

Southern Methodist University

SMU Scholar

---

Civil and Environmental Engineering Theses and  
Dissertations

Civil Engineering and Environmental  
Engineering

---

Spring 5-14-2022

## Equity Of Urban Neighborhood Infrastructure: A Data-Driven Assessment

Zheng Li

*Southern Methodist University*, [zli1@smu.edu](mailto:zli1@smu.edu)

Follow this and additional works at: [https://scholar.smu.edu/engineering\\_civil\\_etds](https://scholar.smu.edu/engineering_civil_etds)



Part of the [Civil and Environmental Engineering Commons](#)

---

### Recommended Citation

Li, Zheng, "Equity Of Urban Neighborhood Infrastructure: A Data-Driven Assessment" (2022). *Civil and Environmental Engineering Theses and Dissertations*. 20.

[https://scholar.smu.edu/engineering\\_civil\\_etds/20](https://scholar.smu.edu/engineering_civil_etds/20)

This Dissertation is brought to you for free and open access by the Civil Engineering and Environmental Engineering at SMU Scholar. It has been accepted for inclusion in Civil and Environmental Engineering Theses and Dissertations by an authorized administrator of SMU Scholar. For more information, please visit <http://digitalrepository.smu.edu>.

EQUITY OF URBAN NEIGHBORHOOD INFRASTRUCTURE:  
A DATA-DRIVEN ASSESSMENT

Approved by:

---

Professor Barbara Minsker  
Civil and Environmental Engineering  
Dissertation Committee Chairperson

---

Professor Dinesh Rajan  
Electrical and Computer Engineering

---

Professor Eric C. Larson  
Computer Science

---

Professor Janille Smith-Colin  
Civil and Environmental Engineering

---

Professor Jessie Marshall Zarazaga  
Civil and Environmental Engineering  
Hunt Institute for Engineering and Hu-  
manity

---

Professor Xinlei Wang  
Department of Statistical Science

EQUITY OF URBAN NEIGHBORHOOD INFRASTRUCTURE:  
A DATA-DRIVEN ASSESSMENT

A Dissertation Presented to the Graduate Faculty of the  
Lyle School of Engineering  
Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Civil and Environmental Engineering

by

Zheng Li

M.S., Civil and Environmental Engineering, University of Illinois at  
Urbana-Champaign

B.S., Water Supply and Drainage Engineering, Tongji University

May 14, 2022

Copyright (2022)

Zheng Li

All Rights Reserved



Li, Zheng

M.S., Civil and Environmental Engineering, UIUC, 2014  
B.S., Water Supply and Drainage Engineering, Tongji University, 2012

Equity of Urban Neighborhood Infrastructure:  
A Data-Driven Assessment

Advisor: Professor Barbara Minsker

Doctor of Philosophy conferred May 14, 2022

Dissertation completed May 06, 2022

Neighborhood infrastructure, such as sidewalks, medical facilities, public transit, community gathering places, and tree canopy, provides essential support for safe, healthy, and resilient communities. This thesis proposes, develops, and implements an innovative approach to thoroughly examine the presence and condition of neighborhood infrastructure. It demonstrates the necessity of considering multiple infrastructure types when studying neighborhood infrastructure and its equity. This thesis provides an automated assessment framework as well as case studies among four major metropolitan cities across the United States, which expands the research opportunities for future infrastructure-related research. Chapter 1 introduces the concept of neighborhoods infrastructure and describes the need and benefits of studying neighborhoods infrastructure. It also highlights the importance of including multiple infrastructure types while trying to fully understand neighborhood infrastructure. In Chapter 2, a generalized data-driven framework is developed and presented at the street level. It addresses the methodological challenge of considering multiple infrastructure types and provides quantitative condition measures. The infrastructure equity is also measured using statistical inference based on the overall infrastructure condition. In Chapter 3, the background of four major cities is introduced including considered infrastructure deserts, assessment criteria, neighborhood demographic, and historical information. In Chapter 4, The infrastructure assessment framework is implemented for 12 types of neighborhood infrastructure in Dallas, Texas. The results show significant infrastructure inequities

across income levels for most types of infrastructure. Statistical inference predicts (with 95% confidence) that low-income neighborhoods are 2.0 to 3.5 times more likely to have highly deficient infrastructure (8 or more deficient infrastructure types) than high-income areas and 1.4 to 2.4 times more likely to have highly deficient infrastructure than middle-income neighborhoods.

Chapter 5 continues to explore infrastructure equity by considering the neighborhood's racial and ethnic demographic composition. The results show significant infrastructure inequities across neighborhoods with different race-ethnicity for most types of infrastructure. Statistical inference also indicates (with 95% confidence) that predominantly Black and Hispanic neighborhoods are 1.44 to 2.56 times and 1.95 to 3.63 times likelier to have highly deficient infrastructure (8 or more deficient infrastructure types out of 12) than areas with no predominant race-ethnicity, respectively. Furthermore, chapter 5 also reveals the legacy of historical discriminatory policy (redlining) and its long-term impacts on neighborhood infrastructure: neighborhoods marked as "less desirable" for financial services during the 1930s still experience significantly higher infrastructure deficiencies nowadays. Chapter 5 expands and deepens the perspectives of understanding infrastructure equity quantitatively using racial-ethnic and historical information.

Chapter 6 generalizes and automates the infrastructure assessment framework using a Web-based platform called "Clowder". The implementation of the assessment directly can be executed using a Web browser with available data. A user-friendly graphical interface is also provided for users to manage and set up the assessment steps based on their needs. Chapter 6 also applies the generalized assessment framework to four major cities across the United States: Dallas, TX, New York City, NYC, Chicago, IL, and Los Angeles, CA. The results show different levels of inequity among studied cities. It indicates infrastructure equity to be region-specific and thus it is essential to understand the specificity among regions to better examine, plan, and redevelop communities. Finally, the dissertation concludes with Chapter 7 where major conclusions, limitations of the thesis, and future work are discussed.

## TABLE OF CONTENTS

LIST OF FIGURES . . . . .	x
LIST OF TABLES . . . . .	xii
ACKNOWLEDGMENTS . . . . .	xiii
CHAPTER	
1. Introduction . . . . .	1
1.0.1. Background and Motivation . . . . .	2
2. Infrastructure Assessment Framework . . . . .	6
2.0.1. Introduction . . . . .	6
2.0.2. Methodology . . . . .	6
2.0.2.1. Compute overall infrastructure deficiency by neighborhood	8
2.0.2.2. Test for existence of infrastructure deserts . . . . .	10
2.0.2.3. Estimate infrastructure inequity risk with cumulative logit models . . . . .	10
3. Case Studies . . . . .	13
3.0.1. Dallas, TX . . . . .	13
3.0.2. Infrastructure Data . . . . .	14
3.0.3. Neighborhood's Income . . . . .	14
3.0.4. Neighborhood's Race-ethnicity and Age . . . . .	16
3.0.5. Historical Redlining Practices . . . . .	16
3.0.6. Los Angeles, CA, New York City, NY, and Chicago, IL . . . . .	18
3.0.6.1. Infrastructure datasets . . . . .	18
3.0.6.2. Income, race-ethnicity, historical redlining, and car ownership	19
4. Do Infrastructure Deserts Exist? Assessment and Statistical Modeling of Neighborhood Infrastructure in Dallas, Texas . . . . .	21

4.0.1. Introduction . . . . .	21
4.0.2. Individual and Overall Infrastructure Condition . . . . .	21
4.0.3. Infrastructure Deserts . . . . .	24
4.0.4. Relative Risk and Infrastructure Inequity . . . . .	25
4.0.5. Robustness of Statistical Model . . . . .	27
4.0.6. Conclusions . . . . .	29
5. Are Infrastructure Deserts Correlated with Neighborhood Race and Ethnicity? .	30
5.0.1. Introduction . . . . .	30
5.0.2. Methodology . . . . .	31
5.0.2.1. Cumulative logit model . . . . .	31
5.0.2.2. Overall infrastructure deficiency $\sim$ race-ethnicity . . . . .	32
5.0.2.3. Overall infrastructure deficiency $\sim$ income, race-ethnicity .	33
5.0.2.4. Gamma statistic . . . . .	34
5.0.3. Results and Discussion . . . . .	35
5.0.3.1. Individual infrastructure condition . . . . .	35
5.0.3.2. Infrastructure inequity across race-ethnicity groups . . . . .	36
5.0.3.3. Infrastructure inequity across income and race-ethnicity groups	40
5.0.3.4. Impacts of historical redlining on infrastructure condition .	41
5.0.3.5. Infrastructure inequity by neighborhood age . . . . .	43
5.0.4. Discussion . . . . .	45
5.0.5. Conclusions . . . . .	48
6. Do neighborhood infrastructure deficits differ by city? A comparative infrastruc-	
ture equity study among Los Angeles, New York City, Chicago, and Dallas .	50
6.0.1. Introduction . . . . .	50
6.0.2. Methodology . . . . .	51

6.0.2.1. Clowder . . . . .	52
6.0.2.2. Prepare assessment configuration JSON file . . . . .	53
6.0.2.3. Upload necessary datasets to Clowder . . . . .	54
6.0.2.4. Execute infrastructure assessment framework as a Clowder extractor . . . . .	55
6.0.2.5. Statistical models . . . . .	56
6.0.3. Results . . . . .	57
6.0.3.1. Comparison by individual infrastructure type . . . . .	58
6.0.3.2. Relative risk among Los Angeles, New York City, Chicago, and Dallas . . . . .	62
6.0.3.3. Infrastructure equity from the view of historical redlining regions . . . . .	63
6.0.4. Discussion . . . . .	64
6.0.4.1. Car ownership and access-related infrastructure deficiencies	64
6.0.4.2. Infrastructure inequity and racial segregation . . . . .	66
6.0.5. Conclusions . . . . .	67
7. Summary, limitations, and future work . . . . .	71
7.0.1. Discussion . . . . .	71
7.0.2. Limitations and Future Work . . . . .	72
7.0.2.1. Limitations of the framework . . . . .	72
7.0.2.2. Limitations of future infrastructure-related applications . .	74
7.0.2.3. Future work . . . . .	75
APPENDIX . . . . .	77
A.1. Supplementary Tables . . . . .	77
A.2. Crosswalk detection . . . . .	81
A.3. Pseudo-code of the method used to compute 12 deficient infrastructure types	84

A.4. Full cumulative logit model considering income, race-ethnicity, and city . . .	90
A.5. Estimated parameters of cumulative logit model for comparative study between Los Angeles, New York City, Chicago, and Dallas . . . . .	91
BIBLIOGRAPHY . . . . .	93

## LIST OF FIGURES

Figure	Page
2.1. Overview of the infrastructure equity framework. . . . .	7
3.1. Historical Home Owner Loan Corporation Redlining Areas (1937). . . . .	17
4.1. Individual infrastructure substandard percentage by income. . . . .	22
4.2. Overall infrastructure deficiency. . . . .	23
4.3. Infrastructure deserts in Dallas, TX. . . . .	24
4.4. Deficient infrastructure as a percentage of neighborhoods. . . . .	25
4.5. Relative risk analysis. . . . .	25
4.6. Robustness analyses with continuous income. . . . .	28
5.1. Percentage of deficient infrastructure by neighborhood income level and infrastructure type. . . . .	36
5.2. Overall infrastructure deficiency histogram and by neighborhood's race-ethnicity.	37
5.3. Relative risk of neighborhoods with different race-ethnicity . . . . .	39
5.4. Probability of having highly deficient infrastructure given neighborhood income and race-ethnicity composition. . . . .	42
5.5. Box-whisker plot of overall infrastructure deficiency among HOLC rated neighborhoods. . . . .	43
5.6. Box-whisker plot of infrastructure deficiency versus neighborhood average built-year during each decade for (a) all neighborhoods combined and (b) neighborhoods by predominant race-ethnicity. . . . .	45

5.7. Distributions of 2017 City of Dallas bond projects infrastructure-related investments and percentage of identified infrastructure deserts by city council districts. . . . .	48
6.1. Overview of the Cloud-based infrastructure assessment framework. . . . .	52
6.2. Infrastructure assessment framework configuration file (JSON). . . . .	54
6.3. Histograms of computed overall infrastructure deficiency in Los Angeles, Chicago, Dallas, and New York City. . . . .	58
6.4. Overall infrastructure deficiency by neighborhood’s income level. . . . .	59
6.5. Overall infrastructure deficiency by neighborhood’s race-ethnicity. . . . .	60
6.6. Percentage of individual deficient infrastructure at three levels of income by cities. . . . .	61
6.7. Percentage of individual deficient infrastructure at different neighborhood’s race-ethnicity by cities. . . . .	61
6.8. Top 5 Ranked deficient infrastructure types represented as percentages across the city. . . . .	62
6.9. Top 5 Ranked deficient infrastructure types represented as percentages within neighborhoods with at least 5 deficient infrastructure types. . . . .	63
6.10. The estimated relative risks of having highly deficient infrastructure given neighborhood’s income, race-ethnicity, and city. . . . .	64
6.11. The distribution of overall infrastructure deficiency across historical redlining regions. . . . .	65
6.12. Heatmap of neighborhoods given car ownership levels and overall infrastructure deficiency. . . . .	66
6.13. The distributions of infrastructure deficiency, car ownership percentage by boroughs in New York City. . . . .	70
6.14. Maps of Los Angeles, Chicago, Dallas, and New York City show the existence of racial segregation. . . . .	70
7.1. Three types of crosswalks in City of Dallas. . . . .	81
7.2. Checking crosswalk presence at an intersection using a 30m search radius. . .	82



## LIST OF TABLES

Table	Page
3.1. Substandard criterion for neighborhood infrastructure types . . . . .	15
3.2. Substandard criteria for Los Angeles, New York City, Chicago, and Dallas (10 infrastructure types) . . . . .	19
4.1. Estimated coefficients of the cumulative logit model . . . . .	26
4.2. Cumulative logit model parameters using continuous income . . . . .	27
5.1. Estimated coefficients of the cumulative logit model (race-ethnicity only) . . .	38
5.2. Estimated coefficients of the cumulative logit model for income and race-ethnicity	41
5.3. Trend analysis using Gamma statistic between overall infrastructure deficiency	44
5.4. Gamma statistic between overall infrastructure deficiency and predominant neighborhood race-ethnicity by average built-year in decades. . . . .	46
A.1. Dataset information for individual infrastructure type . . . . .	77
A.2. Descriptive statistics for the substandard percentage ( $\mu$ ) of individual infrastructure type . . . . .	78
A.3. Description of the Shapefile consisting information of all assessed infrastructure types . . . . .	79
A.4. Infrastructure Data Source Table for Los Angeles, New York City, Chicago, and Dallas (10 infrastructure types) . . . . .	80
A.5. Estimated coefficients of the full cumulative logit model . . . . .	92

## ACKNOWLEDGMENTS

The completion of this dissertation was a long and arduous process of self-discovery and resiliency that was only possible thanks to the love and support from my family, friends, advisors and of course, from my beloved wife, Jiayu Chen.

To Professor Barbara Minsker, my advisor, I will always be grateful for recognizing my potential to be a researcher and for taking a chance on me to become your first Ph.D. student at SMU. It was your caring personality and your humane approach to influencing me to grow as an independent researcher, but more importantly, as a person knowing how to be mindful at work, stay positive and enjoy every moment in life.

To Professor Dinesh Rajan, Professor Eric Larson, Professor Janille Smith-Colin, Professor Jessie Zarazaga, and Professor Xinlei Wang, I thank you for your guidance and contribution to the work. I am very grateful for your dedication and all the efforts you devoted while advising my research.

To my research group colleagues, Dr. Bardia Heidari, Arefeh Safaei, and Farbod Tavakkoli for providing many valuable comments and assistance to improve the work throughout these years. I truly cherish our friendships.

To my family and wife. It was your unconditional support, encouragement, and love what provided me with the energy and motivation I needed to complete my Ph.D. studies.

For the infrastructure assessment framework and its first case study implementation Chapter (1-4), I wish to acknowledge the contributions of the many faculty members, graduate students, community members, and undergraduate research assistants who contributed

to this project, including Dr. Bardia Heidari, Dr. Dinesh Rajan, Dr. Doric Earle, Dr. Eric Larson, Dr. Jo Guldi, Dr. Kenneth Berry, Dr. May Yuan, Dr. Michael Hites, Dr. Tammy Leonard, Dr. Eva Csaky, Dr. Timothy Bray, Arefeh Safaei Moghadam, Azucena Milan, Elizabeth Killingsworth, Farbod Tavakkoli, Gurveer Kuar, Javad Jomehpour, Jennifer Ebinger, Jerry Hawkins, Cyndy Lutz, Jessica MacKinnon, Luigi Marini, Melanie Ferguson, Mohamed Elsaied, Myriam Zakhem, Owen Wilson-Chavez, Raghuram Srinivas, and Sylvia George-Williams.

For the study of role of race-ethnicity in infrastructure deserts (Chapter 5), I wish to acknowledge the contributions of the many faculty and graduate students who contributed to this project, particularly: Dr. Janille Smith-Colin, Dr. Jessie Zarazaga, Farbod Tavakkoli, and Arefeh Safaei.

For multi-city comparative analysis (Chapter 6), I would like to thank Chicago Metropolitan Agency for Planning (CMAP) for providing pavement condition data.

Finally, the completion of this degree would not have been possible without the financial support of funding from the National Science Foundation (NSF) through award number 1835877.

## Chapter 1

### Introduction

This thesis develops and implements a generalizable data-driven framework for assessing the condition and equity of neighborhood infrastructure. Neighborhood infrastructure is a system of relatively small-scale physical structures and service facilities (e.g., sidewalks, tree canopy, and medical facilities) that play an essential role in improving resident's lives, health, safety, and social justice [1, 2]. Understanding neighborhood infrastructure is necessary for effective community growth, quality of life, and prioritization of future infrastructure investments.

While efforts have been made to study the condition and impact of individual types of neighborhood infrastructure, the focus and scope of such efforts remain relatively singular (i.e., limited to one or a few infrastructure types) and fail to treat infrastructure as a complex, interconnected system. For example, many daily activities involve multiple infrastructure elements. Leisure walking experiences could be affected by sidewalk condition, crosswalk presence at intersections, pavement condition, and street tree cover. Thus to truly understand the overall infrastructure condition in a neighborhood, it is essential to rethink the assessment and consider multiple infrastructure types simultaneously.

To address this need, a data-driven framework for assessing multiple infrastructure types is developed and implemented in four U.S. cities to illustrate the framework's advantages: Dallas, TX; New York, NY; Chicago, IL, and Los Angeles, CA. Systematically examining neighborhood infrastructure presence and condition identifies socially disadvantaged neighborhoods suffering severe infrastructure deficits, known as "infrastructure deserts." The equity of the regional distribution of neighborhood infrastructure across social-economic status is examined, along with residues from historic discriminatory mortgage lending practices. Furthermore, the generalizable framework is implemented in a data cyberinfrastructure called Clowder, enabling comparison and insights on how the level of infrastructure equity varies

across different cities. This Chapter introduces the motivation behind the proposed framework, the primary research contributions, and an overview of the remainder of this thesis.

### 1.0.1 Background and Motivation

The forms of neighborhood infrastructure can be physical structures (such as sidewalks, crosswalks, pedestrian trails, street lights, street tree canopy) or facilities (such as hospitals) located or operated within or near a neighborhood to provide community services. The services include human development support (such as health clinics, financial facilities), public services (such as transportation, schools, libraries, internet), and shared space for social gatherings and recreational activities (such as parks, trails, community centers). Neighborhood infrastructure is socially, economically, and operationally linked with the neighborhood and is considered critical for communities' growth and wellbeing [1]. Moreover, neighborhood infrastructure types can be highly diverse and vary from community to community depending on geophysical, socio-cultural, and economic factors that influence a neighborhood's growth. Therefore, the estimation of impacts, changes, and future development of neighborhood infrastructure requires a thorough and in-depth understanding of infrastructure conditions and the community's social-economic setting.

Previous studies have shown the importance of neighborhood infrastructure for human health, community growth, and community safety. For example, neighborhood infrastructure, particularly sidewalks, streets, and access to local destinations such as grocery stores, parks, and recreation facilities, have impacts on obesity [3–5]; related chronic health outcomes [6, 7]; health behaviors [7–10]; mental health outcomes [6, 9, 11]; and social well-being outcomes [10, 12, 13]. Pedestrian-friendly streets, open green spaces, and well-maintained neighborhood infrastructure (such as sidewalks, crosswalks, healthcare, food stores, and community centers) not only promote healthy activities such as walking and bicycling [12] but also enhance social interactions [14], social cohesion and social capital [13]. These factors facilitate the organic growth of community attitudes toward healthy and active lifestyles [14–16]. Furthermore, studies have shown the positive influence of well-established neighborhood infrastructure (such as sidewalks, crosswalks, improved street lighting) on perceived safety from crime or traffic-related events [17–19]. For example, improved street

lighting can reduce crime-related fear by intensifying community surveillance [17]. Similarly, well-designed crosswalks and sidewalks help reduce pedestrian-vehicle crashes [19].

Conversely, the lack of quality neighborhood infrastructure makes the community more susceptible to natural disasters and chronic economic crises. Lack of neighborhood infrastructure has also been considered a critical indicator of social injustice based on three main allocation principles [20, 21]: equity, equality, and need. Equity calls for fairness and equal treatment for equals [22]. Equality means that everyone receives the same public service [23], which usually leads to more harmonious social relationships [20, 24]. The concept of need is consistent with the idea that those needing more service should receive more rather than less [23]. Each of these three principles operates in a specific domain. For neighborhood infrastructure, including public utilities, parks, and facilities, equality is often impossible to achieve in the sense of equal access because of the variation in community development and terrain. The need is also likely tied to population and distributed geographically (Lucy, 1981). Therefore, analyzing the equitable distribution of infrastructure services is a better approach for understanding the present condition and future investments. Thus the author primarily focuses on the equity aspect of neighborhood infrastructure distribution in this work.

One way of studying infrastructure equity, as suggested by the U.S. Department of Transportation, is to compare the infrastructure characteristics or condition of neighborhoods with high concentrations of socially vulnerable populations (such as low-income households, minorities, and car-free households) to those in adjacent neighborhoods or to regional averages [25]. Following this guideline, many researchers have evaluated individual infrastructure conditions and discovered infrastructure inequities across the spectrum of neighborhood infrastructure types. Studies [26–36] have shown economic and ethnic disparities in individual types of neighborhood-scale infrastructure, including walkability, street trees, public transportation, parks, pedestrian crosswalks, and trails. Grocery stores and farmer markets, among categories of neighborhood infrastructure have also been widely studied in the realm of "food deserts" and have also shown substantial inequities across different social-economic and racial groups [37, 38]. While these studies show the importance of individual infrastructure types and their impacts on communities, the presence and impact of multiple deficient

types of physical infrastructure are not yet known. A systematic condition assessment must be established at the community level to evaluate multiple neighborhood infrastructure types and examine their aggregated impacts on the community.

Previous studies have attempted to assess neighborhood infrastructure using a variety of measures. Inspecting infrastructure surface exteriors and identifying defects is the most common method to assess infrastructure such as streets and sidewalks [39, 40]. Researchers have also used proximity as the condition measure to assess the coverage of infrastructure’s service such as parks [30, 33, 34, 41–44], healthcare [45–48], public transportation system (bus stops, rail stations) [49–51], and fresh food supplies [52–55].

In addition, measures derived from field audits, secondary data sources, or satellite imagery/videos allow the assessment of ground-based or hard-to-measure infrastructure such as neighborhood street walkability [39, 56–59], street tree canopy [60–62] internet speed [63], and street condition [40, 64]. However, the integrated physical assessment of multiple neighborhood infrastructure types faces methodological challenges and limitations in effective implementation [65]. Due to the difficulty of gathering neighborhood-scale data of multiple infrastructure types on a large scale, most previous studies have primarily focused on either small-scale studies of one or several infrastructure types or large-scale studies at the city or regional scale of only a single infrastructure type.

This study fills the gap between these two scales by providing quantitative infrastructure condition assessment at the city scale and systematically combining multiple infrastructure types to show overall infrastructure condition using a data-driven framework. The results obtained from the framework can be overlaid with the neighborhood’s social-economic (such as income level) and social-demographic information (such as racial-ethnicity) to further explore infrastructure inequities. The framework can also be used in other cities via an automated cyberinfrastructure that enables nationwide comparative studies on neighborhood infrastructure among different cities. In the following chapters, Chapter 2 presents the structure of the generalized framework that addresses multiple infrastructure types and shows how to statistically model infrastructure inequity. Chapter 3 provides background information on the four cities that are examined as case studies of infrastructure equity. Chapter 4 applies the framework to the city of Dallas as the first case study and reveals the existence

of "infrastructure deserts," as well as highlighting infrastructure inequity across neighborhoods with different median income levels. Chapter 5 further explores the roles of race and ethnicity, historical discriminatory housing policies, and neighborhood age on infrastructure equity, in addition to neighborhood's income characteristics in Dallas. In Chapter 6, the assessment process is automated and the framework is integrated into a Cloud-based data platform called Clowder. Using this cyberinfrastructure, the assessment framework is generalized with application to three other case studies in major U.S. cities and identification of the variability of infrastructure equity in different regions. Finally, in Chapter 7 the author concludes the thesis by highlighting primary topics, limitations, and recommendations for future work.



## Chapter 2

### Infrastructure Assessment Framework

#### 2.0.1 Introduction

This chapter develops a new framework for assessing overall infrastructure conditions from available data and statistically modeling the relative risk of infrastructure inequity across neighborhood social-economic characteristics. As noted in Chapter 1, previous studies have either assessed a single type of infrastructure or several infrastructure types on a relatively small spatial scale: A systematic assessment of multiple infrastructure types across an entire city has yet to be implemented. To accomplish this objective, this study collects street-level data on multiple neighborhood infrastructure types and develops an innovative data-driven framework to comprehensively assess the condition of all types. Furthermore, given neighborhood social-economic characteristics, a risk model is developed to identify any infrastructure inequity across the city.

#### 2.0.2 Methodology

Fig. 2.1 shows the generalized data-driven framework developed in this work to assess multiple neighborhood infrastructure types at the neighborhood level quantitatively and to explore infrastructure inequity in urban settings. The framework consists of three primary components: 1) compute neighborhood infrastructure deficiency by aggregating the presence and condition of each infrastructure type from street to neighborhood level; 2) compare infrastructure deficiency across income levels to identify the existence of infrastructure deserts; and 3) identify infrastructure inequity using statistical models. Each component is discussed in more detail in subsection 2.0.2.1 through 2.0.2.3 below.

The framework given in Fig. 2.1 has several benefits: 1) It integrates street-level condition assessment with neighborhood-level social-economic characteristics such as income; 2) it can add new infrastructure types and still maintains its robustness; 3) The framework is highly

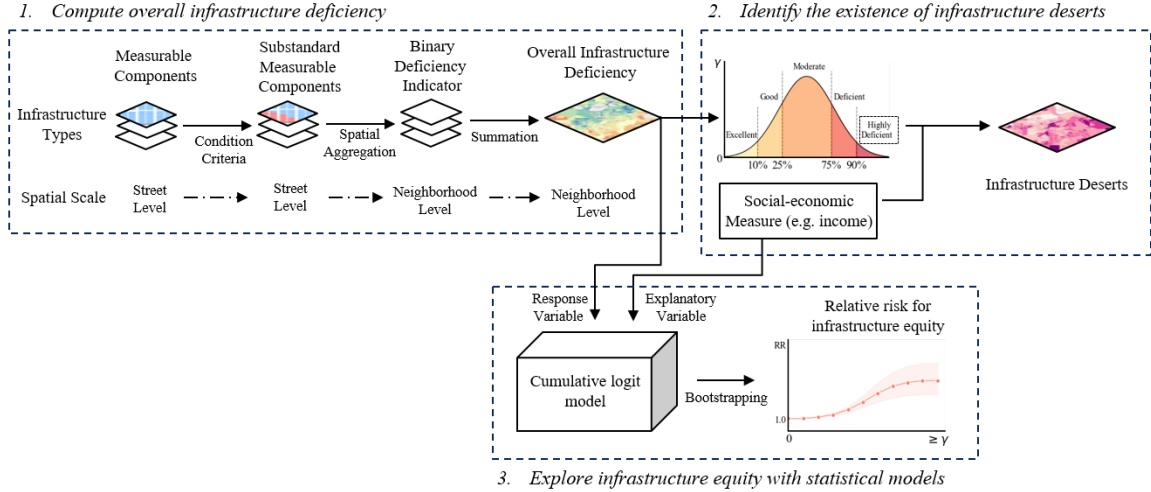


Figure 2.1: Overview of the infrastructure equity framework.

generalized and can be applied to other cities or regions with available data. The first step of the framework is to determine a proper spatial representation of neighborhoods. Ideally, the chosen representation should naturally represent the boundary residency as a neighborhood. To aid in generalization, the framework is applied with a consistent spatial representation of neighborhoods as Census block groups for two main reasons. First, the size of a block group (typically ranges from 500 to 1000 housing units) and the cartographic representation approximates the overall size and geometry of a neighborhood. Second, the Census block group is the smallest administrative boundary for which Census Bureau freely publishes sample data [66]. Therefore, the Census block group seamlessly aligns with the U.S. Census Bureau’s social-economic attributes.

Although the Census block group is one of the most popular geographic boundaries used to represent residential neighborhoods, the correlation between administrative (U.S. Census Bureau) and actual neighborhood boundary and the relevance of those boundaries in neighborhood infrastructure distribution remain inconsistent and unclear [67–69]. Because the administrative boundary was initially designed for data collection, tabulation, and dissemination of small-area data [66] instead of segregating residential neighborhoods. One remedy to this problem would be to use perceived, resident-defined neighborhood bound-

aries because it may better represent the neighborhood and neighborhood-based measures such as access to destinations, walking routes, or the number of residences. For example, Nextdoor, a hyperlocal social network service for neighborhoods, offers a more reliable and accurate neighborhood geometry using a crowd-sourcing mechanism that allows users to sketch or modify the neighborhood they currently live in [70]. However, despite better geographic representation, resident-defined boundaries can be affected by neighborhood reputation and can introduce bias in neighborhood-based studies; For example, residents might report living in positively perceived neighborhoods but exclude stigmatized areas [67]. Besides, resident-defined neighborhood boundaries do not have the spatial compatibility of the social-economic measures embedded in administrative boundaries. Therefore, connecting resident-defined boundaries to social-economic measures often results in improper assumptions or extra spatial interpolation, which introduces more bias to the system.

