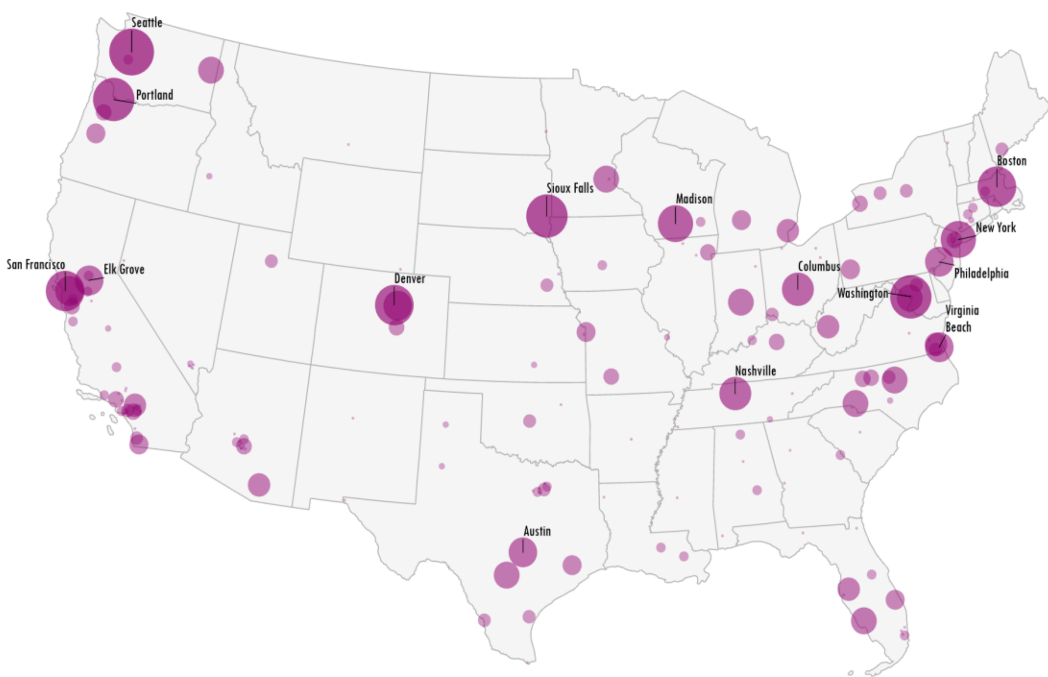

Figure 3. Municipality Scores (those with scores of 7 or more are labeled)

Table 5. Pearson Correlations (n=178)

	Mean, SD^	1	1a	1b	2	3	4	5	5a	5b	6	7	7a	7b	8	9
1.Cross controls	0.10	-														
1a. audible	0.08	**0.90**	-													
1b. flashing beacon	0.04	**0.60**	**0.56**	-												
2. Crosswalk	0.11	**0.47**	**0.40**	*0.20*	-											
3. Curb ramp	0.10	**0.44**	**0.37**	0.12	**0.65**	-										
4. Ground obs.	0.18	0.13	0.07	0.06	**0.25**	**0.28**	-									
5. Sidewalks	0.20	**0.34**	**0.35**	0.11	**0.53**	**0.53**	0.17	-								
5a. Condition	0.11	**0.37**	**0.42**	*0.21*	**0.57**	**0.49**	0.08	**0.69**	-							
5b. Material	0.08	0.17	*0.20*	0.04	**0.53**	**0.50**	0.17	**0.60**	0.62	-						
6. Slope	0.26	**0.26**	*0.23*	0.01	**0.39**	**0.43**	0.18	**0.52**	**0.37**	**0.42**	-					
7. Streets	0.54	*0.24*	*0.20*	0.13	**0.29**	**0.27**	**0.32**	**0.35**	*0.21*	0.16	**0.43**	-				
7a. Lanes	0.22	**0.36**	**0.38**	0.17	**0.29**	**0.32**	*0.21*	**0.51**	**0.43**	**0.28**	**0.39**	**0.49**	-			
7b. Speed	0.32	**0.29**	*0.23*	0.11	**0.29**	**0.37**	0.18	0.40	0.23	0.23	0.46	0.63	0.57	-		
8. Vertical obs.	0.22	0.31	0.32	0.099	**0.41**	**0.49**	0.274	0.399	0.337	0.37	0.47	0.36	0.33	0.38	-	
9. Score	2.58, 3.16	**0.62**	**0.59**	**0.32**	**0.69**	**0.70**	**0.41**	**0.75**	**0.66**	**0.59**	**0.68**	**0.65**	**0.69**	**0.67**	**0.66**	-
10. Med. income	29090, 6903	*0.22*	*0.20*	0.11	**0.28**	**0.34**	0.03	**0.29**	0.17	0.11	0.17	0.06	0.18	0.17	0.17	**0.28**
11. Population size	423177, 739484	**0.27**	**0.30**	0.04	*0.23*	0.14	0.10	**0.28**	**0.31**	0.01	*0.19*	*0.20*	**0.25**	0.14	0.22	**0.31**
12. % disability	12.14, 2.81	-0.08	-0.06	0.01	-0.10	-0.13	-0.03	-0.11	-0.08	-0.06	0.02	-0.04	-0.13	-0.12	-0.04	-0.11
13. Latitude	36.83, 5.37	0.07	0.07	0.08	0.19	0.22	0.08	0.15	0.10	*0.24*	0.17	0.00	0.05	0.09	0.18	0.19
14. Longitude	-97.25, 18.05	-0.01	-0.03	0.02	0.01	-0.05	0.01	0.07	0.09	-0.01	0.08	-0.02	0.15	0.05	0.10	0.06

^Included where meaningful, not for binary measures. **p<0.001** *p<0.01*

There are also relationships between characteristics of municipalities and score. There is a significant positive relationship between score and median income (0.281, p<0.001), population size (0.310, p<0.001), and latitude (0.185, p<0.05). Northern cities tend to have higher scores (see figure 3). While disability rate is correlated with income (-0.586, p<0.001) and latitude (0.176, p<0.05), there is no correlation with score. There is a significant negative correlation between disability and inclusion of information about curb cuts, albeit at a higher p-value (-0.132, p<0.1). Given other correlations, this relationship is likely spurious. It is notable because it is the opposite of what we might hope – as the rate of disability goes up, the likelihood of inclusion of information about curb ramps goes down.

Next, we ran a linear regression to understand how municipality characteristics come together to predict score. Fitted to a linear regression, median income, population size, latitude, and longitude predict about 21% of the variation in score (see table 6). The coefficients for median income, population, and latitude follow the same patterns observed in the Pearson correlations.

Table 6. Linear Regression Results (n=178)

	coefficient (standard error)	β (p-value)
Intercept	-3.219 (1.994)	---- (0.108)
Median Income	0.000 (0.000)	0.281 (0.000)
Population	0.000 (0.000)	0.291 (0.000)
Latitude	0.094 (0.040)	0.160 (0.020)
Longitude	0.020 (0.012)	0.115 (0.105)

R^2: 0.205
F-statistic: 11.185
Standard Error of Estimate: 2.852

Table 7. Regression Residuals for Municipalities with Scores of 6 or More

Municipality	Score	Residual	# of σ
Seattle, WA	13	7.389	2.591
Sioux Falls, SD	12	8.713	3.055
Portland, OR	12	8.445	2.961
Washington, DC	12	6.299	2.209
Denver, CO	11	6.941	2.433
Boston, MA	11	6.696	2.348
San Francisco, CA	11	5.694	1.996
Madison, WI	10	6.580	2.307
New York, NY	10	-3.497	-1.226
Nashville, TN	9	5.762	2.020
Columbus, OH	9	5.325	1.867
Oakland, CA	8	5.337	1.871
Aurora, CO	8	5.044	1.768
Elk Grove, CA	8	4.908	1.721
Austin, TX	8	4.420	1.550
Virginia Beach, VA	8	4.094	1.435
Philadelphia, PA	8	3.781	1.326
Cape Coral, FL	7	5.526	1.938
Spokane, WA	7	4.415	1.548
San Antonio, TX	7	4.264	1.495
Indianapolis, IN	7	3.597	1.261
Raleigh, NC	7	3.450	1.210
Charlotte, NC	7	3.321	1.164
Minneapolis, MN	7	3.305	1.159
Arlington, VA	7	-0.835	-0.293
Tucson, AZ	6	4.832	1.694
Tampa, FL	6	4.242	1.487
Charleston, NC	6	3.872	1.357
Detroit, MI	6	3.709	1.300
Norfolk, VA	6	3.635	1.275
Rancho Cucamonga, CA	6	3.463	1.214

Notably, every city with a score over 6 (except for New York and Arlington, VA) is an outlier with the predicted value at least one standard deviation below the actual score (see table 7). According to the model, we would expect New York to have a higher score – 13.5 instead of 10. The model predicted Arlington, VA fairly accurately (7.8

compared to an actual score of 7). Sioux Falls is the largest outlier with a score over 3 standard deviations larger than expected (the model predicts a score of about 3 compared to 12 in actuality). Portland, Seattle, Denver, Boston, Madison, Washington, DC, Nashville, and San Francisco all have scores above 9 that are at least 2 standard deviations higher than expected based on location, population size, and median income.

Discussion

As shown above, open data on key environmental accessibility features such as curb ramps, cross controls, and cross walks is rare across most municipalities in the United States. Indeed, cities with robust data on these features are outliers in our regionally stratified sample of the largest cities in the country. Missing data on the accessibility of US municipalities is a significant barrier to environmental access and safe routing for persons with disabilities. Not only does this lack of information preclude safe route planning for disabled pedestrians, leave them to fend for themselves with insufficient apps like Google Maps as their guide, but it also reduces the effectiveness of the Americans with Disabilities Act. Information is power, and public infrastructure evaluations would give advocates the tools to push for enforcing the infrastructural requirements of the ADA. In the following pages, we look at possible reasons for data missingness and the consequences of this data gap.

Formats

Early on in this research, we asked a software engineer at Google why their routing application (or any mainstream routing application for that matter) could compile the data for routing cyclists but not wheelchair riders. He gave a simple and sensible answer – a wheelchair is not a mode of transportation. According to Google Maps and the American Community Survey, modes of transportation include walking, biking, taking a taxi, driving, public transit, flying, and other (see figure 4).

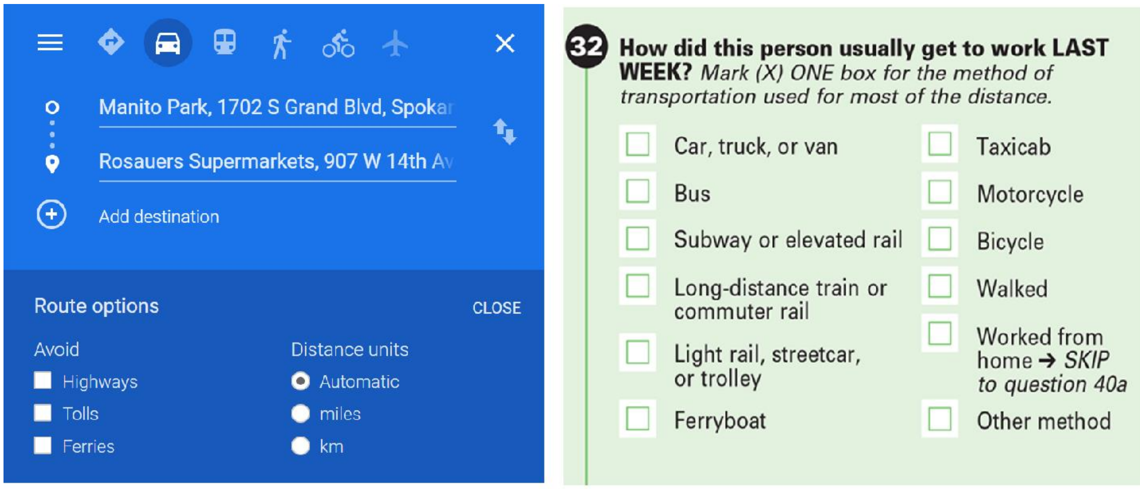

Figure 4. Modes of Transportation

Why are these the only modes of transportation? What is the difference between riding a bicycle or a wheelchair? According to the 2009 National Household Transportation Survey in the United States, fewer than 11% of daily trips were made by walking and only about 1% by bicycle (Kuzmyak and Dill 2012). The American Community Survey suggests these numbers have not changed very much in 10 years. In 2010, 0.5 percent of workers biked, and 2.8 percent walked to work. In 2018, these patterns were the same (considering the margin of error) - 0.6 percent biked, and 2.7 percent walked to work (ACS 2010/2018, Florida 2019). In contrast, 12.6 percent of the U.S. population has a disability and 7 percent of the population has an ambulatory disability (over 20 million people). These commuting and demographic data suggest that there are likely at least 10 times more wheelchair riders than cyclists. Yet, the municipalities in our sample are over 4 times more likely to have information about bicycling infrastructure than curb ramps (74 compared to 18).

While the focus of this research was not on modes of transportation and we did not do a thorough review of public transit datasets, our results do point to trends based on generally accepted ideas that municipal governments and corporate enterprises have about what constitutes a relevant mode of transportation. Information for drivers (car, bus, taxi) was more likely to be included than information for cyclists (96 municipalities compared to 74). Information for cyclists was almost twice as likely to be included as information for pedestrians. Sidewalk information was found in 36 municipalities. And half as many municipalities included information about accessibility (curb ramps, 18). These findings reveal that how we define and delimit modes of transportation influences the data collected and planning priorities.

Blank Spots

The Americans with Disabilities Act was a hard-won piece of legislation, which banned infrastructural discrimination based on disability. This 30-year-old legislation requires municipalities to implement accessible design in transportation infrastructure including curb ramps, sidewalks, and cross controls. However, implementation has largely been pushed along by complaints and lawsuits (squeaky wheels) rather than municipal initiatives.

In Mimi Onuoha's, visual art installation, *The Library of Missing Datasets,* she describes missing data as the "blank spots that exist in spaces that are otherwise data-saturated" (2020). She suggests that what we ignore reveals more than what we give attention to, and that these blank spots can illuminate subtle biases and indifferences.

Why are data missing? Drawing from Onuoha, we suggest four possibilities: (1) Those who have the resources to collect the data lack the incentive to do so (and possibly, have incentives not to (Hamraie, 2018)[12]); (2) the act of collection involves more work than the benefit the presence of data is perceived to give; (3) The data resist quantification; or (4) there are advantages to nonexistence ([2016] 2019).

In application to municipal data in our sample, the first three explanations likely apply. First, if municipalities collect data that reveals their compliance or non-compliance with ADA standards, it would either force an immediate fix or provide thorough documentation for a citizen complaint. We are not suggesting mal intent, only that the incentive is not there for those in the best position - city governments - to collect the data. This relates to reason three - the act of collection involves more work than the perceived benefit - particularly for those who do not depend on accessible infrastructure. Again, this is likely not done maliciously but rather ignorantly. According to a recent assessment of ADA implementation in New England, the top reasons given for non-compliance were lack of personnel (41 percent) and lack of knowledge (36 percent) (Brault et al. 2019). Further, the data resist quantification - not because they are unquantifiable (e.g. slope of curb ramp) but because they do not easily fit into our ideas about modes of transportation and officials are often ignorant or overwhelmed by the requirements.

These trends reveal biases in what data we deem worthy of collection, and the unforeseen consequences for inequality. It is difficult to fully extrapolate why every city with a high score was an outlier in our predictive model. However, there is evidence that

[12] Consider the example of police departments compiling data on officer involved shootings.

these municipalities either had a culture of accessibility or had collected data in response to high profile ADA compliance complaints or lawsuits. For example, the model predicted a score of 4 for Boston compared to the actual score of 11. Boston was also the location of the first ever Disability Pride March in 1990. Similarly, San Francisco also had a score of 11 which was about 6 points higher than expected. In the 1960s at Berkeley, the independent living movement was launched with attention to curb cuts for wheelchair riders. San Francisco was also the location of one of the most noteworthy protests of the disability rights movement, the 504 Sit-in in 1977. Recently a class action lawsuit was filed against the city and county – Kirola v. City & County of San Francisco in 2017. While the suit was not successful due to insufficient evidence, it did expand the city's understanding of their responsibilities under ADA. Sioux Falls, the greatest outlier with a score of 12 instead of the expected 3, undertook ADA compliance after a resident, Charles Santee, filed a 26 page complaint with the Federal Highway Administration, alleging that the city's infrastructure was dangerous to wheelchair riders and non-compliant with the ADA (K. Smith 2016). The Federal Highway Administration investigated and agreed with the complaint, giving the city 90 days to respond. The very thorough data collection on their open data portal is direct result of that action. Other high-profile cases include *Reynoldson et al. v. City of Seattle* (2017), *Denny v. City & County of Denver* (2016), and *Hines et al. v. City of Portland* (2018). All these actions were settled with the allocation of financial resources toward evaluating the infrastructure, sharing the results with the public, and coming up with a transition plan.

Conclusion

The lack of data on accessible transportation features has a profound effect on the safety, inclusion, and independence of people with disabilities. Further, this data gap likely has an impact on the overall quality of life for entire urban communities. Universal design principles have consistently been found to have far-reaching and unexpected social benefits (Hamraie 2017; 2018). This finding is known as the ‘curb cut effect’, a phrase coined by Angela Glover Blackwell, a long-time advocate for social, racial, and economic justice. The term points to these unforeseen benefits (and beneficiaries) of efforts at increasing equality and opportunity for any ‘second-class’ citizens – women, minorities, disabled persons, etc. The phrase is a reference to the mass introduction of curb cuts in American cities, and their unexpected use by parents pushing strollers, pedestrians rolling their groceries, and children cycling to school on sidewalks. In an ironic twist, the same tech companies who have largely ignored the needs of disabled pedestrians in routing have recently begun to pay attention to accessible infrastructure as an asset to their nascent robotic delivery experiments, which have thus far actually succeeded in making wheelchair travel *more* difficult, by presenting error-prone autonomous moving sidewalk obstacles (Ackerman 2019; Girma 2020; Hsu 2019).

Various scholars have pointed to the ways that data are not objective, are embedded in long histories of social prejudice written into the formats and codify bias into complex algorithms (D’Ignazio and Klein 2020; Koopman 2019; Perez 2019). Data are infrastructure even when they are about it – embedded, out of sight, only obvious when they fail (Star 2016) – but fail for whom? And who decides?

Koopman writes that resistance can be conducted from the ground up and encourages questioning the most mundane:

> *By inserting ourselves into those fissures, we can perhaps redirect enough of that hot heat so that tomorrow might be transformed. Where do we resist? In the forms, formats, and information. How do we assemble the competencies to resist these forms and formats? By learning how to reformat them. By understanding how to redesign them. By interrogating the manifold technologies with which they have been designed and redesigned* (2019b, 195).

From here we might imagine and begin to build a different world. Accessibility mappers at the Critical Design Lab posit that we must do more than show the environment as it is but use that information to imagine it differently (Hamraie 2018).

The fight for disability rights, like other civil rights movements, has been one of zigzagged, sometimes stalled progress. The United States is a country of strong ideals and incomplete promises. Legislation and decrees are but tools – not the final goal. Infrastructure, including data infrastructure, is embedded and powerful beneath the surface. While the ADA was landmark legislation, its implementation has fallen far short of ushering in universal design in American cities. A first step to understanding where infrastructure is non-compliant might be a modification to the ADA which requires a plan for publishing and maintaining infrastructure evaluation data – namely on sidewalks, curb ramps, cross walks, and cross controls. In the field of data science more widely, we need to institute better audit practices for identifying our biases and assumptions. Starting from the subjectivity of data, it should be common practice to question who decides what is counted, and who benefits? Further, we ought to take up the credo of the disability rights movement in every research project – “nothing about us without us.”

CHAPTER III

OUTLIER BIAS: AI CLASSIFICATION OF CURB RAMPS, OUTLIERS, AND CONTEXT

In 2018, three years after Uber started experimenting with autonomous vehicles (Metz and Conger 2020), Elaine Herzberg was walking her bicycle across the street in Tempe, Arizona, and was struck and killed by one of Uber's self-driving cars (Wakabayashi 2018). The algorithms detected her but classified her as a vehicle, then a bicycle, then other, and on and on until it was too late. How did this happen? The algorithm trained to recognize pedestrians wasn't taught to look for jaywalkers and had never seen anyone walking a bicycle (Whittaker et al. 2019; Marshall and Davies 2019). Delivery robots on city sidewalks have similarly failed to recognize pedestrians. Haben Girma, who is blind and walks with a service animal, was blocked on the sidewalk by a delivery robot (Girma 2020). Emily Ackerman, a power wheelchair rider, had a similar experience with a robot in Pittsburg. As she writes, "The robot was sitting motionless on the curb cut on the other side of Forbes Avenue. It wasn't crossing with the rest of the pedestrians, and when I reached the curb, it didn't move as the walk signal was ending. I found myself sitting in the street as the traffic light turned green, blocked by a non-sentient being incapable of understanding the consequences of its actions" (Ackerman 2019). These examples point to artificial intelligence's (AI) potentially-deadly inability to deal with context, nuance, and outliers.

While their unintended consequences can range from merely annoying to fatal, these technologies also promise to increase the mobility, freedom, and inclusion of people with disabilities in urban environments. Autonomous vehicles, which use deep learning to sense the environment in real time (Y. Li et al. 2020), would give people who cannot drive the ability to go wherever they like. The same AI sensing technology powers delivery robots, which could similarly improve quality of life by bringing prescriptions or groceries to people who otherwise hire someone to do those errands. Other scholars have used AI tools to examine the accessibility of the transportation environment, including: sidewalk assessment (Bolten et al. 2015; H. Li et al. 2018; Luo et al. 2019; V. Smith, Malik, and Culler 2013), ADA compliance evaluation (Abbott et al. 2018; Ai and Tsai 2016; Goldchain 2017; Saha et al. 2019), and crosswalk identification in aerial images (Ahmetovic et al. 2017; Berriel et al. 2017; Ghilardi, Jacques Jr, and Manssour 2018; Shioyama et al. 2001).

At the core of this contradiction between the benefits of AI and harms are two timeless questions – what kind of world do we want to live in? And who has the right to the city? AI has amplified certain visions of the world and the city[13]. These sociotechnical imaginaries – "collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understanding of forms of social life and social order attainable through, and supportive of, advances in science and technology" (Jasanoff and Kim 2015, 4) – are widespread, but not yet completely cemented in.

In this paper, I argue that the key strength of algorithms – powerful classification and prediction on big datasets – is also a harmful weakness. The related areas of big data and algorithms have been critiqued from all sides for their biases – harmful and

[13] See Zuboff's (2015; 2019; 2021) work on the relationships between capitalism, surveillance, and big data/AI.

systematic errors – but this literature largely overlooks the harms that arise from AI's inability to handle nuance, context, individuality, and exception. Using the case of curb ramp classification across 2 algorithms, multiple data sources, and 9 cities – I examine how error varies across context in ways that are not systematic, and thus cannot just be coded in. I propose a kind of rich or thick description of data error, which is slow, tedious, and subjective, but direly needed if we truly intend to develop equitable AI.

AI & Science

Big data and AI have shifted many disciplines to a new knowledge paradigm, data-driven science (Hey 2009). This paradigm is undergirded by the belief that big enough data are exhaustive, objective, and that with the right algorithmic mediation, they can reveal reality without human subjectivity (Hey 2009; Prensky 2009; Anderson 2008). Big data - data high in volume, velocity, and variety (Laney 2001) - are inextricably linked with AI, because the deluge of information that is "big data" cannot be processed without algorithmic mediation (Kwan 2016). AI encompasses any case of a machine *learning* to imitate human behavior. Data (information), features (things to look for), and algorithms (processes) [14] are needed for computers to learn.

Machine learning and deep learning are nested subsets of AI (see figure 5). While AI encompasses all tasks that require something like human intelligence, machine

[14] An algorithm is simply the steps taken to perform a task (Onuoha and Nucera 2018). For example, we probably all have personal algorithms for how we clean our rooms. However, for the more complex manifestations that are spoken about in the news, infrastructure provides an apt metaphor. Like infrastructure, algorithms are embedded out of view (Star 2016) and attract attention only when they fail (Hommels 2016). These complex algorithms are sociotechnical or "assemblages of institutionally situated code, practices, and norms with the power to create, sustain, and signify relationships among people and data through minimally observable, semiautonomous action" (Ananny 2016, 93)

learning includes tasks that humans cannot really do, such as combing through billions of records for patterns or trends and predicting outcomes (Onuoha and Nucera 2018). Unlike classic statistical analysis (such as linear regression analysis), in machine learning, a human provides the tools (e.g., data, concepts, parameters) and the machine chooses the exact algorithm. Deep learning algorithms, in contrast with the more general machine learning algorithms, involve the machine learning from itself over many iterations. It is near impossible to understand the mechanisms for improvement, which is to say, what the program is *actually learning* in deep learning applications.

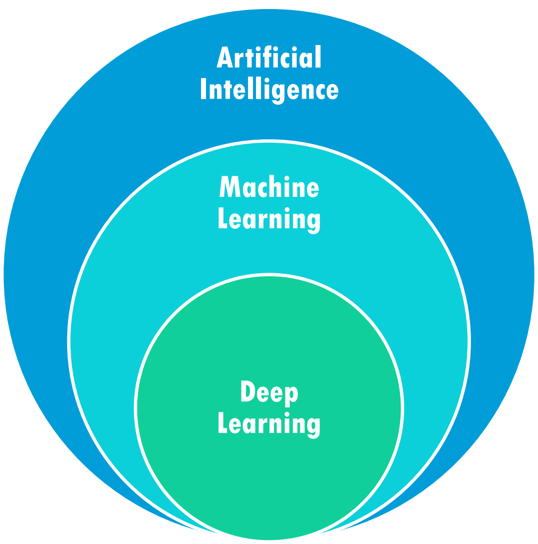

Figure 5. Artificial Intelligence

AI & Error

Machine learning with big data has brought a revolution in data science, and rightfully so. In classification and prediction tasks across a range of fields, high levels of accuracy can be achieved – where accuracy is simply the ratio of correct predictions or classifications out of the total possible. But this rapid and massive adoption of new techniques necessitates a matching increase in scrutiny and review, if we are to build responsibly. As machine learning applications have increasingly been used in ways that impact all areas of social life, critics have pointed to the systematic errors or biases in these predictions or classification which unfairly harm some populations. That is, a machine learning algorithm can achieve a high level of global accuracy while achieving very low accuracy among certain population segments. This type of error causes real-world harm. For example, facial recognition software less accurately identifies female and non-white faces (Buolamwini and Gebru 2018), people of color are more likely to be labeled criminals by policing algorithms (Selbst 2017) and more likely to receive longer jail sentences (Julia Angwin 2016), and families in poverty are more likely to be targeted by algorithms for child welfare investigations (Eubanks 2018).

These kinds of systematic and harmful errors could have been caught before they did harm by simply looking for patterns in model error. Google's ethical AI team released a paper in 2019 that amply covers a framework for doing just that. Their framework, "model cards for model reporting" suggests conducting "benchmarked evaluation in a variety of conditions, such as across different cultural, demographic, or phenotypic groups" (Mitchell et al. 2019). While the standardization of such ethical reporting is more complex in practice, the solutions are straightforward.

However, there is another class of errors which are potentially more pernicious because they are not so cleanly systematic. I call these kinds of errors *outlier bias* because they systematically exclude or harm any person or event that falls outside of a social or algorithmic conception of 'normal.'[15] People with disabilities are an expansive constituency group and the kinds of algorithmic errors that negatively impact this group could not be coded out of an algorithm easily. An algorithm used for employment screening, for example, has been shown to disproportionately harm people with disabilities (Engler 2019; Whittaker et al. 2019). Even recognizing this, weeding it out is almost impossible. As Shari Trewn, an accessibility researcher at IBM states, "The way that AI judges people is with who it thinks they're similar to—even when it may never have seen anybody similar to them—is a fundamental limitation in terms of fair treatment for people with disabilities" (Engler 2019). The encounters with autonomous vehicles and robots outlined above fit this kind of non-systematic outlier error within machine learning.

The problem with outlier error is that it is intrinsic to what machine learning is good at doing – powerful and globally accurate prediction and classification. Pattern and prediction – the major domains of machine learning - are hindered by and incompetent at understanding disorder, irregularity, heterogeneity, and uniqueness. As Onuoha and Nucera succinctly state in *A People's Guide to AI,* "…computers tend to be pretty bad at the things that we're good at, like understanding context and nuance…if you try to ask a computer to clean your room, it won't know what you mean, because it doesn't know what it means to 'clean', the equipment needed to 'clean' or even what a room is" (2018, 34).

[15] There is a rich body of work in the disabilities literature critiquing the use of normal and abnormal as socially constructed categories for distinguishing between people with or without disabilities (Kafer 2013; Linton 1998).

Despite claims of objectivity, values undergird AI and its applications. Living in this world with AI requires a more careful consideration of context and error. Such consideration cannot be conducted with numbers and formulas, but rather requires conscientious and reflexive work that is more akin to ethnography or thick description. As Ponterotto outlines, "thick description involves accurately describing and interpreting social actions within the appropriate context in which the social action took place" (2006, 542). I posit replacing "social action" with algorithm error in Ponterotto's definition, meaning we ought to describe the error within the context in which it took place to better understand the mechanisms and potential consequences. Accuracy is, by definition, "freedom from mistake or error" and "*conformity* to truth or to a standard" (Merriam-Webster, Inc 2021). If we are to robustly understand nuance and context, a step away from sole dependence on measures of conformity is essential.

In the following pages, I apply these ideas to one case – curb ramp classification in nine urban environments across the United States. I use two machine learning algorithms for classification, and examine error and accuracy in the results through traditional error and accuracy metrics - something akin to geographic and algorithmic benchmarking, and something like thick description of error. This work contributes not only to infrastructure classification, AI, and data ethics, but a more robust consideration of the role of geographers and qualitative scientists within data-driven science.

Methods

Study Area & Data

Figure 6. Sample Locations

AI methods require training data. I created a training set based on the open data about curb ramps for nine cities across the United States (see figure 6): Arlington, Boston, Denver, Indianapolis, Nashville, San Francisco, Seattle, Spokane and Washington DC. Based on previous work assessing the open data practices across U.S. municipalities (Deitz, Lobben, and Alfarez Forthcoming), these nine cities were the only ones with both adequate data on curb ramp locations (for labeling), LiDAR point cloud data that were open and spatially/temporally proximate to the curb ramp data, and aerial imagery. These data were sourced from the USGS (USGS 2020) (see table 8). The LiDAR point clouds had densities ranging from 4 to 45 points per meter. Data on the location of streets and landmarks came from the Census' TIGER/line shapefile program (US Census Bureau

2020). I classified ground points in the LiDAR data using the standard algorithm in ArcGIS. This algorithm has a tolerance for slope variation that allows it to capture gradual changes in topography (ESRI 2020). With the street data, I created a point feature for intersections by finding all locations where two streets crossed.

Table 8. Data Source and Year of Measurement*

municipality	aerial imagery	LiDAR	curb ramp locations
Arlington	2016	2018	(Arlington County, VA 2020)
Boston	2016	2013	(City of Boston 2014)
Denver	2015	2013	(City of Denver 2018)
Indianapolis	2014	2016	(City of Indianapolis 2018)
Nashville	2014	2016	(City of Nashville 2019)
San Francisco	2016	2018	(City of San Francisco 2020)
Seattle	2015	2016	(City of Seattle 2020)
Spokane	2015	2015	(City of Spokane 2020)
Washington, DC	2015	2014	(City of Washington, DC 2010)

**all TIGER/line data are from 2020*

I then created a tessellation of hexagons across each city. Each hexagon has an approximate area of 50,000 square feet (see figure 7). I selected one of these hexagons per city as my training sample for a total of nine hexagons. This was done to reduce data size and to have a balanced sample[16] area from each city. I created a second sample of 10,000 square foot hexagons as the validation set for each city.

Within these nine hexagons, I created a second hexagon tessellation of equal areas of about 50 square feet. This was done for multiple reasons – to capture the areal nature of a curb ramps in real life, to create a balanced sample of labeled points, and to add spatial contiguity to the balanced sample. While a curb ramp is not necessarily a discrete

[16] By balanced, I mean an equal number of ramp features and non-ramp features.

entity (but rather a sloped and blended area connecting paths), most curb ramps do not extend beyond 50 square feet[17]. These smaller hexagons were clipped to the LiDAR labeled ground locations because a curb ramp is not going to be found, for example, on the top of a building or in a tree (see figure 8).

Figure 7. Training and validation samples

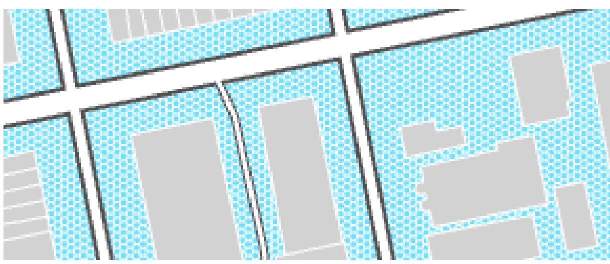

Figure 8. Ground hexagon samples

Next, I created a sample of the smaller hexagons with equal numbers of ramps and non-ramps across the training and validation areas of each city. In total, this resulted in 25,822 hexagons for training, and 5,054 for validation. These hexagon locations were used to clip the point data from the LiDAR point cloud. The point dataset had 1,117,930 points for training and 251,131 for validation. Within the point training and validation

[17] These choices show the interaction between data and algorithmic constraints or needs and how those limit the kinds of results that are possible.

set, I calculated the distance to the nearest: street, intersection, and landmark. I also summarized the DEM and aerial image raster information at the point (see figure 9).

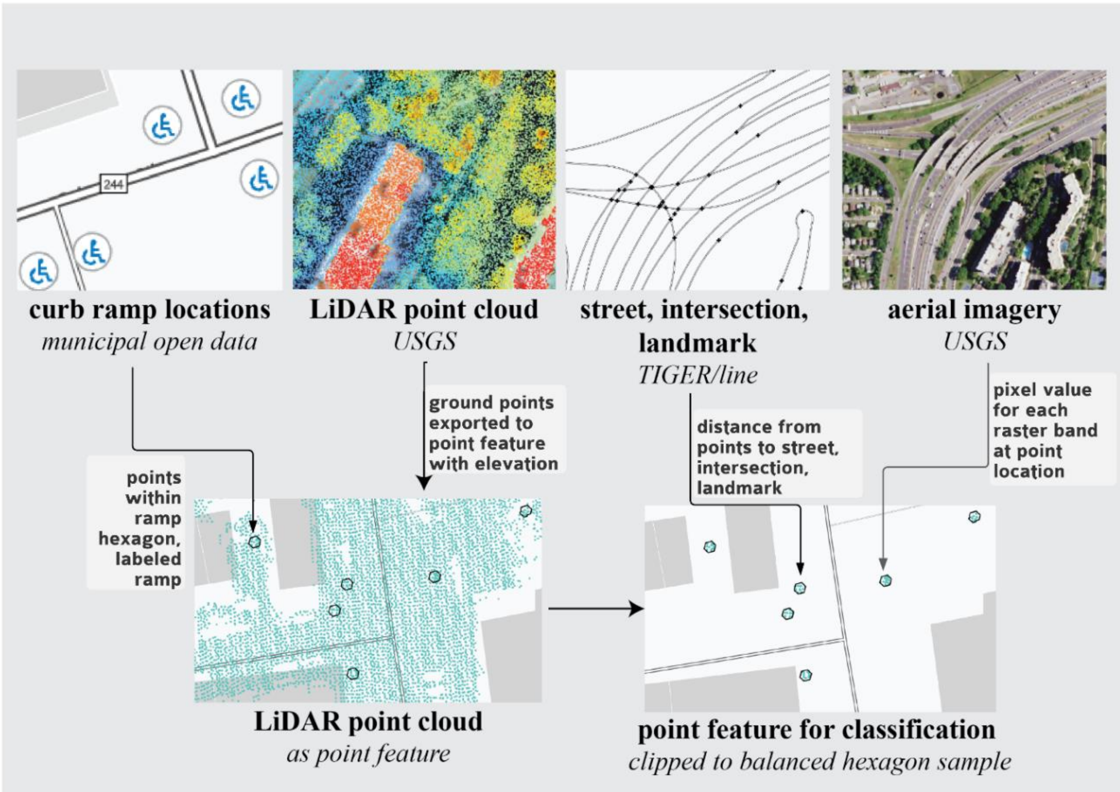

Figure 9. Data Cleaning

Machine Learning

To test the impact of algorithm choice on the results across different contexts, I ran a random forest regression model and a Point CNN model (see figure 10). The random forest is a classic machine learning algorithm, and the Point CNN is a newer deep learning algorithm.