#### *2.0.2.1 Compute overall infrastructure deficiency by neighborhood*

The first step of the framework (shown in Step 1 of Fig. 2.1) examines each infrastructure type's condition and computes the overall neighborhood infrastructure deficiency. At the street level, metrics for measuring infrastructure condition vary across different infrastructure types and may vary in different cities. A neighborhood-level binary deficiency indicator ( $\gamma$ ) is used to aggregate from street level measures within the neighborhood to represent individual infrastructure deficiency. To compute the binary infrastructure deficiency indicator for one infrastructure type, any quantitatively measurable components are identified, equivalent to the condition metric, within a neighborhood. Undoubtedly, there could be various types of measurable components in the neighborhood, depending on the infrastructure type. The most commonly used measurable component is the physical appearance of the structure. For instance, street cracks and uneven surfaces usually indicate inadequate pavement conditions, so street segments with surface information provide measurable components to evaluate pavement conditions [40, 64]. For example, Pavement Condition Index (PCI) is a score-based measure to classify pavement condition, pavement segment are considered as poor if its PCI is less than 55 out of 100. Therefore, the neighborhood-level binary deficiency indicator is determined by the percentage of "poor" pavement segments according to PCI

scores.

Furthermore, if the infrastructure is a facility located in or outside of the neighborhood that operates as a service provider (e.g., hospitals, grocery stores), then the number of residential households not in proximity represents the how inadequate such accessible services are to residents and can be considered measurable infrastructure components. After identifying all measurable infrastructure components within the neighborhood, a substandard criterion is developed for each infrastructure type based on published studies or design manuals and used to compute any substandard measurable components.

Here a fraction number ( $\mu$ ) is defined to represent substandard measurable components as a percentage of measurable infrastructure components within the same neighborhood (see in Equation 2.1). It is computed as:

$$\mu = \frac{\text{substandard measurable infrastructure components}}{\text{total measurable infrastructure components}} \quad (2.1)$$

As shown in Equation 2.2, the binary infrastructure deficiency indicator ( $\theta$ ) equals 1 if at least half of the measurable components are substandard ( $\mu \geq 0.5$ ), denoting a deficient infrastructure type in a neighborhood. Otherwise  $\theta = 0$ .

$$\theta = \begin{cases} 0 & \text{if } \mu \geq 0.5 \\ 1 & \text{if } \mu < 0.5, \end{cases} \quad (2.2)$$

The benefit of using such indicators is to normalize all infrastructure measurements to the same scale of 0 or 1, which allows multiple binary indicators to be combined mathematically in later steps. The above procedure is repeated until  $\gamma$  is obtained for all infrastructure types and then the overall infrastructure deficiency ( $\gamma$ ) of a neighborhood is computed as the summation of  $\theta$  (Equation 2.3):

$$\gamma = \sum_{i \in \text{all infrastructure types}} \theta_i \quad (2.3)$$

The summation of multiple deficiency indicators into a single metric represents the overall neighborhood condition. Thus,  $\gamma$  ranges from zero to the total number of infrastructure

types considered. If a neighborhood does not have any deficient infrastructure types, it gets  $\gamma = 0$ . Finally, to aid in interpretation, a categorical representation of overall infrastructure deficiency is created based on the percentile of the resulting  $\gamma$ . As such, the resulting overall infrastructure deficiency values are defined as (1) *Excellent* ([0% 10%]), (2) *Good* ([10% 25%]), (3) *Moderate* ([25%-75%]), (4) *Deficient* ([75%-90%]) and, (5) *Highly Deficient* ([90%-100%]).

### 2.0.2.2 Test for existence of infrastructure deserts

The next step is to find neighborhoods that are both economically disadvantaged and significantly lacking in neighborhood infrastructure relative to wealthier neighborhoods. These areas are labeled "infrastructure deserts," analagous to "food deserts," which are defined as low-income neighborhoods with insufficient access to healthy food sources [55, 71, 72]; and "transit deserts," which are transit-dependent areas that lack adequate public transit service [29, 54]. The introduction of "infrastructure deserts" presents a more comprehensive and integrated perspective of neighborhood weakness in physical assets and community services. The category of Highly Deficient infrastructure condition from the previous step is chosen as the quantitative representation of neighborhoods as significantly more deficient in infrastructure presence and condition. Such areas that are low income are identified as infrastructure deserts.

To define neighborhood income category, neighborhoods are classified into three groups (*low*, *middle*, and *high*) using tertiles of annual median household income [73–75]. A few studies use annual median family income as a income variable [74, 76]. However annual median household income has richer historical data than family income since 2013 and has been used to interpolate missing income for some neighborhoods. [66] Moreover, both measures are highly correlated and should not significantly bias the resulting spatial patterns.

### 2.0.2.3 Estimate infrastructure inequity risk with cumulative logit models

Lastly, to account for any uncertainty within the observed data, statistical models can further explain the relationship between neighborhood infrastructure condition and income, as well as the significance of infrastructure inequity. Since the overall infrastructure deficiency is computed as an ordinal integer according to Equation 2.3, the cumulative logit model

(also called proportional odds model) [77] is appropriate for this case as it was designed for a response variable that takes values in a set of ordered categories (multiple ordinal responses). This model was initially proposed by [78] as an extension of the logistic regression model for binary responses.

In this study, the model relates a response variable  $\gamma$ , consisting of ordered categories (e.g., overall infrastructure deficiency), to a categorical explanatory variable (e.g., neighborhood income characteristics) with  $k + 1$  levels and represented by  $x$ , a vector of  $k$  dummy variables that represent  $k$  different levels (the remaining level is chosen as the reference level). The model has the following generalized representation:

$$\text{logit}[Pr(Y \leq j|x)] = \alpha_j + \beta^T x; \quad j = 1, \dots, J - 1 \quad (2.4)$$

where  $Pr(Y \leq j|x)$  is the cumulative probability of the event ( $Y \leq j$ ),  $\alpha_j$  are the unknown intercept parameters, and  $\beta^T = (\beta_1, \beta_2, \dots, \beta_k)$  is a vector of regression coefficients used for all response categories.  $J$  is the total number of response categories. *logit*, also known as the log-odds transformation, is the inverse function for the standard logistic cumulative distribution function:

$$\text{logit}(t) = \log \frac{t}{1 - t} \quad (2.5)$$

The model assumes the same effects  $\beta$  for each *logit*. Thus the regression coefficient vector,  $\beta$ , does not depend on  $j$ , implying that the log-odds ratio is proportional to the difference between two  $x$  values and shares the same proportionality constant regardless of  $j$ . This is also called the proportional odds assumption. The validity of this assumption can be checked based on a  $\chi^2$  score test [79].

Applying this model with overall infrastructure deficiency as a response variable and income level as a single explanatory variable results in:

$$\text{logit}[Pr(Y \leq j|x)] = \alpha_j + \beta_M x_M + \beta_H x_H; \quad j = 1, \dots, J - 1 \quad (2.6)$$

Where  $\gamma$  is the computed overall infrastructure deficiency with each value of integer representing one category,  $x_M, x_H$  are two dummy variables:  $x_M = 1$  if the income level is middle,

otherwise  $x_M = 0$ ;  $x_H = 1$  if income level is high, otherwise  $x_H = 0$ ;  $x_M = x_H = 0$  if the income level is low, serving as the reference level.  $J$  is the total number of infrastructure types considered ( $J = 12$  in this case study).  $\beta_M, \beta_H$  are regression coefficients for the dummy variables of the categorical covariate with three levels (*low, middle, high*).

Once the model is fitted properly and validated by performing a likelihood ratio test (*LRT*) between fitted model and the same model except using a multinomial link. With the null hypothesis that proportional odds assumption holds,  $p$ -value of greater than 0.05 indicates that the data does not show gross violation of the assumption, a relative risk measure of deficient infrastructure types ( $RR_{xj}$ ) is computed between different income levels to draw statistical conclusions. In particular, the relative risk of low-income neighborhoods having "more deficient ( $> j$ )" infrastructure types compared to neighborhoods with income-level denoted as  $x$  is written as:

$$\begin{aligned}
 RR_{xj} &= \frac{Pr(\gamma > j | low - income)}{PR(\gamma > j | x)} \\
 &= \frac{1 - Pr(\gamma \leq j | low - income)}{1 - Pr(\gamma \leq j | x)} \\
 &= \frac{1 + e^{\alpha_j} + \beta x}{1 + e^{\alpha_j}} \\
 &j = 1, \dots, J - 1, x \in \{M, H\}
 \end{aligned} \tag{2.7}$$

Relative risk offers adequate measures to compare overall infrastructure condition across different neighborhood income levels. Given the values of  $j$  and  $x$ , if the relative risk value ( $RR_{xj}$ ) is larger than one; then low-income neighborhoods have a higher risk of having more than  $j$  number of deficient infrastructure types than neighborhoods with income level  $x$ , also showing evidence of infrastructure inequity. To obtain the confidence intervals for relative risk at each  $j$ , a bootstrapping method [80] is used with 20,000 iterations to compute the upper (97.5%) and lower (2.5%) confidence level of the relative risk estimates. All of the statistical computations described herein are executed with the statistical software R [81]. The cumulative logit model is fit using the function *polr* from package *MASS* [82]. All coefficients were exported and visualized using Python.

## Chapter 3

### Case Studies

This chapter gives overviews and background information on the cities that are investigated in the following chapters, as well as corresponding datasets. The author implements the infrastructure framework in a total of four cities in the United States: Dallas-TX, Los Angeles-CA, Chicago-IL, and New York City-NY. Chapter 4 and Chapter 5 will primarily discuss infrastructure inequities discovered in Dallas, TX, the first case study that was examined with the most detail. Chapter 6 presents the generalization of the assessment framework by showing comparative results for the most critical analyses completed in Dallas, adding Los Angeles, New York City, and Chicago.

#### 3.0.1 Dallas, TX

This chapter chose one of the metropolises in the United States, Dallas, TX, USA, as the first case study for the methodology developed previously because of its ongoing economic development, significant infrastructure issues, and plans for redevelopment activities in the future to address infrastructure issues. Dallas is one of the Rockefeller Foundation’s 100 Resilient Cities; its resilience strategy was released in 2018 [83], which includes equity and neighborhood infrastructure investment as core goals. Currently, Dallas has the highest level of income inequality in the United States (U.S.) [84,85] and one of the highest rates of increase in urban heat among major US cities [86,87]. Furthermore, Dallas has the 4th highest number of pedestrian fatalities among U.S. counties in 2016 [88]. The city also rated significantly lower than the national average in street and infrastructure maintenance, according to a community survey [89]. These statistics highlight the existing neighborhood infrastructure issues and make Dallas an ideal location to study neighborhood-scale infrastructure equity.



### 3.0.2 Infrastructure Data

To assess neighborhood infrastructure in Dallas, a total of 12 neighborhood infrastructure types with available data were considered (pavement, sidewalk, crosswalk, noise wall, public transportation, trails access, medical facilities access, food access, community gathering places access, bank access, street tree canopy, and internet service). Measurable data for each infrastructure type were identified based on multiple types of data (tabular data, spatial lines, or spatial points) and related references and guidelines as shown in Table 3.1. The table compiles measurable components for all neighborhood infrastructure types considered and corresponding substandard criteria. For noise walls, since only households near highways could potentially be affected by the presence or absence of noise walls, the measurable components are restricted to only residential households within 200 feet away from major highways instead of all residential households [90]. Please refer to Appendix A.1, A.2, and A.3 for data source, steps on how to apply substandard criteria to each infrastructure type.

### 3.0.3 Neighborhood's Income

To represent neighborhood income, the annual median household income of Census block groups in the Dallas region was obtained from the 2018 U.S. Census table B19013. For block groups with missing income records, the average between historical information at the block group level (linear regression using the past five years' income records, as available) and current-year income at the tract level is used to fill in missing data. This method offers a better estimation for missing income data because it accounts for currency inflation over the years and impacts of nearby neighborhoods within the same Census tract. After filling in missing income records, a total of 790 neighborhoods across Dallas had complete income and infrastructure condition data. The neighborhood income was then categorized as low-income (347 neighborhoods), middle-income (205 neighborhoods), and high-income (238 neighborhoods) using tertiles across all of Dallas County. The cutoffs between income levels were \$44,100 for the 33<sup>rd</sup> percentile and \$70,200 for the 66<sup>th</sup> percentile. Income level is also used as a continuous variable in Chapters 5 and 6.

Table 3.1: Substandard criterion for neighborhood infrastructure types.

Infrastructure Type	Assessment Unit	Substandard Criteria	Criteria Reference	Source Reference
Pavement	Street segment	Pavement Condition Index (PCI) < 55	[64]	City of Dallas GIS Service
Sidewalks	Sidewalk Segment	Any existence of obstruction, damage, or missing segments	[39]	City of Dallas Public Works
Internet	Residential Unit	Internet speed less than 200 kbps in at least one direction	[91]	Federal Communication Commission
Crosswalks	Street Intersection	Missing crosswalks at Intersections with traffic lights or school zones	[92]	Google Static Map API
Noise wall	Residential Unit	Within 200 feet of the highway and no noise walls present	[90]	Google Streetview Static API
Street Tree Canopy	Street Segment	Average percentage of street segment covered by tree canopy less than 25%	[60]	The Trust for Public Land
Public Transportation Access	Residential Unit	Not within 5-min walking distance (0.4 km) of the bus stop or 10-min walking distance (0.8 km) of the rail station	[51, 93]	City of Dallas GIS Services
Medical Service Access	Residential Unit	Not within 2 miles (3.2 km) of major hospitals or 1-mile (1.6 km) of walk-in clinics or urgent care	[94, 95]	NCTCOG's Regional Data Center
Bike & Pedestrian Trails	Residential Unit	Not within 10-min walking distance (0.8 km)		City of Dallas GIS Services
Gathering Place	Residential Unit	Including parks, libraries, farmer markets and community centers. Not within 10-min walking distance (0.8 km)		Google Static Map
Food Access	Residential Unit	Nearby food stores not within a 1-mile distance (1.6 km)	[96]	City of Dallas GIS Services NCTCOG's Regional Data Center
Bank Access	Residential Unit	Nearby bank branches are not within a 1-mile distance (1.6 km)		City of Dallas GIS Services

### 3.0.4 Neighborhood’s Race-ethnicity and Age

Neighborhood-level race-ethnicity data is obtained from the American Community Survey’s Table B03002 for 2018. To combine both racial and ethnic information, which are documented as separate attributes by the Census Bureau, four race-ethnicity categories are derived from the ratio of each population group in each Census block group (hereafter denoted as a neighborhood): predominantly non-Hispanic White , predominantly non-Hispanic Black, predominantly Hispanic, and no predominant race-ethnicity. The “predominant” classification of each group is assigned if more than 60% of residents identified with that group [74]. For easier interpretation of the results, the author hereafter simplifies category names to *predominantly White*, *predominantly Black*, *predominantly Hispanic*, and *no predominant race-ethnicity*, respectively. To provide historical perspectives on infrastructure equity, each neighborhood’s average year of construction is computed by averaging the built-year of all residential buildings within the neighborhood. The year of construction for each building is obtained from the building’s footprint map in the city’s parcel database [97].

### 3.0.5 Historical Redlining Practices

To further explore the impacts of historical practices on infrastructure equity, discriminatory policies by the Home Owners’ Loan Corporation (HOLC) are considered. Redlining is a common term for the HOLC’s denial of credit insurance, healthcare, loans, mortgages, and other financial services based on a neighborhood’s demographic makeup. Redlining gets its name from the red outlines drawn around “high-risk” neighborhoods in maps created in the 1930s by HOLC, a New Deal agency formed to refinance mortgages during the Great Depression. Neighborhoods were labeled into four categories, shown in Fig. 3.1 for Dallas, to indicate the perceived level of risk for government-backed mortgages: *Best*, *Still desirable*, *Definitely declining*, and *Hazardous*.

Neighborhoods received HOLC mortgages based on percentages of their homes’ appraised values by category (80% for Grade A, 60%-80% for Grade B, 15% for Grade C, and 0% for Grade D). The assignment of the grades was driven by neighborhoods’ racial and ethnic composition, with the majority of “grade D” (redlined) areas having primarily Black households. Consequently, redlining has raised the level of racial and wealth inequity and caused

long-term impacts on real estate wealth accumulation that persist today [98, 99].

With the help of modern mapping software, Nelson [100] digitized and created downloadable online Web maps as ArcGIS Shapefile or GeoJSON) for the majority of the redlining areas across the United States from hand-drawn and scanned maps. It is noteworthy that when these redlining maps were first created in the 1930s, their boundaries did not follow the Census administrative boundaries and their total areas are much smaller than current city boundaries (Fig. 3.1). To spatially combine the maps with infrastructure data, a subset of neighborhoods (Census block groups) are identified whose centroids are within the redlining area. This resulted in 218 neighborhoods considered as historical redlining areas.

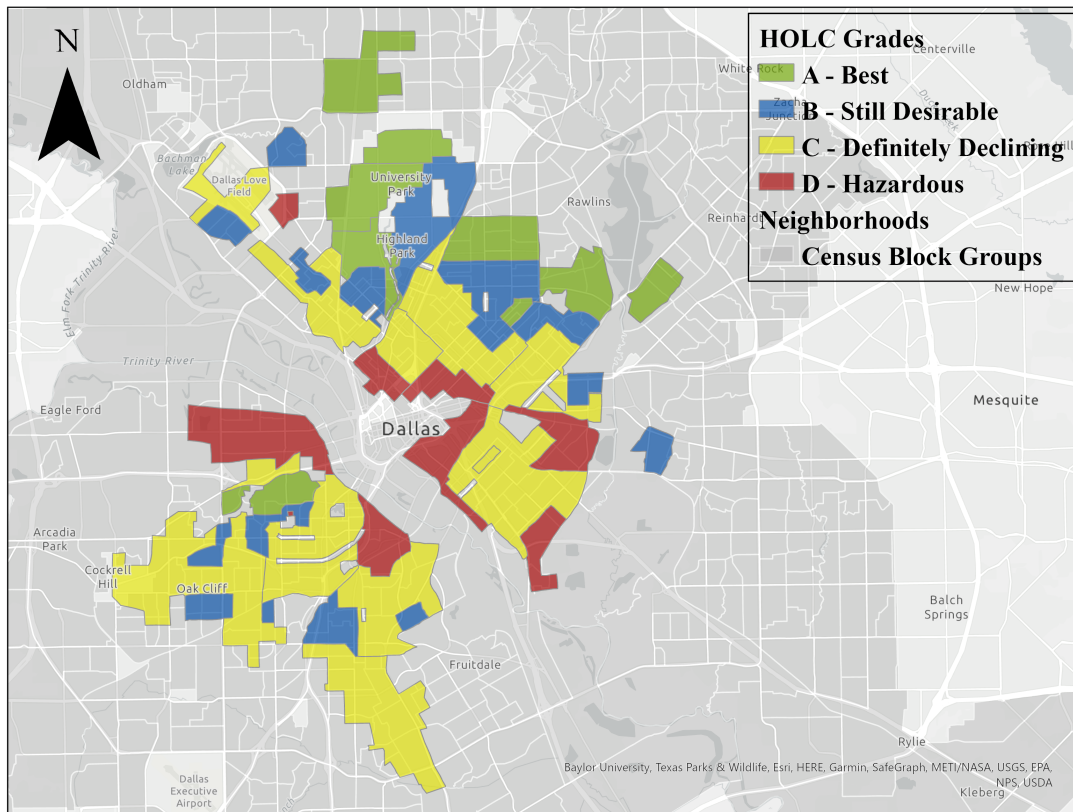


Figure 3.1: Historical Home Owner Loan Corporation Redlining Areas (1937).

Note that available infrastructure data only show a snapshot of current conditions and the lack of digitized historical infrastructure condition data hinders a full understanding of how the city’s overall infrastructure is changing in history. Nonetheless, the author may still find insightful associations between neighborhood infrastructure conditions and quantitative historical indicators such as redlining maps and neighborhoods’ average built-year. This can reveal the need for more laborious future studies to digitize and analyze the full set of available historical data.

### 3.0.6 Los Angeles, CA, New York City, NY, and Chicago, IL

In addition to Dallas, the author also apply the generalized framework to three other cities in Chapter 5: Los Angeles-CA, Chicago-IL, and New York City-NYC. The selection of cities is made such that: (1) they are geographically located in 4 regions (west, south, central north, and east) across the nation, which helps to represent the infrastructure condition in each major region of the country. (2) Those cities offer platforms such as Open Data Portal to allow most of the infrastructure-related datasets to be publicly accessible for infrastructure-related research [29, 44, 96, 101–103].

#### 3.0.6.1 Infrastructure datasets

To properly compare infrastructure conditions between cities, the author should assess common infrastructure types in each city. Despite the availability of data for a total of 12 infrastructure types in Dallas, due to difficulties in acquiring tree canopy and crosswalk data for other cities, the author consider 10 common infrastructure types across the four cities: *pavement*, *sidewalk*, *bank access*, *trail access*, *medical facility access*, *gathering place access*, *food access*, *internet service*, *noise walls*, and *public transportation access*. It is worth noting that certain datasets such as pavement and sidewalks do not have identical condition attributes since each city measures streets and collects data differently. To ensure that the assessment is as consistent as possible, the substandard criteria are adjusted accordingly for each city without significantly altering the criterion. The resulting substandard criteria for the four cities are also shown in Table 3.2.

Table 3.2: Substandard criteria for Los Angeles, New York City, Chicago, and Dallas (considering 10 infrastructure types).

Infrastructure Type	Los Angeles	New York City	Dallas	Chicago
Pavement	Pavement Condition Index (PCI) < 55	"Poor" rating based on National Performance Management Measures for Accessing Pavement Condition	Pavement Condition Index (PCI) < 55	"Poor" rating based on assessment of NYC streets by the agency
Sidewalk	Any existence of missing segments on residential streets (Condition data not available)	Any existence of missing segments on residential streets (Condition data not available)	Any existence of obstruction, damage or missing segments	Any existence of missing segments on residential streets (Condition data not available)
Internet	Internet speed less than 200 kbps in at least one direction	Internet speed less than 200 kbps in at least one direction	Internet speed less than 200 kbps in at least one direction	Internet speed less than 200 kbps in at least one direction
Noise Wall	Within 200 feet of highway and no noise walls present	Within 200 feet of highway and no noise walls present	Within 200 feet of highway and no noise walls present	Within 200 feet of highway and no noise walls present
Bank Access	Not within 1 miles distance (1.6 km)	Not within 1 miles distance (1.6 km)	Not within 1 miles distance (1.6 km)	Not within 1 miles distance (1.6 km)
Public Transportation Access	Not within 5-min walking distance (0.4 km) of bus stop or 10-min walking distance (0.8 km) of rail station	Not within 5-min walking distance (0.4 km) of bus stop or 10-min walking distance (0.8 km) of rail station	Not within 5-min walking distance (0.4 km) of bus stop or 10-min walking distance (0.8 km) of rail station	Not within 5-min walking distance (0.4 km) of bus stop or 10-min walking distance (0.8 km) of rail station
Medical Facility Access	Not within 2 miles (3.2 km) of major hospitals or 1-mile (1.6 km) of walk-in clinics	Not within 2 miles (3.2 km) of major hospitals or 1-mile (1.6 km) of walk-in clinics	Not within 2 miles (3.2 km) of major hospitals or 1-mile (1.6 km) of walk-in clinics or urgent care	Not within 2 miles (3.2 km) of major hospitals or 1-mile (1.6 km) of walk-in clinics
Bike&Pedestrian Trails	Not within 10-min walking distance (0.8 km)	Not within 10-min walking distance (0.8 km)	Not within 10-min walking distance (0.8 km)	Not within 10-min walking distance (0.8 km)
Gathering Place Access	Not within 10-min walking distance (0.8 km), including: parks, community centers, farmers markets, and libraries	Not within 10-min walking distance (0.8 km), including: parks, community centers, farmers markets, and libraries	Not within 10-min walking distance (0.8 km), including: parks, community centers, farmers markets, and libraries	Not within 10-min walking distance (0.8 km), including: parks, community centers, farmers markets, and libraries
Food Access	Nearby food stores (including farmers markets) not within 1 mile distance (1.6 km)	Nearby food stores (including farmers markets) not within 1 mile distance (1.6 km)	Nearby food stores (including farmers markets) not within 1 mile distance (1.6 km)	Nearby food stores (including farmers markets) not within 1 mile distance (1.6 km)

### 3.0.6.2 Income, race-ethnicity, historical redlining, and car ownership

Neighborhood income is computed using annual median household income from the same Census Table as mentioned in Section 3.0.3. Categorized 3-level income (*low*, *medium*, and *high*) is used to visualize the comparisons between cities but the continuous income record is used to fit the statistical models given in Chapter 6. Similar to income, neighborhood race-ethnicity is represented as the four predominant groups mentioned in Section 3.0.4 and data are obtained from the same Census Table B03002. For historical redlining areas, the redlining regions were downloaded from the same data source provided by [100]. In Chapter 6, the author also include car ownership (Table B25044 from American Community Survey) as a

mobility indicator to discuss the impacts of highly deficient infrastructure on neighborhoods with less vehicle access.

## Chapter 4

### Do Infrastructure Deserts Exist? Assessment and Statistical Modeling of Neighborhood Infrastructure in Dallas, Texas

#### 4.0.1 Introduction

This chapter applies the infrastructure assessment framework defined in Chapter 2 to the City of Dallas as the first case study described in Chapter 3. Categories of infrastructure deficiency were then allocated to each neighborhood as follows: *Excellent* ( $\gamma \leq 3$ ), *Good* ( $\gamma = 4$ ), *Moderate* ( $5 \leq \gamma \leq 6$ ), *Deficient* ( $\gamma = 7$ ), and *Highly deficient* ( $\gamma \geq 8$ ). Following the definition of infrastructure deserts, low-income neighborhoods with highly deficient infrastructure ( $\gamma \geq 8$ ) were then identified across the city. The results reveal the existence of infrastructure deserts, low-income areas with significantly more deficient infrastructure types than higher-income areas, and show a significant pattern of infrastructure inequity. The author discusses the detailed findings in the following subsections.

#### 4.0.2 Individual and Overall Infrastructure Condition

Fig. 4.1 shows the percentage of neighborhoods with deficiencies for each individual type of infrastructure by income level. This distribution of deficient infrastructure exhibits three distinct patterns by infrastructure type: 1) For *crosswalks*, *internet service*, *medical facility*, *noise walls*, and *food access*, the share of neighborhoods with deficient infrastructure is much higher in low-income neighborhoods than others, showing a decreasing trend with increasing income; 2) For *pavement*, *sidewalks*, *community gathering places*, *trail access*, and *street tree canopy*, the share of deficient infrastructure does not show much difference across the three income groups; 3) For *public transit*, an increasing trend exists with deficient infrastructure versus income level.

Some of the results are consistent with previous findings, which show inequities across community's socio-economic status for individual infrastructure, such as *crosswalk* [59, 104],



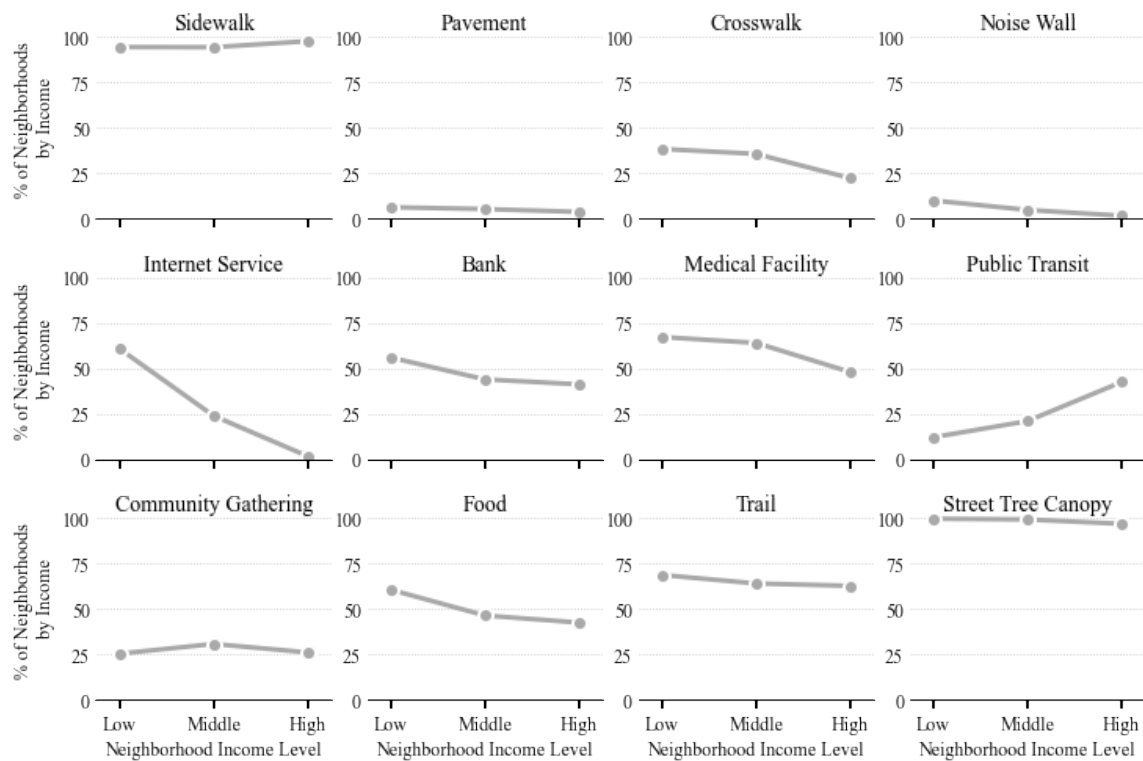


Figure 4.1: The percentage of individual deficient infrastructure types across neighborhood income levels .

*internet service* [63], and *food access* [73]. However, high-income neighborhoods experience more deficiency than low-income neighborhoods for *public transit* and, to some extent, *sidewalks*. This finding is not consistent with the literature [26] and may be due to the higher percentage of vehicle ownership in high-income neighborhoods.

These types of mixed relationships between infrastructure types and neighborhoods' socio-economic status introduce challenges to studying infrastructure equity by individual infrastructure type. This illustrates the need to consider multiple infrastructure types simultaneously and to develop a multi-infrastructure framework with an overall infrastructure deficiency metric.