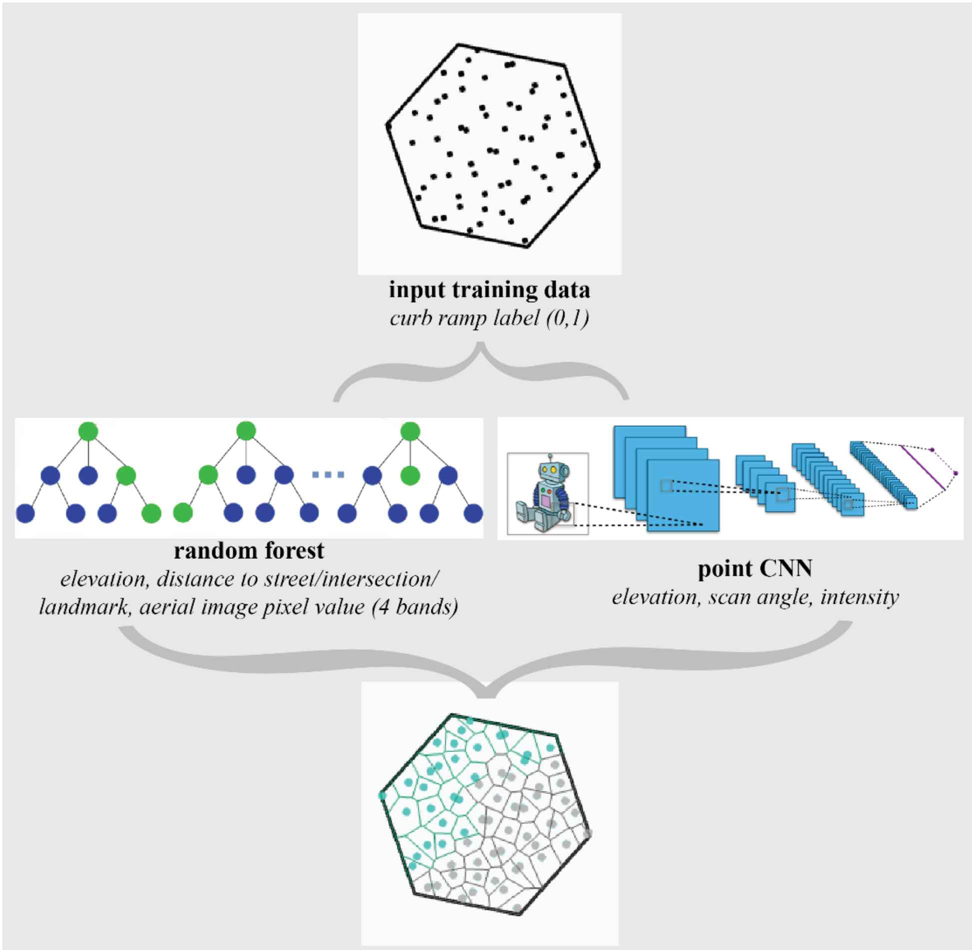

Figure 10. Ramp point classification

Random Forest Classification

Random forest classification is one of the oldest machine learning methods. It was developed by Tin Kam Ho in 1995 and improved by Leo Breiman in the early 2000s (Tin Kam Ho 1995; Breiman 2001). The random forest regression method simply involves running multiple models or decision trees through a set of training and prediction data and allowing them to vote on the answer for each location. I ran the random forest using the Scikit Learn library for python in Google Colab. The variable to predict or classify was a binary ramp value, and the potential explanatory variables were distance to street, distance to intersection, distance to landmark, elevation (from DEM), and raster value across all 4 aerial imagery bands. The random forest model is spatially unaware, in the sense that it does not account for or weight for neighboring features. The distance measures and raster values were added as predictors to manually give the models spatial information.

Point CNN

Point CNN is a deep learning model using convolutional neural networks (Yangyan Li [2017] 2021; Yangyan Li et al. 2018). Convolutional Neural Networks, or CNNs, are a promising advancement in object detection using AI, and have been used extensively on autonomous vehicles. At the basic level, a CNN scans over the input data and breaks it into smaller parts. The CNN runs through alternating convolutions and further subsampling. This process changes the information from sparse to dense iteratively. Over the process, the feature size is reduced but information is added (see top of figure 11). An apt metaphor for understanding a CNN is the act of playing with a cotton ball. Picture taking a round and condensed cotton ball out of a bag and absent mindedly and iteratively pulling it apart and pushing it back together. If this continues, we might reach a point where all the once uniform parts are separated into individual strands giving us a richer idea of the cotton balls composition. Classic CNNs work better on uniform data such as the gridded pixels in photographs (raster images). Point CNN is especially suited for LiDAR data because it accounts for the irregular and unordered nature of point clouds (see bottom of figure 11). Point CNN first weights and permutes the input features through a process called x-conv. X-conv involves a series of operations on processed point cloud blocks including sampling and normalization with K-Nearest Neighbors. That is, using a subsample of points, this process finds the nearest neighbors or most similar points and recursively aggregates them to represent a smaller set of points richer in information. Following the x-conv step, point CNN then conducts a set of classic convolutions.

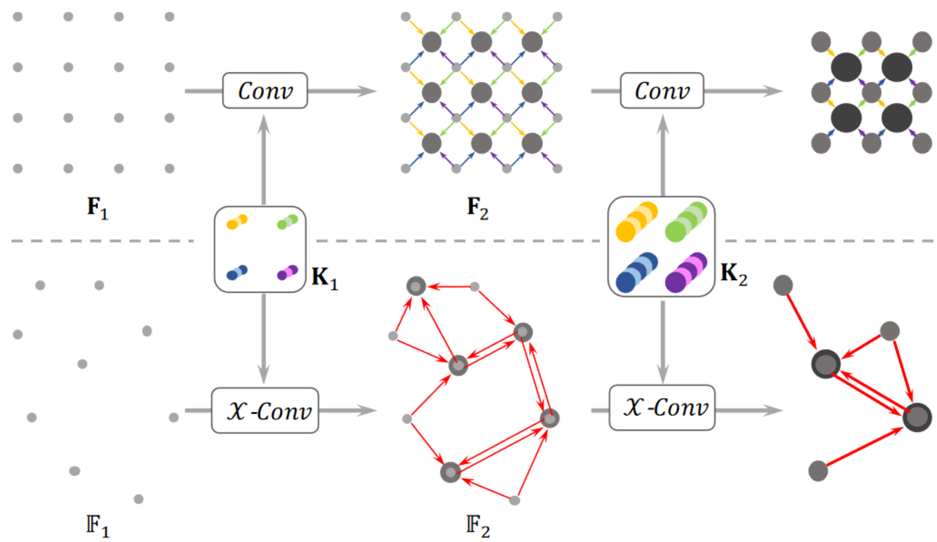

Figure 11. CNN and Point CNN Convolution Process (Yangyan Li et al. 2018)

I ran Point CNN using the ArcGIS deep learning frameworks and Jupyter notebooks. I set up my model with one hundred epochs, early stopping[18], and a one cycle learning rate. The one cycle or cyclical learning rate method lets the learning rate cycle between reasonable boundaries and has been shown to achieve improved classification accuracy over fewer iterations without a need for further tuning (L. N. Smith 2017). Extra features (beyond x, y, and z) used for learning included LiDAR intensity, scan angle, and return number. The validation loss stopped decreasing at 11 epochs.

[18] Meaning the algorithm should run for 100 epochs or until the validation loss stopped increasing.

Error & Outlier Analysis

I analyzed the success of these models on the classification task according to classic accuracy metrics, contextual benchmarking, and a qualitative process of ground-truthing Google street view imagery.

Performance metrics

As stated above, machine learning results are predominately evaluated in terms of overall accuracy and by class (see table 9). Overall accuracy is simply the proportion of correct classifications over all possible classification locations. Precision is used to understand the proportion of positive identifications that were correctly classified. Recall accounts for the number of false negative classifications or the number of actually positive classifications that were correctly identified. I calculated all three performance metrics for each model across the entire sample.

Table 9. Performance metrics

Accuracy	Precision	Recall
$\frac{true\ positives\ (tp) + true\ negatives (tn)}{sample\ size}$	$\frac{tp}{tp + false\ positive\ (fp)}$	$\frac{tp}{tp + false\ negative\ (fn)}$

I also looked at the rate of agreement between models and the original training data. There were 8 possible combinations of results (see table 10). The first two were that all were correct (all 0, all 1: all0, all1). The second two were that all were incorrect (all false negative, all false positive: allFN, allFP). The next four were combinations of only one model or the other making a false classification (Point CNN false negative, Point CNN false positive, random forest ramp false positive, random forest ramp false negative: (pcnnFP, pcnnFN, rfFP, rfFN).

Table 10. Possible Classification Result Combinations

short name	ramp	rf	pcnn[19]
all0	0	0	0
all1	1	1	1
allFN	1	0	0
allFP	0	1	1
pcnnFP	0	0	1
pcnnFN	1	1	0
rfFP	0	1	0
rfFN	1	0	1

Contextually Benchmarking

I then looked at error in a way similar to that proposed in "model cards for model reporting." That is, "benchmarked evaluation in a variety of conditions" (Mitchell et al. 2019). In this case, my conditions were geographic area and model. I calculated the overall accuracy by city and model as well as the agreement between models by city.

Qualitative Analysis

Finally, I used satellite imagery (Google street view) and cartographic visualizations to better understand the error and look for potential non-systematic outliers. I mapped the model predictions on an aerial image and found the locations in Google street view. Pairing the results, aerial imagery, and street view imagery allows for thicker description of the performance measures and contextual benchmarking. This method allows for the identification of outliers, surprise trends, and the act of showing rather than just telling and reporting opens the process up to other interpretations and observations. The results of this qualitative work are presented in the discussion section below.

[19] For brevity, I will sometimes refer to the original label from the municipal open data as just "ramp", the random forest ramp classification as "rf", and the Point CNN ramp classification as "pcnn".

Performance Metrics

Across all cities, the random forest model performed much better than the Point CNN in terms of accuracy – 88% compared to 56% (see table 11). The precision on the random forest model was slightly better than the recall because the number of false positive classifications was slightly smaller than the number of false negatives (89.8% compared to 89.4%). The recall on the Point CNN model classifications is much better than precision (71.9% compared to 59.9%), meaning there were more false positives than false negatives values.

Table 11. Classification Accuracy

	validation (n)	accuracy	precision	recall
Random Forest (rf)	251,131	88.0%	89.8%	89.4%
Point CNN (pcnn)		56.0%	59.9%	71.9%

The Point CNN included information about LiDAR intensity, scan angle, and return number, but there is no way of calculating the importance of each variable. Importance metrics in random forests are generated by calculating each time the variable is responsible for a split or decision in each decision tree. In the random forest, distance to street intersection and distance to street (two closely related variables) were the most important variables (see table 12). Street intersection was responsible for a split 47% of the time and distance to street 17% of the time. Elevation was the next most important variable (9.6%), followed by distance to landmark (7.1%), and then the aerial image bands.

Table 12. Variable Importance, Random Forest

Variable	Responsible for a Split in Random Forest (%)
Distance to landmark	07.1
Distance to street intersection	47.4
Distance to street	17.1
Elevation	09.6
Aerial Image	
band one	04.6
band two	04.6
band three	05.1
band four	04.4

Contextual Benchmarking

Algorithmic context

Across all the models, there were 8 different combinations of possible results (including the initial ramp label used for training from the municipal open data). Just over half of the points (50.1%) were classified correctly by both models (12.9% all 0, and 37.2% all 1, see table 13). Only 6.1% were classified incorrectly by both models (1.8% all false negative and 4.3% all false positive). The point CNN model falsely classified almost a quarter of the points as ramp (23.4) and falsely classified 14.4% as non-ramp – these were points that the random forest model correctly classified. The random forest model classified 5.8% of the points incorrectly when the point CNN classified them correctly (1.6% FP and 4.2% FN).

Table 13. Classification results across all models

	ramp, rf, pcnn	total points	%
all 0	0, 0, 0	32,456	12.9%
all 1	1, 1, 1	93,413	37.2%
all FN	1, 0, 0	4,601	01.8%
all FP	0, 1, 1	10,733	04.3%
pcnn FP	0, 0, 1	58,943	23.4%
pcnn FN	1, 1, 0	36,188	14.4%
rf FP	0, 1, 0	4,011	01.6%
rf FN	1, 0, 1	10,786	04.2%

Spatial context

The classification accuracy varied widely across cities and models (see table 14). The average accuracy for the random forest model was 88% across all cities. Arlington (91%), Denver (96%), San Francisco (90%), Seattle (94%), and Spokane (92%) performed better than average. Boston (87%), Indianapolis (74%), Nashville (85%), and Washington, DC (85%) performed worse than average. The classification accuracy of the point CNN model was higher than average in Boston (62%), Denver (67%), Nashville (58%), Spokane (75%), and Washington, DC (60%). It was lowest in Arlington (41%), followed by Indianapolis (53%), San Francisco (45%), and Seattle (45%). Since the data are balanced between ramp and non-ramp classes, a model could predict every point as either ramp or non-ramp and achieve better accuracy than Arlington, San Francisco, and Seattle. In summary, Denver and Spokane performed better than average across both models, and Indianapolis performed worse.

Table 14. Accuracy across models and cities

location	rf	pcnn
Overall	*88.0%*	*56.0%*
Arlington	91.2%	41.1%
Boston	87.2%	61.9%
Denver	96.3%	66.7%
Indianapolis	73.8%	52.8%
Nashville	85.0%	58.0%
San Francisco	90.4%	45.4%
Seattle	93.5%	45.0%
Spokane	92.0%	74.8%
Washington, DC	84.8%	59.9%

**above average values highlighted.*

Spatial & algorithmic context

In the same way that accuracy varied by city, the agreement between algorithms also varied across cities (see table 15). Arlington (38%), Boston (32%), Denver (48%), Indianapolis (30%), Nashville (48%), and San Francisco (42%) had above average agreement between models on non-ramp classifications (all0). The remaining cities – Seattle (39%), Spokane (70%), and Washington, DC (47%) had above average model agreement on positive ramp classification (all1). All models falsely classified ramp locations as non-ramp (allFN) more than average in Arlington (6%), Boston (4%), Indianapolis (10%), Nashville (6%), and San Francisco (6%). In Spokane (5%) and Washington, DC (6%) a slightly higher proportion of points were falsely labeled as ramps across models (allFP). The random forest model performed worse than average by falsely labeling ramp points as ramps (rfFN) in Washington, DC (7%). On average, 2% of the points were falsely classified as ramps by only the random forest model – this was higher in Arlington (3%), Boston (5%), Indianapolis (15%), Nashville (7%), and San Francisco (4%). The performance of the PointCNN in false negative and positive classifications was widely varying across cities. The rate of false negative classifications was higher in Arlington (53%), Boston (28%), Denver (24%), Indianapolis (34%), Nashville (23%), and San Francisco (49%). The number of false positive classifications by the Point CNN was nearly double average in Seattle (49%) and slightly higher than average in Washington, DC (26%).

Table 15. Classification results across models by city

location	all 0	all 1	all FN	all FP	rf FN	rf FP	pcnn FN	pcnn FP
overall	*12.9%*	*37.2%*	*1.8%*	*4.3%*	*4.3%*	*1.6%*	*14.4%*	*23.5%*
Arlington	38.4%	0.1%	6.3%	0.0%	0.0%	2.5%	52.6%	0.1%
Boston	31.7%	23.5%	4.1%	2.0%	2.1%	4.5%	28.2%	3.8%
Denver	47.5%	17.9%	0.6%	1.8%	0.4%	0.9%	24.1%	6.9%
Indianapolis	30.3%	6.4%	9.7%	0.5%	1.3%	14.7%	34.2%	2.8%
Nashville	48.1%	2.3%	6.4%	0.9%	0.5%	7.1%	22.5%	12.1%
San Francisco	41.6%	0.1%	5.9%	0.0%	0.0%	3.7%	48.8%	0.0%
Seattle	3.5%	38.9%	0.1%	3.7%	2.5%	0.2%	1.9%	49.3%
Spokane	1.8%	70.0%	0.2%	4.8%	2.7%	0.3%	4.6%	15.6%
Washington, DC	5.2%	46.6%	0.9%	6.2%	7.1%	1.0%	7.5%	25.5%

**above average values highlighted.*

Discussion

The results of this study confirm that machine learning algorithms are generally good at learning classifications. The random forest model was more accurate than the deep learning point CNN model at curb ramp classification in this context[20].The random forest classified ramp locations with 88% accuracy and the Point CNN achieved only 56% accuracy. The point CNN model's classifications were only slightly better than assigning every point as a ramp. This is a surprising result because point CNN is a spatially aware machine learning model and tends to be better suited for spatial applications such as this.

These models likely differed in performance in part because they had different information to learn from. The random forest model had information about streets and intersections, which proved to be important variables for random forest decisions or

[20] With three times more data, the Point CNN achieved 69% accuracy, suggesting that giving it more information might improve performance. These results are not reported here because scaling up makes it more difficult to understand error.

splits. Distance to street intersection accounted for 47% of the random forest decisions and distance to street accounted for 17%.

The importance of street design in what the random forest learned about classifying curb ramps is also apparent by looking at maps of the cities that performed best compared to the others. The classification accuracy in Denver was 96% - the city, or at least the portion of the city used for training and validation is almost perfectly gridded (see figure 12). The random forest also performed better than average in Seattle, Spokane, Arlington, and San Francisco. These cities are less perfectly gridded than Denver but certainly more than Washington, DC, Nashville, Boston, or Indianapolis.

Figure 12. Street design by city

While this result seems obvious – machine learning was better at prediction on municipalities with predictable or uniform design – the implications are important. If AI only works on cases that are similar, predictable, and dominant, this has serious social implications for human uniqueness, variability, and individuality. Disability scholars have long pointed to the harmful effects of categories like "normal" and "abnormal" (Goggin, Steele, and Cadwallader 2017; Saltes 2013; Shakespeare 2007). AI further produces and reflects a normative vision of the world – where outliers are at best ignored and at worst disadvantaged.

In Denver, the random forest produced 167 false positive predictions and 62 false negatives (2.7% and 1.0% of the city's validation points respectively). Figure 13 shows a few of the outliers that were not correctly classified in the random forest model. In gray are areas where the random forest incorrectly classified points as ramp (*a*, *c* in figure 13). In image *a*, it appears that the location is misaligned to where there is a ramp (see image a.1.), and rather, located at the path from someone's house (see image a.2.). This path has most of the characteristics of a curb ramp but no gradual slope which makes it not functionally a curb ramp. It appears that the random forest classified the part that is like a ramp as a curb ramp, and the edge where there is no transition as non-ramp (see green area of the hexagon in quadrant a). Functionally, a location that is almost a ramp in this way is missing the key functionality – transition from a higher to lower path. Location c points to a location where there is a ramp, but it appears to be misaligned in the data. Quadrant b in figure 13 is a location with a ramp that was labeled by all models as non-ramp. This could be because it looks different than a typical ramp with the parallel dips running down the slope. Quadrant d is the most interesting for considering outliers. These locations were labeled falsely by all models as non-ramp. However, the area to the left is similar to quadrant a.2, in that it has all the characteristics of a ramp, but no transition to the street – it is just a path to the curb. Like a.2, this location should not be classified as a ramp because it is functionally not adequate. The hexagon area to the right in quadrant d is functionally a curb ramp but transitions very gradually, and ends in a kind of gravel surface.

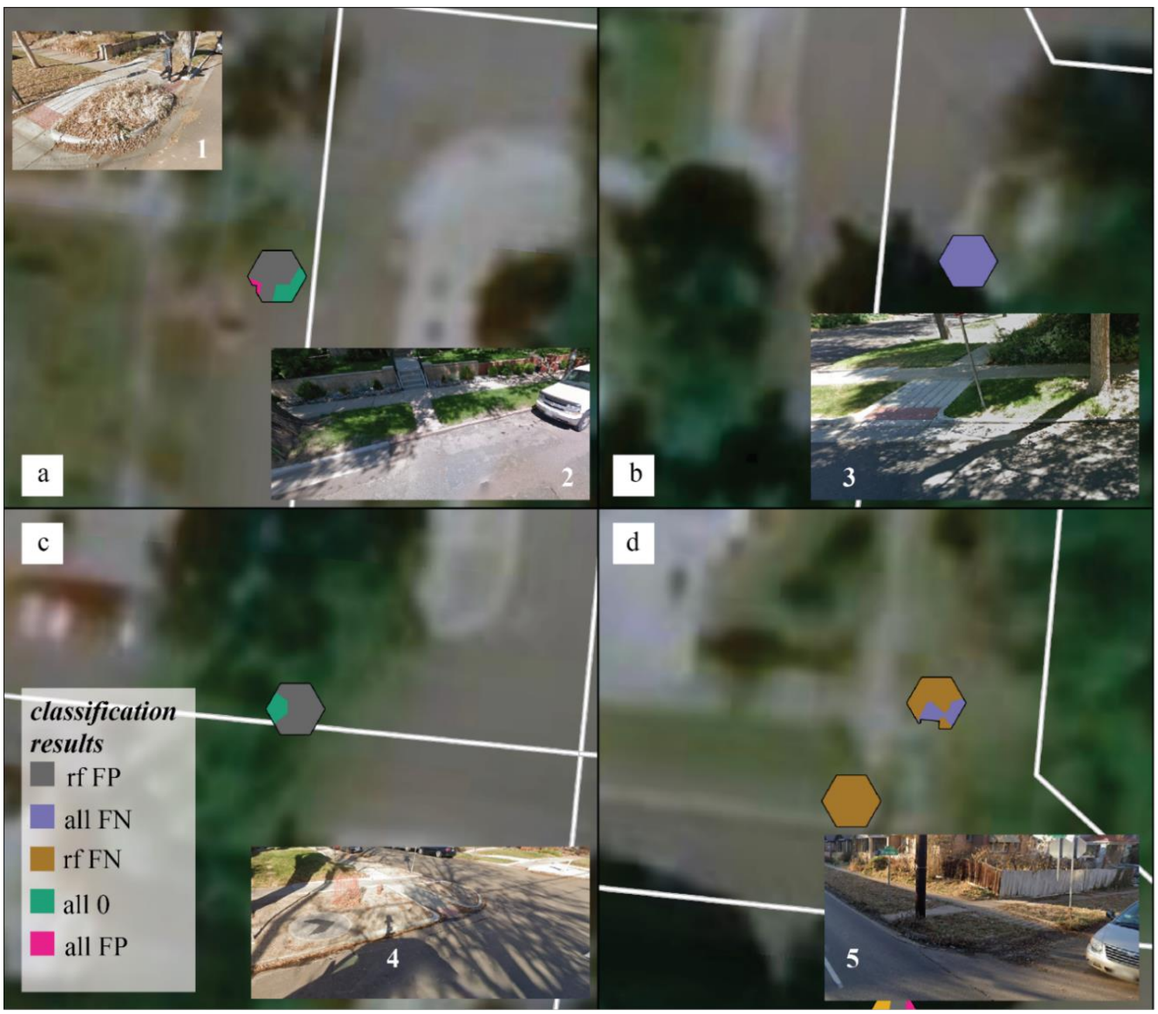

Figure 13. Denver Ground Truth Each of the four quadrants (a, b, c, d) shows the hexagon data area with colors representing the kinds of point prediction that occurred at that location (I aggregated similar predictions back to a polygon from a point). The base of the map is an aerial satellite image. The images labeled 1-5 are from Google street view at that location.

In Arlington, the behavior of the Point CNN model is the most surprising, and appears to have influenced the above average rate of agreement on non-ramp classifications and above average all false negative predictions. Figure 14 shows 5 locations where points were labeled falsely as non-ramps (green hexagons). Locations a, b, and e are all pedestrian curb ramps with red warning material. These ramps feed into a crosswalk and appear to have gradual elevation changes from sidewalk into the street. There is a very small portion of location e that was falsely labeled as non-ramp by the random forest (gray). In this case, the random forest was likely correct in the classification (recall that representing a ramp as a hexagon is not a perfect choice, but one made under the data and algorithm constraints). Locations d and c are slightly more difficult. Location d is a connector between streets and pedestrian pathways; however, it does not have an elevation change. The areas in purple mean that both models labeled the area as non-ramp. That the random forest labeled any points in that location as non-ramp is surprising; perhaps it picked up the warning material color or proximity to the street intersection. Location c appears to be misaligned; however, like location d, the points that the random forest model labeled as ramp (green) do not make a lot of sense visually unless the model was picking up the brightness of the crosswalk. These results are interesting because they suggest a disconnect between what a pedestrian curb ramp is by definition and function. Further, ramps a, b, e, and c all have similar design characteristics (red warning material, gradual slope change) so it is unclear why the random forest was correct about locations a, b, and e but mostly incorrect on c. This location is an outlier in some way that is not visually apparent.

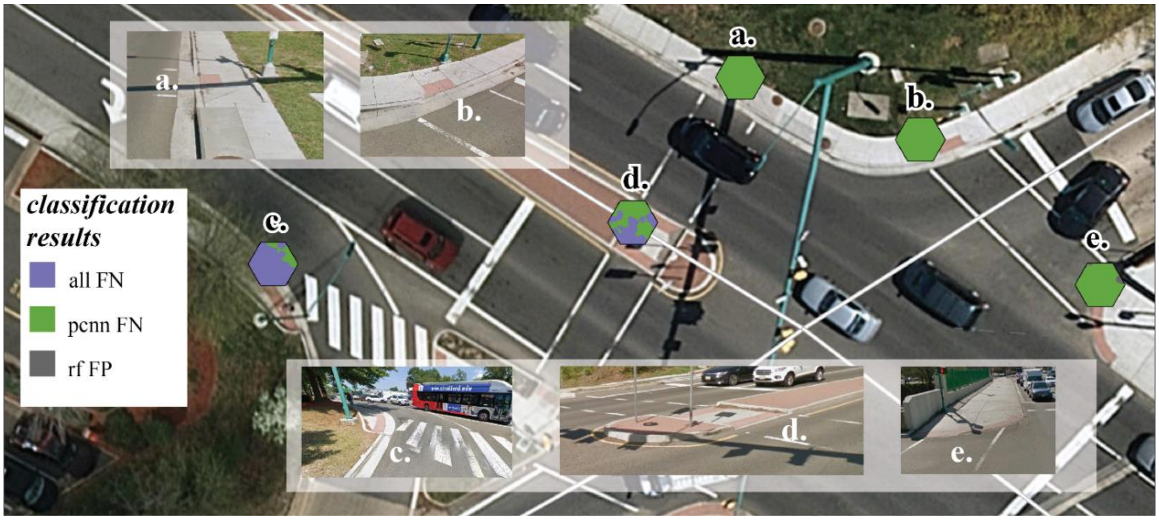

Figure 14. Arlington Ground Truth This is one intersection with 5 ramp locations (hexagons) and Google street view imagery showing a more the view on the ground. Letters a-e connect the imagery to the hexagon locations.

In contrast to Arlington, in Seattle, the rate of false positive classifications from the Point CNN model alone (pcnnFP) is much higher than average (49% of points). In figure 15, the orange areas (*a, c,* and *d*) mostly had agreement between models in correctly identifying ramp locations. Those ramps are slightly weathered, but have distinct transitions from flares to landing. Location *e* was falsely labeled as a ramp by the point CNN and location b was falsely labeled by both models (pink). These locations are not ramps, but narrow driveways that gradually change in elevation towards the street. Narrow driveways are a very common feature in Seattle's design. These locations could functionally be used by people with disabilities as curb ramps but might be less safe because of the likelihood of cars entering and exiting. Recall, machine learning algorithms look for features that are most similar to each other. In this case, narrow driveways fit the "normal" characteristics of a curb ramp and are functionally like pedestrian curb ramps. At the same time, they could be more dangerous because use will be shared with automobiles.

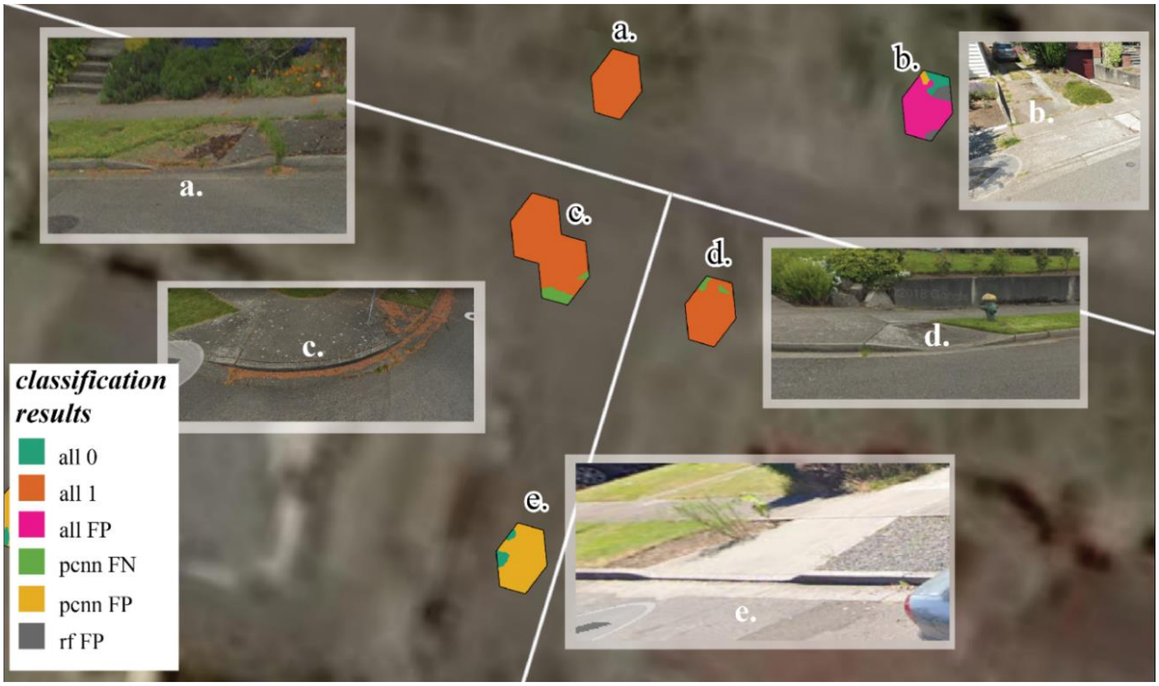

Figure 15. Seattle Ground Truth This visual is set up like figure 14.

These results reveal the diversity of model error. This error is not bias in that it is not systematic. Machine learning is an excellent tool for classification and prediction. It can achieve accuracy and process large data far better than any human. This tool also makes systematic errors which are likely to disproportionately affect categories of people and there is a rich body of work on addressing bias (see for example Crawford 2017; Friedman and Nissenbaum 1996; Eubanks 2018). But another, equally pernicious kind of error cannot just be calculated out or benchmarked. This kind of error, the kind of error that led to the death of Elaine Herzberg and the stalled mobility of Haben Girma and Emily Ackerman, cannot be solved with computation nor code, but requires the slow and messy human work of observation, description, and contextualization to unravel.

Conclusion

This case focused on classifying curb ramps, an infrastructural feature that is very important to the safe travel of people with disabilities, and which has improved the lives of many others (e.g. people walking with children in strollers). But this case also reveals something very important about the relationships between AI and the rights of people with disabilities. Notions of normality have long been tied to power and dominance at the exclusion of people with disabilities. Machine learning is simply faster and more efficient at categorizing and identifying normality. This means that anyone who falls outside of the norm – who is an outlier in statistical terms based on how and what environmental features they require for access, is likely to experience unforeseen harms. As scholar Jutta Treviranus recounted of her tinkering with autonomous vehicle models, "When I presented a capture of my friend to the learning models, they all chose to run her over…I was told that the learning models were immature models that were not yet smart enough to recognize people in wheelchairs…When I came back to test out the smarter models they ran her over with greater confidence" (Whittaker et al. 2019, 12). Behind the scenes, the model was likely trained to recognize wheelchair riders as pedestrians, but even then, it would learn a "normal" or average profile of a wheelchair rider. What then of the safety of a wheelchair rider with glow sticks on their wheels and a cat in their lap? These kinds of outliers and exceptions cannot be coded in.

There is a glut of work and reporting revealing bias in AI and proposing solutions, but bias itself is socially constructed, contextual, and varying. Who gets to decide what is fair and who benefits? Even if we were to agree on what bias is, under what conditions does it exist? And as these results highlight, what happens when the bias is bias towards a heterogenous group such as people with disabilities? Importantly, these outliers represent lives at stake, not simply observations in a dataset. I propose that, to address this kind of error or bias, we need to look to qualitative methodologies. Consider replacing social action with machine learning or AI: “Thick description involves accurately describing and interpreting social actions within the appropriate context in which the social action took place” (Ponterotto 2006, 542). To rephrase, what AI needs is descriptions and interpretations in context. In this paper, I have made a first attempt at this kind of description through ground truthing model error. There is much more work to be done in this area.

Academics are abuzz with discussions about the impacts of the data-driven paradigm and AI on their discipline and on science itself. This is important work, but we also need to consider how academic disciplines might shape AI. For example, defining the field of geoAI, has been a core concern of geographers in recent years. Definitions have mostly focused on the application of AI to geographic problems (Hu et al. 2019; Janowicz et al. 2020; W. Li 2020; VoPham et al. 2018). But what is desperately needed is to bring a distinctly geographic perspective to the interpretation and use of AI. Geographers are, at the most basic level, scientists of context, and contextualization is a key weakness of AI. Future work should analyze AI, bias, and error in context, in order to better understand the implications of these technologies on society and to imagine more just uses.

In the context of these results, I return to the questions - what kind of world do we want to live in? And who has the right to the city? Failing to think critically about these questions means that dominant narratives or imaginaries win the day. As Jasanoff and Kim write, "[w]hat we ‘see’ in familiar surroundings looks right, epistemically as well as normatively. So the socially conditioned eye can take for granted that all-male orchestras or all-black passengers on the backseats of buses, or even scenes of filth and abject poverty simply represent the rightful order of things" (2015, 14). As Turnbull writes, “[i]n the long run, social and cultural complexity cannot be winnowed away; it’s all there is” (2003, 227). Valuing outliers is not the domain of AI, but I argue that we can still mobilize the power of machine learning if we also work to value exception and outliers through qualitative approaches. This is a different kind of imaginary about what technology can do and what we would like it to do. This work is hard and slow. It requires open eyes, reflexivity, and wading through messes of big data, but to fail to try to contextualize and understand AI is to fail to take responsibility for shaping the world we live in.

CHAPTER IV

REPLICABILITY, RELEVANCE, AND STUDY DESIGN

The phrase "let the data speak for themselves" has been bandied about with giddy excitement in recent years as big data, powerful algorithms for processing big data, and computational power for both are increasingly available on regular desktop computers (Anderson 2008; Hey 2009; Prensky 2009). In response to this, numerous scholars have cautioned that increasing data size and analysis complexity does not mean that data can suddenly speak for themselves (see for example, (Barocas and Selbst 2016; Boyd and Crawford 2012; D'Ignazio and Klein 2020; Kwan 2016). I take a different approach. Considering AI as a learning collaborator in the research process – I ask, if data and AI want to speak, what do they have to say and what are the implications for scientific replicability and social relevance?

I answer these larger questions by looking at the impact of different data inputs and the subsequent machine learning classifications of curb ramp locations in Seattle, WA. On the way to thinking about replicability and relevance, I consider key aspects of the scientific method and the impact of new data and algorithms on these processes. Specifically, I look at the implications for scientific knowledge of abandoning careful and reflexive conceptualization, operationalization, and measurement. I argue that because of complexity, big data and AI do in fact "speak" and careful consideration of what they have to say is important.

Big Data and Study Design

Conceptualization, operationalization, and measurement are vital aspects of research design. Traditionally, researchers first articulate shared meanings of abstract concepts (conceptualization), look for variables in the world to capture those definitions (operationalization), and then attach data to the variables (measurement) (see figure 16). For example, I could start by defining disability as the result of barriers in the environment (conceptualization). These barriers could be operationalized in various ways such as inaccessible physical environments or negative attitudes towards people with disabilities. I then might measure the accessibility of physical environments by mapping curb ramp locations and route connectivity. As the example illustrates, this is a narrowing process.

In spatial data science, cartographic representation is intertwined with measurement (Dodge, Kitchin, and Perkins 2011; Fairbairn et al. 2001; Robinson et al. 2017). Geographic information is always uncertain, vague, and imperfect. That is, “observations are imperfect in the sense that they can never fully or correctly reflect all aspects of reality” (Duckham et al. 2001). This is not dissimilar to the uncertainties inherent in quantitative data writ large. While geographers are aware of the uncertainty and error in geographic information, the data-driven paradigm in geography[21] and its dependence on accuracy and precision metrics has mostly abandoned these concerns to

[21] I am using data-driven geography to refer to several other trends in geography that have come about due to big data and machine learning. For example, algorithmic geography (Kwan 2016) and geoAI (Hu et al. 2019; Janowicz et al. 2020; W. Li 2020; VoPham et al. 2018).

focus on the potentialities of new knowledge using big data and complex machine learning algorithms.