Fig. 4.2 (a) shows a histogram of overall infrastructure deficiency as a percentage of whole neighborhoods; infrastructure deficiency categories are also represented by color. Fig. 4.2 (b) shows the distributions of overall infrastructure deficiency by income level; the y-axis

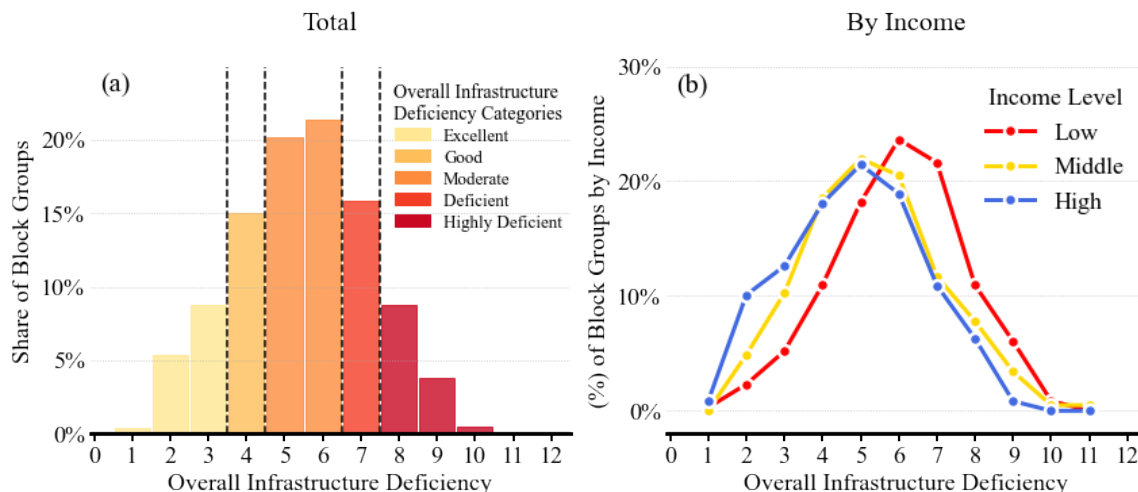


Figure 4.2: The percentage of individual deficient infrastructure types across neighborhood income levels .

represents the number of neighborhoods as a percentage of neighborhoods with the same income level. The results suggest that the overall infrastructure deficiency ranges from 1 to 11, meaning that all neighborhoods have at least one deficient infrastructure type. None of the neighborhoods is deficient in all infrastructure types (12 types in total). The results also show that the majority of neighborhoods have between 4 and 7 deficient infrastructure types.

Overall, 14% of the neighborhoods (114) are classified as *Excellent* for their overall infrastructure condition, while 13% of neighborhoods (104) are *Highly deficient*. As suggested in Fig. 4.2 (b), middle-income neighborhoods show a similar distribution to high-income neighborhoods, except that there are more high-income neighborhoods with very few infrastructure deficits (*Excellent*). However, the figure clearly shows that low-income neighborhoods exhibit higher overall infrastructure deficiency than other neighborhoods, as the distribution is horizontally shifted towards the direction of *Highly deficient* (8 or more deficient infrastructure types). This pattern reveals evidence of inequitable infrastructure between low-income neighborhoods and others.

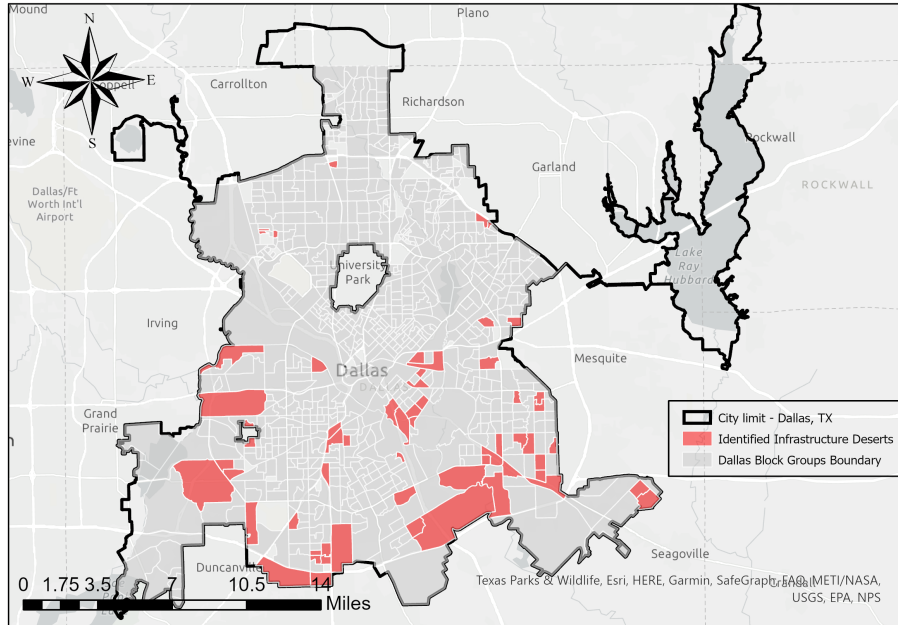


Figure 4.3: Infrastructure deserts identified in Dallas based on infrastructure assessment framework.

#### 4.0.3 Infrastructure Deserts

Fig. 4.3 shows the map of infrastructure deserts (low-income neighborhoods with highly deficient infrastructure ( $\gamma \geq 8$ )) in Dallas. A total of 62 neighborhoods were identified as infrastructure deserts. The infrastructure deserts have a clear spatial pattern overlapping with low-income areas located in the city's southern region [105], as opposed to upper-income areas that are more prevalent in the north. As further comparison of infrastructure deserts versus other areas, Fig. 4.4 shows individual deficient infrastructure types as a percentage of neighborhoods citywide versus within infrastructure deserts. It suggests that more than half of the neighborhoods have inadequate infrastructure among street tree canopy, sidewalk, noise wall, trail access, medical facility access, and food access. However, substantially more neighborhoods suffer from these deficiencies within infrastructure deserts. Besides, more than half of neighborhoods within infrastructure deserts have deficient crosswalks and access to banks, internet services, and gathering places. However, street tree canopy and sidewalks are the most widespread deficient infrastructure types.

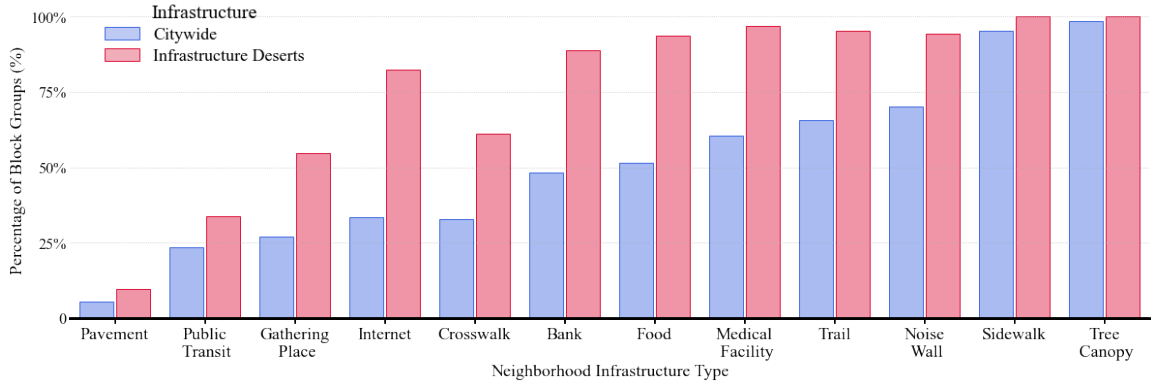


Figure 4.4: Deficient infrastructure as a percentage of block groups by individual infrastructure type.

#### 4.0.4 Relative Risk and Infrastructure Inequity

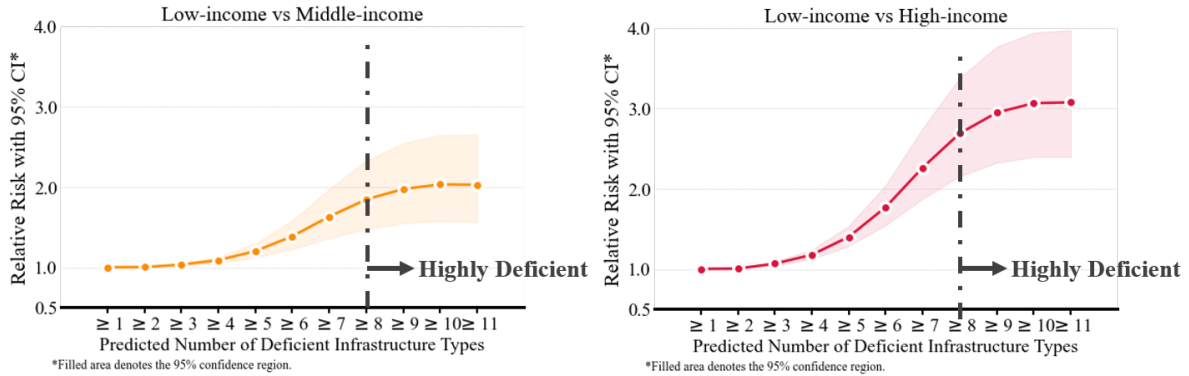


Figure 4.5: Relative risk: Computed relative risk is shown as circles, and shaded regions denote the upper (95%) and lower (5%) confidence limits. (a) The relative risk of overall infrastructure deficiency between low-income and middle-income areas; (b) Relative risk of overall infrastructure deficiency between low-income and high-income areas.

The estimated parameters for the fitted cumulative logit model are shown in Table A.5. The positive coefficients for  $(\beta_M, \beta_H)$  indicate a tendency for overall infrastructure deficiency

Table 4.1: Estimated coefficients of the cumulative logit model. The model assumption (proportional odds) is validated by performing a likelihood ratio test (16 degrees of freedom) between fitted model and the same model except using a multinomial link. With the null hypothesis that proportional odds assumption holds, p-value of 0.678 indicates that the data does not show gross violation of the assumption.

	Value	Std. Error	<i>t</i> value
Coefficients			
$\beta_M$	0.714	0.157	4.558
$\beta_H$	1.124	0.153	7.364
Intercepts ( $\alpha_j$ )			
$\alpha_1$	-6.196	0.582	-10.641
$\alpha_2$	-3.430	0.178	-19.235
$\alpha_3$	-2.380	0.134	-17.796
$\alpha_4$	-1.436	0.113	-12.713
$\alpha_5$	-0.534	0.103	-5.209
$\alpha_6$	0.427	0.102	4.197
$\alpha_7$	1.450	0.121	12.022
$\alpha_8$	2.651	0.182	14.585
$\alpha_9$	4.645	0.452	10.283
$\alpha_{10}$	6.261	1.002	6.248
Residual Deviance	3088.505	AIC	3223.505

to become smaller (less deficient) for middle-income and high-income neighborhoods compared to low-income neighborhoods. The estimated coefficient for the middle-income neighborhoods ( $\beta_M$ ) is 0.714, and the estimated coefficient for high-income neighborhoods ( $\beta_H$ ) is 1.124. These mean that the tendency of overall infrastructure deficiency toward less deficient appears to be stronger for high-income neighborhoods than middle-income neighborhoods. To test the model assumption of proportional odds with these parameters, a likelihood ratio test (16 degrees of freedom) was performed between the fitted model and the same model with a multinomial link. With the null hypothesis that proportional odds assumption holds, a p-value of 0.678 was computed, which indicates that the data do not show gross violation of the assumption. Fig. 4.5 shows the resulting relative risks: (1) between low-income and high-income neighborhoods; (2) between low-income and middle-income neighborhoods. The *x*-axis denotes the overall infrastructure deficiency to be equal or greater than displayed

ticks. The  $y$ -axis represents the value of relative risk estimates where mean results are plotted as lines and 95% confidence levels denoted by the shaded regions. As indicated in Fig. 4.5, the positive values of relative risk for both scenarios suggest that low-income neighborhoods show a greater risk of having "more" deficient infrastructure than middle and high-income neighborhoods. Furthermore, the relative risk increases for both scenarios as overall infrastructure deficiency increases. More specifically, low-income neighborhoods are 2.04 ~ 3.53 times more likely to have highly deficient infrastructure ( $\gamma \geq 8$ ) than high-income neighborhoods; and 1.42 ~ 2.44 times more likely to have highly deficient infrastructure ( $\gamma \geq 8$ ) than middle-income neighborhoods. Such substantial differences suggest significant infrastructure inequities across income levels for most types of infrastructure.

#### 4.0.5 Robustness of Statistical Model

Table 4.2: Cumulative logit model parameters using continuous income.

		Value	Std. Error	$t$ value
Coefficients				
	Log Income ( $\beta$ )	0.670	0.105	6.404
Intercepts ( $\alpha_j$ )				
	$\alpha_1$	-12.941	1.292	-10.014
	$\alpha_2$	-10.163	1.163	-8.741
	$\alpha_3$	-9.117	1.153	-7.906
	$\alpha_4$	-8.182	1.146	-7.141
	$\alpha_5$	-7.294	1.139	-6.406
	$\alpha_6$	-6.349	1.132	-5.609
	$\alpha_7$	-5.340	1.130	-4.725
	$\alpha_8$	-4.145	1.136	-3.648
	$\alpha_9$	-0.215	1.207	-1.780
	$\alpha_{10}$	-0.532	1.502	-0.354
Residual Deviance	3104.692		AIC	3126.692

To further confirm the association between neighborhood income and overall infrastruc-

ture deficiency, the model was refit using continuous (log) income instead of categorical income levels (Table A.5). The use of log income helps linearize the exponentially growing trends and remain unbiased compared to linear income [106]. Table 4.2 shows the estimates of model parameters. As the log income increases, the positive estimated coefficient shows that the overall infrastructure deficiency has a trend to be "less" deficient, which corroborates the previous findings using categorical income data. Fig. 4.6 shows the predicted probability of overall infrastructure deficiency by different income percentiles (5<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, 95<sup>th</sup>). Note that the probability curve shifts to the direction of "more" deficient with decreased neighborhood income, again showing a tendency to have more deficient infrastructure types for lower-income neighborhoods. This trend agrees with the earlier findings that lower-income neighborhoods have a significantly higher risk of greater infrastructure deficiency than other neighborhoods and, meanwhile, show the model’s robustness using either continuous or categorical income data.

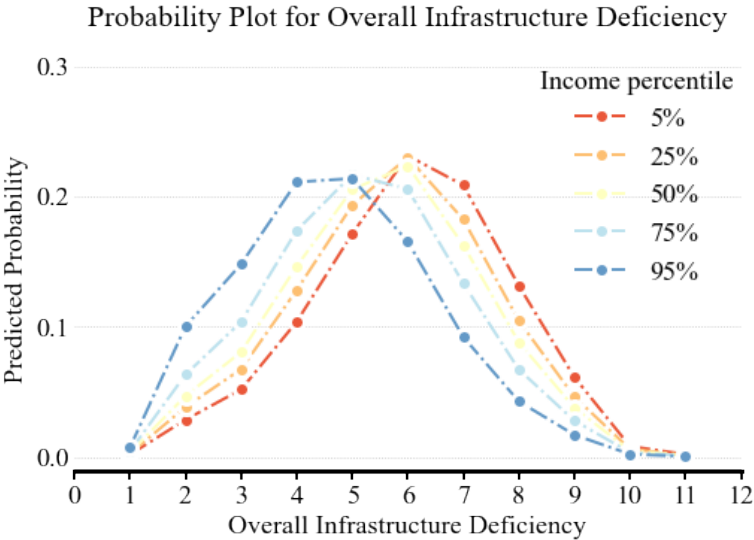


Figure 4.6: Infrastructure deserts identified in Dallas based on infrastructure assessment framework.

#### 4.0.6 Conclusions

Given a wide variety of physical attributes within a neighborhood and their inter-dependent interactions, assessing neighborhood infrastructure conditions can be highly challenging. It involves a substantial set of neighborhood infrastructure condition indicators that are multidimensional and heavily data-dependent. To the author's knowledge, there is a lack of approaches or frameworks in the existing neighborhood infrastructure-related literature that consider the diversity of neighborhood infrastructure and study multiple types of infrastructure together. This chapter contributes a novel approach to assessing neighborhood infrastructure conditions by systematically measuring multiple infrastructure types and statistically analyzing infrastructure inequity across neighborhood income characteristics. A critical strength of this chapter is the systematic and street-level assessment of multiple neighborhood infrastructure types. The introduction of binary infrastructure indicators and overall infrastructure deficiency effectively integrates multiple infrastructure types and provides a straightforward and intuitive neighborhood-level representation of infrastructure issues. Furthermore, with the new concept of "infrastructure deserts" – low-income areas with substantially higher infrastructure deficiency – the case study in Dallas, TX showed the existence of infrastructure deserts and infrastructure inequity throughout low-income areas. The statistical analyses also show that the observed infrastructure inequities between low-income and higher-income neighborhoods are statistically significant. In the next chapter, the author continues to explore infrastructure equity in Dallas from another perspective, considering neighborhoods' racial and ethnic composition and indicators that reflect the influence of historical housing policies.



## Chapter 5

### Are Infrastructure Deserts Correlated with Neighborhood Race and Ethnicity?

#### 5.0.1 Introduction

This chapter assesses and investigates the condition and equity of neighborhood infrastructure from the perspective of socio-demographic indicators and historical practices. Prior studies have shown that *predominantly White* neighborhoods have better access to food stores, medical resources, and community facilities than predominantly Black neighborhoods [103, 107, 108]. However, none of those studies have examined the role of race and ethnicity in neighborhoods as integrated, multi-infrastructure systems. This chapter extends the infrastructure equity framework developed in Chapter 2 to consider race, ethnicity, and historical data, in addition to their connection with neighborhoods' income characteristics.

The infrastructure equity study presented in Chapter 3 considered neighborhood median annual household income as the sole explanatory variable for infrastructure equity measure. However, several other socio-demographic indicators may correlate with neighborhood infrastructure conditions, such as race and ethnicity. Areas with majority Non-White populations receive inequitable resources and present racial disparity in various aspects including less access to infrastructure facilities [health care services [109, 110], parks [107], urban green space [103], and energy resources [111]], higher exposure to environmental risks such as flooding and air pollution [112, 113], and historically disproportionately fewer infrastructure investments and redevelopments [101, 114, 115].

These findings are drawn from a variety of case studies that primarily emphasize one or a few infrastructure types, but previous work has not considered how racial-ethnic characteristics of a neighborhood relate to systematic measures of infrastructure condition across multiple infrastructure types. This chapter addresses this gap by extending the infrastructure equity framework developed in Chapter 2 to consider the role of neighborhood race-

ethnicity in infrastructure inequities. Furthermore, the author examines historical influences on current neighborhood infrastructure conditions, including historical redlining regions and neighborhood age.

More specifically, the following major research questions are addressed: I) Do infrastructure inequities exist among neighborhoods with different racial and ethnic populations, as well as income levels? That is, do neighborhoods with certain predominant race-ethnicities and incomes have higher infrastructure deficits than others? II) Do historical discriminatory lending policies or neighborhood ages correlate with current infrastructure conditions?

### 5.0.2 Methodology

To answer the research questions posed above, several cumulative logit models were built to examine the relative risk of having deficient infrastructure across neighborhoods with different income and race-ethnicity combinations.

For the first analysis, the author explores the relationship between overall infrastructure deficiency and neighborhood's race-ethnicity by only considering a single explanatory variable (race-ethnicity group) in a cumulative logit model. Next, the author further investigates how both income and race-ethnicity interact with infrastructure deficiency via model selection from a full cumulative logit model containing both explanatory variables. In the third analysis, neighborhood age and historical information are used to reveal any legacies from discriminatory practices in past decades on current infrastructure equity. A statistical metric, Gamma statistic, is also used to show the statistical significance of any observed trends. For any trends observed in the descriptive analysis, the author computes Gamma statistic to provide their statistical significance. The following subsections provide details on each of these steps in the methodology.

#### 5.0.2.1 *Cumulative logit model*

Since the overall infrastructure deficiency ( $\gamma$ ) is computed as an ordinal integer, the author continues to use the cumulative logit model (also called proportional odds model) [77] in this study, which is designed for a response variable with values in a set of ordered categories. Chapter 2 discussed the several benefits of using the cumulative logit model to study the relationships between explanatory variables and ordinal categorical responses.

### 5.0.2.2 Overall infrastructure deficiency $\sim$ race-ethnicity

To understand the role of race-ethnicity in infrastructure equity, the author first builds a model only between overall infrastructure deficiency and neighborhood race-ethnicity characteristics. Applying the model with overall infrastructure deficiency as a response variable and race-ethnicity as a single explanatory variable results in:

$$\text{logit}[\text{Pr}(\gamma \leq j|x)] = \alpha_j + \beta_W x_W + \beta_H x_H + \beta_B x_B; \quad j = 1, \dots, J - 1 \quad (5.1)$$

where  $\gamma$  is the computed overall infrastructure deficiency for one neighborhood and  $\alpha_j$  is the intercept coefficients, where  $j$  is the number of deficient infrastructure types;  $x_W, x_H, x_B$  are three dummy variables.  $x_W = 1$  if the neighborhood is *predominantly White*, otherwise  $x_W = 0$ ;  $x_H = 1$  if the neighborhood is classified as *predominantly Hispanic*, otherwise  $x_H = 0$ ; similarly,  $x_B = 1$  if the neighborhood is classified as *predominantly Black*.  $x_W = x_H = x_B = 0$  if the neighborhood has *no predominant race-ethnicity*, also serving as the reference level in the model.  $J$  is the maximum observed number of deficient infrastructure types considered ( $J = 11$  out of 12 in this chapter).  $\beta_W, \beta_H, \beta_B$  are regression coefficients for the dummy variables of the categorical race-ethnicity covariate with four classes (*predominantly White, predominantly Hispanic, predominantly Black, no predominant race-ethnicity*).

After the model is fit, the model is validated using the Chi-square test by performing a likelihood ratio test (LRT) between the fitted model and the same model except using a multinomial link. With the null hypothesis that proportional odds assumption holds, a  $p$ -value of greater than 0.05 indicates that the data do not show a gross violation of the assumption. As suggested in Chapter 2, a relative risk measure of deficient infrastructure types ( $RR_{xj}$ ) is computed across the neighborhood race-ethnicity groups to draw statistical conclusions about infrastructure equity. In particular, the relative risk of neighborhoods with predominant race-ethnicity groups (denoted as  $x$ ) having "more deficient ( $> j$ )" infrastructure types compared to neighborhoods with *no predominant race-ethnicity* is written

as:

$$\begin{aligned}
RR_{xj} &= \frac{Pr(\gamma > j|x)}{PR(\gamma > j|no\ predominant\ race-ethnicity)} \\
&= \frac{1 - Pr(\gamma \leq j|x)}{1 - Pr(\gamma \leq j|no\ predominant\ race-ethnicity)} \\
&= \frac{1 + e^{\alpha_j}}{1 + e^{\alpha_j + \beta_x}} \\
j &= 1, \dots, J - 1, x \in \{W, H, B\}
\end{aligned} \tag{5.2}$$

Relative risk offers adequate measures to compare overall infrastructure conditions across neighborhoods with different predominant race-ethnicity. Given the values of  $j$  and  $x$ , if the relative risk value ( $RR_{xj}$ ) is larger than one, then neighborhoods with predominant race-ethnicity group  $x$  have a higher risk of having more than  $j$  deficient infrastructure types than neighborhoods with no predominant race-ethnicity, indicating evidence of infrastructure inequity. To obtain the confidence intervals for relative risk at each  $j$ , a bootstrapping method [80] is conducted with 20,000 iterations to compute the upper (97.5%) and lower (2.5%) confidence levels of the relative risk estimates.

### 5.0.2.3 Overall infrastructure deficiency $\sim$ income, race-ethnicity

Next, a second cumulative logit model is built including both income and race-ethnicity versus infrastructure deficiency. Unlike the implementation in Chapter 4, where income was grouped into three levels (*low*, *middle*, *high*), the neighborhood's average annual median household income is used as a continuous variable with a logarithm transformation. To establish the final model, the full model (including all interaction terms as shown in Equation 5.3) is first fit and then backward model selection [77] is performed to remove any insignificant terms using Akaike Information Criterion (AIC) as the dropping criteria. To better interpret the results, the author estimates and compares the probability of having highly deficient infrastructure ( $\gamma \geq 8$  for this case study) given different combinations of income and race-ethnicity composition. This allows statistical identification of the neighborhoods that have the highest number of deficient infrastructure types, which is one of the two criteria used to

identify infrastructure deserts in Chapter 2.

$$\begin{aligned} \text{logit}[Pr(\gamma \leq j|x)] &= \alpha_j + \beta_I x_I + \beta_W x_W + \beta_H x_H + \beta_B x_B + \\ &\quad \beta_{IW} x_I x_W + \beta_{IH} x_I x_H + \beta_{IB} x_I x_B; \\ &\quad j = 1, \dots, J - 1 \end{aligned} \tag{5.3}$$

#### 5.0.2.4 Gamma statistic

Gamma statistic [116] is a correlation statistic for ordinal data based on the number of concordant and discordant pairs among two variables. The use of Gamma statistics for correlation analysis is recommended when data have tied observations [77]. Given  $n$  observations, the number of concordant pairs  $P$  among two variables  $X$  and  $Y$  is:

$$P = |\{i, j\} : \{1 \leq i \leq j \leq n, (x_i - x_j)(y_i - y_j) > 0\}| \tag{5.4}$$

Similarly, the number of discordant pairs  $Q$  can be written as:

$$Q = |\{i, j\} : \{1 \leq i \leq j \leq n, (x_i - x_j)(y_i - y_j) < 0\}| \tag{5.5}$$

The *Gamma statistic* denoted as  $\tau$  in this thesis is therefore defined as:

$$\tau = \frac{P - Q}{P + Q} \tag{5.6}$$

Like other correlation measures, *Gamma statistic* treats the variables symmetrically, and it has a range of  $-1 \leq \tau \leq 1$ . The absolute value of  $\tau$  equals 1 when the relationship between  $X$  and  $Y$  is perfectly linear. When  $\tau = 1$ , the linear relationship has a monotone increasing trend, versus a monotone decreasing trend with  $\tau = -1$ . It is noted that independence implies  $\tau = 0$ , but the converse is not true because a U-shaped joint distribution can also lead to  $\tau = 0$ .

This chapter uses *Gamma statistic* to examine the associations between neighborhood

infrastructure deficiency and any ordinal metric. For example, the statistic identifies whether infrastructure deficiency trends across current race-ethnicity categories are statistically significant. To compute the *Gamma statistic*, a contingency table is first created between two ordinal factors: One of the ordinal factors, in this study, is always overall infrastructure deficiency. Another factor is either the redlining Home Owner Loan Corporation (HOLC) grades or race-ethnicity categories depending on the analysis (called comparing factor). Finally, the author does not report Gamma statistics if the subgroup has zero observations across any categories in the comparing factor.

### 5.0.3 Results and Discussion

Using the overall infrastructure deficiency computed from the infrastructure assessment framework, the author overlays it with the neighborhood's race-ethnicity. Several cumulative models mentioned above are fitted to investigate the role of race-ethnicity and historical lending policies to infrastructure equity in addition to income characteristics. In the result section, the correlation between individual infrastructure conditions and race-ethnicity is firstly examined. Secondly, statistical inference is conducted considering only race-ethnicity and overall infrastructure deficiency to describe the relative risks of neighborhoods with certain predominant race-ethnicity groups. Furthermore, a more complicated statistical model is introduced including both race-ethnicity and income characteristics to examine their interaction effects. Finally, this thesis shows how the historical redlining policies influence the current infrastructure conditions by comparing the neighborhood's age and groups of neighborhoods that were inside historical redlining regions.

#### 5.0.3.1 Individual infrastructure condition

The author first looks at the percentage of neighborhoods with deficiencies for each type of infrastructure by race-ethnicity groups, as shown in Fig. 5.1. For consistent interpretation, the author uses the order of (*predominantly White, no predominant race-ethnicity, predominantly Hispanic, and predominantly Black*) in all figures to describe the distribution of infrastructure conditions. Three primary patterns are seen, arranged as three separate rows in Fig. 5.1. In the first row, a declining trend is observed across race-ethnicity, meaning that *predominantly White* neighborhoods have on average the worst *public transit access* to

all neighborhoods. This finding is consistent with the finding in Chapter 4 that high-income neighborhoods, which are more likely to be *predominantly White*, have a higher percentage of substandard *public transit access*. In the second row of Fig. 5.1, the author observes five

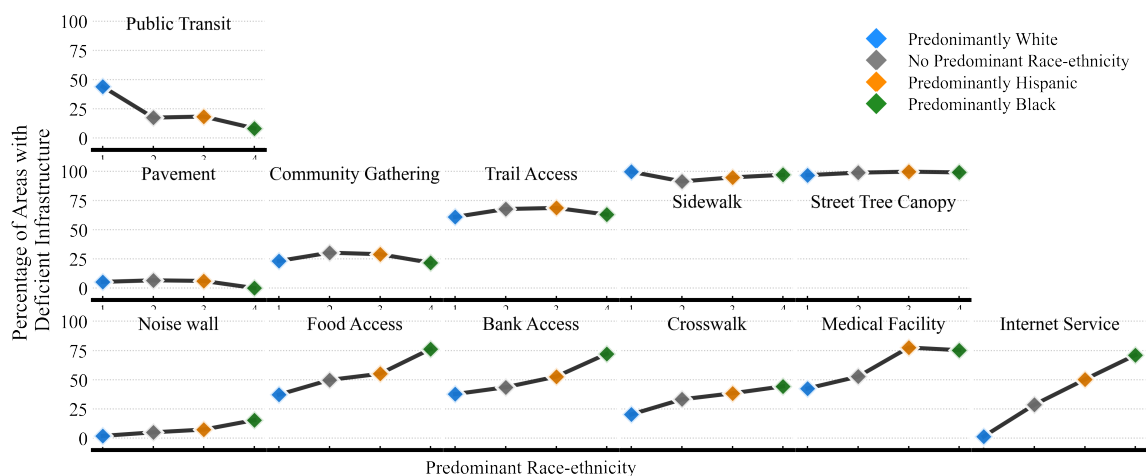


Figure 5.1: Percentage of deficient infrastructure by neighborhood income level and infrastructure type.

infrastructure types (*pavement*, *gathering place access*, *trail access*, *sidewalk*, and *street tree canopy*) that have no clear trend in levels of deficiencies. Finally, *predominantly Hispanic* and *Black* neighborhoods tend to have higher percentages of deficiencies in six infrastructure types shown in the third row (*noise wall*, *food access*, *bank access*, *crosswalk*, *medical facility access*, and *internet service*). This means that *predominantly White* neighborhoods have lower percentages than the other neighborhoods for the same infrastructure types, indicating an “increasing” trend across the charts in the bottom row.