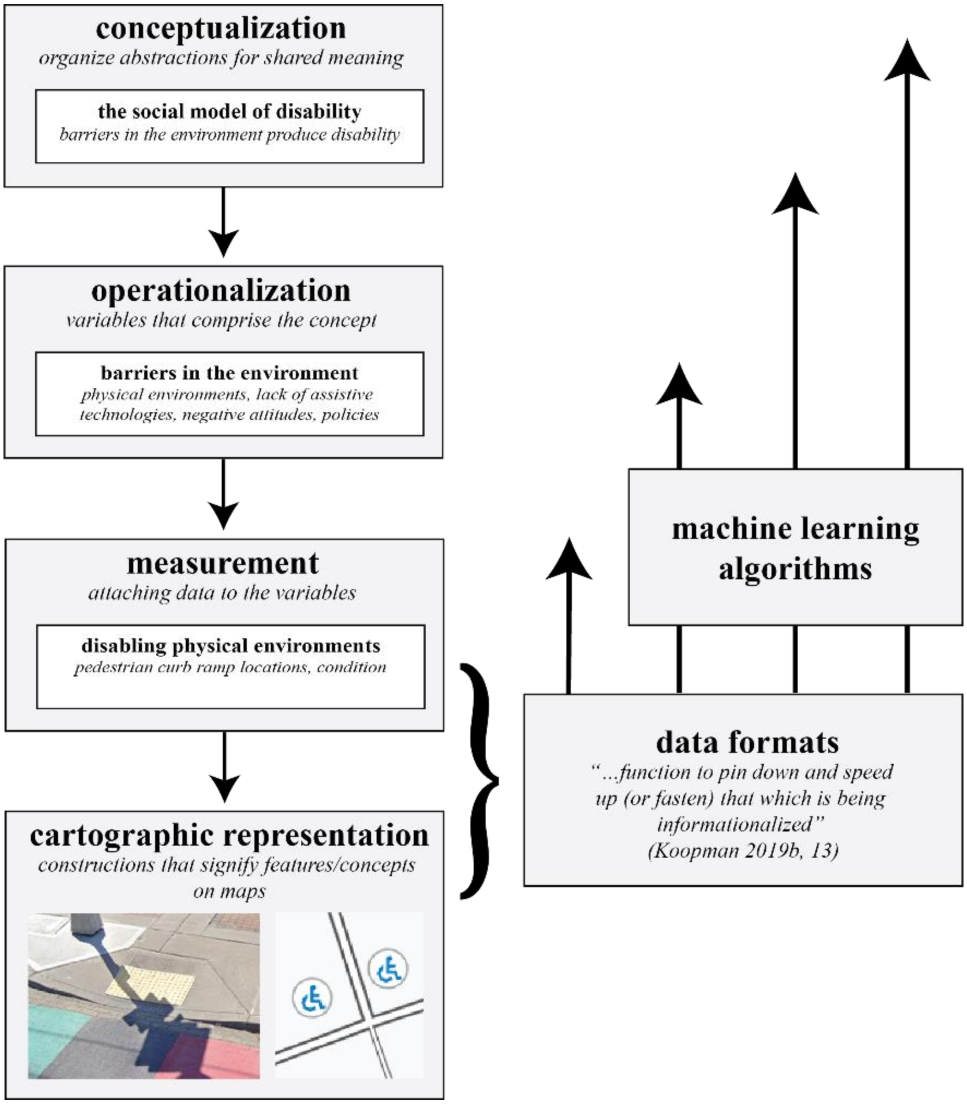

Figure 16. Study design and the data-driven paradigm

The data-driven paradigm, suggests that big data and complex algorithms can provide meaning without all the slow, messy, difficult, and reflexive work of conceptualization, operationalization, and measurement – "the data can speak for themselves." Data (measurement and representation) come first in this paradigm (see figure 16). But data formats are the product of conceptualization, operationalization, and measurement *somewhere* we just don't necessarily know where or when. For example, bureaucrats, not social engineers, have largely mainstreamed American gender as a binary choice through the birth certificate and thus shaped identity in unintentional ways (Koopman 2019b; 2019a). Or drawing from the example above (figure 16), if I start with my data on curb ramps – which is already imperfectly represented as a point – and work my way up to then defining the social model of disability[22] as the presence or absence of curb ramps, much has been lost in translation intellectually. While, putting it this way makes it seem absurd, consider for a moment how the robust meaning of gender has been mostly distilled to the binary male/female categories that we check on census forms or driver's license applications. Data-determined study design is a big enough problem but one that is exacerbated by the mediation of *learning* algorithms.

In order to learn, machine learning algorithms need training and validation data. Training data are data with labeled features and attributes for grouping similar features together. The AI only works if it is given information to learn from. Validation data are data that the AI does not see information about but can check its predictions against as it learns. With enough training and validation data, carefully specified parameters, and sufficient computing power, machine learning algorithms are highly efficient at accurate classifications and predictions.

[22] The social model of disability situates disability as the product of disabling environments and attitudes. These barriers, which include physical environments, lack of assistive technologies, negative attitudes, and public policies transform impairments into disabilities. This model contrasts with the medical model of disability which sees disability as purely a physical condition intrinsic to the individual.

The key weakness of machine learning algorithms is that they learn to maximize accuracy (as ratio of correct predictions to sample size) and thus do not function well when prediction classes are imbalanced. For example, curb ramps make up 1% or less of the ground area in Seattle. So, a machine learning model could achieve 99% accuracy by not classifying anything as a ramp. This problem is more widespread than that because most of the things that are interesting to predict or classify are rare (e.g. identity theft, crime, natural disasters, disease, terrorist attacks).

Various methods have been proposed for improving and creating balance within imperfect training data (Tajbakhsh et al. 2020; Haixiang et al. 2017; Platanios et al. 2020; More and Rana 2017). The focus of that literature is to streamline big data processing and improve machine learning accuracy metrics. The problem with these techniques is that they involve cleaning or clustering data in ways that require pre-existing knowledge of the feature of interest. These strategies cannot be easily replicated in new cases where nothing is known about the feature.

Replicability and Relevance

Barriers to scientific replication and reproducibility arise from data and researcher uncertainties that propagate into conceptualization, measurement, analysis, and communication (Kedron et al. 2019). Reproducibility – using the same data and methods to get the same results – is important to science and politically challenging but the solutions are technically simple. Research reproducibility could be achieved if researchers were incentivized to and provided the infrastructure for sharing all study data and carefully documenting their analysis and code through an open-source platform (Kedron et al. 2019; S. Li et al. 2016).

Replicability is the ability to use similar data and methods to get similar results in a similar case. Data-driven science, which suggests starting from the data and algorithms and exploring what might be found, poses unique challenges to replicability. One challenge is that studies using big data and algorithmic mediations are largely black-boxed – the processes are too complex, the data are too big, and the mechanisms for learning are unknown (Pasquale 2015; Latour 1988). Relatedly, without careful study design it is impossible to really know what a similar case might be (Lund 2014). In the example given above of curb ramp locations, is it really a case of analyzing accessibility or is it a case of transportation modeling?

Asserting what a study is a case of is reflexive work that is at odds with the reverence for objectivity that is inherent in most data driven work. This reflexive work is closely related to relevance which I posit is inherent to replicability. Social relevance, from a scientific perspective is "an influence or benefit (realized or expected) from the results of research activity to the research community or to society at large" (Scaratti et al. 2017). Considering social relevance contributes to understanding what a similar case might be and why the work is being done. These are the kinds of questions that data and AI cannot and should not answer for us.

In the following pages, I examine the impact of unbalanced and imperfect data on the results of machine learning classifications through one case - curb ramp locations in the city of Seattle. This case contributes to understanding social relevance and scientific replicability as challenges tied to conceptualization in machine learning and data-driven science.

Methods

Data & Cleaning

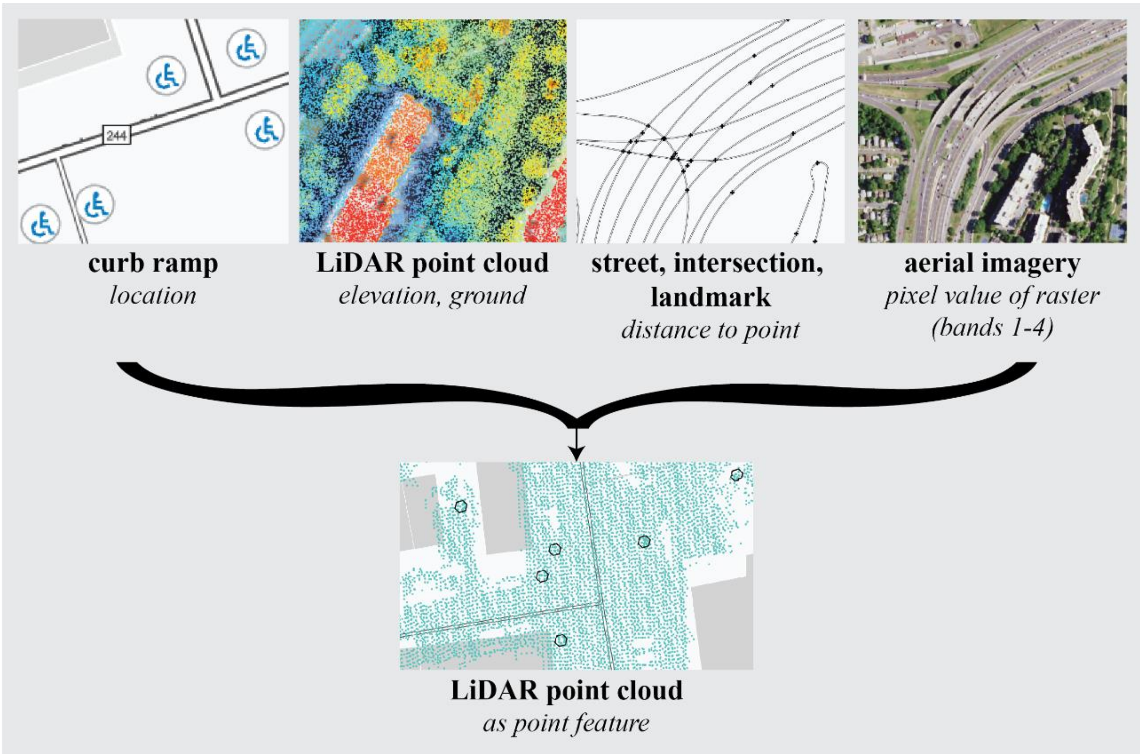

Figure 17. Data sources

In service of replicability, this analysis draws only on open data – data that are freely available for anyone to access. Information about the location of curb ramps came from the city of Seattle's open data portal and was current at the time of download (March 2020) (City of Seattle 2020)[23]. Aerial imagery came from the USGS national map download center (USGS 2020). The aerial image was taken on August 7, 2015. The LiDAR point cloud data came from the Department of Natural Resources in Washington State (Washington State Department of Natural Resources (DNR) 2021) and were

[23] About 18 other cities collect and maintain open data on curb ramps (see Chapter 2).

captured in 2016. The street and landmark layers are TIGER/line shapefiles (US Census Bureau 2020). I derived street intersections from the street layer. Aerial imagery, LiDAR data, and the TIGER/line shapefiles have near national coverage in the United States.

I classified ground points in the LiDAR point cloud and those locations became the sample features. This was done to reduce the data size and reduce the noise in training. Non-ground points such as buildings and trees would also appear in the aerial imagery and they would add unnecessary information to the models – all curb ramps are on the ground. I then joined attributes of the other data to those points (see figure 17). These included the closest distance to a street, intersection, or landmark and the pixel values of the raster for each of the 4 bands at the point location. Labeling points in the sample as curb ramp locations was slightly more complicated due to the representation of curb ramps in Seattle's open data.

The curb ramp feature was represented as a point in the open data. A curb ramp is not a discrete point in real life (see figure 18). The ADA requires ramps to be at least 3' wide and the flares, landing, and potential warning material are all parts of a curb ramp that take up more space. The area that curb ramps cover is varying and non-uniform. Therefore, I created a tessellation of 50ft^2 hexagons across the city and labeled those with a ramp point within them as ramp locations. As I will describe below, this was done to retain as much of the ramp locations in the data as possible and for reasons related to sampling and the needs of the machine learning algorithm used.

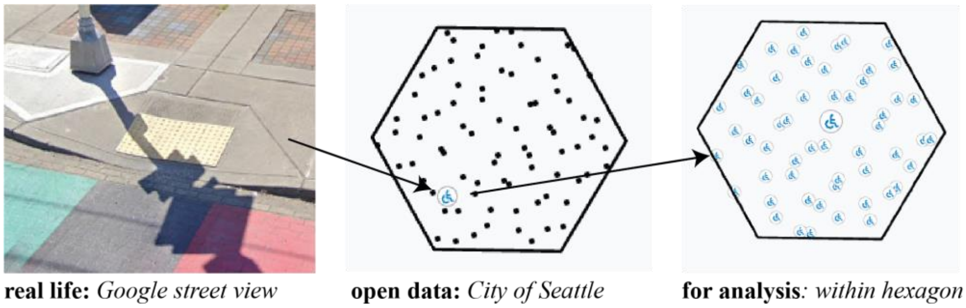

Figure 18. Curb ramp representation based on available data and algorithm data requirements

Learning algorithms require training and validation data. I created two large hexagon areas – 50,000 square feet for training and 10,000 square feet for validation (20% of training area) (see figure 19). No ramp locations in the training or validation data were more than 60 feet from a street. I clipped the data further to only points within 60 feet of a street, reducing the training data size from 21,083,268 to 12,825,712 points.

Figure 19. Train and test samples

Sampling

Machine learning algorithms learn from their success. If a feature class is rare in a dataset, a machine learning algorithm can achieve high accuracy by ignoring the rare class. Balanced samples are needed for machine learning algorithms to work. In this context, balance means equal representation in the data of feature classes (e.g. ramp, non-ramp). To create balanced samples, I used the hexagon tessellation that I created for labeling ramp points. I used the hexagon areas to balance the data for two reasons: First, I did not want to lose any ramp points since they were already rare in the data (only 1% of the ground area). Second, features that are near in space are more similar to those that are further away (Tobler 1970), so clusters of ramp points with scattered non-ramp points would introduce a different kind of data imbalance. I randomly deleted non-ramp areas by generating random numbers until there were equal proportions of ramp and non-ramp areas[24]. There are many advanced algorithmic techniques for balancing samples that likely would have achieved better classification accuracy, but the goal of this project was to look at the relationship between data and algorithms in a few controlled contexts without introducing further complexity.

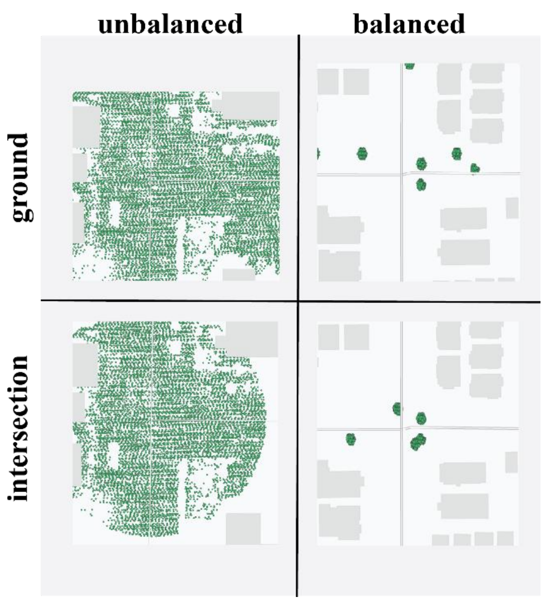

Figure 20. Training and validation samples

[24] The balanced samples had slightly more ramp points than non-ramp points because of differences in point density within the hexagons (see table 16 and table 17).

I created four types of samples across the training and validation datasets (see figure 20, tables 16 and 17). These included balanced and unbalanced points across the entire ground surface within 60 feet of a street and a subset of those that were within 20 meters of a street intersection. All but 2.7% of the ramp locations were within 20 meters of an intersection.

Table 16. Training samples

sample	total points	ramp points	% ramp
ground (unbalanced)	12,825,712	130,286	01.02%
randomly balanced	235,908	130,143	55.17%
intersection (unbalanced)	3,590,404	126,752	03.53%
randomly balanced	228,941	126,609	55.30%

Table 17. Validation samples

sample	total points	ramp points	% ramp
ground (unbalanced)	2,979,367	19,711	00.66%
balanced	35,694	19,669	55.10%
intersection (unbalanced)	754,753	19,039	02.52%
randomly balanced	35,259	18,987	53.85%

Random Forest Classification

The random forest machine learning algorithm was used to classify curb ramp points and test the impact of various sampling strategies. The random forest regression method simply involves running multiple models or decision trees through a set of training and prediction data and allowing them to vote on the answer for each location (Breiman 2001; Tin Kam Ho 1995). Each model or decision tree creates a prediction for each potential location and the majority prediction or vote becomes the final result. The variable to classify was a binary ramp value. The potential predictor variables were: distance to street, distance to street intersection, distance to landmarks, elevation, and raster image value for each of the 4 bands in the aerial imagery (see figure 17).

I evaluated the results based on overall accuracy, precision, the number of false positive classifications, and the number of false negative classifications. Accuracy is simply the proportion of correct classifications over all possible classification locations. Precision is used to understand the proportion of positive identifications that were correctly classified.

Iteration I

10 different models were run over 2 iterations. The first iteration included 3 combinations of the ground unbalanced and balanced training and validation data and parallel combinations for the street intersection data (see figure 21).

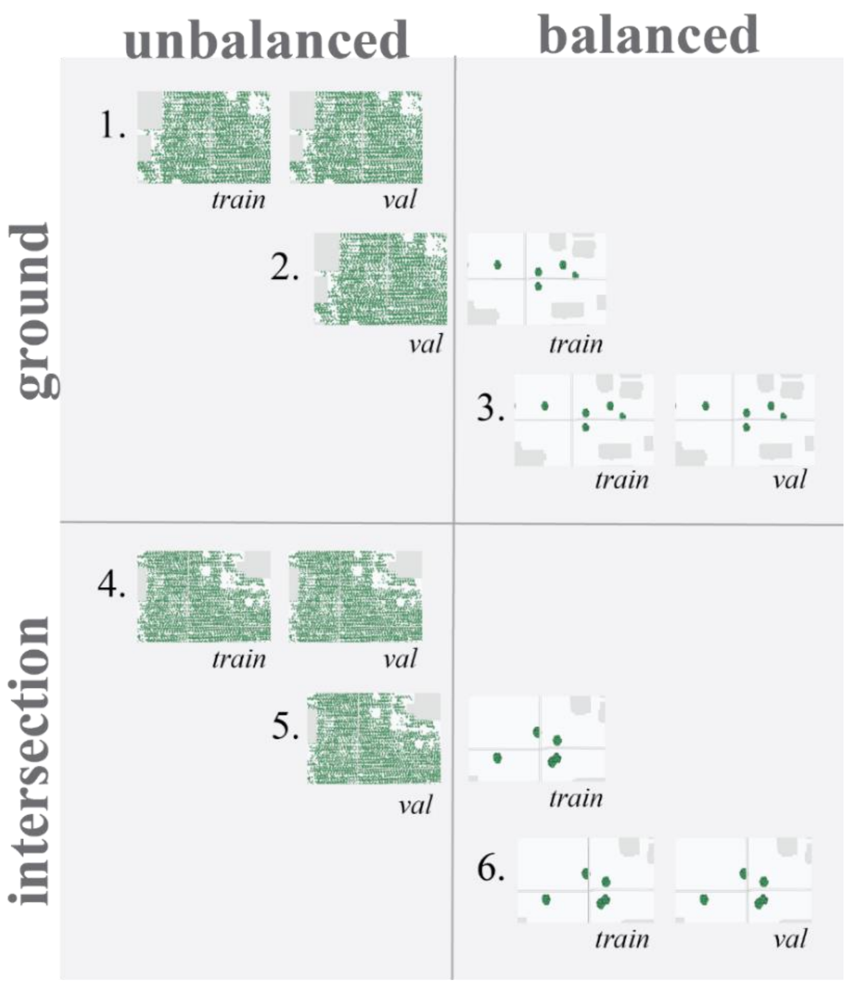

Figure 21. Iteration I Models

I did not train on the unbalanced data and validate on the balanced data because that is not a replicable test in a case where nothing is known about the area to classify[25]. That is, data balancing requires knowledge of the feature distributions in the validation set and pre-existing knowledge removes the need to test classification unless the goal is to test it for model development. In the same way, models 3 and 6 are not actually replicable but were run to compare the results with those of unbalanced sets.