### 5.0.3.2 Infrastructure inequity across race-ethnicity groups

Fig. 5.2(a) shows the histogram of overall infrastructure deficiency as a percentage of city-wide neighborhoods and Fig. 5.2(b) shows the distribution of overall infrastructure deficiency by predominant race-ethnicity groups. The  $y$ -axis represents the number of neighborhoods

as a percentage of neighborhoods with the same race-ethnicity category. It can be seen that *predominantly Black* neighborhoods have the highest average infrastructure deficiency, *predominantly Hispanic* neighborhoods have the second-highest average deficiency, and *predominantly White* neighborhoods have the lowest average deficiency score. Neighborhoods with *no predominant race-ethnicity* have an infrastructure deficiency score between *predominantly Hispanic* and *predominantly White* neighborhoods (second lowest). Furthermore, the race/ethnicity distributions in Fig. 5.2(b) are skewed to the right, indicating that the same pattern holds for the highest numbers of infrastructure deficiencies *predominantly Black* > *Hispanic* > *no predominant race-ethnicity* > *White*.

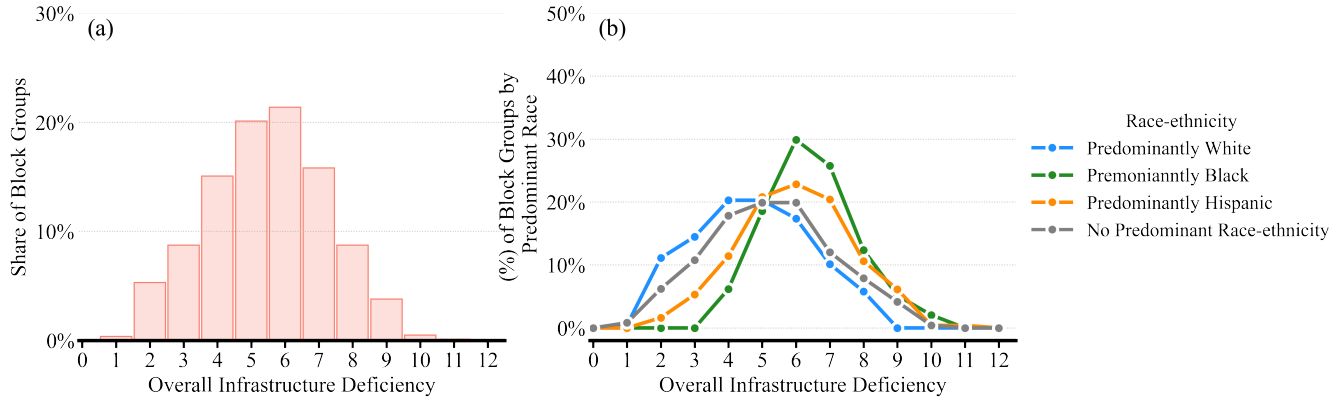


Figure 5.2: Overall infrastructure deficiency. (a) Histogram of overall infrastructure deficiency as a percentage of block groups, (b) Histogram of overall infrastructure deficiency as a percentage of block groups by neighborhood’s race-ethnicity.

To statistically investigate the inequities shown in Fig. 5.2(b), a simple cumulative model between overall deficiency and race-ethnicity (Eqn. 5.1) is fit with the parameters estimated as in Table 5.1. The positive coefficients for  $(\beta_W, \beta_H, \beta_B)$  indicate a tendency for overall infrastructure deficiency to become lower (less deficient) for neighborhoods that have predominant race-ethnicity groups in the order *predominantly White*, *Hispanic*, and *Black* and the values indicate the strength of the trend. For example, the estimated coefficient



for *predominantly Hispanic* neighborhoods ( $\beta_H$ ) is -0.739 and the coefficient for *predominantly Black* neighborhoods ( $\beta_B$ ) is -1.159, suggesting that these neighborhoods are likely to have more deficient infrastructure types compared to neighborhoods with *no predominant race-ethnicity* and *predominant Black* neighborhoods are likelier than *predominantly Hispanic* neighborhoods. The magnitude of the estimated coefficients also indicates that the tendency of overall infrastructure deficiency toward more deficient (higher deficiency) appears to be stronger for *predominantly Black* neighborhoods than *predominantly Hispanic* neighborhoods. On the other hand, the estimated coefficient for *predominantly White* neighborhoods ( $\beta_W$ ) is 0.714, showing that *predominantly White* neighborhoods have a tendency towards fewer deficiencies compared to neighborhoods with *no predominant race-ethnicity*. These statistics also align with the patterns observed in Fig. 5.2(b) and described above.

Table 5.1: Estimated coefficients of the cumulative logit model (race-ethnicity only). Likelihood Ratio Test (27 degrees of freedom) was conducted between the fitted model and the same model with a multinomial link. With the null hypothesis that proportional odds assumption holds, a p-value of 0.075 indicates that the data do not show a significant violation of the assumption.

	Value	Std. Error	<i>t</i> value
Coefficients			
Race-ethnicity			
$\beta_W$	0.53	0.17	3.13
$\beta_B$	-1.16	0.21	-5.55
$\beta_H$	-0.74	0.16	-4.54
Intercepts ( $\alpha_j$ )			
$\alpha_1$	-5.51	0.59	-9.38
$\alpha_2$	-2.72	0.18	-14.93
$\alpha_3$	-1.65	0.14	-11.81
$\alpha_4$	-0.69	0.12	-5.51
$\alpha_5$	0.24	0.12	2.00
$\alpha_6$	1.22	1.13	9.41
$\alpha_7$	2.26	0.15	15.01
$\alpha_8$	3.46	0.20	16.94
$\alpha_9$	5.46	0.46	11.82
$\alpha_{10}$	7.08	1.01	7.03
Residual Deviance	3069.116	AIC	3085.116

Fig. 5.3 shows the resulting relative risks (Eqn. 5.2) of neighborhoods with deficiencies for the three categories of predominant race-ethnicity (*predominantly White*, *predominantly Hispanic*, and *predominantly Black*) versus neighborhoods with *no predominant race-ethnicity*. The  $x$ -axis denotes the overall infrastructure deficiency to be equal or greater than the displayed ticks. The  $y$ -axis represents the value of relative risk estimates where mean results are plotted as lines and 95% confidence levels are denoted by the shaded regions. The values of relative risk for *predominantly Black* and *Hispanic* neighborhoods across all possible infrastructure types scenarios are greater than 1, suggesting that *predominantly Black* and *Hispanic* neighborhoods show a greater risk of having "more" deficient infrastructure than neighborhoods with *no predominant race-ethnicity*. For *predominantly White* neighborhoods, the values of relative risk are less than 1, indicating that *predominantly White* neighborhoods have a smaller risk of high infrastructure deficiency compared to neighborhoods with *no predominant race-ethnicity*.

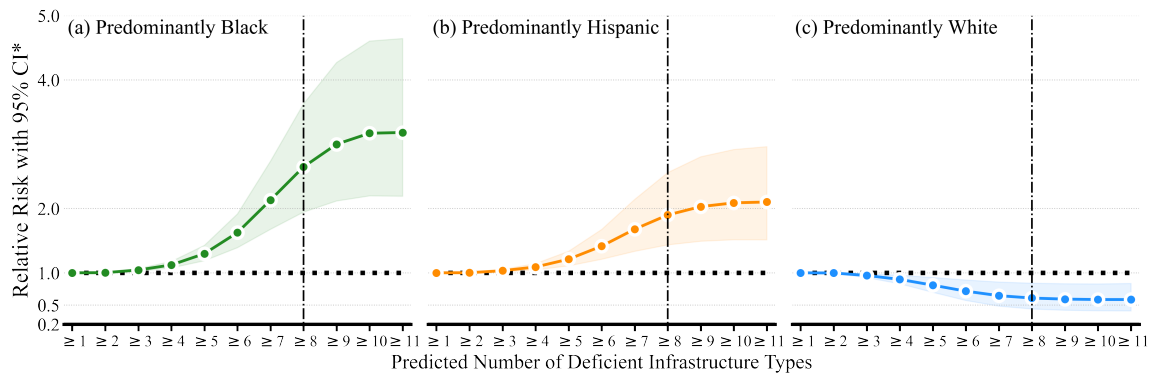


Figure 5.3: Relative risk: Computed relative risk is shown as circles, and shaded regions denote its upper (97.5%) and lower (2.5%) confidence limits. 95% confidence intervals of the three cases were obtained using bootstrapping with 20,000 simulations. (a) The relative risk of overall infrastructure deficiency between *predominantly Black* and No Predominant Race neighborhoods, (b) Relative risk of overall infrastructure deficiency between *predominantly Hispanic* and *no predominant race* neighborhoods, (c) Relative risk of overall infrastructure deficiency between *predominantly White* and *no predominant race-ethnicity* neighborhoods.

Furthermore, as the overall infrastructure deficiency increases (more deficient infrastructure types), the relative risk for *predominantly Black* and *Hispanic* neighborhoods also increases. Specifically, *predominantly Black* neighborhoods are 2.0 ~ 3.6 times; and *predominantly Hispanic* neighborhoods are 1.4 ~ 2.6 times more likely to have highly deficient infrastructure ( $\gamma \geq 8$ ) compared to neighborhoods with *no predominant race-ethnicity*. On the contrary, the relative risk of *predominantly White* neighborhoods decreases as infrastructure increases, with a maximum confidence interval of 0.4 ~ 0.8 times less likely to have highly deficient infrastructure compared to *no predominant race-ethnicity* neighborhoods. Such substantial differences indicate significant infrastructure inequities across race-ethnicity for most types of infrastructure.

### 5.0.3.3 Infrastructure inequity across income and race-ethnicity groups

To examine relative risk for both income and race-ethnicity, a full model with both income and race-ethnicity is fitted with the estimated parameters shown in Table 5.2. After backward model selection, no extra terms were dropped and the full model equals the final model, as expressed in Eqn. 5.3.  $\alpha_j$  is the intercept coefficient,  $\beta_I$  is the regression coefficient for continuous income variable  $x_I$ ;  $\beta_W, \beta_H, \beta_B$  are regression coefficients for the dummy variables of the categorical race-ethnicity covariate with four levels (*predominantly White, predominantly Hispanic, predominantly Black, no predominant race-ethnicity*).  $\beta_{IB}, \beta_{IH}, \beta_{IW}$  are coefficients for income, race-ethnicity interactions.

To understand the relationship between infrastructure deficiency and income, race-ethnicity combinations, under the full model the author computes the probability of having highly deficient infrastructure based on neighborhoods' income and race-ethnicity status, and the results are shown in Fig. 5.4. Based on the predicted probability, *predominantly Black* neighborhoods have the overall highest probability of highly deficient infrastructure compared to other neighborhoods with similar annual income. *Predominantly Hispanic* neighborhoods are the second-highest group and *predominantly White* neighborhoods have the lowest probability compared to all other neighborhoods with the same income level except for 95% percentile income. In addition, the probability gaps among race-ethnicity are reduced as neighborhoods' income increases, as also shown in Fig. 5.4.

Table 5.2: Estimated coefficients of the cumulative logit model for income and race-ethnicity.

		Value	Std. Error	<i>t</i> value
Coefficients				
Income				
	$\beta_I$	0.40	0.25	1.59
Race-ethnicity				
	$\beta_W$	17.57	4.35	4.04
	$\beta_B$	-2.04	4.75	-0.43
	$\beta_H$	-5.66	5.00	-1.13
Interaction				
	$\beta_{IW}$	-1.50	0.39	-3.87
	$\beta_{IH}$	0.10	0.46	0.22
	$\beta_{IB}$	0.46	0.47	0.99
Intercepts ( $\alpha_j$ )				
	$\alpha_1$	-9.85	2.77	-3.55
	$\alpha_2$	-7.05	2.71	-2.60
	$\alpha_3$	-5.97	2.71	-2.20
	$\alpha_4$	-4.98	2.71	-1.84
	$\alpha_5$	-4.04	2.71	-1.49
	$\alpha_6$	-3.04	2.71	-1.22
	$\alpha_7$	-1.99	2.71	-0.74
	$\alpha_8$	-0.79	2.71	-0.29
	$\alpha_9$	1.21	2.74	0.44
	$\alpha_{10}$	2.83	2.88	0.98
Residual Deviance	3035.912		AIC	3069.912

At the higher incomes (95% income quantiles), *predominantly Hispanic*, *predominantly White*, and neighborhoods with *no predominant race-ethnicity* have very comparable probabilities risks of deficient infrastructure. Although *predominantly Black* neighborhoods still have higher risks than the other neighborhoods, the probability is decreased for wealthier neighborhoods. Note that these findings are consistent with the results in Chapter 4 for neighborhood income characteristics.

#### 5.0.3.4 Impacts of historical redlining on infrastructure condition

Fig. 5.5(a) shows the infrastructure deficiency within historical redlining areas (218 out of the 790 block groups) for each HOLC grade. Fig. 5.5(b) shows the distribution of infras-

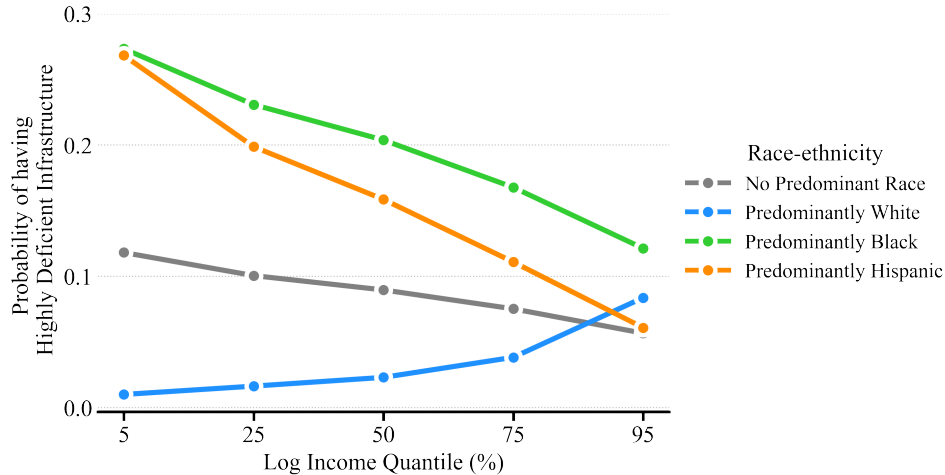


Figure 5.4: Probability of having highly deficient infrastructure given neighborhood income and race-ethnicity composition.

structure deficiency in redlining areas for each HOLC grade by neighborhood race-ethnicity groups. An increasing trend is seen, with infrastructure deficiency becoming higher from grades A to D. This trend is found to be statistically significant with a positive *Gamma statistic* of 0.222.

According to the HOLC, areas marked with grades C and D were considered to be the most undesirable for mortgages. This practice negatively influenced infrastructure maintenance and rehabilitation through the decades, where its long-term effect is reflected as the uptrend of overall infrastructure deficiency from grade A to grade D shown in Fig. 5.5(a). Despite re-development activities in the central core of the city since redlining, neighborhoods that were graded “worse” in the past decades still persist with higher infrastructure deficits in the present than other neighborhoods.

Fig. 5.5(b) also highlights different levels of racial inequity in historical redlining areas. In areas with grades B, C, and D, *predominantly Black* neighborhoods have significantly higher infrastructure deficiencies than other neighborhoods, but this inequity is greatest in redlining areas with grade D, which have the highest *gamma statistic* in Table 5.3. From these results, it can be seen that how the injustice of historical housing policies created racial inequity whose legacy remains in the infrastructure of *predominantly Black* neighborhoods

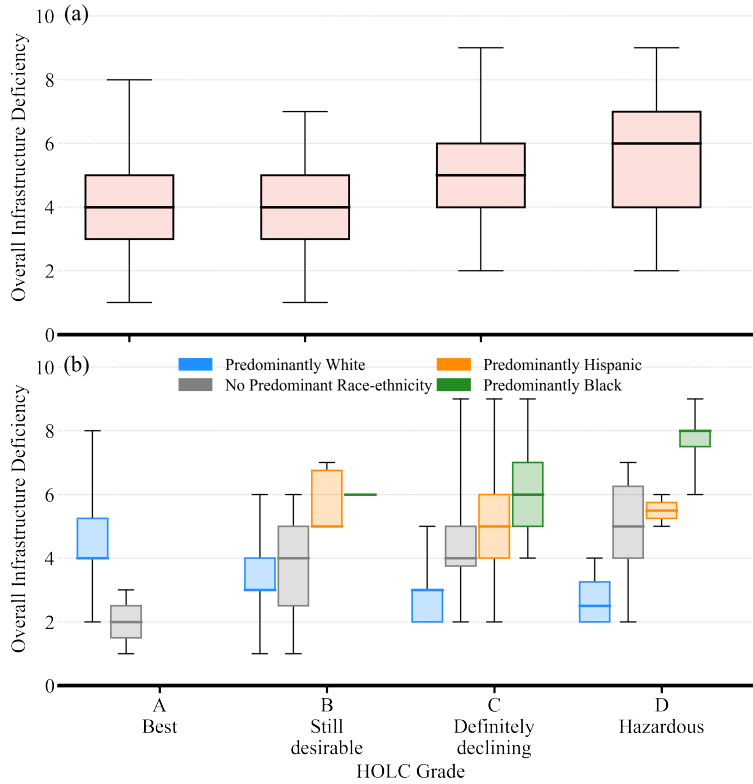


Figure 5.5: Box-whisker plot of overall infrastructure deficiency among HOLC rated neighborhoods. (a) Box-whisker plot for each HOLC redlining grade. (b) Box-whisker plot for each HOLC redlining grade by neighborhoods’ race-ethnicity (denoted by different colors). Upper and lower whiskers indicate the maximum and minimum value of the population, upper and lower boxes indicate the 1st and 3rd quartile respectively. The centerline across the box indicates the median of the population

today.

### 5.0.3.5 Infrastructure inequity by neighborhood age

In addition to historical redlining, a neighborhood’s age could be another important indicator of infrastructure condition, given how policies and development practices have changed over time. Fig. 5.6(a) shows a box-whisker graph indicating decades of neighborhood average built-year ( $x$ -axis) and quantiles of infrastructure deficiencies across the neighborhoods built in each decade ( $y$ -axis). Fig. 5.6(a) indicates that neighborhoods built from the 1930s to 1960s have more infrastructure deficiencies, on average than neighborhoods built during

Table 5.3: Trend analysis using Gamma statistic between overall infrastructure deficiency and 1) redlining HOLC grades and 2) neighborhood race-ethnicity within each HOLC grade. Gamma statistic is not reported if any of the neighborhood race-ethnicity categories have missing records. Significance level: \* :  $p$ -value < 0.1; \*\* :  $p$ -value < 0.01; \*\*\* :  $p$ -value < 0.001. (Note: W: *predominantly White*; N: *no predominant race-ethnicity*; H: *predominantly Hispanic*; B: *predominantly Black*)

Categories of Areas	Number of Observations (n)				Gamma Statistic	Significance Level
	A	B	C	D		
Redlining	23	43	131	21	0.222	***
HOLC Grade	W	N	H	B		
A	20	3	0	0	-	-
B	21	11	10	1	0.612	***
C	19	36	57	19	0.565	***
D	4	8	2	7	0.859	***

other periods. It is worth noting that neighborhoods built earlier than this timespan have relatively lower infrastructure deficiencies; this might be due to redevelopment projects and programs for older neighborhoods [117].

Fig. 5.6(b) shows a further breakdown between infrastructure deficiency and average neighborhood built-year by race-ethnicity. The results reveal that *predominantly Black* neighborhoods have worse average infrastructure conditions than *predominantly White* neighborhoods across all decades when *predominantly Black* neighborhoods were built. Furthermore, *predominantly Hispanic* neighborhoods also have higher average infrastructure deficiencies than *predominantly White* neighborhoods except during the 1960s, when the average deficiencies are equivalent. Overall, the race-ethnicity categories shown in Fig. 5.6(b) have upward trends that are statistically significant according to the *Gamma statistic* given in Table 5.4. While the infrastructure inequities are most severe in neighborhoods built from the 1930s to the 1950s, when HOLC policies were in place, *predominantly White* neighborhoods still have significantly better infrastructure conditions than *predominantly Hispanic* and *Black* neighborhoods even in areas with average built-years during the 21st century. These inequities also persist in neighborhoods built in most of the decades except the 1970s

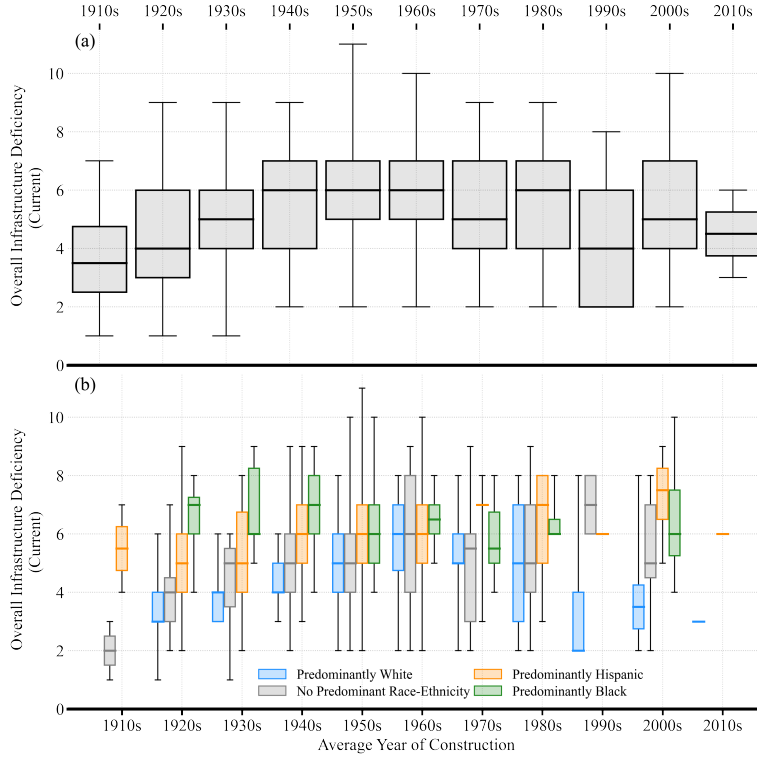


Figure 5.6: Box-whisker plot of infrastructure deficiency versus neighborhood average built-year during each decade for (a) all neighborhoods combined and (b) neighborhoods by predominant race-ethnicity. (Denoted by different colors). Upper and lower whiskers indicate the maximum and minimum value of the population, upper and lower boxes indicate the 1st and 3rd quartile respectively. The centerline across the box indicates the median of the population

and the 1990s. These results imply that historical plans or policies in different decades might play important roles in affecting neighborhood infrastructure and lead to today's inequities.

#### 5.0.4 Discussion

Dallas has a long history of racial and wealth segregation and the findings show that this segregation persists today [118]. The city facilitated the underdevelopment of minority neighborhoods through the process of zoning, which clusters landfills, liquor stores, and industrial activities in disenfranchised communities [119]. The segregation of housing based on neighborhoods' racial markup was reinforced through the introduction of HOLC redlin-



Table 5.4: Trend analysis using Gamma statistic between overall infrastructure deficiency and predominant neighborhood race-ethnicity by average built-year in decades. Gamma statistic is not reported if any of the subgroups with missing records. Significance level: \* :  $p\text{-value} < 0.1$ ; \*\* :  $p\text{-value} < 0.01$ ; \*\*\* :  $p\text{-value} < 0.001$ . (Note: W: *predominantly White*; N: *no predominant race-ethnicity*; H: *predominantly Hispanic*; B: *predominantly Black*)

Neighborhoods' Average Built Decade	Number of Observations (n)				Gamma Statistic	Significance Level
	W	N	H	B		
1910s	0	2	2	0	-	-
1920s	12	15	24	8	0.594	***
1930s	9	11	18	6	0.526	***
1940s	17	34	67	13	0.515	***
1950s	52	42	74	38	0.366	***
1960s	44	33	23	12	0.134	***
1970s	31	38	14	10	0.246	***
1980s	11	41	10	4	0.289	***
1990s	18	5	2	0	-	-
2000s	12	19	8	6	0.568	***
2010s	1	0	1	0	-	-

ing practices. Basic infrastructure services such as modern plumbing, electricity, and trash pickups were disproportionately planned provided to neighborhoods with minorities [119]. Despite the city introducing revitalization plans to dissipate such disparities [120], neighborhood infrastructure still reflects the legacy of these historic decisions and practices.

One of the major findings of this paper is that *predominantly Black* and *Hispanic* neighborhoods have a statistically higher risk of highly deficient infrastructure compared to other neighborhoods. The risk analysis shows that neighborhoods with a higher average level of household income have a lower probability of highly deficient infrastructure, but not for *predominantly White* neighborhoods. Such interactions also indicate a highly correlated relationship between wealth and a neighborhood's racial demographic. While the infrastructure inequities across race-ethnicity are higher in neighborhoods that were built during the redlining period, even neighborhoods built during the 21st century show clear race-ethnicity biases. These findings indicate a pressing need for decision-makers to adopt policies and investment

strategies that target *predominantly Black* and *Hispanic* neighborhoods for infrastructure improvements.

It is also worth noting that *predominantly Black* neighborhoods, despite having overall higher infrastructure deficiency, still have better public transit than predominantly White neighborhoods. Many low-income communities in southern Dallas are *predominantly Black* neighborhoods. Having access to public transit is more critical for low-income communities that have lower rates of car ownership, which means investments in public transportation can be more beneficial. On the other hand, in northern Dallas, many high-income neighborhoods do not heavily rely on public transportation and may opt out of public transit routes or bus stops to lower taxes. However, lack of public transit in high-income areas can still create difficulties for those who cannot drive (e.g., youth or elderly) and may prevent those without cars from entering these areas. This finding highlights the importance of considering diverse neighborhood needs for building a more equitable city.

The existence of infrastructure inequity not only influences vulnerable neighborhoods with a lack of infrastructure resources but also draws attention to highly deficient areas for optimizing investments. Bond programs, separate from the city's annual budget, focus on improving capital funding for city assets, including neighborhoods infrastructure, such as facilities, streets, libraries, and parks [121]. 2017 Bond investments data is acquired from the city open data portal (<https://www.dallasopendata.com>) and extracted projects that are infrastructure-related based on their included project description. The investments are summarized for each city council district as shown in Fig. 5.7. The vertical bars indicate the infrastructure-related funding allocation of 2017 bond projects by city council district (14 districts in total), the solid line represents the distribution of identified infrastructure deserts within each council district (i.e., the percentage of the total number of neighborhoods that are infrastructure deserts). It can be seen that the allocation of bond projects deviates from the distribution of infrastructure deserts, which serves as a proxy of community needs for infrastructure improvements due to the high deficiency. As a result, the most recent bond investments do not resolve infrastructure inequity and infrastructure gaps remain unclosed. For example, Districts 8 and 5, which have the highest percentages of infrastructure deserts, will be unable to catch up with other districts if the investments aren't distributed more

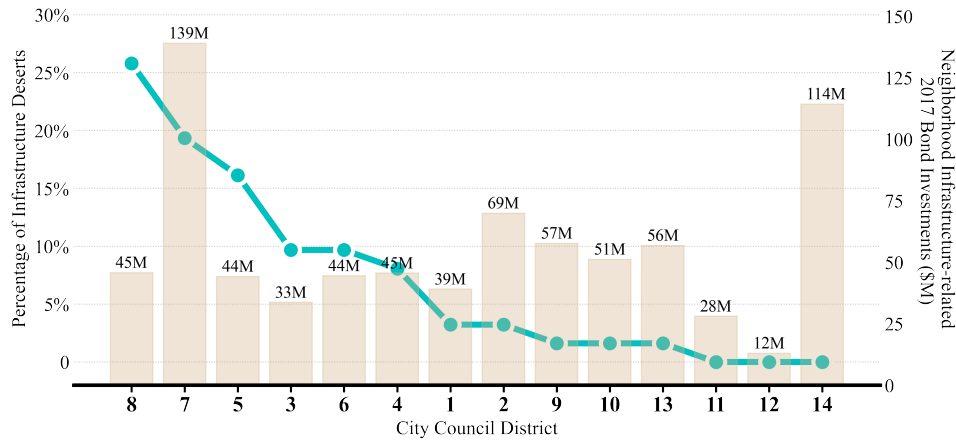


Figure 5.7: Distributions of 2017 City of Dallas bond projects infrastructure-related investments and percentage of identified infrastructure deserts by city council districts

equitably. This result, again, highlights the important role of understanding infrastructure equity to guide the city’s future plans and policies.

### 5.0.5 Conclusions

Studying infrastructure equity involves comparing infrastructure characteristics or conditions in neighborhoods with a high concentration of socially vulnerable populations compared to adjacent neighborhoods or the regional average [25]. In Chapter 4, the author identified infrastructure deserts, low-income areas with highly deficient neighborhood infrastructure, and showed increasing trends of infrastructure inequity across decreasing median neighborhood income in a case study in Dallas, Texas. This chapter expands that work to explore the relationships between overall infrastructure deficiency and neighborhood race-ethnicity overall and by neighborhood age and historical redlining policies. To the author’s knowledge, such analyses connecting multiple infrastructure types to neighborhood characteristics are rarely discussed in previous literature due to a lack of available data at neighborhood scales.

This chapter delivers comprehensive insights on infrastructure inequity across income, race-ethnicity, and statistically shows the various levels of inequity among neighborhoods with different income and race compositions. Statistical inference indicates that *predominantly Black* and *Hispanic* neighborhoods are 1.4 to 2.6 times and 2.0 to 3.6 times likelier

to have *highly deficient* infrastructure (8 or more deficient infrastructure types out of 12) than areas with *no predominant race-ethnicity*, respectively. Furthermore, these disparities reflect the legacy of historical discriminatory housing policies (“redlining”) and their long-term impacts on neighborhood infrastructure. Neighborhoods marked as “less desirable” for mortgages during the 1930s still experience significantly more infrastructure deficiencies today.

## Chapter 6

### Do neighborhood infrastructure deficits differ by city? A comparative infrastructure equity study among Los Angeles, New York City, Chicago, and Dallas

#### 6.0.1 Introduction

This chapter continues to assess and explore the existence of infrastructure deserts by applying the infrastructure equity assessment framework to multiple major cities across the country. In the previous chapters, The author investigated and demonstrated the inequities in neighborhood infrastructure via the first case study conducted in the City of Dallas, Texas. The findings raise the question of whether infrastructure equity in other cities shows similar patterns. In order to efficiently address this question and enable future comparisons in more cities, an automated system is created that allows researchers, planners, and engineers to more easily implement infrastructure assessment. The automated framework is implemented using a Cloud-based platform called “Clowder,” which enables a comparative analysis among four major cities: Los Angeles, CA; Dallas, Texas; Chicago, IL, and New York City, NY.

Previous infrastructure-related studies have primarily focused on developing new approaches or improving existing methods to achieve better evaluation performance of single types of infrastructure [29, 34, 44, 96, 108, 122, 123]. The generalization of these approaches has not yet been discussed. However, the benefits of creating generalized tools for research not only increase the applicability of the research but also create opportunities to foster other innovative ideas within the field. For example, the Economic Research Service published the Atlas guide in 2015 to create food desert indicators for many cities across the country [124]. This guide raises awareness of food access inequities across the nation and can support research related to food access.

With the development of Cloud-based storage and computing platforms, integrating the framework with a Cloud-based platform supports easier publishing and sharing the tools

among researchers [125]. Currently, the implementation of infrastructure assessment requires users to gather all of the data and assess substandard criteria manually. The process is repeated for each infrastructure type and becomes a time-demanding task to complete for multiple cities and time periods without any automation. Furthermore, the size and number of datasets collected continue to grow as more infrastructure types are added, which will inevitably increase the efforts to transfer, share, and manage the data and analyses. Such concerns could be preventing similar studies [31, 62, 113, 123] from being effectively and efficiently implemented in other cities. Thus, generalizing the framework and integrating it with a Cloud-based platform will allow the framework to be more accessible and effective for testing results with other case studies.

### 6.0.2 Methodology

Clowder (<https://clowderframework.org>) [125] is a Cloud-based open-source data management platform developed by the National Center for Supercomputing Applications. More details on Clowder are provided in Section 6.0.2.1. The integration between the infrastructure assessment framework described previously and Clowder delivers three main benefits. First, the framework can be executed anywhere via any Internet-connected browser at any time without any program installation. Second, Clowder offers functionalities for users to store, manage, and share collected and analyzed datasets on the platform. Finally, using Clowder, the framework can more easily be generalized to other cities and regions. Fig. 6.1 shows the flow chart of the Cloud-based assessment framework: Users first upload the configuration file and required datasets to Clowder. Then, an “infrastructure assessment extractor” reads the configuration file and analyzes the uploaded data to compute the overall infrastructure deficiency rating as mentioned in Equation from Chapter 2. Finally, a map of overall infrastructure deficiency and the relative risk results are generated as the outputs. All of the computations are performed on the Cloud without any installation on the local machine and the data can also be stored and shared on Clowder for future use. The implementation of the automated infrastructure assessment framework in Clowder consists of three major steps: 1) prepare the assessment configuration file as a JSON file; 2) upload datasets to Clowder according to configuration; 3) execute the infrastructure assessment framework as

the Clowder “extractor”. Each step is described in more detail in subsections 6.0.2.2 through 6.0.2.4 below.

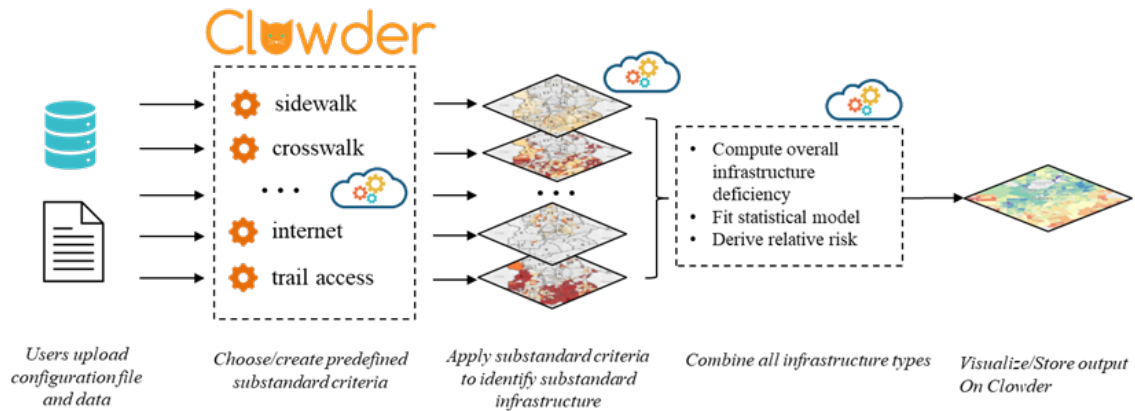


Figure 6.1: Overview of the Cloud-based infrastructure assessment framework.

### 6.0.2.1 Clowder

Clowder is a Web-based, open-source, customizable and scalable data management framework to support any data format and multiple research domains. It allows flexible metadata representation which supports both user-defined and machine-defined metadata from either a Web user interface or Web service Application Programming Interface (API). Clowder also has a cluster of extraction services that process newly added data to extract metadata and ways to write Javascript-based widgets to visualize the contents of files and datasets. These services include previews of CSV files, interactive Web maps for GIS layers, and image thumbnails for video data.

Clowder is built to simplify the ingestion and curation of data and to be extensible to support the long tail of research data and many research domains. Since its inception, Clowder has been leveraged and augmented to support scientific needs within numerous NSF-funded projects, including DataNet SEAD, Sustainable Environment through Actionable Data, <http://sead-data.net> [126, 127]; DIBBs Brown Dog [128–130]; XSEDE [131]; and IML-

CZO [132, 133]. Clowder has supported communities such as biology, geoscience, materials science, crop science, urban science, social science [134], and the humanities. In addition, Clowder’s rich Web interfaces let users upload, curate, and share raw data along with complex metadata and extractors that process uploaded data, which allows users to better manage project data and results.

#### *6.0.2.2 Prepare assessment configuration JSON file*

A JSON file is required for users to describe and define all of the needed assessment parameters including considered infrastructure types, substandard criteria, corresponding datasets, and intended output characteristics. The structure of the JSON file is shown in Fig. 6.2, where the first few arguments include essential ArcGIS layers for the assessment such as neighborhood boundaries (Census block groups in this case), income, and residential parcels. The considered infrastructure type information is structured as an expandable dictionary where users can decide how many infrastructure types are to be included.

Within each infrastructure type, the users choose a substandard criterion from one of two predefined options that correspond with the two primary criteria for quantifying substandard infrastructure (see Chapter 2). The first substandard criterion is directly assessed based on the dataset’s existing attributes. For example, the Pavement Condition Index (PCI) is commonly used to assess the pavement condition . Under the created dictionary for pavement, the user should set the attribute corresponding to PCI in the condition attribute and specify less than 55 as the criterion argument for substandard pavement segment.

The second substandard criterion uses the proxy relationship between neighborhood residential households and infrastructure facilities to define whether resident’s access to services is sufficient. For example, the travel distance between the nearest grocery store to each residential household is one of the most common measures to indicate accessibility to food stores. A pre-filled template of commonly used substandard criteria is also provided to guide users in easily configuring their infrastructure types properly. Once all of the considered infrastructure types have been configured, the user then uploads the JSON file to Clowder, where it serves as the input to the assessment framework. On the extractor page after clicking “submit for extractions” in the “dataset” page, Clowder also provides a Web-based



user-friendly Graphical User Interface that guides users to create a new JSON configuration file or import existing file from local.

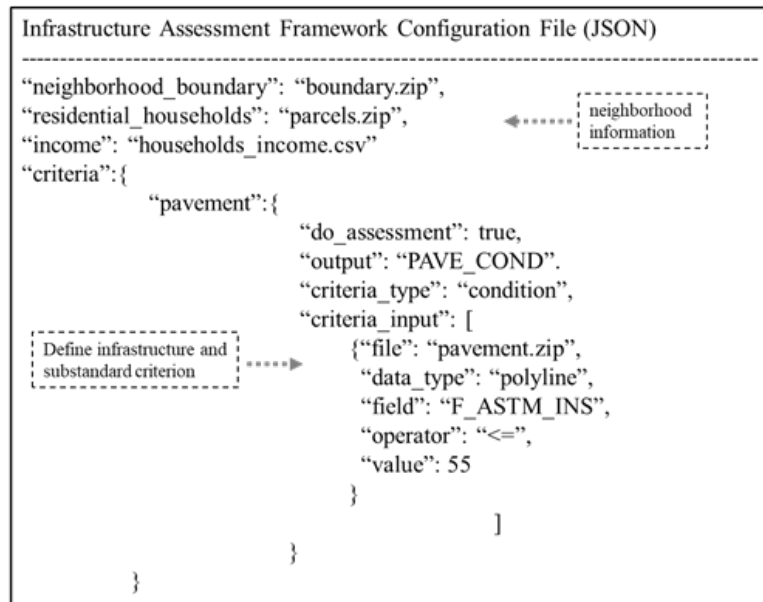


Figure 6.2: Infrastructure assessment framework configuration file (JSON). The first three arguments represent neighborhood boundary, residential households, and annual median income. A common attribute of Block Group ID is required when preparing the datasets. Each infrastructure type is defined as a separate dictionary (see an example of “pavement”) with its data source and substandard criterion specified inside the dictionary.

### 6.0.2.3 Upload necessary datasets to Clowder

In addition to choosing substandard criteria in the configuration file, users also need to note the name of the dataset files to the “file” argument under “criteria\_input” of the configuration file. Once users manually upload all required datasets to Clowder, they can click the “submit for extraction” button shown along with the datasets page. A list of available extractors will be displayed and choosing “Infrastructure Assessment” triggers and runs the infrastructure assessment framework. The name and directory of the uploaded datasets should follow the definitions in the configuration file. The uploaded format depends on the

information within the dataset. If data contains spatial information such as location, lines, or areas, uploading these data as a zipped ArcGIS Shapefile is recommended. Otherwise, CSV files can be used to store tabular records if no geospatial information is used. To ensure that the assessment performs properly, all of the spatial data should have the same Coordinate Reference System (CRS), for example set CRS to EPSG 3857 to represent WGS 84/Pseudo-Mercator.

#### 6.0.2.4 *Execute infrastructure assessment framework as a Clowder extractor*

Once all of the datasets have been successfully uploaded to Clowder along with the configuration JSON file, the infrastructure assessment framework is ready to be executed as a Clowder extractor. “Extractors” are one of the functionalities built into the Clowder platform to allow users to run customized operations on Clowder datasets. Extractors are executable structured scripts that are compiled and stored on the Clowder platform. The extractor will be executed when the user clicks the “submit” button on the page showing a list of available extractors, the corresponding script will run and execute predefined operations. There are several pre-built Clowder extractors that perform tasks such as creating preview images for pdf and other data files, text extractions from documents, and information tagging for video clips. The extractor can also be self-built by Clowder users in programming languages such as Python, R, or Matlab to customize other needs.

The current Clowder infrastructure assessment framework consists of Python scripts to execute spatial operations such as measuring distances and evaluating spatial relationships. An open-sourced Python library "*GeoPandas*" is used to support all geometry implementation between datasets such as spatial intersection, addition, and deletion. Beyond these scripts, extractors have been built to reproduce all of the steps in the infrastructure assessment framework described in Chapters 4 and 5 and deploy them on Clowder for public use. Users can see the list of available extractors by clicking “Submit for Extractors” on their dataset page.

The infrastructure assessment framework consists of two main “extractors .” The first is a Python-based extractor that imports datasets, applies substandard infrastructure criteria, and calculates overall infrastructure deficiency. The second extractor is R-based and fits sta-

tistical models (described in Chapters 4, 5, and 6) using the overall infrastructure deficiency from the first extractor and computing relative risk across demographic indicators such as income and race-ethnicity, etc. The computation of relative risk along with 95% confidence intervals is performed in the R-based extractors using the bootstrapping method [80, 135]. Once users submit the job by clicking “submit”, the extractor reads configuration file and executes operations with uploaded datasets in the background, and generates results (e.g., new shapefiles and graphics) in the same Clowder storage directory as the datasets. During execution, the progress of the assessment is tracked and can be seen by viewing the metadata of the implemented datasets.

#### 6.0.2.5 Statistical models

To compare the infrastructure condition among cities, the author fits cumulative logit models similar to those discussed in Chapters 2, 4, and 5. In addition to income and race-ethnicity, city is also included as one of the explanatory variables. Thus, the full model can be written as:

$$\begin{aligned}
 \text{logit}[Pr(\gamma \leq j|x)] = & \alpha_j + \beta_I x_I + \beta_R^T \mathbf{x}_R + \beta_C^T \mathbf{x}_C + \\
 & x_I \beta_{IR}^T \mathbf{x}_R + x_I \beta_{IC}^T \mathbf{x}_C + \beta_{RC}^T \mathbf{x}_R \mathbf{x}_C + \\
 & x_I \beta_{IRC}^T \mathbf{x}_R \mathbf{x}_C; \\
 & j = 1, \dots, J - 1
 \end{aligned} \tag{6.1}$$

Where  $\gamma$  is the computed overall infrastructure deficiency with each value of integer representing one category,  $x_I$  is a continuous income variable from annual median household income. Vector  $x_R = [x_{Rh}, x_{Rw}, x_{Rb}]$  represents three dummy variables that indicate the neighborhood’s race-ethnicity:  $x_{Rh} = 1, x_{Rw} = 0$ , and  $x_{Rb} = 0$  if the race-ethnicity is *predominantly Hispanic*;  $x_{Rh} = 0, x_{Rw} = 1$ , and  $x_{Rb} = 0$  if the race-ethnicity is *predominantly White*; Similarly, if the race-ethnicity is *predominantly Black*,  $x_{Rh} = 0, x_{Rw} = 0$ , and  $x_{Rb} = 1$ . If there is *no predominant race-ethnicity* group presented,  $x_{Rh} = 0, x_{Rw} = 0$ , and  $x_{Rb} = 0$ , serving as the reference level. Vector  $x_C = [x_{Cn}, x_{Cd}, x_{Cc}]$  represents four city locations with Los Angeles as the reference level, where  $x_{Cn} = 1, x_{Cd} = 0$ , and  $x_{Cc} = 0$  if it is

New York City;  $x_{Cn} = 0, x_{Cd} = 1$ , and  $x_{Cc} = 0$  if it is Dallas;  $x_{Cn} = 0, x_{Cd} = 0$ , and  $x_{Cc} = 1$  if it is Chicago; and  $x_{Cn} = 0, x_{Cd} = 0$ , and  $x_{Cc} = 0$  for Los Angeles.  $J$  is the total number of infrastructure types considered ( $J = 10$  in this comparative study). Single value  $\beta_I$  is the regression coefficient for income;  $\beta_R = [\beta_{Rh}, \beta_{Rw}, \beta_{Rd}]$ ,  $\beta_C = [\beta_{Cn}, \beta_{Cd}, \beta_{Cc}]$  are regression coefficients for race-ethnicity and city respectively.  $\beta_{IR}, \beta_{IC}, \beta_{RC}$ , and  $\beta_{IRC}$  are the interaction coefficients between income, race-ethnicity, and city variables. The expanded form of the model can be found in Appendix A.4.

To fit the model parameters, the author performs backward model selection [77] to remove any insignificant terms using Akaike Information Criterion (AIC) [136] as the dropping criteria. Once the model is properly fit and passes the proportional odds assumptions mentioned in Chapters 2 (section 2.0.2.3) and Chapter 5 (section 5.0.2.1), the relative risk of highly deficient infrastructure in neighborhoods of each predominant race-ethnicity group compared to the baseline (neighborhoods with no predominant race-ethnicity groups) for a given income level is computed (percentiles of continuous income). Recall from Chapter 2 that the category of “highly deficient” is derived based on the distribution of the city’s overall infrastructure deficiency (90th quantile and above), which differs for each of the four cities. Thus the risk is a measure of equity relative to other residents of the same city. The estimated parameters can be used compute the relative risk as the ratio of the probabilities of highly deficient infrastructure (refer to Chapter 5 section 5.0.2.2) for two compared scenarios (e.g., In Chicago, the probability of having highly deficient infrastructure for *predominantly Black* neighborhoods against neighborhoods with *no predominant race-ethnicity*) given income, race-ethnicity, and city. Similarly, relative risk is computed at various income levels (5%, 50%, and 90% quantiles of the distribution) to show how neighborhood income affects infrastructure equity across the different race-ethnicity and cities.

### 6.0.3 Results

The author applied the Cloud-based framework described above to Los Angeles, New York, Chicago, and Dallas, including data on 10 infrastructure types that were available in all cities: *pavement, sidewalk, public transit access, trail access, food access, bank access, medical facility access, gathering place access, internet service, and noise wall*. With the

available Census data, the author identified 2637 Census block groups in Los Angeles, 6182 block groups in New York City, 2139 block groups in Chicago, and 790 block groups in Dallas. The estimated parameters is shown in Appendix A.5 and the following sections focus showing the comparison of overall infrastructure deficiency and individual deficient infrastructure by income, race-ethnicity across four cities. The differences of top severe infrastructure types and estimated relative risks from statistical model between cities are also discussed. The author also highlights the comparisons the infrastructure condition within historical redlining areas.

6.0.3.1 Comparison by individual infrastructure type

Fig. 6.3 shows the histogram of overall infrastructure deficiency by city. The author observes that overall infrastructure deficiency is distributed differently among the cities: Dallas has the worst average infrastructure condition and New York City has the best average infrastructure condition. Los Angeles, New York City, and Chicago’s average infrastructure condition ranges from 1 to 2, while Dallas’ average infrastructure condition is 6, which is the highest of all. Fig. 6.4 shows the infrastructure deficiency by income level. Fig. 6.4 shows that

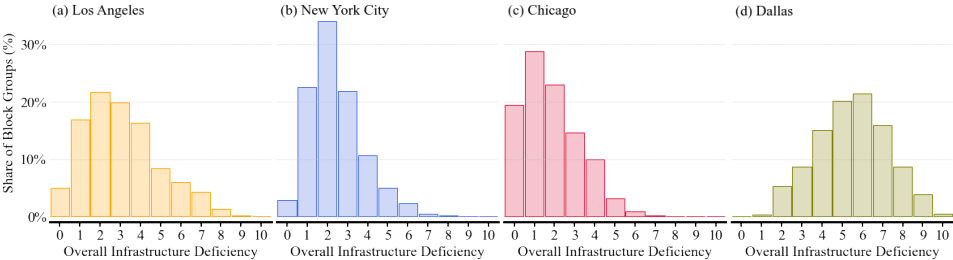


Figure 6.3: Histograms of computed overall infrastructure deficiency in Los Angeles, Chicago, Dallas, and New York City. A total of 10 infrastructure types is considered.

low-income neighborhoods have the worst overall infrastructure than other neighborhoods in both Dallas and Chicago, consistent with the earlier findings in Chapter 2. However, the

author also observed a “reversed” pattern in Los Angeles, where high-income neighborhoods have overall higher infrastructure deficiency than low-income neighborhoods. This might be due to the location of many high-income areas in northern Los Angeles that are isolated geographically from the city center and have low population density. This could lead to a higher likelihood of access deficiencies, which account for six of the ten infrastructure types examined in these cities. New York City, on the other hand, does not reveal any obvious infrastructure inequities across income characteristics compared to the other cities.

Fig. 6.5 shows the infrastructure deficiency by neighborhood race-ethnicity. Inequities across race-ethnicity are not apparent in Los Angeles and New York City, with all race-ethnicity groups showing similar deficiency patterns. However, Chicago and Dallas show strong signals revealing infrastructure inequity across race-ethnicity, with *predominantly Black* and *predominantly Hispanic* neighborhoods having relatively higher overall infrastructure deficiency.

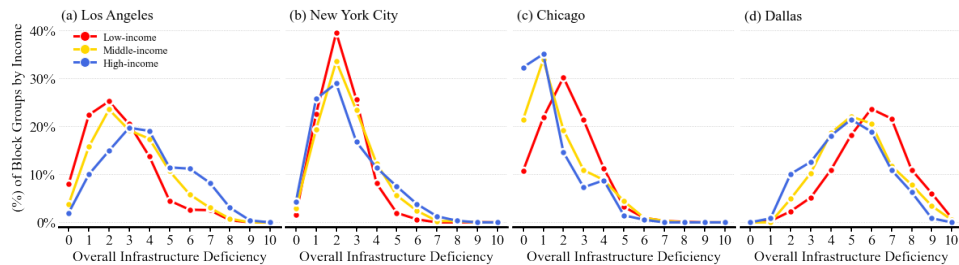


Figure 6.4: Overall infrastructure deficiency by neighborhood’s income level.

Fig. 6.6 shows how the percentage of each deficient infrastructure type changes with income. This figure reveals three primary patterns: 1) an overall increasing or decreasing trend across income and race-ethnicity groups (e.g., high-income neighborhoods have better internet service and worse gathering place access compared to lower-income areas); 2) the deficient infrastructure condition is relatively independent of income levels and does show any specific trends (e.g., pavement condition, sidewalks, trail access, and noise wall have

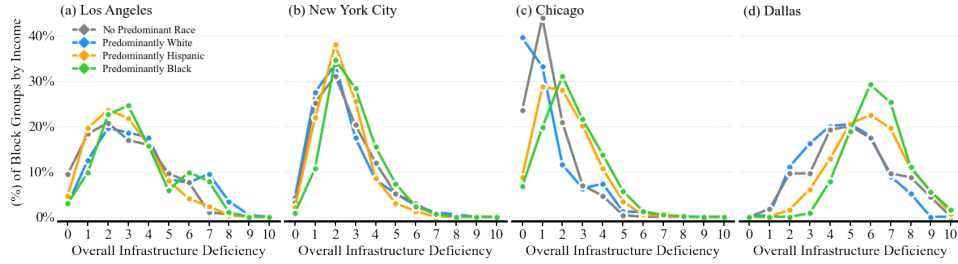


Figure 6.5: Overall infrastructure deficiency by neighborhood’s race-ethnicity.

little fluctuation across neighborhoods with increasing income characteristics); and 3) mixed trends with increasing income level among cities. The figure also shows that higher-income neighborhoods have more deficient food access in Los Angeles, New York City, and Chicago but not in Dallas. Higher-income neighborhoods have more deficient medical facility access in New York City, Chicago, and Dallas, but not in Los Angeles.

Fig. 6.7 shows how the individual types of deficient infrastructure change with race-ethnicity groups. For all 10 infrastructure types, *predominantly Black* neighborhoods have more deficient infrastructure than other neighborhoods. This trend is greatest for internet service, food access, bank access, and medical facility access in the city of Dallas. Similar to the findings by income, pavement and noise walls do not show any substantial changes among different race-ethnicity groups. However bank access and internet services show a consistent trend for all four cities: *predominantly Black* and *Hispanic* neighborhoods have a higher percentage of deficient infrastructure.

Next, Fig. 6.8 shows the five infrastructure types with the highest percentage of deficient infrastructure in each city. It can be seen that noise walls are the most prevalent deficiencies in all cities. Although public transit is the fourth-highest deficient infrastructure in New York City, the percentage of neighborhoods is 15%, significantly less than Los Angeles (44%). Note that sidewalk is assessed in this chapter only by evaluating the number of missing segments; damaged sidewalk data were only available in Dallas and were thus not included in this comparative analysis. This adjustment explains why sidewalk is not showing in the top five deficient infrastructure types in Dallas, despite its prevalence in Chapter 4. To further

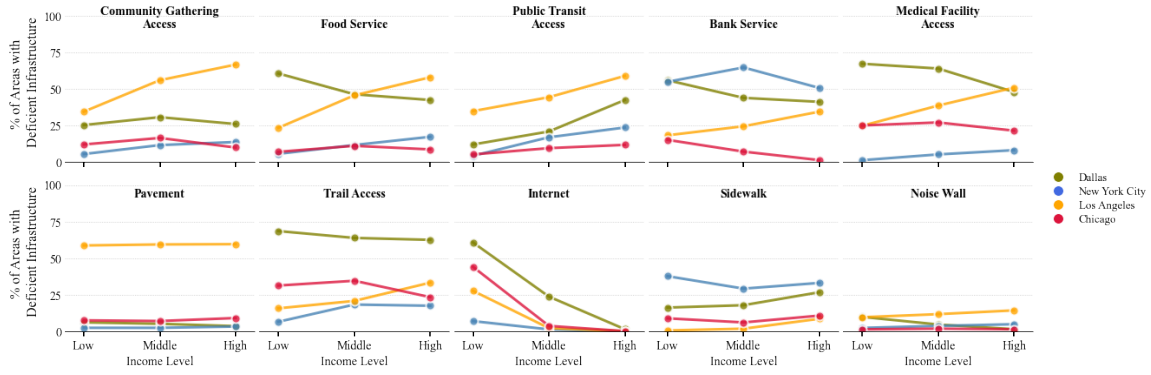


Figure 6.6: Percentage of individual deficient infrastructure at three levels of income (as low, middle, high defined in chapter 2) by city: Los Angeles, Chicago, Dallas, and New York City.

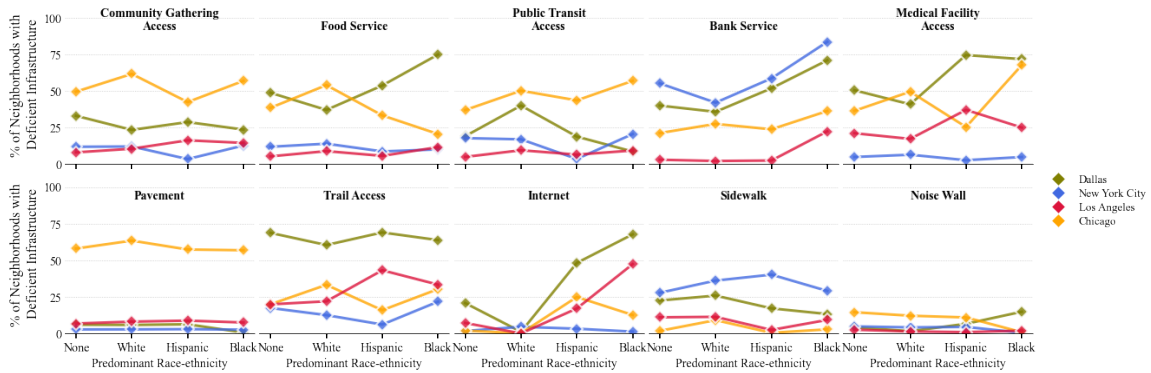


Figure 6.7: Percentage of individual deficient infrastructure at different neighborhood's race-ethnicity by city: Los Angeles, Chicago, Dallas, and New York City.

investigate conditions in neighborhoods with the most severe infrastructure problems (defined here as five or more deficient infrastructure types) , Fig. 6.9 shows the five most prevalent deficient infrastructure in these areas. Compared to citywide deficient infrastructure in Fig. 6.8, at least 78% of these highly neglected areas in all four cities have deficient food access. Public transportation access is another infrastructure type that ranks relatively lower in the citywide deficient infrastructure types but is deficient in up to 90% of the areas that have at least five deficient infrastructure types. In addition, the author sees that in Chicago, areas with five or more deficient infrastructure types have 100% deficient noise walls and



trail access. In New York City, areas with five or more deficient infrastructure types have 100% deficient bank access. These results highlight the difference in deficient infrastructure types between citywide regions and only regions that have severe infrastructure issues.

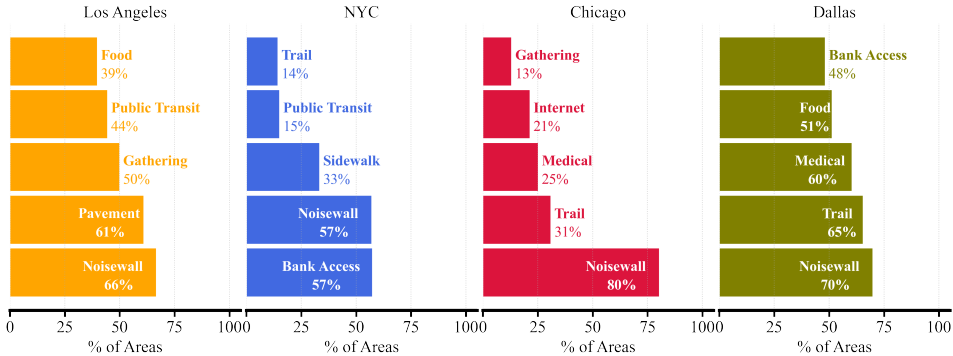


Figure 6.8: Top 5 Ranked deficient infrastructure types represented as percentages across the city

6.0.3.2 Relative risk among Los Angeles, New York City, Chicago, and Dallas

Fig. 6.10 shows the computed relative risk of highly deficient infrastructure from the cumulative logit model. To show the impacts of income, relative risks are computed at three income levels: lower (5% quantile), average (50% quantile), and higher (95% quantile). Colored squares show the relative risk of Black and Hispanic neighborhoods having highly deficient infrastructure compared to neighborhoods with no predominant race, assuming the average income across the city. It is noted that each city has slightly different cutoffs for defining highly deficient areas because the definition of highly deficient is based on quantiles of citywide infrastructure deficiency. Thus the results give relative measures of equity for each city across neighborhood characteristics.

According to Fig. 6.10, *predominantly Black* neighborhoods are most likely to have highly deficient infrastructure compared to *predominantly White* neighborhoods in all four cities. *Predominantly Hispanic* neighborhoods are more likely to have highly deficient infrastructure

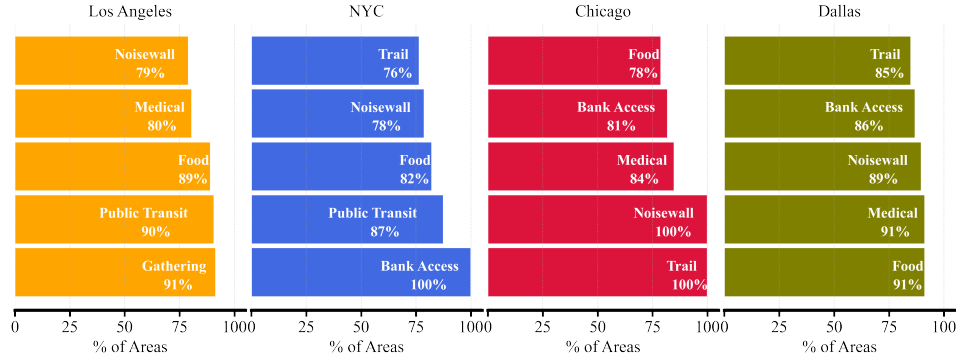


Figure 6.9: Top 5 Ranked deficient infrastructure types represented as percentages within neighborhoods with at least 5 deficient infrastructure types.

in Los Angeles, Chicago, and Dallas. Such risks are much higher in Dallas (up to 3.2 times for *predominantly Black* neighborhoods and 2.6 times for *predominantly Hispanic* neighborhoods) and Chicago (up to 5.3 time for *predominantly Black* neighborhoods and 3.7 times for *predominantly Hispanic* neighborhoods).