Iteration II

Using the results of the first 6 models, I ran 4 more combinations. These combinations were validated on a sample that had been reduced by the results of iteration I. Specifically, I deleted all points that were not classified as ramp by model 2 and 5 (the models with low numbers of false negative classifications). Considering the case of curb ramp locations, a false positive classification (ramp classification when no ramp is present) is worse than a false negative classification. That is because it would be more dangerous for someone who needs curb ramps for safe travel to expect to find one and not than for them to be pleasantly surprised by a curb ramp. However, in terms of data loss, where the goal is to retain ramp locations and reduce non-ramp locations, false negative classifications are worse.

Before deleting the points that were falsely classified as non-ramp by model 2 or 5, I ground truthed the locations using Google street view imagery. As I report in the results, only three of these locations were actually curb ramps. Models 7-10 were

[25] It is also not useful information. A model trained on unbalanced data and validated on balanced data will likely perform even worse than a model trained and validated on unbalanced data.

validated on the points classified as curb ramp by model 2 *or* 5 and trained on the 4 original training samples (see table 18 and figure 21) [26].

Table 18. Training and Validation Samples

Iteration	#	Training	Validation
1	1	No change (OG)	OG
	2	Randomly balanced (RB)	OG
	3	RB	RB
	4	Points within 20 feet of intersections (INT20)	INT20
	5	INT20 randomly balanced (INT20r)	INT20
	6	INT20r	INT20r
2	7	RB	Ramp classification on #2 or #5 (2&5)
	8	OG	2&5
	9	INT20	2&5
	10	INT20r	2&5

Results

The validation sample had almost 3 million points or was 23% the size of the training sample (see table 16). 19,711 were ramp points in the unchanged dataset. Only 3.4% of the ramp points (672) were not within 20 meters of a street intersection. The randomly balanced sample is made up of about 55% ramp points and the balanced intersection sample is made up of 54% ramp points. The percentage of ramp points can serve as a basic benchmark for the classification models. Since the original sample had only 0.66% ramp points, for example, a classification model validating on the unbalanced original data (models 1 and 2, see table 18) could result in 99.34% accuracy while classifying every point as non-ramp.

[26] I ran a third iteration using the ramp classifications from model 7 and 10. The results were not useful. The models trained on balanced data predicted almost every point was a ramp and the model trained on unbalanced data predicted almost every point as a non-ramp.

Relationship between ramp and predictor variables

Across the training and validation samples there were 149,997 ramp points. The average distance from ramp locations to a landmark was 591.67 meters (see table 19). The average distance to a street was 4.85 meters and the average distance to a street intersection was 10.38 meters. The average elevation was 300.51 meters. The average pixel values of the aerial imagery were 109.21, 116.50, 112.71, 121.49 from band one to four.

Table 19. Predictor variables at ramp locations

	minimum	maximum	mean (standard deviation)
distance to (m)			
landmark	24.51	1254.74	591.67 (268.86)
street	0.00	14.14	4.85 (2.24)
intersection	0.04	114.36	10.38 (7.38)
elevation	79.88	429.90	300.51 (62.22)
aerial image			
band 1	22.36	233.94	109.21 (45.66)
band 2	45.00	234.40	116.50 (36.78)
band 3	60.00	230.76	112.71 (28.37)
band 4	38.00	218.47	121.49 (39.85)

Between the training and validation data there were 15,805,079 point features, and all Pearson correlations between the predictor variables used in the random forest were significant ($p<0.001$, see table 20). Distance to intersection had the largest correlation with the binary ramp value (-0.114). Distance to landmark was most correlated with elevation (0.368). Distance to street was most correlated with distance to intersection (0.184). Lastly, the first three raster bands were highly correlated with each other.

Table 20. Pearson Correlations

	ramp	Dist2lm	Dist2st	dist2int	DEM	b1_AER	b2_AER	b3_AER	b4_AER
ramp									
Dist2lm	-0.001	1							
Dist2st	-0.030	-0.003	1						
dist2int	-0.114	.057	.184	1					
DEM	0.017	.368	.007	-.073	1				
b1_AER	0.027	-.057	-.091	-.022	-.159	1			
b2_AER	0.027	-.065	-.091	-.022	-.163	.982	1		
b3_AER	0.031	-.075	-.124	-.037	-.160	.950	.954	1	
b4_AER	-0.008	-.015	.090	.055	-.059	.491	.551	.354	1

Random Forest Classification

Iteration I

In the first iteration or set of 6 different training/validation data combinations, model 1 performed best based on classification accuracy, followed by 4, 3, 2, 6, and lastly 5 (see table 21). Based on precision in ramp classification, model 6 performed best. While 97% accuracy for model 4 seems good, ramps only make up about 2.5% of the data (see table 16) and the random forest algorithm classified every point as a non-ramp (hence, 0 false positives). The worst performing model in terms of accuracy had the best precision on ramps (model 6).

Table 21. Random forest classification results (iteration I)

#	accuracy (%)	precision	fp	fn
1	99.29%	0.08	1,450	19,589
2	86.44%	0.04	402,817	1,155
3	92.09%	0.08	1,667	1,155
4	97.48%	0.00	0	19,039
5	55.37%	0.05	334,825	2,001
6	71.09%	0.69	7267	2,927

1.1.1.1 Ground truth false negatives (model #2 & #5)

There were 1,021 points that I originally labeled as ramps (using Seattle's open data and the 50ft^2 hexagon tessellation) that were not labeled as ramp by model 2 or 5 (the two models validated on unbalanced data with the lowest number of false negative classifications). These 1,021 points were spatially clustered and aggregated to 22 polygon areas with a total area of almost 97 square feet.

Of these locations, 10 were adjacent to points labeled as ramps meaning that the model captured a ramp at that location just in a slightly different location. Seven were not actually curb ramp locations (see a-g in figure 22). Among these seven were two connecting paths (*a, b*), a staircase (*c*), 3 driveways (*e*-*g*), and one that was just a sidewalk (*d*). Five locations were true false negatives (*h*-*l*). In two of those cases, the open data point was not aligned with the ground truth location (*i*, *k*).

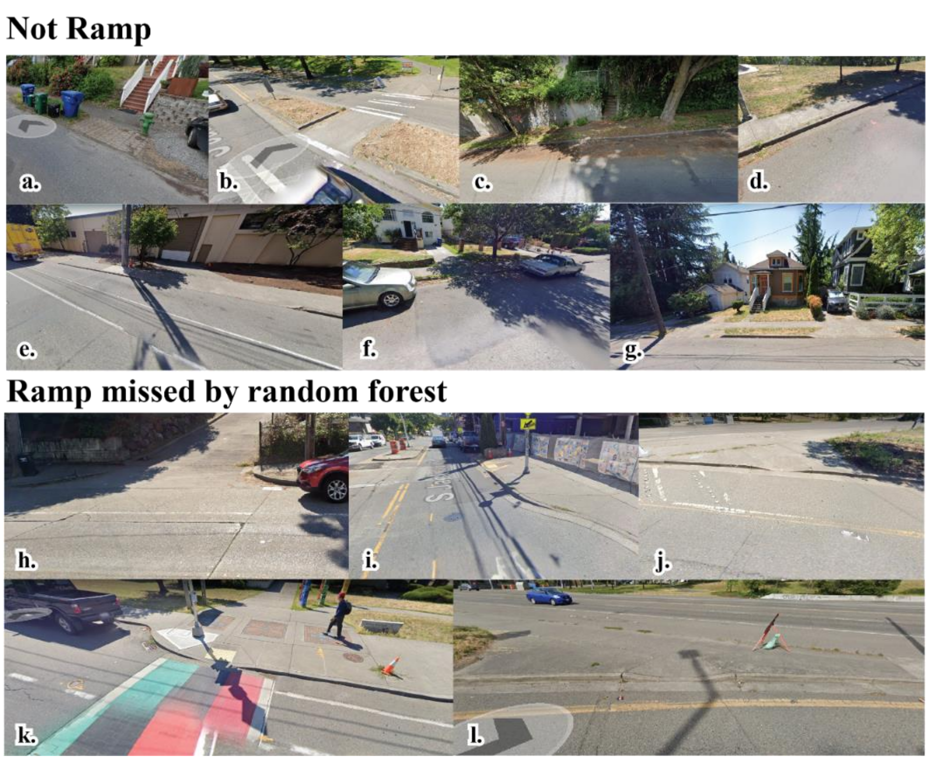

Figure 22. False negatives from model 2 & 5 This figure shows the Google street view image for all areas falsely labeled by model 2 or 5 as non-ramp.

These results suggest that only 3 ramp locations would be lost by deleting every point not classified as a ramp by either model 2 or 5 (*h, j*, and *l* in figure 22). Doing this reduced the data size and increase the proportion of points with ramp labels. This new validation set had 449,045 points (about half as many as the intersection validation data). 18,690 of these points or 4.2% were ramps.

Iteration II

Even model in iteration II was validated with the thinned sample from iteration I. The models were trained using the randomly balanced original data (7), original data (8), near intersection data (9), and the randomly balanced near intersection data (10). The two models trained on randomly balanced data (7 and 10), classified almost every point as a ramp location and had low overall accuracy. The two models trained on unbalanced data (8 and 9), classified almost every point as non-ramp and achieved fairly high accuracy.

Table 22. Random forest classification results (iteration II)

#	accuracy (%)	precision	fp	fn
7	14.14	0.05	385,399	152
8	95.83	0.12	60	18682
9	95.82	0.22	77	18668
10	39.72	0.06	268883	1816

Discussion

This research is *replicable*[27] if another researcher could use similar open data sources, follow the same steps, and achieve similar classification accuracy on curb ramp locations. The balancing strategy for validation data used in models 3 and 6 is replicable. However,

[27] Recall that replicability is using similar data and similar methods to get a similar result.

a relevant study would allow another researcher to use similar data, follow the same steps and classify curb ramp locations with similar accuracy in an area where little or nothing is already known about their locations. Across the 10 models run with different combinations of training and validation data, there is a tradeoff between replicability and relevance and model accuracy (see figure 23). That is, the random forest model, which prioritizes global accuracy, performs well with balanced data (models 3 and 6) but balancing data requires pre-existing knowledge of the feature of interest. So, model 3 and 6 were highly accurate in classification but irrelevant.

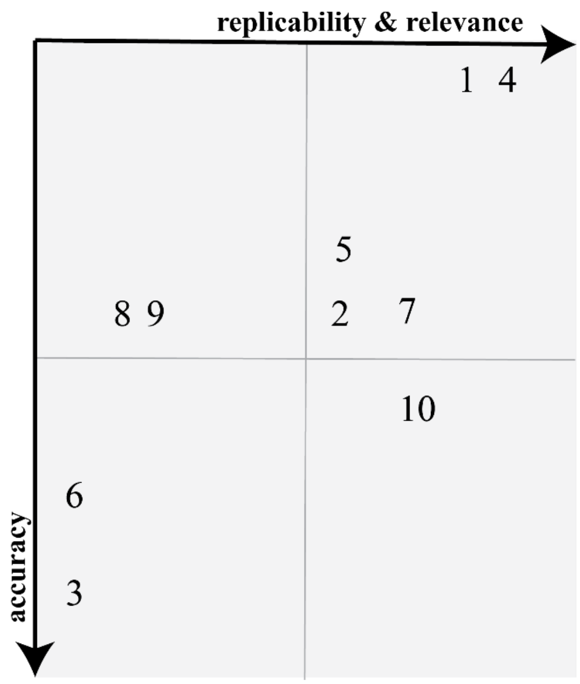

Figure 23. Models mapped by accuracy and replicability/relevance

This tradeoff between relevance and accuracy exists because curb ramps make up only a very low percentage of Seattle's total ground surface area. Models 1 and 4 would have been preferable from a replicability and relevance perspective because they require the least algorithmic mediation. A model validated on the ground surface could achieve 99.34% accuracy by classifying every point as non-ramp. This is approximately the accuracy of model 1 (99.29%) which nearly classified all points as non-ramp (see figure 24). Model 4 performed in a similar way. Models 2 and 5, which were trained on balanced data but validated on unbalanced data (a replicable strategy), achieved mediocre accuracy by labeling groupings of points (mostly surrounding intersections) as ramp locations. These models erred towards labeling many non-ramp locations as ramp locations (false positive) rather than incorrectly classifying true ramp locations (false negative). In fact, only three locations falsely labeled non-ramp by models 2 or 5 were actually ramp locations (see figure 22).

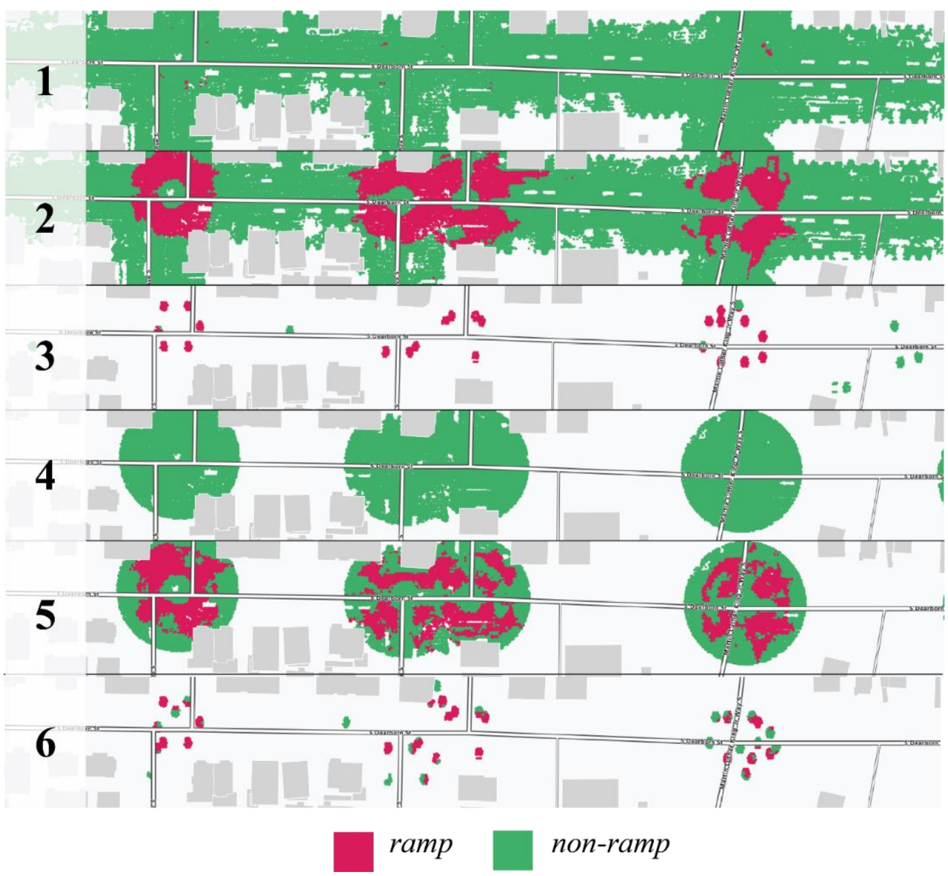

Figure 24. Prediction results (models 1-6)

In the second iteration all models were validated on the same dataset (all points classified as ramp by models 2 or 5). The patterns in false negative and false positive classifications based on sample balance increased. The models trained on balanced data (7 and 9), falsely classified most points as ramp locations. The models trained on unbalanced data falsely classified most of the study area as non-ramp (see figure 25).

Figure 25. Iteration 2 ramp classifications

Recall that a curb ramp is not a point in space but an area with varying characteristics. Some curb ramps have warning material at the bottom and some do not. The width, slope, size of landing, and size of flare all vary widely as do the materials that curb ramps are made from and the kinds of streets or paths that they connect. Curb ramps were represented in Seattle's open data portal as points because that is a relevant representation for holding information about municipal maintenance needs and routing in the city's accessmap.io application. A curb ramp is also not a hexagon, but I represented it as a hexagon as a result of a combination of aesthetic choices, constraints in the data, and needs of the algorithm[28]. The algorithm did not represent curb ramps in the same way that the city of Seattle nor I had but that does not necessarily mean that the representation is wrong.

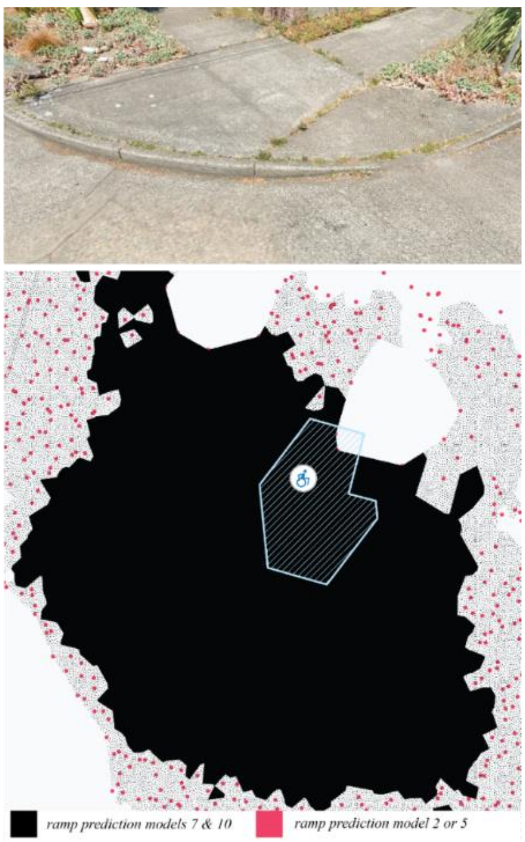

Figure 26. Google street view of location (top), model predictions overlaid by original ramp point and hexagon (bottom)

---

[28] See Demerritt (2001) and D'Ignazio and Klein (2020) for a thorough look at the micro decisions (including aesthetic ones) made in complex data-driven research projects.

See for example, figure 26 which shows the actual location from Google street view, overlaid by the model classifications, hexagon used for training, and original curb ramp point from Seattle's open data. The area classified as ramp by models 7 and 10 (in black) appears to cover the entire area where the sidewalks meet for crossing (see figure 27).