In most situations, neighborhoods with higher income have decreased risks as indicated by light blue arrows pointing downwards in Fig. 6.10, which agrees with the findings from Chapter 4. However, for New York City, higher income shows the reverse trend, increasing neighborhoods’ risk of highly deficient infrastructure. Similarly, risk increases as income decreases, as indicated by the red arrows pointing upwards for Los Angeles, Chicago, and Dallas. New York City, however, has the least inequities across both income and race-ethnicity characteristics as the value of relative risk (squares in Fig. 6.10) is close to one and has the least variation across higher and lower income neighborhoods compared to the other cities. These patterns are consistent with the histograms in Fig. 6.3 shown previously.

### 6.0.3.3 Infrastructure equity from the view of historical redlining regions

Finally, Fig. 6.11 shows the overall infrastructure deficiency across redlining grades from A to D. In Dallas, the author observes a trend of infrastructure becoming worse from areas with “excellent” HOLC ratings (Grade A) to areas classified as “hazardous” (Grade D) that were not eligible for mortgages. Los Angeles, however, shows the opposite pattern where grade A areas have higher deficiencies and areas that had lower grades have much lower

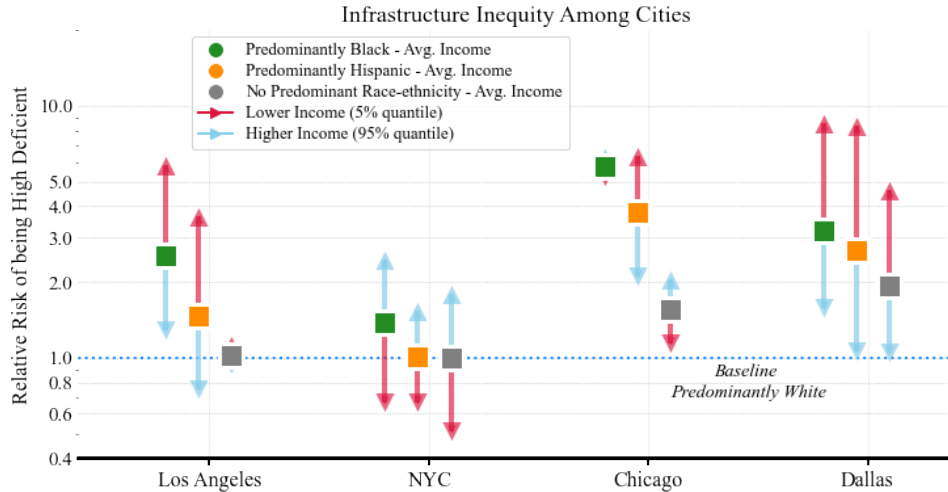


Figure 6.10: The estimated relative risks of having highly deficient infrastructure given neighborhood’s income, race-ethnicity, and city. Baseline reference is a neighborhood with an income level at 50% quantile of the distribution and no predominant race-ethnicity. The reach of red arrow and purple arrow indicates the relative risk if neighborhoods are at 5% (lower income) and 95% (higher income) of income quantiles respectively.

deficiency. For New York City and Chicago, the pattern is less obvious compared to Dallas and Los Angeles.

#### 6.0.4 Discussion

##### 6.0.4.1 Car ownership and access-related infrastructure deficiencies

The previous analyses do not consider the potential impacts of car ownership, which can be used as an estimate of residents’ extended mobility due to private vehicle transport. The U.S. Census Bureau provides summarized statistics on the number of vehicles owned per household at the scale of Census block groups. The percentage of car ownership can then be computed for each block group as the ratio between the number of households that own at least one car to the total number of households within the same block group. The data can be acquired from American Community Survey Table B25044.

In current infrastructure assessment framework, car ownership is not included because the private vehicle itself is not part of the infrastructure system. On the other hand, residents with sufficient resources can determine whether they need vehicles based on their personal

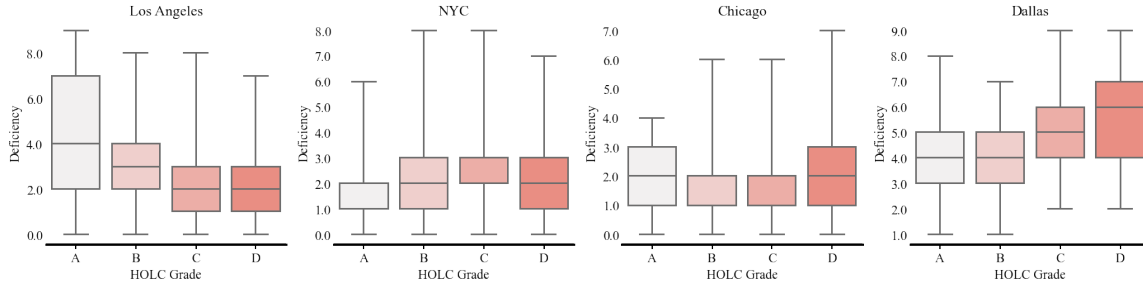


Figure 6.11: The distribution of overall infrastructure deficiency across historical redlining regions. The assignment of HOLC grades is 80% of mortgage value for Grade A, 60%-80% for Grade B, and 15% for Grade C, or was not eligible for any mortgages for Grade D.

lifestyles, which may potentially be affected by the overall infrastructure condition. One possible impact of being in a neighborhood without sufficient infrastructure support is that residents may have no access to community facilities (for example, grocery stores, banks, clinics, parks, and so on). Under these circumstances, neighborhoods with high car ownership may be able to greatly mitigate such negative influences because the majority of residents can still travel further with vehicles to reach similar facilities.

Fig. 6.12 shows heatmaps of infrastructure deficiency and car ownership across the four studied cities. It shows that Dallas, Los Angeles, and Chicago have relatively high car ownership compared to New York City. Despite the lack of access-related infrastructure across all four cities, the cities with high car ownership may experience less impact as opposed to cities that have relatively low car ownership such as New York City which has been known for having low car ownership in some boroughs.

To investigate this finding further, the author divides New York City into smaller areas following borough administrative boundaries (Bronx, Queens, Manhattan, Brooklyn, and Staten Island). The histogram of each borough's infrastructure deficiency, together with its car ownership, is shown in Fig. 6.13. Among the five boroughs, Manhattan has the lowest car ownership percentage (20%) but it has the best average infrastructure condition. Even though people do not have cars, the borough only has up to four deficient infrastructure types, which indicates that Manhattan is a highly walkable area where people have access to many facilities and resources.

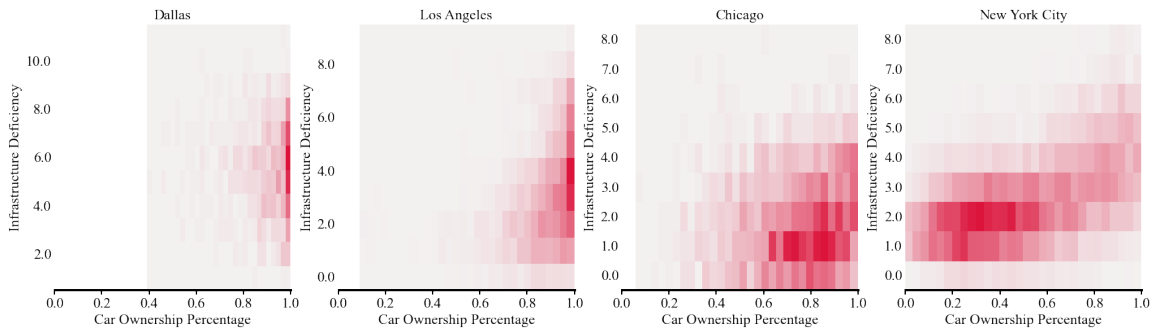


Figure 6.12: Heatmap of neighborhoods given car ownership levels and overall infrastructure deficiency.

On the other hand, Staten Island has the highest car ownership (average 80%) and the worst average infrastructure deficiency among all boroughs. However, neighborhoods within Staten Island may not be heavily affected by such infrastructure deficits due to their high car ownership percentage. This analysis shows that despite neighborhoods receiving a high infrastructure deficiency score, having high car ownership can, to some degree, reduce the negative impacts. However, having high car ownership is not equivalent to having high mobility across the neighborhoods. A fraction of Staten Island residents have no cars and can still struggle if they live in areas with high infrastructure access deficits. Therefore, future public transit infrastructure planning should consider the role of car ownership and provide transportation alternatives for those in low-transit areas who lack cars. This also pinpoints a possible direction of future research to better understand the societal and behavioral impacts of infrastructure deficits on communities.

#### 6.0.4.2 Infrastructure inequity and racial segregation

Fig. 6.14 shows the spatial distribution of predominant race-ethnicity groups across four cities. Except for New York City, the author observes various degrees of racial segregation in the rest of the cities. In Los Angeles, large areas of the city are *predominantly White* neighborhoods (shown in blue), while the majority of the remaining neighborhoods are *predominantly Hispanic* and are located on both the south and north side of Los Angeles. In Chicago, the location of each race-ethnicity group is more spatially separated than LA, with

most of the *predominantly White* neighborhoods in the northern region, *predominantly Black* neighborhoods in the southern region, and *predominantly Hispanic* neighborhoods between them. A similar pattern of segregation can be seen in the City of Dallas, where *predominantly Black* and *predominantly White* neighborhoods are clustered in the southern and northern parts of the city, respectively, and predominantly Hispanic neighborhoods are on the west and east sides of the city.

It is interesting to note that Chicago and Dallas both experience the highest levels of infrastructure inequity (Fig. 6.5 and Fig. 6.10) and the greatest racial segregation, as indicated by the highly distinct race/ethnicity areas in Fig. 6.14. While the findings do not try to establish a definitive connection between the two phenomena, it seems possible that the higher cost of living in LA and NYC could have driven more integration of historically segregated areas, thereby improving neighborhood infrastructure (e.g., through the forces of redevelopment and gentrification). This hypothesis requires further research with more cities, as well as tracking infrastructure and race/ethnicity patterns in the past compared to today. Unfortunately identifying such trends would be difficult given a lack of available historical infrastructure data. However, historical aerial maps, city records, and Census data could yield some clues as to the generalizability of these findings.

### 6.0.5 Conclusions

Evaluating infrastructure equity across multiple cities poses many challenges due to the difficulties of identifying a collection of measurable infrastructure types with sufficient comparable datasets for each location. In addition, in some scenarios (e.g., different sidewalk assessment approaches in this study), researchers may need to adjust the approach to derive comparable data, which requires more time and effort. Fortunately, infrastructure assessment framework developed in Chapter 2 is sufficiently generalizable and automated to ease this process.

This chapter describes how the framework generalized for multiple cities through implementation in a Cloud-based platform called Clowder. The platform also allows users to upload, store, and share data with other researchers, planners, city staff, etc. The ease and rapid implementation of the Cloud-based framework allows users to analyze the city of their

interests more efficiently with available datasets.

This chapter applied the new Cloud-based framework to four major cities located across the United States: Los Angeles, New York City, Chicago, and Dallas, and observed varying levels of infrastructure inequity across both income and race-ethnicity characteristics for 10 infrastructure types. The statistical analyses show that the prevalence of infrastructure inequity across all cities and the levels of such inequity are more obvious when there is pre-existing racial segregation within the region. In all four cities, the statistical analyses show that *predominantly Black* and *Hispanic* neighborhoods are at a higher risk of having highly deficient infrastructure compared to predominantly White neighborhoods and those with no predominant race/ethnicity.

Although neighborhoods with higher income generally have reduced risks, some cities (e.g., LA and portion of NYC) exhibit a pattern where higher-income neighborhoods have more deficient infrastructure than lower-income neighborhoods. This situation is likely geographically dependent, with more wealthy areas being more sparsely populated and located further from “clusters” of infrastructure facilities (banks, grocery stores, hospitals, and so on). However, these access issues may cause few difficulties for residents of Los Angeles and New York City due to a high percentage of car ownership in wealthy areas,

Several limitations must also be acknowledged. The number of neighborhood infrastructure types considered in this chapter is primarily determined by their data availability. This may introduce bias to the overall infrastructure deficiency due to the absence of other important neighborhood infrastructure types within the framework such as crosswalks, street tree canopy, and water and sewer pipes. Despite considering more infrastructure types can offer a more comprehensive, broader view of overall infrastructure condition, the associated cost of collecting all measurable data is challenging to achieve for multiple cities. In addition, the existing datasets do not represent the condition of the city at the same period which may also introduce bias to the resulting patterns. However, it is important to realize that this study provides a comprehensive comparison of infrastructure conditions across cities rather than highlighting street-level issues. The findings also raise the awareness for future improvements/investments as neighborhoods with high infrastructure deficits are identified. Further investigation is needed to study the relationship between the city’s geographic, and demo-

graphic information and neighborhood infrastructure, this leads to a more efficient process of choosing significant infrastructure types and data preparation. With the growing availability of fine-grained infrastructure-related data and a better understanding of neighborhood infrastructure, the framework can provide a consistent, national-level analysis.



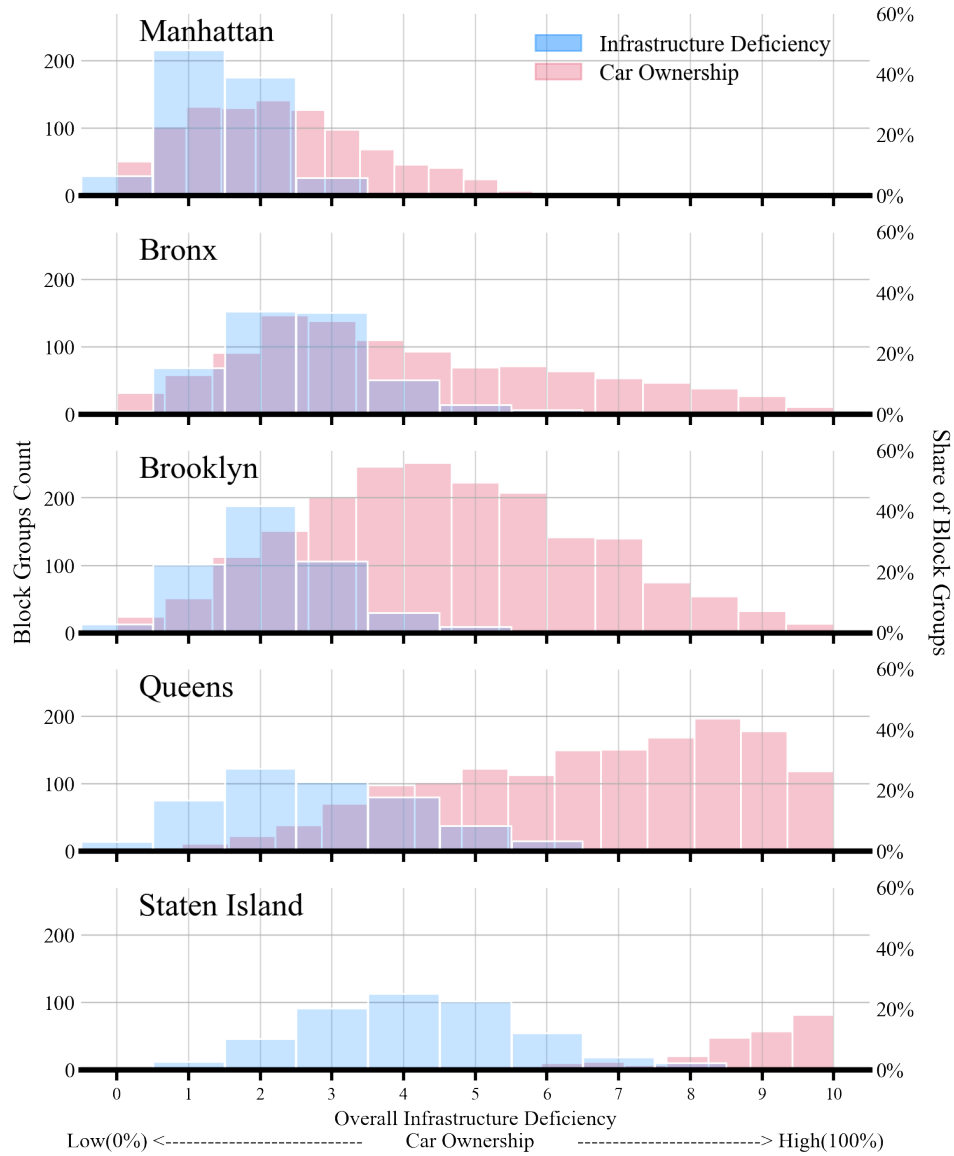


Figure 6.13: The distributions of infrastructure deficiency, car ownership percentage by boroughs in New York City.

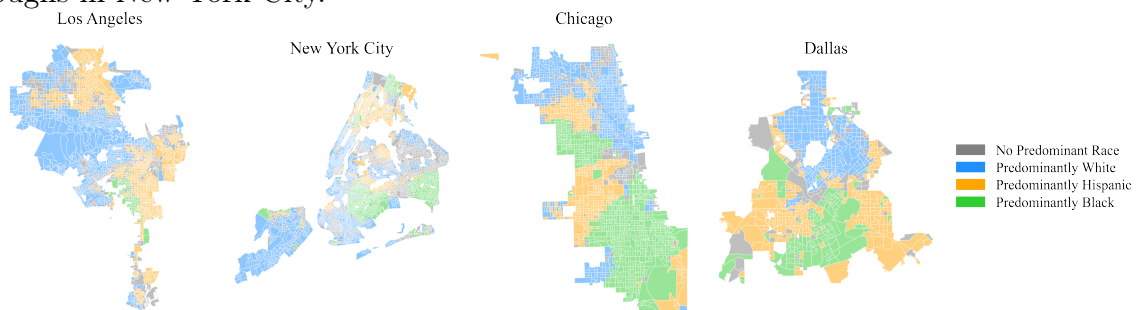


Figure 6.14: Maps of Los Angeles, Chicago, Dallas, and New York City show the existence of racial segregation.

## Chapter 7

### Summary, limitations, and future work

#### 7.0.1 Discussion

Assessing the condition of neighborhood infrastructure is essential to understanding infrastructure equity. Past studies of neighborhood infrastructure primarily focus on individual types of infrastructure and rarely provide a quantitative solution for multiple infrastructure types. Undoubtedly, fully assessing neighborhood infrastructure is a challenge due to the variety of infrastructure types and the difficulty of obtaining neighborhood-scale measurements across a city. This thesis bridges the gap between understanding infrastructure equity and the need for a systematic, data-driven approach to assessing multiple neighborhood infrastructure types.

More specifically, the author first develops a generalized assessment framework that considers multiple infrastructure types. The framework is implemented in the City of Dallas as a case study that reveals the existence of infrastructure deserts and widespread infrastructure inequity across the neighborhood income and race/ethnicity. To address these types of inequities, long-term investments are needed to improve infrastructure in low-income and Black/Hispanic areas. Investment prioritization based on asset conditions and economic impacts [137] is one popular approach for infrastructure management that could be used to foster healthier and more equitable communities.

A notable finding from this study is that sidewalk and street tree canopy deficits are more widespread across the whole city than other neighborhood infrastructure types, which suggests immediate opportunities for improvement, particularly with increasing urban warming under climate [138] resulting in calls for massive tree planting [86, 139].

Chapter 5 delivers comprehensive insights on infrastructure inequity across income and race-ethnicity, statistically showing varying levels of inequity among neighborhoods with dif-

ferent income and race compositions. The statistical inference indicates that *predominantly Black* and *Hispanic* neighborhoods are 1.4 to 2.6 times and 2.0 to 3.6 times likelier to have *highly deficient* infrastructure (8 or more deficient infrastructure types out of 12) than areas with no predominant race-ethnicity, respectively. Furthermore, these disparities reflect the legacy of historical discriminatory housing policies (“redlining”) and their long-term impacts on neighborhood infrastructure. Neighborhoods marked as “less desirable” for mortgages during the 1930s still experience significantly higher infrastructure deficiencies today.

Lastly, in Chapter 6, the author automates the assessment framework and integrate it with a Cloud-based platform called Clowder so that it becomes more accessible and easy to implement for researchers and others. The platform allows users to upload, store, and share data with other users, thus providing ready sharing among teams. The ease and quick implementation of the Cloud-based framework also allows users to analyze any city of interest more efficiently with available datasets.

The author demonstrates this generalization of the framework through application to three other case studies across the United States: Los Angeles, New York City, and Chicago. One important contribution of Chapter 6 is the comparative analysis of infrastructure equity across multiple cities considering 10 common infrastructure types. The statistical analyses show that the prevalence and levels of infrastructure inequity across all cities appear to be greater when there is pre-existing racial segregation within the region.

Further analyses show that predominantly Black and Hispanic neighborhoods are at a higher risk of having highly deficient infrastructure in all four cities. Although neighborhoods with higher income usually have reduced risks, some higher-income neighborhoods (e.g., in Los Angeles) have more deficient infrastructure than lower-income neighborhoods. This is likely due to access deficiencies in sparsely populated and geographically remote high-income areas, which are mitigated by high car-ownership levels.

## 7.0.2 Limitations and Future Work

### 7.0.2.1 *Limitations of the framework*

This thesis has several limitations. First, the spatial representation of neighborhoods is challenging and has been addressed in multiple ways [65, 67–69, 140]. Despite the pop-

ularity that Census tracts or block groups have received, there is no definitive argument claiming which is the best spatial neighborhood boundary among all available options such as Census tracts, block groups, or zip codes [67–69]. Past studies have shown that the types of geographic boundaries used to aggregate data can affect variance, standard deviations, correlation, and regression analyses [67]. However, since this thesis aims to explore spatial patterns of infrastructure conditions at the city level from a relative risk perspective, choosing the Census block groups as the representation of neighborhoods provides ready access to social-economic characteristics. Further research is needed to more deeply explore whether spatial boundaries of neighborhood analyses introduce biases in the results.

A second limitation is uncertainties in the criteria for measuring substandard infrastructure components. Every criterion was developed based on prior studies, practical design guidelines, or community surveys. However, access measures developed using GIS procedures may fail to account for the actual quality of and access to infrastructure (e.g., healthcare facilities) [65, 141]. For example, residents may access facilities that are not necessarily near their neighborhoods, potentially due to social networks, transportation availability, or perceptions of crime and safety [141]. Hence relying only on proximity without considering social aspects of neighborhoods can result in misinformation on infrastructure availability. Similarly, the weighting scheme (currently equally weighted) for multiple infrastructure types could be modified to better represent neighborhoods’ needs or city preferences. The choices of weights and thresholds may be different from city to city. Exploring the sensitivity of outcomes to these assumptions is recommended for future research.

Another issue is the limits of accuracy and validity of using fixed travel distances to measure infrastructure accessibility. Recently, an increase in GIS implementation has led to improvements in measuring the accessibility of activity locations [142–146]. The gravity model-based method [147–149] calculates accessibility based on zones as a function of activity opportunity attractiveness and the travel distance between other zones and the individual’s resident zones. It becomes one of the most popular methods to measure accessibility because of the ease of interpretation and robustness of model extensions [150, 151].

Nonetheless, fixed distance approaches, such as those implemented in this study, remain favorable in many infrastructure-related studies due to their simple intuition and easy

implementation. However, the choice of "proper" distance is mostly empirical and lacks theoretical justification. For instance, the critical distance used in assessing healthcare services is 2-mile (3.2km) for major hospitals and 1-mile (1.6km) for walk-in clinics and urgent care [94,95]. However, many factors could affect people's accessibility to these destinations, such as travel behaviors, transportation mode, and city development, resulting in different values of suggested critical distances for accessibility assessment [152]. Despite these inevitable uncertainties, the criteria chosen for this case study are sufficient for a comparative assessment of infrastructure equity across multiple infrastructure types. Future research is needed to perform sensitivity analyses on the impacts of these assumptions.

Finally, a full and complete assessment of neighborhood infrastructure should involve six primary categories: connective infrastructure, protective infrastructure, socio-economic structures, water and sanitation lifelines, energy lifelines, and communication lifelines [1]. In this study, 12 infrastructure types were considered that included four of the six categories, excluding energy and water sanitation lifelines. With additional data availability, more infrastructure types such as stormwater drains, water supply and wastewater pipes, and street lights will undoubtedly add to the story of complex, interdependent dynamics among neighborhood infrastructure. The framework proposed in this study can easily be expanded to include other infrastructure types as data are available, providing the capacity to measure conditions of a wide range of infrastructure types systematically.

#### *7.0.2.2 Limitations of future infrastructure-related applications*

Although the assessment framework is generalizable to be applied in other cities, it still poses challenges to ensuring consistent substandard infrastructure criteria and measurements across different regions, potentially limiting the number of infrastructure types that can be considered. It has been a common practice for each city to have its own ways of collecting and managing infrastructure-related datasets. This typically leads to inconsistent condition attributes for the same or similar aspects of structures or facilities. For instance, for sidewalk inspection, the city of Dallas provides attributes of each inspected sidewalk segment and the level of inspection provides data on sidewalk segment obstruction or damage.

On the other hand, although Chicago also has a sidewalk dataset, the included attributes

do not provide such detailed information. The information gap between two different measuring practices may prevent the calculated substandard metrics from accurately describing conditions between regions. Although in this thesis, the author minimized such inconsistency by transforming datasets into comparable formats, this issue will inevitably be raised again as the framework is continued to be applied to more case studies across the country. Therefore, researchers have to be aware of the datasets being used in the analysis and try to avoid any inconsistent condition representations when preparing the data. To make such cross-city comparisons more viable, establishing standard practices and metadata for assessing neighborhood infrastructure and storing the data would be of great value for future research.

The current framework also relies on the quality of collected data and does not have built-in outliers detection to automatically detect and exclude abnormal records from the infrastructure data. Although the aggregation process from street level to neighborhoods using a 50% cutoff mitigates the impacts of outliers, it is still possible for errors to occur if outliers are prevalent and closely located in a few areas. Manual inspection of the data quality is recommended to minimize the impacts of such errors such as outliers and abnormal data points.

Finally, the binary deficiency indicator (defined as  $\mu$  in Eqn. 2.2) treats the infrastructure condition as binary values, which may not fully reflect the magnitude of infrastructure deficits. For instance, a neighborhood having 99% substandard sidewalks is equivalent to a neighborhood that has 51% substandard sidewalks because both have more than 50% of sidewalks found to be substandard. To address this issue, the percentage of substandard infrastructure could be used directly in the framework. The trade-off of this change would be a less intuitive and more complex measure overall infrastructure condition and the integration of multiple infrastructure types as a single metric may require more interpretation.

### *7.0.2.3 Future work*

Despite the limitations noted above, this thesis takes the first step to considering neighborhood infrastructure as an integrated system that involves multiple infrastructure types and assesses infrastructure conditions with data-driven approaches. The findings have im-

portant policy implications and lessons for cities and developers that are promoting equitable infrastructure. Much progress has been made on this front in Dallas, with the Dallas Sidewalk Replacement Program [153], Urban Forest Management Plan [154], and other initiatives to improve neighborhood infrastructure. However, as the findings of this thesis suggest, infrastructure inequities persist across income and race-ethnicity lines and planners and policymakers should address these issues to close the "infrastructure gap." In addition to prioritized investments in disadvantaged neighborhoods, community engagement is also needed to better understand the impact of the lack of infrastructure on residents and develop smart and effective strategies for promoting neighborhood infrastructure that better meets neighborhood needs. For example, new designs of infrastructure, such as complete streets [155–157], may better meet resident needs than the installation of previous standards.

It is also beneficial to consider a more complete inventory of historical investments to show whether the investments address existing infrastructure issues and are moving towards a more equitable future. The framework also needs to be adaptive to accommodate the evolution of certain infrastructure types. For example, with advancements in high-speed internet and cellular networks, many residents spend more time on virtual activities such as virtual grocery shopping and banking. This may decrease the actual need for residents to visit physical stores and banks, physical access for those facilities may become less important.

Another opportunity for future research would be to develop relative risk models that include other parameters, such as redlining or neighborhood age. In terms of infrastructure data sources, identifying noise walls would be considerably easier if a machine learning model were trained to identify noise walls from the street images labeled in all four cities considered in this study. However, as long as some residents lack high-speed or any internet access, these measures should remain a consideration. Finally, the author hopes this proposed Cloud-based framework gathers more interest in supporting neighborhood infrastructure and helps to establish a nationwide understanding of infrastructure equity via more case studies across the country.

## APPENDIX

### A.1 Supplementary Tables

Table A.1: Dataset information for individual infrastructure type.

	Dataset Source	Data Year	Notes
Pavement	City of Dallas REST Service <sup>1</sup>	2018	Polyline
Crosswalk <sup>2</sup>	Object detection using Google Satellite images on residential intersections	2019	Point
Noise Wall	Annotated dataset using Google StreetView images along state highways	2019	Point
Internet Service	Federal Communication Commission <sup>3</sup> Broadband width map	2016	point
Bank Access	Bank branches locations from NCTCOG <sup>4</sup> regional data center	2019	Point
Medical Facility Access	Major hospitals, urgent care or clinics from NCTCOG data center and Yelp search listings	2018	Point
Public Transportation Access	Bus stops, rail stations locations from Dallas Area Rapid Transit (DART)	2018	Point
Gathering Place Access	Public parks, libraries, farmer markets and community centers extracted from NCTCOG data center, tax parcel data	2019	Point
Food Access	Food stores (grocery stores, wholesale) locations from NCTCOG data center	2019	Point
Trail Access	Bike pedestrian trails from Dallas GIS Service website	2019	Polyline
Street Tree Canopy	Tree coverage from Smart growth for Dallas <sup>5</sup>	2018	Polygon
Sidewalk	City of Dallas REST Service - Public Works	2019	Polygon

1:<https://gis.dallascityhall.com/wwgis/rest/services/>

2:Crosswalks' locations are predicted using object detection model (YOLOv3), which determines if a satellite image of intersection contains crosswalks

3:<https://www.fcc.gov/reports-research/maps/>

4:[https://web.tplgis.org/smart\\_growth\\_dallas/](https://web.tplgis.org/smart_growth_dallas/)

5:[https://web.tplgis.org/smart\\_growth\\_dallas/](https://web.tplgis.org/smart_growth_dallas/)



Table A.2: Descriptive statistics for the substandard percentage ( $\mu$ ) of individual infrastructure type.

Infrastructure Type	Census Block Groups (n)	Standard Deviation	Min	25%	50%	75%	Max
Pavement	790	0.150	0.000	0.136	0.246	0.346	0.794
Crosswalk	790	0.281	0.000	0.000	0.333	0.500	1.000
Noise Wall	70	0.446	0.000	0.013	1.000	1.000	1.000
Internet Service	790	0.195	0.100	0.100	0.300	0.500	0.700
Bank Access	790	0.439	0.000	0.000	0.458	1.000	1.000
Medical Facility Access	790	0.427	0.000	0.066	0.835	1.000	1.000
Public Transportation Access	790	0.322	0.000	0.000	0.138	0.473	1.000
Gathering Place Access	790	0.349	0.000	0.000	0.114	0.552	1.000
Food Access	790	0.416	0.000	0.029	0.529	1.000	1.000
Trail Access	790	0.361	0.000	0.341	0.765	1.000	1.000
Street Tree Canopy	790	0.131	0.286	0.763	0.856	0.943	1.000
Sidewalk	790	0.145	0.101	0.768	0.874	0.933	1.000

Table A.3: Description of the Shapefile consisting information of all assessed infrastructure types.

Filename	Infrastructure_assessment_Dallas.zip
File Format	ArcGIS Shapefile (zipped)
Attributes contains in the shapefile	
BLOCKGROUP	12 digits Census block Group ID.
Income3	Categorical income class based on tertiles: Low, Middle, High.
IncomeLog	Log value of annual household median income.
Overall_IF	Overall infrastructure deficiency - integer.
IF_5	Categorical overall infrastructure deficiency: Excellent, Good, Moderate, Deficient, Highly Deficient.
PCNG_PAVE	Percentage of substandard pavement segments.
PCNG_SDWK	Percentage of residential street segments that has substandard sidewalks.
PCNG_CRWK	Percentage of intersections that do not have crosswalk present.
PCNG_MEDL	Percentage of residential households that don't have access <sup>1</sup> to medical service facilities.
PCNG_GATH	Percentage of residential households that don't have access to gathering places.
PCNG_BANK	Percentage of residential households that don't have access to local bank branches.
PCNG_INTT	Percentage of residential households with substandard internet service.
PCNG_TRIL	Percentage of residential households that don't have access to bicycle & pedestrian trails.
PCNG_TRAN	Percentage of residential households that don't have access to bus stops nor rail stations.
PCNG_TREE	Percentage of residential street segments with substandard tree canopy percentage (below 25%).
PCNG_NSWL	Percentage of residential households near highways that do not have noise wall present.
Geometry	Geometry of census block group.

<sup>1</sup>:Based on corresponding substandard criteria table (see corresponding Table for more details).

Table A.4: Infrastructure Data Source Table for Los Angeles, New York City, Chicago, and Dallas (considering 10 infrastructure types).

Infrastructure Type	Los Angeles	New York City	Dallas	Chicago
Pavement	LA Open Data <sup>1</sup> - "Road Surface Condition Map"	NYC Open Data <sup>2</sup> - "Street Pavement Rating"	City of Dallas GIS Service <sup>3</sup> - "pavement condition"	Chicago Metropolitan Agency for Planning <sup>4</sup> (CMAP)
Sidewalk	LA Geohub <sup>5</sup> - "Sidewalks"	NYC Open Data - "Sidewalk"	City of Dallas Department of Public Works, City of Dallas GIS Service - "sidewalks"	Chicago Data Portal <sup>6</sup> - "Sidewalks"
Internet	Federal Communications Commission <sup>7</sup> - "Fixed Broadband Deployment"	Federal Communications Commission - "Fixed Broadband Deployment"	Federal Communications Commission - "Fixed Broadband Deployment"	Federal Communications Commission - "Fixed Broadband Deployment"
Noise wall	Google Streetview imagery	Google Streetview imagery	Google Streetview imagery	Google Streetview imagery
Bank Access	LA Geohub - "Banking and Finance"	NYC-Tax-Parcels <sup>8</sup> - "code [06,07,K7]"	City of Dallas GIS Service - "tax Appraisal parcels"	Cook County Open Data <sup>9</sup> - "historical parcels - 2019"
Public Transportation Access	LA Geohub - "rail lines and stop benches"	NYC Open Data - "subway stations, bus stop shelters"	City of Dallas GIS Service - "rails, busstops"	Chicago Data Portal - "CTA_Rail Lines, CTA_Bustops"
Medical Facility	LA Geohub - "hospitals and medical centers"	NYC Tax Parcels - "code [I, I1, I5]"	NCTCOG's Regional Data Center <sup>10</sup> ; Yelp search <sup>11</sup> - "urgnet care"	Chicago Data Portal - "Hospitals", "Neighborhood health clinics"
Bike & Pedestrian Trails	LA Geohub - "trails"	NYC Open Data - "Parks Trails"	CMAP data hub - "Bikeway inventory System (BIS)"	City of Dallas GIS Service - "trails"
Gathering Places	LA Open Data - "Road Surface Condition Map"	NYC Open Data - "library", NYC Tax Parcels - "code [Q1, P5]"	City of Dallas GIS Service - "parks, community centers, farmet markets, libraries", Google Static Map API	Chicago Data Portal - "parks, community service centers, farmer markets, libraries"
Food Access	LA Geohub - "grocery stores, farmer markets"	Open NY <sup>12</sup> - "Retail Food Stores"	NCTCOG's Regional Data Center	Chicago Data Portal - "grocery stores, farmer markets"

1:LA Open Data. <https://data.lacity.org>

2:NYC Open Data. <https://opendata.cityofnewyork.us>

3:City of Dallas GIS Service. <https://gis.dallascityhall.com/shapefileDownload.aspx>

4:CMAP data hub. <https://datahub.cmap.illinois.gov/dataset>

5:LA Geohub. <https://geohub.lacity.org>

6:Chicago Data Portal. <https://data.cityofchicago.org>

7:FCC Fixed Broadband Deployment.. <https://broadbandmap.fcc.gov>

8:NYC Tax Parcels. <https://gis.ny.gov/gisdata/inventories/details.cfm?DSID=1300>

9:Cook County Open Data. <https://datacatalog.cookcountyil.gov>

10:NCTCOG's Regional Data Center. <https://www.nctcog.org/regional-data>

11:Yelp Search. <https://www.yelp.com>

12:Open NY. <https://data.ny.gov>

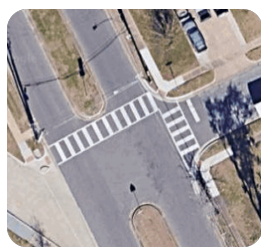
## A.2 Crosswalk detection

The recommended locations for crosswalk installation depend on many factors such as speed limit, street type, and traffic volume [92]. To identify crosswalk deficiencies, the first step is to find residential intersections (called “proposed intersections”) that should have crosswalks installed according to available design guidelines. Next, a crosswalk detection model is trained and executed on satellite images at each “proposed intersection”. Finally, each proposed intersection is evaluated to identify whether it is deficient (i.e., lacking a crosswalk). This information is then passed into the infrastructure assessment framework described in Chapter 2.

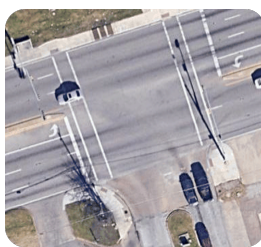
In Step 1, each intersection is classified as a “proposed intersection” if it meets any of the following conditions based on the design guidelines [92]:

- 
1. The street speed limit is not less than 40 mph
  2. The intersections contain traffic lights
  3. The intersections are within school zones
- 

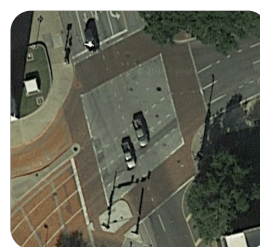
For Step 2, an object detection model called YOLOv3 [158] was trained to identify any crosswalks from the satellite images. YOLOv3 is a deep convolutional neural network for detecting objects and their positions on the image as bounding boxes. The output of the model gives the coordinates of detected crosswalks. It has been shown to have the benefits of both fast prediction and good performance. To train the YOLOv3 model, images of 120



**Zebra stripes style**



**Parallel lines style**



**Brick style**

Figure 7.1: Three types of crosswalks in City of Dallas.

intersections with crosswalks were manually collected as training data. The data include three types of crosswalks in the City of Dallas: zebra stripes, parallel lines style, and brick style (Fig. 7.1). Because zebra stripes crosswalks are the most commonly observed crosswalk type in the city, parallel line crosswalks are less common, and brick style crosswalks are only found in downtown areas, the percentage of the three types in the ground truth dataset is 70%, 20%, and 10% respectively. After adding false images with no crosswalks, a total of 417 images are used for training. The trained model achieved an overall  $f1$ -score (Equation 7.1) of 0.8.

$$f1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7.1)$$

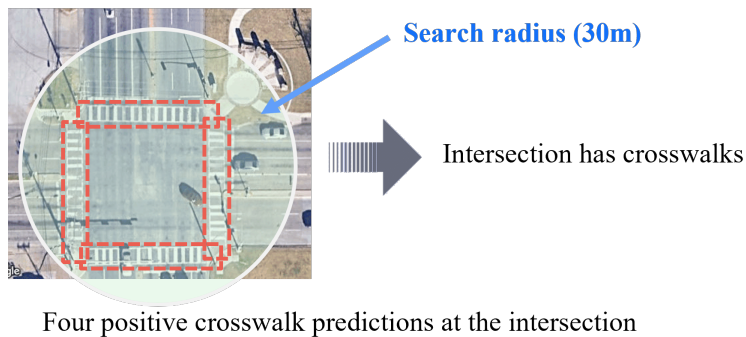


Figure 7.2: Checking crosswalk presence at an intersection using a 30m search radius.

It is worth noting that more than one crosswalk may be at one intersection. For the purposes of identifying deficiencies, the author defines crosswalk existence as at least one crosswalk existing at the intersection. To avoid duplicate counting at intersections with more than one crosswalk, a 30m-radius circle (shown in Fig. 7.2) is drawn around the intersection to check whether it contains any positive predictions (crosswalk identified) from the YOLOv3 model. If there is at least one positive crosswalk detection within the radius, the intersection

is considered to have crosswalks. Otherwise, the intersection is noted as lacking crosswalks if there is no positive crosswalk prediction within the radius. For the Dallas case study, a total of 2972 “proposed intersections” were found and 1728 intersections were identified as having crosswalks.

### A.3 Pseudo-code of the method used to compute 12 deficient infrastructure types

---

**Algorithm 1** Pavement

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2:

**for** each *Neighborhood*( $i$ ) **do**

Find all pavement segments  $C_i$  within/intersect with the neighborhood boundary.

Initialize *Measurable Components*  $M = 0$ ;

Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each segment  $C_{ij}$  **do**

Calculate segment length  $L_{ij}$ .

Count *MeasurableComponents* in length  $M = M + L_{ij}$ .

**if** segment  $C_{ij}$ 's *Pavement Condition Index*  $\leq 55$  **then**  $M_{std} = M_{std} + L_{ij}$

**end if**

Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

---

**Algorithm 2** Sidewalks

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2:

**for** each *Neighborhood*( $i$ ) **do**

Find residential street segments  $C_i$  within/intersect with the neighborhood boundary.

Initialize *Measurable Components*  $M = 0$ ;

Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each street segment  $C_{ij}$  **do**

Calculate segment length  $L_{ij}$ .

Count *MeasurableComponents* in length  $M = M + L_{ij}$ .

**if** segment  $C_{ij}$ 's has no sidewalks on both side **then**  $L_{missing} = L_{ij}$

**else if** **then**  $L_{missing} = 0$

**end if**

**if** segment  $C_{ij}$  has sidewalk on at least one side **then** Calculate the portion length ( $L_{std}$ ) of segment that has been obstructed or damaged sidewalks;

$M_{std} = M_{std} + \max(L_{missing}, L_{std})$

**end if**

Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

---

**Algorithm 3** Noise Wall

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2:

**for** each *Neighborhood*( $i$ ) **do**

Find residential households  $C_i$  located within 200 feet (61m) from major highways.

Initialize *Measurable Components*  $M = 0$ ;

Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each segment  $C_{ij}$  **do**

Count *Measurable Components*  $M = M + 1$ .

**if** no noise walls existed within 200 feet (61m) from  $C_{ij}$  **then**  $M_{std} = M_{std} + 1$

**end if**

Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---



---

**Algorithm 4** Crosswalks

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2:

**for** each *Neighborhood*( $i$ ) **do**

    Within Neighborhood boundary, find all crosswalk intersections  $C_i$  intersections that are either:

    1) Intersections between residential streets

    2) Intersections between school zones.

    Initialize *Measurable Components*  $M = 0$ ;

    Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each crosswalk intersection  $C_{ij}$  **do**

        Create a search buffer region (34m radius)  $b_{ij}$  given its coordinates.

        Count Measurable Components  $M = M + 1$

**if** no crosswalks existed within  $b_{ij}$  **then**  $M_{std} = M_{std} + 1$

**end if**

        Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

---

**Algorithm 5** Street Tree Canopy Coverage

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2:

**for** each *Neighborhood*( $i$ ) **do**

    Final all street segments  $C_i$  within the neighborhood

    Create street buffer polygons  $C'_{ij}$  (use city-wide median width: 6.5 feet or 2 meter radius)

    Initialize *Measurable Components*  $M = 0$ ;

    Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each street polygon  $C_{ij}$  **do**

        Count Measurable Components  $M = M + 1$

        Compute the area of street polygon  $A_{ij}$ .

        Compute the area of the tree canopy  $A_{ij}^t$  within  $C'_{ij}$ .

        Compute the street tree canopy percentage as  $p_{ij} = \frac{A_{ij}^t}{A_{ij}}$ .

**if**  $p_{ij} \leq 0.25$  **then**  $M_{std} = M_{std} + 1$

**end if**

        Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

---

**Algorithm 6** Pedestrian & bicycle trail access

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2: Break the pedestrian & bicycle trails into points using 600-meter intervals.

Use points to create service area  $S$  for pedestrian & bicycle trails (0.8 km travel distance).

Step-3:

**for** each *Neighborhood*( $i$ ) **do**

Find all residential households  $C_i$  within the neighborhood.

Initialize *Measurable Components*  $M = 0$ ;

Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each residential household  $C_{ij}$  **do**

Count Measurable Components  $M = M + 1$

**if**  $C_{ij}$  is spatially outside of  $S$  **then**  $M_{std} = M_{std} + 1$

**end if**

Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

---

**Algorithm 7** Medical Facility Access

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2: Create service area  $S_1$  for major hospitals (2-mile or 3.2 km travel

Create service area  $S_2$  for walk-in clinics and urgent care (1-mile or 1.6 km travel distance).distance).

Step-3:

**for** each *Neighborhood*( $i$ ) **do**

Find all residential households  $C_i$  within the neighborhood.

Initialize *Measurable Components*  $M = 0$ ;

Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each residential household  $C_{ij}$  **do**

Count Measurable Components  $M = M + 1$

**if**  $C_{ij}$  is spatially not in  $S_1$  nor  $S_2$  **then**  $M_{std} = M_{std} + 1$

**end if**

Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

---

**Algorithm 8** Public transportation access

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2: Create service area  $S_1$  rail stations (0.8 km travel distance).

Create service area  $S_2$  for bus stops (0.4 km travel distance).

Step-3:

**for** each *Neighborhood*( $i$ ) **do**

Find all residential households  $C_i$  within the neighborhood.

Initialize *Measurable Components*  $M = 0$ ;

Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each residential household  $C_{ij}$  **do**

Count Measurable Components  $M = M + 1$

**if**  $C_{ij}$  is spatially not in  $S_1$  nor  $S_2$  **then**  $M_{std} = M_{std} + 1$

**end if**

Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

---

**Algorithm 9** Food access

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2: Create service area  $S$  for fresh food stores (1-mile or 1.6 km travel distance).

Step-3:

**for** each *Neighborhood*( $i$ ) **do**

Find all residential households  $C_i$  within the neighborhood.

Initialize *Measurable Components*  $M = 0$ ;

Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each residential household  $C_{ij}$  **do**

Count Measurable Components  $M = M + 1$

**if**  $C_{ij}$  is spatially not in  $S$  **then**  $M_{std} = M_{std} + 1$

**end if**

Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

---

**Algorithm 10** Bank access

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2: Create service area  $S$  for bank branches (1-mile or 1.6 km travel distance).

Step-3:

**for** each *Neighborhood*( $i$ ) **do**

Find all residential households  $C_i$  within the neighborhood.

Initialize *Measurable Components*  $M = 0$ ;

Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each residential household  $C_{ij}$  **do**

Count Measurable Components  $M = M + 1$

**if**  $C_{ij}$  is spatially not in  $S$  **then**  $M_{std} = M_{std} + 1$

**end if**

Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

---

**Algorithm 11** Gathering place access

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2: Create service area  $S_1$  for parks (1-mile or 1.6 km travel distance).

Create service area  $S_2$  for libraries (1-mile or 1.6 km travel distance).

Create service area  $S_3$  for community centers (1-mile or 1.6 km travel distance). Create service area  $S_4$  for farmers' markets (1-mile or 1.6 km travel distance).

Step-3:

**for** each *Neighborhood*( $i$ ) **do**

Find all residential households  $C_i$  within the neighborhood.

Initialize *Measurable Components*  $M = 0$ ;

Initialize *Substandard Measurable Components*  $M_{std} = 0$ .

**for** each residential household  $C_{ij}$  **do**

Count Measurable Components  $M = M + 1$

**if**  $C_{ij}$  is spatially not in  $S_1$  and  $S_2$  and  $S_3$  and  $S_4$  **then**  $M_{std} = M_{std} + 1$

**end if**

Calculate *Substandard Measurable Components Percentage* ( $\mu_i$ ) =  $\frac{M_{std}}{M}$

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

---

**Algorithm 12** Internet Service

---

Step-1: Initiate  $\theta$  as an empty array with size  $N \times 1$  ( $N =$  the number of neighborhoods).

Step-2: Uniformly dis-aggregate data from census tract level into neighborhood level (Note: internet data is only available at Census tract level)

Step-3:

**for** each *Neighborhood*( $i$ ) **do**

    Find attribute (*pcat<sub>all</sub>*) $S_i$  that represents the households with over 200 kbps in at least one direction

    Convert  $S_i$  into percentage measure  $s_i$  by taking the average of the range

    Calculate *Substandard Measurable Components Percentage* ( $\mu_i = 1 - s_i$ )

**if**  $\mu_i \geq$  **then**  $\theta_i = 1$

**else if** **then**  $\theta_i = 0$

---

#### A.4 Full cumulative logit model considering income, race-ethnicity, and city

$$\begin{aligned} \text{logit}[Pr(\gamma \leq j|x)] &= \alpha_j + \beta_I x_I + \beta_R^T \mathbf{x}_R + \beta_C^T \mathbf{x}_C + \\ &\quad x_I \beta_{IR}^T \mathbf{x}_R + x_I \beta_{IC}^T \mathbf{x}_C + \beta_{RC}^T \mathbf{x}_R \mathbf{x}_C + x_I \beta_{IRC}^T \mathbf{x}_R \mathbf{x}_C; \quad (7.2) \\ &\quad j = 1, \dots, J - 1 \end{aligned}$$

where:

$$\begin{aligned}
\beta_{\mathbf{R}}^T \mathbf{x}_{\mathbf{R}} &= \beta_H x_H + \beta_W x_W + \beta_B x_B \\
\beta_{\mathbf{C}}^T \mathbf{x}_{\mathbf{C}} &= \beta_N x_N + \beta_D x_D + \beta_C x_C \\
\beta_{\mathbf{IR}}^T \mathbf{x}_{\mathbf{R}} &= \beta_{I \times H} x_H + \beta_{I \times W} x_W + \beta_{I \times B} x_B \\
\beta_{\mathbf{IC}}^T \mathbf{x}_{\mathbf{R}} &= \beta_{I \times N} x_N + \beta_{I \times D} x_D + \beta_{I \times C} x_C \\
\beta_{\mathbf{RC}}^T \mathbf{x}_{\mathbf{RC}} &= \beta_{N \times H} x_N x_H + \beta_{N \times W} x_N x_W + \beta_{N \times B} x_N x_B + \\
&\beta_{D \times H} x_D x_H + \beta_{D \times W} x_D x_W + \beta_{D \times B} x_D x_B + \\
&\beta_{C \times H} x_C x_H + \beta_{C \times W} x_C x_W + \beta_{C \times B} x_C x_B \\
\beta_{\mathbf{IRC}}^T \mathbf{x}_{\mathbf{RC}} &= \beta_{I \times N \times H} x_N x_H + \beta_{I \times N \times W} x_N x_W + \beta_{I \times N \times B} x_N x_B + \\
&\beta_{I \times D \times H} x_D x_H + \beta_{I \times D \times W} x_D x_W + \beta_{I \times D \times B} x_D x_B + \\
&\beta_{I \times C \times H} x_C x_H + \beta_{I \times C \times W} x_C x_W + \beta_{I \times C \times B} x_C x_B
\end{aligned} \tag{7.3}$$

#### A.5 Estimated parameters of cumulative logit model for comparative study between Los Angeles, New York City, Chicago, and Dallas

Table A.5: Estimated coefficients of the full cumulative logit model.

	Value	Std. Error	<i>t</i> value
Coefficients			
$\beta_I$	-1.87	0.14	-13.74
$\beta_W$	0.82	2.31	0.36
$\beta_B$	-8.13	3.12	-2.61
$\beta_H$	-8.79	2.03	-4.32
$\beta_N$	-14.58	1.81	-8.07
$\beta_D$	-22.38	2.75	-8.15
$\beta_C$	-16.44	3.18	-5.17
$\beta_{I \times W}$	-0.07	0.20	-0.36
$\beta_{I \times B}$	0.65	0.29	2.26
$\beta_{I \times H}$	0.76	0.19	4.09
$\beta_{I \times N}$	1.35	0.16	8.28
$\beta_{I \times D}$	1.80	0.26	7.03
$\beta_{I \times C}$	1.67	0.29	5.77
$\beta_{N \times W}$	-12.50	2.66	-4.70
$\beta_{N \times B}$	9.97	3.41	2.92
$\beta_{N \times H}$	4.85	2.47	1.97
$\beta_{D \times W}$	10.36	3.90	2.66
$\beta_{D \times B}$	4.06	3.79	1.07
$\beta_{D \times H}$	0.80	4.11	0.20
$\beta_{C \times W}$	-3.71	4.13	-0.90
$\beta_{C \times B}$	3.58	4.22	0.85
$\beta_{C \times H}$	-4.54	4.25	-1.07
$\beta_{I \times N \times W}$	1.13	0.23	4.81
$\beta_{I \times N \times B}$	-0.87	0.31	-2.78
$\beta_{I \times N \times H}$	-0.40	0.22	-1.79
$\beta_{I \times D \times W}$	-0.85	0.35	-2.46
$\beta_{I \times D \times B}$	-0.38	0.36	-1.04
$\beta_{I \times D \times H}$	-0.09	0.39	-0.24
$\beta_{I \times C \times W}$	0.37	0.37	1.03
$\beta_{I \times C \times B}$	-0.38	0.39	-0.99
$\beta_{I \times C \times H}$	0.34	0.39	0.89
Intercepts ( $\alpha_j$ )			
$\alpha_1$	17.35	1.51	11.50
$\alpha_2$	19.33	1.51	12.81
$\alpha_3$	20.69	1.51	13.70
$\alpha_4$	21.72	1.51	14.37
$\alpha_5$	22.70	1.51	15.01
$\alpha_6$	23.59	1.51	15.59
$\alpha_7$	24.61	1.51	16.24
$\alpha_8$	25.81	1.52	17.02
$\alpha_9$	27.29	1.52	17.92
$\alpha_{10}$	29.58	1.59	18.55
Residual Deviance	40393.78	AIC	40475.78

## BIBLIOGRAPHY

- [1] G. F. f. D. Reduction and Recovery(GFDRR), “Post-Disaster Needs Assessments Guidelines Volume B Community Infrastructure,” Apr. 2017. 1, 2, 74
- [2] S. MacDonnell, D. Embuldeniya, J. Anderson, P. Roberts, F. Ratanshi, and K. Rexe, “Poverty by Postal Code: The Geography of Neighbourhood Poverty, 1981–2001,” 2004. [Online]. Available: <https://search.datacite.org/works/10.13140/rg.2.1.3201.3521> 1
- [3] K. M. Booth, M. M. Pinkston, and W. S. C. Poston, “Obesity and the Built Environment,” *Journal of the American Dietetic Association*, vol. 105, no. 5, pp. 110–117, May 2005, iSBN: 0002-8223. [Online]. Available: <https://search.datacite.org/works/10.1016/j.jada.2005.02.045> 2
- [4] J. E. Holsten, “Obesity and the community food environment: a systematic review,” *Public health nutrition*, vol. 12, no. 3, pp. 397–405, Mar. 2009, iSBN: 1368-9800. [Online]. Available: <https://dx.doi.org/10.1017/S1368980008002267> 2
- [5] M. A. Papas, A. J. Alberg, R. Ewing, K. J. Helzlsouer, T. L. Gary, and A. C. Klassen, “The Built Environment and Obesity,” *Epidemiologic Reviews*, vol. 29, no. 1, pp. 129–143, 2007, iSBN: 0193-936X. [Online]. Available: <https://academic.oup.com/epirev/article/29/1/129/441917> 2
- [6] A. V. D. Roux and C. Mair, “Neighborhoods and health,” *Annals of the New York Academy of Sciences*, vol. 1186, no. 1, pp. 125–145, Feb. 2010, iSBN: 0077-8923. [Online]. Available: <https://search.datacite.org/works/10.1111/j.1749-6632.2009.05333.x> 2
- [7] M. D. Kipke, E. Iverson, D. Moore, C. Booker, V. Ruelas, A. L. Peters, and F. Kaufman, “Food and Park Environments: Neighborhood-level Risks for Childhood Obesity in East Los Angeles,” *Journal of Adolescent Health*, vol. 40, no. 4, pp. 325–333, Apr. 2007, iSBN: 1054-139X. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1054139X06004277> 2
- [8] P. Gordon-Larsen, “Inequality in the Built Environment Underlies Key Health Disparities in Physical Activity and Obesity,” *Pediatrics (Evanston)*, vol. 117, no. 2, pp. 417–424, Feb. 2006, iSBN: 0031-4005. [Online]. Available: <https://search.datacite.org/works/10.1542/peds.2005-0058> 2
- [9] T. Leventhal and J. Brooks-Gunn, “The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes,” *Psychological Bulletin*, vol. 126, no. 2, pp. 309–337, Mar. 2000, iSBN: 0033-2909. 2



- [10] G. R. McCormack and A. Shiell, "In search of causality: a systematic review of the relationship between the built environment and physical activity among adults," *The international journal of behavioral nutrition and physical activity*, vol. 8, no. 1, p. 125, 2011, iSBN: 1479-5868. [Online]. Available: <https://search.datacite.org/works/10.1186/1479-5868-8-125> 2
- [11] C. E. Ross, "Neighborhood Disadvantage and Adult Depression," *Journal of Health and Social Behavior*, vol. 41, no. 2, pp. 177–187, 2000, iSBN: 0022-1465. [Online]. Available: <https://www.jstor.org/stable/2676304> 2
- [12] S. L. Handy, M. G. Boarnet, R. Ewing, and R. E. Killingsworth, "How the built environment affects physical activity: Views from urban planning," *American Journal of Preventive Medicine*, vol. 23, no. 2, Supplement 1, pp. 64–73, Aug. 2002, iSBN: 0749-3797. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0749379702004750> 2
- [13] S. Rogers, K. Gardner, and C. Carlson, "Social Capital and Walkability as Social Aspects of Sustainability," *Sustainability (Basel, Switzerland)*, vol. 5, no. 8, pp. 3473–3483, Aug. 2013, iSBN: 2071-1050. [Online]. Available: [https://explore.openaire.eu/search/publication?articleId=dedup\\_wf\\_001::b6cc19c08abb4a8949402d351557fd93](https://explore.openaire.eu/search/publication?articleId=dedup_wf_001::b6cc19c08abb4a8949402d351557fd93) 2
- [14] L. D. Gunn, S. Mavoa, C. Boulangn, P. Hooper, A. Kavanagh, and B. Giles-Corti, "Designing healthy communities: creating evidence on metrics for built environment features associated with walkable neighbourhood activity centres," *The International Journal of Behavioral Nutrition and Physical Activity*, vol. 14, no. 1, 2017, iSBN: 1479-5868. 2
- [15] S. Srinivasan, L. R. O'Fallon, and A. Dearry, "Creating Healthy Communities, Healthy Homes, Healthy People: Initiating a Research Agenda on the Built Environment and Public Health," *American Journal of Public Health*, vol. 93, no. 9, pp. 1446–1450, Sep. 2003, iSBN: 0090-0036. [Online]. Available: <https://ajph.aphapublications.org/doi/full/10.2105/AJPH.93.9.1446> 2
- [16] J. M. Ulmer, K. L. Wolf, D. R. Backman, R. L. Tretheway, C. J. Blain, J. P. O'Neil-Dunne, and L. D. Frank, "Multiple health benefits of urban tree canopy: The mounting evidence for a green prescription," *Health & Place*, vol. 42, pp. 54–62, Nov. 2016, iSBN: 1353-8292. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1353829216301332> 2
- [17] D. P. Farrington and B. C. Welsh, "Improved street lighting and crime prevention," *Justice Quarterly*, vol. 19, no. 2, pp. 313–342, Jun. 2002, iSBN: 0741-8825. [Online]. Available: <https://doi.org/10.1080/07418820200095261> 2, 3
- [18] J. Hong and C. Chen, "The role of the built environment on perceived safety from crime and walking: examining direct and indirect impacts," *Transportation (Dordrecht)*, vol. 41, no. 6, pp. 1171–1185, Nov. 2014, iSBN: 0049-4488. [Online]. Available: <https://search.proquest.com/docview/1617714568> 2