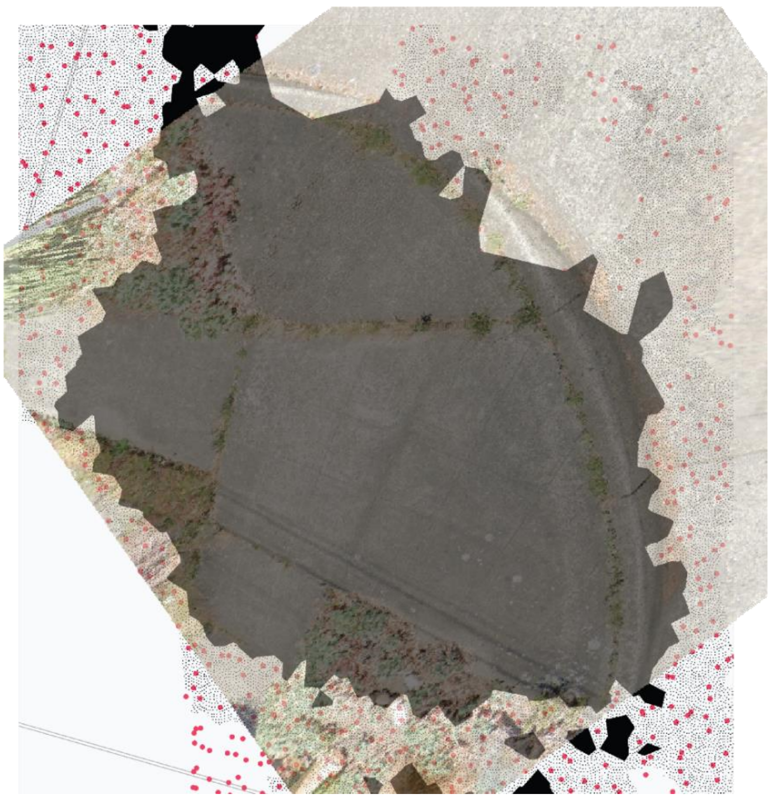

Figure 27. Curb ramp classification overlaid by Google street view image of curb ramp location

I observed this kind of result across many other intersections. For example, figure 28 shows an area where there should have been two curb ramps. The machine learning model came up with two areas that cover the entire angle of the ramp or corner of the intersection (locations *a* and *b* in figure 28). The machine learning models also classified all points across the street (location *c*) as ramp locations. Area *c* is particularly interesting because it is obviously not a pedestrian curb ramp but shares characteristics with pedestrian curb ramps that suggest something about what the models are learning. Location *c* has a slight change of elevation connecting a sidewalk to a street, is concrete, and is near a street intersection.

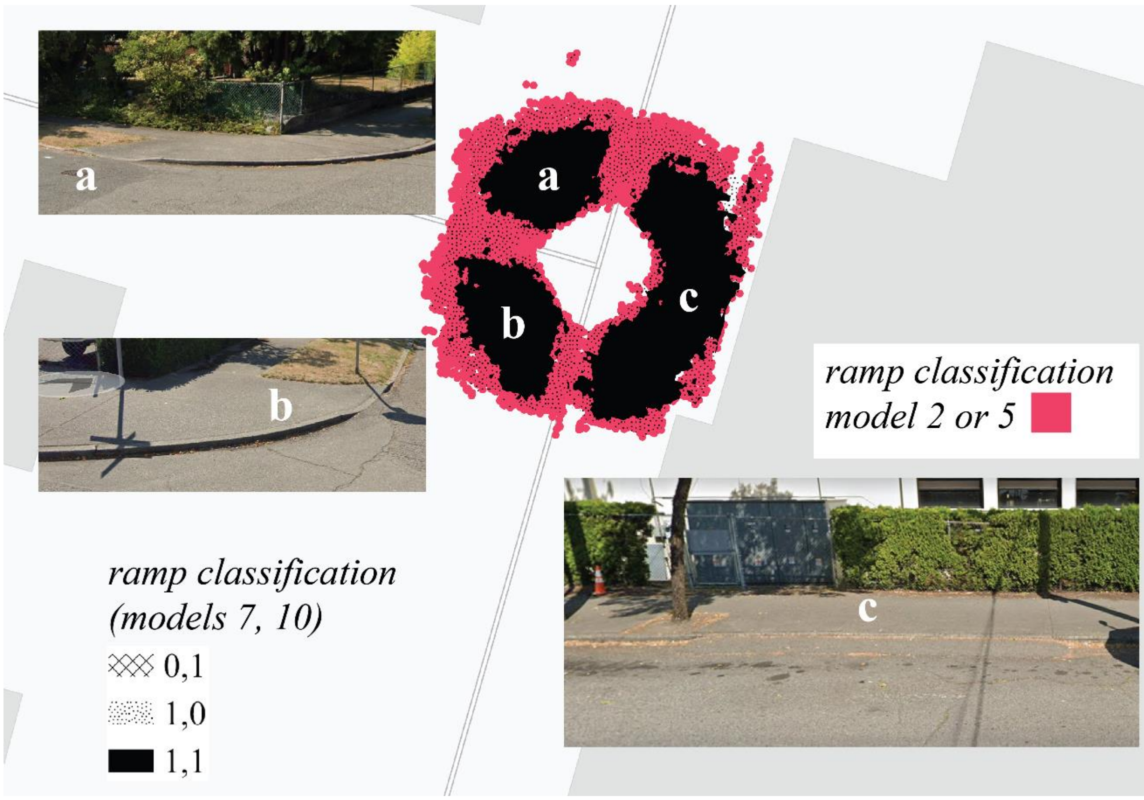

Figure 28. model ramp classifications and Google street view imagery of the location

Similar to location *c* in figure 28, figure 29 shows three locations where pedestrian curb ramps were not present, but the models classified points as ramp. On the right (locations *a* and *b*) are two sides of an alley labeled as curb ramp. Interestingly, a gap of non-ramp classification was left between *a* and *b* where the alley comes through. The area across the street (*c*) is visually similar but all points across the drive have been labeled as ramp locations.

From a data-driven perspective this result is not surprising. In machine learning, the model uses the information and parameters it is given to classify similar groups of observations together. In this case, the model had information about distance to streets, distance to intersections, distance to landmark sites, elevation, and the pixel values of raster imagery. The results show that the model was actually highly effective at finding similar features, but these features were not necessarily curb ramps (see Figures 9-12).

Figure 29. model ramp classifications and Google street view imagery of the location

Returning to the interaction between data, AI, and study design, available data (and the way they were cartographically represented) shaped these results at every step. Consider that in part this study begins with something that can be measured (see measurement, figure 30). Curb ramp locations (which are not point features) were formatted or represented as such in the open data. I then labeled training data using hexagon areas to capture more ramp points – this is also an inexact representation. The variables that were available to train the model to identify curb ramps (see operationalization, figure 30) were not exhaustive, but came to define the feature curb ramp through modeling. The model learned that a curb ramp is typically concrete, near an intersection or near a street, and has a slight elevation change. This is important information, but it is not what I meant by pedestrian curb ramp in the beginning. As Drummond writes, “a measure my capture something of importance but not everything of importance” (2006). Reflexive use of AI requires consideration of what is being measured or what the AI is actually “saying.”

Figure 30. Data-Driven Study Design

Conclusion

These results point to a larger issue in machine learning and automation. Algorithms have been set out in the world to do ambitious things such as identifying terrorists, tagging inappropriate content, and stopping crime before it happens. But terrorist, inappropriate, and crime are all abstract social concepts. AI cannot define concepts like terrorist or crime or inappropriate for us, it can only identify sets of similar characteristics or behaviors (which are limited by the predictor variables available). For example, Deborah Raji was working on a computer vision model for flagging inappropriate images and found that people of color were being flagged at a much higher rate (Strong et al. n.d.).

The model was being trained to learn to identify salacious content from porn imagery and safe content from stock photos. As it turns out, pornography is more diverse than stock imagery. This is an interesting finding for the AI to stumble upon but has little to do with weeding out inappropriate content.

Further, the limitations of using quantitative data to understand the world have gotten lost in the hype over big data and AI. As Harold Koh said, “when you can’t measure what is important, you make important that which you measure” (Mezey 2002). There is nothing inherently wrong with doing that if scientists reflexively consider the limitations and clearly articulate what it is that is being measured. And what it is that the study is a case of. For example, it is impossible to quantify water insecurity but we can count the number of households without access to a flushing toilet, hot and cold running water, or a shower (Deitz and Meehan 2019). In this case, it might not be possible to actually classify curb ramp locations, but it is possible to find features with similar characteristics. This is a relevant outcome for narrowing the locations for ground truthing but irrelevant if the goal is full automation.

At the heart of all of this is the slow and messy art of research design. AI and big data do not remove the need to consider these things. In fact, big data and AI only introduce new challenges to replicability and relevance. Data-driven discovery, when not conducted with care, reflexivity, and consideration will only make bad concepts louder and faster. To understand what is going on inside the black box of algorithmic mediation and big data we direly need to reconsider study design from the perspective of what data want to say. Concepts are always driven by social realities and contexts, even within the data-driven paradigm. Ignoring this only means that problematic conceptualizations will be magnified in ways unknown.

CHAPTER V

CONCLUSION

This research is among the first to center social equity in quantitative analysis and develop new methodologies while also exploring the limitations of data-driven research. Through the careful examination of data on environmental features that promote accessible pedestrian transportation in the United States, I point to the inherently social nature of data, methodologies for filling in missing data, and the limitations of those AI methodologies. Running through all three papers are larger interrogations of the data-driven paradigm which focus on who benefits and who decides?

In response to my first research question, concerning the extent that municipalities collect and maintain open-source data on environmental features that impact accessible travel, I surveyed the open data collections of 178 municipalities across the United States. I found that most municipalities do not collect and maintain data on environmental accessibility features such as curb ramps, crosswalks, pedestrian signals, and sidewalk conditions. While the majority of critiques of data collection focus on privacy (Agre 1994; Elwood and Leszczynski 2011; Nissenbaum 2009; Solove 2010), I suggest that exclusion from big data represents a substantial injustice as well. The lack of information about the accessibility of US municipalities is a significant barrier to environmental access and safe routing for people with disabilities and those cities with robust data on accessibility features were outliers in the sample. I suggest that the lack of data on accessibility occurs both because of our conception of what constitutes a mode of transportation and because those with the resources to collect and maintain the data (municipalities) are incentivized not to. Most of the municipalities with complete or almost complete data on environmental accessibility, also faced high profile ADA compliance lawsuits.

In the second paper, I explore how predictive spatial modelling can be used to generate data on curb ramp locations. Specifically, I use two machine learning algorithms to classify curb ramp locations in 9 urban areas across the United States. In this paper, I argue that the key strength of algorithms – powerful classification and prediction on big datasets – is also a harmful weakness. The related areas of big data and algorithms have been critiqued from all sides for their biases – harmful and *systematic* errors – but this literature largely overlooks the harms that arise from AI's inability to handle nuance, context, individuality, and exception. I examine how error varies across context in ways that are not systematic, and thus cannot just be coded in. I conclude this paper by proposing a kind of rich or thick description of data error, which is slow, tedious, and subjective, but direly needed if we truly intend to develop equitable AI.

The final paper is related to the second and draws from the results of various sampling strategies for curb ramp location classifications in Seattle, WA. This case illuminated key insights about the relationship between scientific replicability and social relevance in the data-driven paradigm. I show how AI and data-driven approaches reverse the process of conceptualization, operationalization, and measurement in study design with consequences for new knowledge generation. This analysis takes the statement that data can speak for themselves seriously and examines what they might have to say. In doing this, I reveal the limitations in data-driven science and the need for careful design and reflexivity.

In framing this research around big questions of: what data are collected, who do these data serve, and what can be done to ameliorate social inequities in data collection, privacy, and use – this research has the potential to help shift practices surrounding big data towards a greater awareness of social impact, and thus, a more just future. This work contributes to several subfields within spatial data science – interpolation, GeoAI, replication and reproducibility (R&R)[29], and data ethics.

[29] Replicable research uses the same or similar methods and data to get the same or similar results. Reproducible research uses the same methods and data to get the same results (Kedron et al. 2019).

REFERENCES CITED

Abbott, Andrew, Alex Deshowitz, Dennis Murray, and Eric C Larson. 2018. “WalkNet: A Deep Learning Approach to Improving Sidewalk Quality and Accessibility” 1 (1): 13.

Ackerman, Emily. 2019. “Lessons From My Standoff With an Autonomous Sidewalk Robot.” *Bloomberg.Com*, November 19, 2019. https://www.bloomberg.com/news/articles/2019-11-19/why-tech-needs-more-designers-with-disabilities.

ADA National Network. 2020. “ADA Title II Action Guide.” May 2020. https://www.adaactionguide.org/.

Agre, Philip E. 1994. “Surveillance and Capture: Two Models of Privacy.” *The Information Society* 10 (2): 101–27. https://doi.org/10.1080/01972243.1994.9960162.

Ahmetovic, Dragan, Roberto Manduchi, James M. Coughlan, and Sergio Mascetti. 2017. “Mind Your Crossings: Mining GIS Imagery for Crosswalk Localization.” *ACM Transactions on Accessible Computing* 9 (4). https://doi.org/10.1145/3046790.

Ai, Chengbo, and Yichang (James) Tsai. 2016. “Automated Sidewalk Assessment Method for Americans with Disabilities Act Compliance Using Three-Dimensional Mobile Lidar.” *Transportation Research Record* 2542 (1): 25–32. https://doi.org/10.3141/2542-04.

Ananny, Mike. 2016. “Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness.” *Science, Technology, & Human Values* 41 (1): 93–117. https://doi.org/10.1177/0162243915606523.

Anderson, Chris. 2008. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.” *Wired*, June 23, 2008. https://www.wired.com/2008/06/pb-theory/.

Arlington County, VA. 2020. “Handicap Ramp Polygons.” Arlington County, VA - Official GIS Open Data Portal. 2020. https://gisdata-arlgis.opendata.arcgis.com/datasets/handicap-ramp-polygons.

Barocas, Solon, and Andrew D. Selbst. 2016. “Big Data’s Disparate Impact.” SSRN Scholarly Paper ID 2477899. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2477899.

Beer, David. 2017. “The Social Power of Algorithms.” *Information, Communication & Society* 20 (1): 1–13. https://doi.org/10.1080/1369118X.2016.1216147.

Berriel, Rodrigo F., Andre Teixeira Lopes, Alberto F. de Souza, and Thiago Oliveira-Santos. 2017. "Deep Learning Based Large-Scale Automatic Satellite Crosswalk Classification." *IEEE Geoscience and Remote Sensing Letters* 14 (9): 1513–17. https://doi.org/10.1109/LGRS.2017.2719863.

Bolten, Nicholas, Amirhossein Amini, Yun Hao, Vaishnavi Ravichandran, Andre Stephens, and Anat Caspi. 2015. "Urban Sidewalks: Visualization and Routing for Individuals with Limited Mobility." In *Proceedings of the 1st International ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics - UrbanGIS'15*, 122–25. Bellevue, WA, USA: ACM Press. https://doi.org/10.1145/2835022.2835042.

Boyd, Danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15 (5): 662–79.

Brault, Matthew, Oce Harrison, Kathy Gips, Jason Angel, and Katherine Blakeslee. 2019. "Results from the Identifying Challenges to Implementing the ADA Survey for Cities and Towns in New England." Boston, MA: Institute for Human Centered Design.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Proceedings of Machine Learning Research*, 81:15.

Bureau of Transportation Statistics. 2018. "Travel Patterns of American Adults with Disabilities." September 6, 2018. https://www.bts.gov/newsroom/travel-patterns-american-adults-disabilities.

Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (1): 2053951715622512. https://doi.org/10.1177/2053951715622512.

City of Boston. 2014. "Pedestrian Ramp Inventory - Analyze Boston." Analyze Boston. 2014. https://data.boston.gov/dataset/pedestrian-ramp-inventory.

City of Denver. 2018. "Curb Ramps (2018)." Denver Open Data Catalog. 2018. https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-curb-ramps-2018.

City of Indianapolis. 2018. "Indianapolis DPW Curb Ratings." Open Data. OpenIndy. March 9, 2018. https://data.indy.gov/.

City of Nashville. 2019. “Sidewalk Inventory for ADA Self-Assessment.” Nashville Open Data Portal. June 12, 2019. https://data.nashville.gov/Transportation/Sidewalk-Inventory-for-ADA-Self-Assessment/vpxc-b5te.

City of San Francisco. 2020. “Curb Ramps | DataSF | City and County of San Francisco.” 2020. https://data.sfgov.org/City-Infrastructure/Curb-Ramps/ch9w-7kih.

City of Seattle. 2020. “City of Seattle Open Data Portal.” Seattle. 2020. https://data.seattle.gov/.

City of Spokane. 2020. “Sidewalk.” September 17, 2020. https://data-spokane.opendata.arcgis.com/datasets/sidewalk.

City of Washington, DC. 2010. “Sidewalk Ramps 2010.” Open Data DC. 2010. https://opendata.dc.gov/.

Collins, Patricia Hill. 2008. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. 1st Edition. New York: Routledge.

Crawford, Kate. 2017. “The Trouble with Bias.” In . https://www.youtube.com/watch?v=fMym_BKWQzk&t=698s.

Deitz, Shiloh, and Katie Meehan. 2019. “Plumbing Poverty: Mapping Hot Spots of Racial and Geographic Inequality in U.S. Household Water Insecurity.” *Annals of the American Association of Geographers* 109 (4): 1092–1109. https://doi.org/10.1080/24694452.2018.1530587.

Demeritt, David. 2001. “The Construction of Global Warming and the Politics of Science.” *Annals of the Association of American Geographers* 91 (2): 307–37. https://doi.org/10.1111/0004-5608.00245.

Department of Justice. 2010. “2010 ADA Standards for Accessible Design.”

D’Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. Cambridge, Massachusetts: The MIT Press.

Dodge, Martin, Rob Kitchin, and Chris Perkins, eds. 2011. *The Map Reader: Theories of Mapping Practice and Cartographic Representation*. 1st edition. Chichester, West Sussex England ; Hoboken, NJ: Wiley.

Drescher, Jack. 2015. “Out of DSM: Depathologizing Homosexuality.” *Behavioral Sciences* 5 (4): 565–75. https://doi.org/10.3390/bs5040565.

Drummond, Chris. 2006. “Machine Learning an Experimental Science (Revisited).” In *AAAI Workshop on Evaluation Methods for Machine Learning*. Boston, MA.

Duckham, M., K. Mason, J. Stell, and M. Worboys. 2001. "A Formal Approach to Imperfection in Geographic Information." *Computers, Environment and Urban Systems*, GISRUK 2000, 25 (1): 89–103. https://doi.org/10.1016/S0198-9715(00)00040-5.

Elwood, Sarah, and Agnieszka Leszczynski. 2011. "Privacy, Reconsidered: New Representations, Data Practices, and the Geoweb." *Geoforum* 42 (1): 6–15. https://doi.org/10.1016/j.geoforum.2010.08.003.

Engler, Alex. 2019. "For Some Employment Algorithms, Disability Discrimination by Default." *Brookings* (blog). October 31, 2019. https://www.brookings.edu/blog/techtank/2019/10/31/for-some-employment-algorithms-disability-discrimination-by-default/.

ESRI. 2020. "Classify LAS Ground (3D Analyst)." ArcGIS Pro Documentation. 2020. https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/classify-las-ground.htm.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

Fairbairn, David, Gennady Andrienko, Natalia Andrienko, Gerd Buziek, and Jason Dykes. 2001. "Representation and Its Relationship with Cartographic Visualization." *Cartography and Geographic Information Science* 28 (1): 13–28. https://doi.org/10.1559/152304001782174005.

Friedman, Batya, and Helen Nissenbaum. 1996. "Bias in Computer Systems." *ACM Trans. Inf. Syst.* 14 (3): 330–47. https://doi.org/10.1145/230538.230561.

Garcia, Mariana, Sharon L. Mulvagh, C. Noel Bairey Merz, Julie E. Buring, and JoAnn E. Manson. 2016. "Cardiovascular Disease in Women: Clinical Perspectives." *Circulation Research* 118 (8): 1273–93. https://doi.org/10.1161/CIRCRESAHA.116.307547.

Garland-Thomson, Rosemarie. 2005. "Feminist Disability Studies." *Signs* 30 (2): 1557–87. https://doi.org/10.1086/423352.

Ghilardi, Marcelo C., Julio Jacques Jr, and Isabel Manssour. 2018. "Crosswalk Localization from Low Resolution Satellite Images to Assist Visually Impaired People," January. https://doi.org/10.1109/MCG.2016.50.

Girma, Haben. 2020. "The Robots Occupying Our Sidewalks." *TechCrunch* (blog). August 11, 2020. https://social.techcrunch.com/2020/08/11/the-robots-occupying-our-sidewalks/.

Goggin, Gerard, Linda Steele, and Jessica Robyn Cadwallader. 2017. "Normality and Disability: Intersections among Norms, Law, and Culture." *Continuum* 31 (3): 337–40. https://doi.org/10.1080/10304312.2017.1275077.

Goldchain, Michelle. 2017. "Project Sidewalk, the Crowd-Sourced Map of D.C.'s Least Accessible Sidewalks, Is Nearly Complete." Curbed DC. October 17, 2017. https://dc.curbed.com/2017/10/17/16488310/project-sidewalk-accessibility-map.

Haixiang, Guo, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and Gong Bing. 2017. "Learning from Class-Imbalanced Data: Review of Methods and Applications." *Expert Systems with Applications* 73 (May): 220–39. https://doi.org/10.1016/j.eswa.2016.12.035.

Hamraie, Aimi. 2017. *Building Access: Universal Design and the Politics of Disability*. 3rd ed. edition. Minneapolis: Univ Of Minnesota Press.

———. 2018. "A Smart City Is an Accessible City." The Atlantic. November 6, 2018. https://www.theatlantic.com/technology/archive/2018/11/city-apps-help-and-hinder-disability/574963/.

Haraway, Donna J. 1996. *Simians, Cyborgs and Women: The Reinvention of Nature*. London: Free Assn Books.

Hey, Tony. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Edited by Stewart Tansley and Kristin Tolle. 1 edition. Redmond, Washington: Microsoft Research.

Hsu, Jeremy. 2019. "Out of the Way, Human! Delivery Robots Want a Share of Your Sidewalk." *Scientific American*, February. https://www.scientificamerican.com/article/out-of-the-way-human-delivery-robots-want-a-share-of-your-sidewalk/.

Hu, Yingjie, Song Gao, Dalton Lunga, Wenwen Li, Shawn Newsam, and Budhendra Bhaduri. 2019. "GeoAI at ACM SIGSPATIAL: Progress, Challenges, and Future Directions." *SIGSPATIAL Special* 11 (2): 5–15. https://doi.org/10.1145/3377000.3377002.

Janowicz, Krzysztof, Song Gao, Grant McKenzie, Yingjie Hu, and Budhendra Bhaduri. 2020. "GeoAI: Spatially Explicit Artificial Intelligence Techniques for Geographic Knowledge Discovery and Beyond." *International Journal of Geographical Information Science* 34 (4): 625–36. https://doi.org/10.1080/13658816.2019.1684500.

Jasanoff, Sheila. 1998. *The Fifth Branch: Science Advisers as Policymakers*. Cambridge, Mass: Harvard University Press.

Jasanoff, Sheila, and Sang-Hyun Kim, eds. 2015. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. 1st edition. Chicago ; London: University of Chicago Press.

Julia Angwin, Jeff Larson. 2016. “Machine Bias.” Text/html. ProPublica. May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Kafer, Alison. 2013. *Feminist, Queer, Crip*. Illustrated edition. Bloomington, Indiana: Indiana University Press.

Kedron, Peter, Amy E. Frazier, Andrew B. Trgovac, Trisalyn Nelson, and A. Stewart Fotheringham. 2019. “Reproducibility and Replicability in Geographical Analysis.” *Geographical Analysis* 0 (0). https://doi.org/10.1111/gean.12221.

Kitchin, Rob. 2014. “Big Data, New Epistemologies and Paradigm Shifts.” *Big Data & Society* 1 (1): 2053951714528481. https://doi.org/10.1177/2053951714528481.

Koopman, Colin. 2019a. “Information before Information Theory: The Politics of Data beyond the Perspective of Communication.” *New Media & Society* 21 (6): 1326–43. https://doi.org/10.1177/1461444818820300.

———. 2019b. *How We Became Our Data: A Genealogy of the Informational Person*. First edition. Chicago: University of Chicago Press.

Kraemer, John D, and Connor S Benton. 2015. “Disparities in Road Crash Mortality among Pedestrians Using Wheelchairs in the USA: Results of a Capture–Recapture Analysis.” *BMJ Open* 5 (11). https://doi.org/10.1136/bmjopen-2015-008396.

Kwan, Mei-Po. 2016. “Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge.” Edited by Mei-Po Kwan and Tim Schwanen. *Annals of the American Association of Geographers* 106 (2): 32–40. https://doi.org/10.4324/9781315266336-4.

Laney, Doug. 2001. “3D Data Management: Controlling Data Volume, Velocity, and Variety.” 949. http://blogs.gartner.com/douglaney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-VolumeVelocity-and-Variety.pdf.

Latour, Bruno. 1988. *Science in Action: How to Follow Scientists and Engineers Through Society*. Revised ed. edition. Cambridge, Mass: Harvard University Press.

Laurie, Emma W., and Ian G. R. Shaw. 2018. “Violent Conditions: The Injustices of Being.” *Political Geography* 65 (July): 8–16. https://doi.org/10.1016/j.polgeo.2018.03.005.

Li, Hanyan, Jack Cebe, Sara Khoeini, Yanzhi “Ann” Xu, Chelsea Dyess, and Randall Guensler. 2018. “A Semi-Automated Method to Generate GIS-Based Sidewalk Networks for Asset Management and Pedestrian Accessibility Assessment.” *Transportation Research Record* 2672 (44): 1–9. https://doi.org/10.1177/0361198118757981.

Li, Songnian, Suzana Dragicevic, Francesc Antón Castro, Monika Sester, Stephan Winter, Arzu Coltekin, Christopher Pettit, et al. 2016. “Geospatial Big Data Handling Theory and Methods: A Review and Research Challenges.” *ISPRS Journal of Photogrammetry and Remote Sensing* 115 (May): 119–33. https://doi.org/10.1016/j.isprsjprs.2015.10.012.

Li, Wenwen. 2020. “GeoAI: Where Machine Learning and Big Data Converge in GIScience.” *Journal of Spatial Information Science*, no. 20 (June): 71–77. https://doi.org/10.5311/JOSIS.2020.20.658.

Li, Y., L. Ma, Z. Zhong, F. Liu, M. A. Chapman, D. Cao, and J. Li. 2020. “Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review.” *IEEE Transactions on Neural Networks and Learning Systems*, 1–21. https://doi.org/10.1109/TNNLS.2020.3015992.

Li, Yangyan. (2017) 2021. *Yangyanli/PointCNN*. Python. https://github.com/yangyanli/PointCNN.

Li, Yangyan, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. 2018. “PointCNN: Convolution On X-Transformed Points.” *ArXiv:1801.07791 [Cs]*, November. http://arxiv.org/abs/1801.07791.

Linton, Simi. 1998. *Claiming Disability: Knowledge and Identity*. NYU Press.

Lund, Christian. 2014. “Of What Is This a Case?: Analytical Movements in Qualitative Social Science Research.” *Human Organization* 73 (3): 224–34. https://doi.org/10.17730/humo.73.3.e35q482014x033l4.

Luo, Ji, Guoyuan Wu, Zhensong Wei, Kanok Boriboonsomsin, and Matthew Barth. 2019. “Developing an Aerial-Image-Based Approach for Creating Digital Sidewalk Inventories.” *Transportation Research Record* 2673 (8): 499–507. https://doi.org/10.1177/0361198119842820.

Marshall, Aarian, and Alex Davies. 2019. “Uber’s Self-Driving Car Didn’t Know Pedestrians Could Jaywalk.” *Wired*, November 5, 2019. https://www.wired.com/story/ubers-self-driving-car-didnt-know-pedestrians-could-jaywalk/.

Merriam-Webster, Inc. 2021. “Dictionary by Merriam-Webster: America’s Most-Trusted Online Dictionary.” 2021. https://www.merriam-webster.com/.

Metz, Cade, and Kate Conger. 2020. “Uber, After Years of Trying, Is Handing Off Its Self-Driving Car Project.” *The New York Times*, December 7, 2020, sec. Technology. https://www.nytimes.com/2020/12/07/technology/uber-self-driving-car-project.html.

Mezey, Naomi. 2002. “Erasure and Recognition: The Census Race and the National Imagination.” *Northwestern University Law Review* 97 (4): 1701–68.

Miller, Harvey J., and Michael F. Goodchild. 2015. “Data-Driven Geography.” *GeoJournal* 80 (4): 449–61. https://doi.org/10.1007/s10708-014-9602-6.

Mills, Kim. 1995. “Disabilities Act: An ‘Unfunded Mandate’ Too Potent To Oppose.” *AP NEWS*, February 26, 1995, sec. Archive. https://apnews.com/e369ecf20fad4877b5cbea7b072ab2ee.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. “Model Cards for Model Reporting.” In *FAT*. https://doi.org/10.1145/3287560.3287596.

More, A., and D. Rana. 2017. “Review of Random Forest Classification Techniques to Resolve Data Imbalance.” *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*. https://doi.org/10.1109/ICISIM.2017.8122151.

Nissenbaum, Helen. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. 1 edition. Stanford, Calif: Stanford University Press.

Onuoha, Mimi. (2016) 2019. *MimiOnuoha/Missing-Datasets*. https://github.com/MimiOnuoha/missing-datasets.

Onuoha, Mimi, and Diana Nucera. 2018. *A People’s Guide to AI*. Michigan: Whitlock.

Open Knowledge Foundation. 2020. “What Is Open Data?” Open Data Handbook. 2020. https://opendatahandbook.org/guide/en/what-is-open-data/.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. 1 edition. Cambridge: Harvard University Press.

Patten, Eileen. 2016. “Racial, Gender Wage Gaps Persist in U.S. despite Some Progress.” *Pew Research Center* (blog). July 1, 2016. https://www.pewresearch.org/fact-tank/2016/07/01/racial-gender-wage-gaps-persist-in-u-s-despite-some-progress/.

Perez, Caroline Criado. 2019. *Invisible Women: Data Bias in a World Designed for Men*. First Printing edition. New York: Harry N. Abrams.

Platanios, Emmanouil Antonios, Maruan Al-Shedivat, Eric Xing, and Tom Mitchell. 2020. "Learning from Imperfect Annotations." *ArXiv:2004.03473 [Cs, Stat]*, April. http://arxiv.org/abs/2004.03473.

Ponterotto, Joseph. 2006. "Brief Note on the Origins, Evolution, and Meaning of the Qualitative Research Concept 'Thick Description.'" *Qualitative Report* 11 (September): 538–49.

Poon, Linda. 2015. "Why Are Wheelchair Users More Likely to Be Killed in Traffic Than Other Pedestrians?" CityLab. November 19, 2015. http://www.citylab.com/politics/2015/11/wheelchair-users-are-more-likely-to-be-killed-in-traffic-than-other-pedestrians/416667/.

Prensky, Marc. 2009. "H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom." *Innovate: Journal of Online Education* 5 (3). https://nsuworks.nova.edu/innovate/vol5/iss3/1.

Richardson, Rashida, Jason Schultz, and Kate Crawford. 2019. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." SSRN Scholarly Paper ID 3333423. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=3333423.

Robinson, Anthony C., Urška Demšar, Antoni B. Moore, Aileen Buckley, Bin Jiang, Kenneth Field, Menno-Jan Kraak, Silvana P. Camboim, and Claudia R. Sluter. 2017. "Geospatial Big Data and Cartography: Research Challenges and Opportunities for Making Maps That Matter." *International Journal of Cartography* 3 (sup1): 32–60. https://doi.org/10.1080/23729333.2016.1278151.

Saha, Manaswi, Kotaro Hara, Jon Froehlich, Michael Saugstad, Hanuma Teja Maddali, Aileen Zeng, Ryan Holland, et al. 2019. "Project Sidewalk: A Web-Based Crowdsourcing Tool for Collecting Sidewalk Accessibility Data At Scale." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–14. Glasgow, Scotland Uk: ACM Press. https://doi.org/10.1145/3290605.3300292.

Saltes, Natasha. 2013. "'Abnormal' Bodies on the Borders of Inclusion: Biopolitics and the Paradox of Disability Surveillance." *Surveillance & Society* 11 (1/2): 55–73. https://doi.org/10.24908/ss.v11i1/2.4460.

Scaratti, Giuseppe, Laura Galuppo, Mara Gorli, Caterina Gozzoli, and Silvio Ripamonti. 2017. "The Social Relevance and Social Impact of Knowledge and Knowing." *Management Learning* 48 (1): 57–64. https://doi.org/10.1177/1350507616680563.

Selbst, Andrew D. 2017. "Disparate Impact in Big Data Policing." *Georgia Law Review* 52 (1): 109–96.

Selin, Cynthia, and Jathan Sadowski. 2015. "Against Blank Slate Futuring: Noticing Obduracy in the City through Experiential Methods of Public Engagement." In *Remaking Participation: Science, Environment and Emergent Publics*. https://asu.pure.elsevier.com/en/publications/against-blank-slate-futuring-noticing-obduracy-in-the-city-throug.

Shakespeare, Tom. 2007. "Disability, Normality, and Difference." In *Psychological Challenges in Obstetrics and Gynecology: The Clinical Management*, edited by Jayne Cockburn and Michael E. Pawson, 51–59. London: Springer. https://doi.org/10.1007/978-1-84628-808-1_6.

Shioyama, Tadayoshi, Haiyuan Wu, Yuna Nishibe, Naoki Nakamura, and Suguru Kitawaki. 2001. "Image Analysis of Crosswalk." *Proceedings 11th International Conference on Image Analysis and Processing*, 168–73. https://doi.org/10.1109/ICIAP.2001.957003.

Siebers, Tobin Anthony. 2008. *Disability Theory*. Ann Arbor: University of Michigan Press.

Simonite, Tom. 2017. "Machines Taught by Photos Learn a Sexist View of Women." *Wired*, August 21, 2017. https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/.

———. 2018. "Google Photos Still Has a Problem with Gorillas." MIT Technology Review. 2018. https://www.technologyreview.com/f/609959/google-photos-still-has-a-problem-with-gorillas/.

Smith, Kelley. 2016. "City Responds to ADA Complaint." September 28, 2016. https://www.dakotanewsnow.com/content/news/City-responds-to-ADA-complaint-395207541.html.

Smith, Leslie N. 2017. "Cyclical Learning Rates for Training Neural Networks." *ArXiv:1506.01186 [Cs]*, April. http://arxiv.org/abs/1506.01186.

Smith, Virginia, Jitendra Malik, and David Culler. 2013. "Classification of Sidewalks in Street View Images." In *2013 International Green Computing Conference Proceedings*, 1–6. https://doi.org/10.1109/IGCC.2013.6604476.

Solove, Daniel J. 2010. *Understanding Privacy*. 2/28/10 edition. Cambridge, Massachusetts London, England: Harvard University Press.

Star, Susan Leigh. 2016. "The Ethnography of Infrastructure:" *American Behavioral Scientist*, July. https://doi.org/10.1177/00027649921955326.

Strong, Jennifer, Tate Ryan-Mosley, Emma Cillekens, and Karen Hao. n.d. "Who Owns Your Face?" In Machines We Trust. Accessed January 31, 2021. https://www.technologyreview.com/2020/08/12/1006678/who-owns-your-face/.

Suresh, Harini, and John V. Guttag. 2020. “A Framework for Understanding Unintended Consequences of Machine Learning.” *ArXiv:1901.10002 [Cs, Stat]*, February. http://arxiv.org/abs/1901.10002.

Tajbakhsh, Nima, Laura Jeyaseelan, Qian Li, Jeffrey Chiang, Zhihao Wu, and Xiaowei Ding. 2020. “Embracing Imperfect Datasets: A Review of Deep Learning Solutions for Medical Image Segmentation.” *ArXiv:1908.10454 [Cs, Eess]*, February. http://arxiv.org/abs/1908.10454.

Tin Kam Ho. 1995. “Random Decision Forests.” In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1:278–82 vol.1. https://doi.org/10.1109/ICDAR.1995.598994.

Tobler, Waldo Rudolph. 1970. “A Computer Movie Simulating Urban Growth in the Detroit Region.” In . https://doi.org/10.1111/j.1467-8306.2004.09402005.x.

Turnbull, David. 2003. *Masons, Tricksters and Cartographers: Comparative Studies in the Sociology of Scientific and Indigenous Knowledge*. Taylor & Francis.

US Census Bureau. 2020. “TIGER/Line Shapefiles.” The United States Census Bureau. 2020. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html.

USGS. 2020. “TNM Download V2.” National Map Viewer. 2020. https://viewer.nationalmap.gov/basic/#/.

VoPham, Trang, Jaime E. Hart, Francine Laden, and Yao-Yi Chiang. 2018. “Emerging Trends in Geospatial Artificial Intelligence (GeoAI): Potential Applications for Environmental Epidemiology.” *Environmental Health* 17 (1): 40. https://doi.org/10.1186/s12940-018-0386-x.

Wakabayashi, Daisuke. 2018. “Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam.” *The New York Times*, March 19, 2018, sec. Technology. https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html.

Washington State Department of Natural Resources (DNR). 2021. “Lidar.” Lidar Portal. January 25, 2021. https://www.dnr.wa.gov/lidar.

Whittaker, Meredith, Meryl Alper, Olin College, Liz Kaziunas, and Meredith Ringel Morris. 2019. “Disability, Bias, and AI,” November, 32.

World Health Organization, ed. 2001. *International Classification of Functioning, Disability and Health: ICF*. Geneva: World Health Organization.

Zuboff, Shoshana. 2015. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization." *Journal of Information Technology* 30 (1): 75–89. https://doi.org/10.1057/jit.2015.5.

———. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. 1 edition. New York: PublicAffairs.

———. 2021. "Opinion | The Coup We Are Not Talking About." *The New York Times*, January 29, 2021, sec. Opinion. https://www.nytimes.com/2021/01/29/opinion/sunday/facebook-surveillance-society-technology.html.