- [19] J. Hwang, K. Joh, and A. Woo, "Social inequalities in child pedestrian traffic injuries: Differences in neighborhood built environments near schools in Austin, TX, USA," *Journal of transport & health*, vol. 6, pp. 40–49, Sep. 2017, iISBN: 2214-1405. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2214140516304169> 2, 3
- [20] M. Deutsch, "Equity, equality, and need: What determines which value will be used as the basis of distributive justice?" *Journal of Social issues*, 1975. 3
- [21] G. F. Wagstaff, "Equity, equality, and need: Three principles of justice or one? An analysis of "equity as desert"," no. 13, pp. 138–152, 1994. 3
- [22] R. A. Musgrave, "theory of public finance; a study in public economy," 1959. 3
- [23] W. Lucy, "Equity and Planning For Local Services," *Journal of the American Planning Association*, vol. 47, no. 4, pp. 447–457, Oct. 1981, iISBN: 0194-4363. [Online]. Available: <https://doi.org/10.1080/01944368108976526> 3
- [24] M. Deutsch, "Distributive justice: A social-psychological perspective," *Yale University Press*, 1985. 3
- [25] D. J. Forkenbrock, J. Sheeley, and N. C. H. R. Program, *Effective Methods for Environmental Justice Assessment*. Transportation Research Board, 2004. 3, 48
- [26] E. Borowski, A. Ermagun, and D. Levinson, "Disparity of Access: Variations in Transit Service by Race, Ethnicity, Income, and Auto Availability," 2018. 3, 22
- [27] T. L. Grant, N. Edwards, H. Sveistrup, C. Andrew, and M. Egan, "Inequitable walking conditions among older people: examining the interrelationship of neighbourhood socio-economic status and urban form using a comparative case study," *BMC public health*, vol. 10, no. 1, p. 677, Nov. 2010, iISBN: 1471-2458. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/21054879> 3
- [28] J. A. Hirsch, G. F. Green, M. Peterson, D. A. Rodriguez, and P. Gordon-Larsen, "Neighborhood Sociodemographics and Change in Built Infrastructure," *Journal of urbanism*, vol. 10, no. 2, pp. 181–197, 2017, iISBN: 1754-9175. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5353850/> 3
- [29] J. J. C. Aman and J. Smith-Colin, "Transit Deserts: Equity analysis of public transit accessibility," *Journal of Transport Geography*, vol. 89, p. 102869, Dec. 2020, iISBN: 0966-6923. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0966692320309467> 3, 10, 18, 50
- [30] A. P. Jones, J. Brainard, I. J. Bateman, and A. A. Lovett, "Equity of access to public parks in birmingham, England," *Environmental Research Journal*, vol. 3, no. 2/3, p. 237, 2009. 3, 4
- [31] K. Neckerman, G. Lovasi, S. Davies, M. Purciel, J. Quinn, E. Feder, N. Raghunath, B. Wasserman, and A. Rundle, "Disparities in urban neighborhood conditions: Evidence from GIS measures and field observation in New York city," 2009. [Online]. Available: <http://hdl.handle.net/20.500.11937/33017> 3, 51

- [32] J. Wang and G. Lindsey, “Equity of Bikeway Distribution in Minneapolis, Minnesota,” *Transportation Research Record*, vol. 2605, no. 1, pp. 18–31, Jan. 2017, iISBN: 0361-1981. [Online]. Available: <https://doi.org/10.3141/2605-02> 3
- [33] J. Wolch, J. P. Wilson, and J. Fehrenbach, “Parks and Park Funding in Los Angeles: An Equity-Mapping Analysis,” *Urban Geography*, vol. 26, no. 1, pp. 4–35, Feb. 2005, iISBN: 0272-3638. [Online]. Available: <https://doi.org/10.2747/0272-3638.26.1.4> 3, 4
- [34] Y. Xiao, Z. Wang, Z. Li, and Z. Tang, “An assessment of urban park access in Shanghai – Implications for the social equity in urban China,” *Landscape and urban planning*, vol. 157, pp. 383–393, Jan. 2017, iISBN: 0169-2046. [Online]. Available: <https://search.datacite.org/works/10.1016/j.landurbplan.2016.08.007> 3, 4, 50
- [35] R. Zandieh, J. M. Martin, J. Flacke, P. Jones, and M. F. A. M. van Maarseveen, “Older adults’ outdoor walking: inequalities in neighbourhood safety, pedestrian infrastructure and aesthetics,” *International journal of environmental research and public health*, vol. 13, no. 12, p. 1179, 2016, iISBN: 1661-7827. 3
- [36] X. Zhu and C. Lee, “Walkability and safety around elementary schools: economic and ethnic disparities,” *American journal of preventive medicine*, vol. 34, no. 4, pp. 282–290, Apr. 2008, iISBN: 0749-3797. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/18374241> 3
- [37] A. Hilmers, D. C. Hilmers, and J. Dave, “Neighborhood disparities in access to healthy foods and their effects on environmental justice,” *American Journal of Public Health*, vol. 102, no. 9, pp. 1644–1654, 2012. 3
- [38] S. Raja, C. Ma, and P. Yadav, “Beyond food deserts: measuring and mapping racial disparities in neighborhood food environments,” *Journal of Planning Education and Research*, vol. 27, no. 4, pp. 469–482, 2008. 3
- [39] A. Frackelton, A. Grossman, E. Palinginis, F. Castrillon, V. Elango, and R. Guensler, “Measuring Walkability: Development of an Automated Sidewalk Quality Assessment Tool,” *Suburban Sustainability*, vol. 1, no. 1, 2013, iISBN: 2164-0866(ONLINE). [Online]. Available: <https://scholarcommons.usf.edu/subsust/vol1/iss1/4> 4, 15
- [40] Y. U. Shah, S. S. Jain, D. Tiwari, and M. K. Jain, “Development of Overall Pavement Condition Index for Urban Road Network,” *Procedia, social and behavioral sciences*, vol. 104, pp. 332–341, Dec. 2013, iISBN: 1877-0428. [Online]. Available: <https://search.datacite.org/works/10.1016/j.sbspro.2013.11.126> 4, 8
- [41] B. B. Cutts, K. J. Darby, C. G. Boone, and A. Brewis, “City structure, obesity, and environmental justice: An integrated analysis of physical and social barriers to walkable streets and park access,” *Social Science & Medicine*, vol. 69, no. 9, pp. 1314–1322, Nov. 2009, iISBN: 0277-9536. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0277953609005395> 4

- [42] A. Rigolon, “A complex landscape of inequity in access to urban parks: A literature review,” *Landscape and Urban Planning*, vol. 153, pp. 160–169, Sep. 2016, iSBN: 0169-2046. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0169204616300846> 4
- [43] E. Talen, “The Social Equity of Urban Service Distribution: An Exploration of Park Access in Pueblo, Colorado, and Macon, Georgia,” *Urban Geography*, vol. 18, no. 6, pp. 521–541, Aug. 1997, iSBN: 0272-3638. [Online]. Available: <https://doi.org/10.2747/0272-3638.18.6.521> 4
- [44] C. C. Weiss, M. Purciel, M. Bader, J. W. Quinn, G. Lovasi, K. M. Neckerman, and A. G. Rundle, “Reconsidering Access: Park Facilities and Neighborhood Disamenities in New York City,” *Journal of urban health*, vol. 88, no. 2, pp. 297–310, Mar. 2011, iSBN: 1468-2869. [Online]. Available: <https://search.datacite.org/works/10.1007/s11524-011-9551-z> 4, 18, 50
- [45] P. Apparicio, M. Abdelmajid, M. Riva, and R. Shearmur, “Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues,” *International journal of health geographics*, vol. 7, no. 1, p. 7, 2008, iSBN: 1476-072X. [Online]. Available: <https://search.datacite.org/works/10.1186/1476-072x-7-7> 4
- [46] B. A. Graves, “Integrative literature review: a review of literature related to geographical information systems, healthcare access, and health outcomes,” *Perspectives in health information management*, vol. 5, p. 11, Jul. 2008. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/18698429> 4
- [47] G. Higgs, “A Literature Review of the Use of GIS-Based Measures of Access to Health Care Services,” *Health services and outcomes research methodology*, vol. 5, no. 2, pp. 119–139, Jun. 2004, iSBN: 1387-3741. [Online]. Available: <https://search.proquest.com/docview/228985732> 4
- [48] F. Wang and W. Luo, “Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas,” *Health & place*, vol. 11, no. 2, pp. 131–146, Jun. 2005, iSBN: 1353-8292. [Online]. Available: <https://search.datacite.org/works/10.1016/j.healthplace.2004.02.003> 4
- [49] A. M. El-Geneidy, P. Tetreault, and J. Surprenant-Legault, “No title,” *Pedestrian access to transit: Identifying redundancies and gaps using a variable service area analysis*, 2010. 4
- [50] M. Foda and A. Osman, “Using GIS for Measuring Transit Stop Accessibility Considering Actual Pedestrian Road Network,” *Journal of Public Transportation*, vol. 13, no. 4, 2010, iSBN: 1077-291X. [Online]. Available: <https://scholarcommons.usf.edu/jpt/vol13/iss4/2> 4
- [51] S. O’Sullivan and J. Morrall, “Walking distances to and from light-rail transit stations,” *Transportation Research Record*, vol. 1538, no. 1, pp. 19–26, 1996. 4, 15

- [52] P. Apparicio, M.-S. Cloutier, and R. Shearmur, “The case of Montréal’s missing food deserts: Evaluation of accessibility to food supermarkets,” *International journal of health geographics*, vol. 6, no. 1, p. 4, Feb. 2007, iSBN: 1476-072X. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/17295912> 4
- [53] C. Gordon, M. Purciel-Hill, N. R. Ghai, L. Kaufman, R. Graham, and G. V. Wye, “Measuring food deserts in New York City’s low-income neighborhoods,” *Health & place*, vol. 17, no. 2, pp. 696–700, 1995, iSBN: 1353-8292. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1353829210001905> 4
- [54] J. Jiao, A. V. Moudon, J. Ulmer, P. M. Hurvitz, and A. Drewnowski, “How to Identify Food Deserts: Measuring Physical and Economic Access to Supermarkets in King County, Washington,” *American journal of public health (1971)*, vol. 102, no. 10, pp. e32–e39, Oct. 2012, iSBN: 1541-0048. [Online]. Available: <https://search.datacite.org/works/10.2105/ajph.2012.300675> 4, 10
- [55] M. v. Ploeg, V. Breneman, T. Farrigan, K. Hamrick, D. Hopkins, P. Kaufman, B.-H. Lin, M. Nord, T. A. Smith, R. Williams, K. Kinnison, C. Olander, A. Singh, and E. Tuckermanty, “Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences: Report to Congress,” Tech. Rep., 2009. 4, 10
- [56] R. Ewing, S. Handy, R. C. Brownson, O. Clemente, and E. Winston, “Identifying and Measuring Urban Design Qualities Related to Walkability,” *Human Kinetics Journals*, vol. 3, no. s1, pp. S223–S240, 2006. 4
- [57] F. Moura, P. Cambra, and A. B. Gonçalves, “Measuring walkability for distinct pedestrian groups with a participatory assessment method: A case study in Lisbon,” *Landscape and urban planning*, vol. 157, pp. 282–296, Jan. 2017, iSBN: 0169-2046. [Online]. Available: <http://dx.doi.org/10.1016/j.landurbplan.2016.07.002> 4
- [58] J. R. Parks and J. L. Schofer, “Characterizing neighborhood pedestrian environments with secondary data,” *Transportation research. Part D, Transport and environment*, vol. 11, no. 4, pp. 250–263, 2006, iSBN: 1361-9209. [Online]. Available: <http://dx.doi.org/10.1016/j.trd.2006.04.003> 4
- [59] S. Su, H. Zhou, M. Xu, H. Ru, W. Wang, and M. Weng, “Auditing street walkability and associated social inequalities for planning implications,” *Journal of transport geography*, vol. 74, pp. 62–76, Jan. 2019, iSBN: 0966-6923. [Online]. Available: <http://dx.doi.org/10.1016/j.jtrangeo.2018.11.003> 4, 21
- [60] S. E. Maco and E. G. McPherson, “Assessing canopy cover over streets and sidewalks in street tree populations,” *Journal of Arboriculture.28 (6): 270-276*, vol. 28, no. 6, pp. 270–276, 2002. 4, 15
- [61] J. Roberts, A. Koeser, A. Abd-Elrahman, B. Wilkinson, G. Hansen, S. Landry, and A. Perez, “Mobile Terrestrial Photogrammetry for Street Tree Mapping and



- Measurements,” *Forests*, vol. 10, no. 8, p. 701, Aug. 2019, iISBN: 1999-4907. [Online]. Available: <https://search.datacite.org/works/10.3390/f10080701> 4
- [62] W.-L. Tsai, L. Yngve, Y. Zhou, K. M. M. Beyer, A. Bersch, K. M. Malecki, and L. E. Jackson, “Street-level neighborhood greenery linked to active transportation: A case study in Milwaukee and Green Bay, WI, USA,” *Landscape and urban planning*, vol. 191, p. 103619, Nov. 2019, iISBN: 0169-2046. [Online]. Available: <https://search.datacite.org/works/10.1016/j.landurbplan.2019.103619> 4, 51
- [63] A. T. Perzynski, M. J. Roach, S. Shick, B. Callahan, D. Gunzler, R. Cebul, D. C. Kaelber, A. Huml, J. D. Thornton, and D. Einstadter, “Patient portals and broadband internet inequality,” *Journal of the American Medical Informatics Association*, vol. 24, no. 5, pp. 927–932, 2017. 4, 22
- [64] U. A. C. o. Engineering-USACE, “Micro Paver 6.5.7 User Manual,” 2014. 4, 8, 15
- [65] K. A. Rollings, N. M. Wells, and G. W. Evans, “Measuring physical neighborhood quality related to health,” *Behavioral sciences*, vol. 5, no. 2, pp. 190–202, Apr. 2015, iISBN: 2076-328X. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/25938692> 4, 72, 73
- [66] U. S. C. Bureau, “American Community Survey; Table B03002;,” 2020. [Online]. Available: <https://data.census.gov/cedsci/> 7, 10
- [67] R. Flowerdew, D. J. Manley, and C. E. Sabel, “Neighbourhood effects on health: Does it matter where you draw the boundaries?” *Social science & medicine (1982)*, vol. 66, no. 6, pp. 1241–1255, Mar. 2008, iISBN: 0277-9536. [Online]. Available: <https://search.datacite.org/works/10.1016/j.socscimed.2007.11.042> 7, 8, 72, 73
- [68] A.-V. D. Roux, “Neighborhoods and health: where are we and where do we go from here?” *Revue d’épidémiologie et de santé publique*, vol. 55, no. 1, pp. 13–21, 2007. 7, 72, 73
- [69] N. Schaefer-McDaniel, M. O. Caughy, P. O’Campo, and W. Gearey, “Examining methodological details of neighbourhood observations and the relationship to health: A literature review,” *Social science & medicine*, vol. 70, no. 2, pp. 277–292, 1978, iISBN: 0160-8002. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0277953609006947> 7, 72, 73
- [70] R. Abhyanker, “Nextdoor neighborhood social network method, apparatus, and system,” *United States Patent*, 2014. 8
- [71] S. Cummins and S. Macintyre, ““Food deserts”—evidence and assumption in health policy making,” *Bmj*, vol. 325, no. 7361, pp. 436–438, 2002, publisher: British Medical Journal Publishing Group. 10
- [72] N. Wrigley, “Food deserts’ in British cities: policy context and research priorities,” *Urban studies*, vol. 39, no. 11, pp. 2029–2040, 2002, publisher: Sage Publications Sage UK: London, England. 10

- [73] C. N. Chau, J. M. Zoellner, and J. L. Hill, “Availability of Healthy Food: Does Block Group Race and Income Matter?” *Journal of hunger & environmental nutrition*, vol. 8, no. 1, pp. 22–38, Jan. 2013, iISBN: 1932-0248. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/19320248.2012.758063> 10, 22
- [74] M. Franco, A. V. D. Roux, T. A. Glass, B. Caballero, and F. L. Brancati, “Neighborhood Characteristics and Availability of Healthy Foods in Baltimore,” *American journal of preventive medicine*, vol. 35, no. 6, pp. 561–567, 2008, iISBN: 0749-3797. [Online]. Available: <https://search.datacite.org/works/10.1016/j.amepre.2008.07.003> 10, 16
- [75] L. M. Powell, S. Slater, D. Mirtcheva, Y. Bao, and F. J. Chaloupka, “Food store availability and neighborhood characteristics in the United States,” *Preventive medicine*, vol. 44, no. 3, pp. 189–195, Mar. 2007, iISBN: 0091-7435. [Online]. Available: <https://search.datacite.org/works/10.1016/j.ypped.2006.08.008> 10
- [76] K. Alaimo, C. M. Olson, and E. A. Frongillo, “Low Family Income and Food Insufficiency in Relation to Overweight in US Children: Is There a Paradox?” *Archives of pediatrics & adolescent medicine*, vol. 155, no. 10, pp. 1161–1167, Oct. 2001, iISBN: 1072-4710. [Online]. Available: <http://dx.doi.org/10.1001/archpedi.155.10.1161> 10
- [77] A. Agresti, *Categorical Data Analysis*. John Wiley & Sons, 2003. 11, 31, 33, 34, 57
- [78] S. H. Walker and D. B. Duncan, “Estimation of the Probability of an Event as a Function of Several Independent Variables,” *Biometrika*, vol. 54, no. 1/2, pp. 167–179, 1967, iISBN: 0006-3444. [Online]. Available: <https://www.jstor.org/stable/2333860> 11
- [79] B. Peterson and F. E. Harrell, “Partial proportional odds models and the LOGIST procedure,” in *SAS Users Group International, thirteenth annual conference, Orlando FL*, 1988. 11
- [80] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112. 12, 33, 56
- [81] R. C. Team, “R: A language and environment for statistical computing,” 2013. 12
- [82] B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, D. Firth, and M. B. Ripley, “Package ‘mass’,” *Cran R*, vol. 538, 2013. 12
- [83] C. of Dallas, “Resilient dallas: Dallas’ pathway to shared prosperity,” 2018. [Online]. Available: <https://100resilientcities.org/wp-content/uploads/2018/06/Resilient-Dallas-Strategy-PDF-1.pdf> 13
- [84] P. Taylor and R. Fry, “The Rise of Residential Segregation by Income,” *Washington, DC: Pew Research Center.*, 2012. 13

- [85] R. Florida, *The New Urban Crisis: How Our Cities Are Increasing Inequality, Deepening Segregation, and Failing the Middle Class-and What We Can Do About It*. Basic Books, 2017. 13
- [86] T. T. Foundation, “Dallas urban heat management study,” 2017. [Online]. Available: [https://texas-trees.blob.core.windows.net/static/Texas%20Trees%20UHI%20Study\\_Final\\_v4\\_title%20change.pdf](https://texas-trees.blob.core.windows.net/static/Texas%20Trees%20UHI%20Study_Final_v4_title%20change.pdf) 13, 71
- [87] D. Habeeb, J. Vargo, and B. Stone, “Rising heat wave trends in large US cities,” *Natural Hazards*, vol. 76, no. 3, pp. 1651–1665, 2015, publisher: Springer. 13
- [88] R. Retting and others, “Pedestrian traffic fatalities by state,” *Governors Highway Safety Association: Washington, DC, USA*, 2017. 13
- [89] C. of Dallas, “city of dallas community survey 2018,” 2018. [Online]. Available: [https://dallascityhall.com/government/Council%20Meeting%20Documents/b\\_2018-community-survey\\_combined\\_050218.pdf](https://dallascityhall.com/government/Council%20Meeting%20Documents/b_2018-community-survey_combined_050218.pdf) 13
- [90] J. L. Rochat and D. Reiter, “Highway Traffic Noise,” *Acoustics Today*, vol. 12, no. 4, 2016. [Online]. Available: <https://acousticstoday.org/wp-content/uploads/2018/08/Highway-Traffic-Noise-Judith-L.-Rochat.pdf> 14, 15
- [91] F. C. Commission, “Internet Access Services: Status as of December 31, 2016,” 2019. 15
- [92] C. V. Zegeer, J. R. Stewart, H. H. Huang, P. A. Lagerwey, J. R. Feaganes, and B. J. Campbell, “Safety effects of marked versus unmarked crosswalks at uncontrolled locations final report and recommended guidelines. United States.” *Federal Highway Administration*, 2005. 15, 81
- [93] A. M. El-Geneidy, P. Tetreault, and J. Surprenant-Legault, “Pedestrian Access to Transit: Identifying Redundancies and Gaps Using a Variable Service Area Analysis,” *Transportation Research Board 89th Annual Meeting*, 2010. 15
- [94] J. S. Ashwood, R. O. Reid, C. M. Setodji, E. Weber, M. Gaynor, and A. Mehrotra, “Trends in retail clinic use among the commercially insured,” *The American Journal of Managed Care*, vol. 17, no. 11, p. e443, 2011. 15, 74
- [95] M. Bersamin, M. Todd, and L. Remer, “Does distance matter? Access to family planning clinics and adolescent sexual behaviors,” *Maternal and child health journal*, vol. 15, no. 5, pp. 652–659, 2011. 15, 74
- [96] C. Gordon, M. Purciel-Hill, N. R. Ghai, L. Kaufman, R. Graham, and G. V. Wye, “Measuring food deserts in New York City’s low-income neighborhoods,” *Health & place*, vol. 17, no. 2, pp. 696–700, 2011. 15, 18, 50
- [97] City of Dallas GIS Service, “Census block group summaries of 2013 tax appraisal data,” 2013. [Online]. Available: [https://gis.dallascityhall.com/downloads/ShpZip/NeighborhoodHealth/NH\\_DallasCountyBlockGroups2010WithAppraisalRollups.zip](https://gis.dallascityhall.com/downloads/ShpZip/NeighborhoodHealth/NH_DallasCountyBlockGroups2010WithAppraisalRollups.zip) 16



- [98] R. O. Self, *American babylon*. Princeton University Press, 2005. 17
- [99] T. J. Sugrue, *The origins of the urban crisis*. Princeton University Press, 2014. 17
- [100] R. K. Nelson, L. Winling, R. Marciano, N. Connolly, and E. L. Ayers, “Mapping inequality: Redlining in new deal America,” *American Panorama: An Atlas of United States History*. University of Richmond: Digital Scholarship Lab, 2020. 17, 19
- [101] E. Flanagan, U. Lachapelle, and A. El-Geneidy, “Riding tandem: Does cycling infrastructure investment mirror gentrification and privilege in Portland, OR and Chicago, IL?” *Research in Transportation Economics*, vol. 60, pp. 14–24, 2016, publisher: Elsevier. 18, 30
- [102] K. Heilmann, “Transit access and neighborhood segregation. Evidence from the Dallas light rail system,” *Regional Science and Urban Economics*, vol. 73, pp. 237–250, 2018, publisher: Elsevier. 18
- [103] D. Liu, M.-P. Kwan, and Z. Kan, “Analysis of urban green space accessibility and distribution inequity in the City of Chicago,” *Urban Forestry & Urban Greening*, vol. 59, p. 127029, 2021, publisher: Elsevier. 18, 30
- [104] B. Bereitschaft, “Equity in Microscale Urban Design and Walkability: A Photographic Survey of Six Pittsburgh Streetscapes,” *Sustainability (Basel, Switzerland)*, vol. 9, no. 7, p. 1233, Jul. 2017, iSBN: 2071-1050. 21
- [105] T. Huynh and L. Kent, “In Greater Dallas Area, Segregation by Income and Race,” *Pew Research Center*. June, vol. 29, 2015. 24
- [106] L. Ermini and D. F. Hendry, “Log income vs. linear income: An application of the encompassing principle,” *Oxford Bulletin of Economics and Statistics*, vol. 70, no. s1, pp. 807–827, 2008, iSBN: 1468-0084. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0084.2008.00531.x> 28
- [107] L. Franzini, W. Taylor, M. N. Elliott, P. Cuccaro, S. R. Tortolero, M. J. Gilliland, J. Grunbaum, and M. A. Schuster, “Neighborhood characteristics favorable to outdoor physical activity: disparities by socioeconomic and racial/ethnic composition,” *Health & place*, vol. 16, no. 2, pp. 267–274, 2010, publisher: Elsevier. 30
- [108] L. Shi, L. A. Lebrun, J. Tsai *et al.*, “Access to medical care, dental care, and prescription drugs: the roles of race-ethnicity, health insurance, and income,” *Southern medical journal*, vol. 103, no. 6, p. 509, 2010. 30, 50
- [109] R. Chunara, Y. Zhao, J. Chen, K. Lawrence, P. A. Testa, O. Nov, and D. M. Mann, “Telemedicine and healthcare disparities: a cohort study in a large healthcare system in New York City during COVID-19,” *Journal of the American Medical Informatics Association*, vol. 28, no. 1, pp. 33–41, 2021, publisher: Oxford University Press. 30

- [110] R. Yearby, “Racial disparities in health status and access to healthcare: the continuation of inequality in the United States due to structural racism,” *American Journal of Economics and Sociology*, vol. 77, no. 3-4, pp. 1113–1152, 2018, publisher: Wiley Online Library. 30
- [111] A. M. Brockway, J. Conde, and D. Callaway, “Inequitable access to distributed energy resources due to grid infrastructure limits in California,” *Nature Energy*, vol. 6, no. 9, pp. 892–903, 2021, publisher: Nature Publishing Group. 30
- [112] M. L. Messenger, A. K. Ettinger, M. Murphy-Williams, and P. S. Levin, “Fine-scale assessment of inequities in inland flood vulnerability,” *Applied geography*, vol. 133, p. 102492, 2021, publisher: Elsevier. 30
- [113] J. L. Servadio, A. S. Lawal, T. Davis, J. Bates, A. G. Russell, A. Ramaswami, M. Convertino, and N. Botchwey, “Demographic inequities in health outcomes and air pollution exposure in the Atlanta area and its relationship to urban infrastructure,” *Journal of Urban Health*, vol. 96, no. 2, pp. 219–234, 2019, publisher: Springer. 30, 51
- [114] J. R. Elliott, P. L. Brown, and K. Loughran, “Racial inequities in the federal buyout of flood-prone homes: a nationwide assessment of environmental adaptation,” *Socius*, vol. 6, p. 2378023120905439, 2020, publisher: SAGE Publications Sage CA: Los Angeles, CA. 30
- [115] S. Heck, “Greening the color line: Historicizing water infrastructure redevelopment and environmental justice in the St. Louis metropolitan region,” *Journal of Environmental Policy & Planning*, vol. 23, no. 5, pp. 565–580, 2021, publisher: Taylor & Francis. 30
- [116] W. H. Kruskal and L. Goodman, “Measures of association for cross classifications,” *Journal of the American Statistical Association*, vol. 49, no. 268, pp. 732–764, 1954. 34
- [117] C. of Dallas, “TARGETED REHAB PROGRAM (TRP),” Tech. Rep., 2021. [Online]. Available: <https://dallascityhall.com/departments/housing-neighborhood-revitalization/Pages/TRP.aspx> 44
- [118] J. Schutze, *The Accommodation: The Politics of Race in an American City*. Deep Vellum Publishing, 2021. 45
- [119] M. Phillips, *White metropolis*. University of Texas Press, 2006. 45, 46
- [120] K. Krupala, “The Evolution of Uneven Development in Dallas, TX,” *Human Geography*, vol. 12, no. 3, pp. 17–30, 2019, publisher: SAGE Publications Sage UK: London, England. 46
- [121] C. of Dallas, “The big picture: 2017 capital bond program,” 2017. [Online]. Available: <https://www.dallascitynews.net/picturedallas> 47

- [122] J. Jiao and M. Dillivan, “Transit Deserts: The Gap between Demand and Supply,” *Journal of Public Transportation*, vol. 16, no. 3, 2013, iISBN: 1077-291X. [Online]. Available: <https://scholarcommons.usf.edu/jpt/vol16/iss3/2> 50
- [123] R. W. Smith, K. Bertolaccini, and N. E. Lownes, “Improving transit access measures in affordable housing funding criteria,” *Transport Policy*, vol. 106, pp. 239–248, 2021. 50, 51
- [124] Economic Research Service (ERS), “Food access research atlas,” *U.S. Department of Agriculture (USDA)*, 2015. [Online]. Available: <https://www.ers.usda.gov/data-products/food-access-research-atlas/> 50
- [125] L. Marini, I. Gutierrez-Polo, R. Kooper, S. P. Satheesan, M. Burnette, J. Lee, T. Nicholson, Y. Zhao, and K. McHenry, “Clowder: Open source data management for long tail data,” ser. Proceedings of the Practice and Experience on Advanced Research Computing, 2018, pp. 1–8. 51
- [126] B. Plale, R. H. McDonald, K. Chandrasekar, I. Kouper, S. Konkiel, M. L. Hedstrom, J. Myers, and P. Kumar, “Sead virtual archive: Building a federation of institutional repositories for long-term data preservation in sustainability science,” 2013. 52
- [127] J. Myers and M. Hedstrom, “Active and social curation: Keys to data service sustainability,” in *position paper, National Data Service (NDS) Consortium Planning Workshop, Boulder, CO*, 2014. 52
- [128] S. Padhy, E. Black, B. Cowdery, L. Diesendruck, M. Dietze, G. Jansen, R. Kooper, P. Kumar, J. Lee, R. Liu *et al.*, “Autocuration cyberinfrastructure for scientific discovery and preservation,” in *2015 IEEE 11th International Conference on e-Science*. IEEE, 2015, pp. 264–267. 52
- [129] S. Padhy, G. Jansen, J. Alameda, E. Black, L. Diesendruck, M. Dietze, P. Kumar, R. Kooper, J. Lee, R. Liu *et al.*, “Brown dog: Leveraging everything towards autocuration,” in *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, 2015, pp. 493–500. 52
- [130] G. Jansen, S. Padhy, and R. Marciano, “Designing scalable cyberinfrastructure for metadata extraction in billion-record archives.” in *iPRES*, 2016. 52
- [131] M. Slavenas, P. Rodriguez, A. Craig, E. Wuerffel, and J. Will, “Image analysis and infrastructure support for data mining the farm security administration: Office of war information photography collection,” in *Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale*, 2016, pp. 1–6. 52
- [132] D. K. Woo, J. C. Quijano, P. Kumar, S. Chaoka, and C. J. Bernacchi, “Threshold dynamics in soil carbon storage for bioenergy crops,” *Environmental science & technology*, vol. 48, no. 20, pp. 12 090–12 098, 2014. 53
- [133] P. Kumar, “Hydrocomplexity: Addressing water security and emergent environmental risks,” *Water Resources Research*, vol. 51, no. 7, pp. 5827–5838, 2015. 53

- [134] L. B. Poole, “The basics of thiols and cysteines in redox biology and chemistry,” *Free Radical Biology and Medicine*, vol. 80, pp. 148–157, 2015. 53
- [135] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994. 56
- [136] H. Akaike, “A new look at the statistical model identification,” *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974. 57
- [137] S. G. R. Quadros and C. D. Nassi, “An evaluation on the criteria to prioritize transportation infrastructure investments in Brazil,” *Transport Policy*, vol. 40, pp. 8–16, 2015, publisher: Elsevier. 71
- [138] C. Aniello, K. Morgan, A. Busbey, and L. Newland, “Mapping micro-urban heat Island using LANDSAT TM and a GIS,” *Computers & geosciences*, vol. 21, no. 8, pp. 965–969, 1995, iSBN: 0098-3004. 71
- [139] City of Dallas, “Branch Out Dallas,” 2021. [Online]. Available: <https://dallascityhall.com/departments/waterutilities/Pages/branch-out-dallas.aspx> 71
- [140] N. Nicotera, “Measuring Neighborhood: A Conundrum for Human Services Researchers and Practitioners,” *American journal of community psychology*, vol. 40, no. 1, pp. 26–51, Sep. 2007, iSBN: 0091-0562. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1007/s10464-007-9124-1> 72
- [141] N. M. Wells and K. A. Rollings, “The natural environment in residential settings: Influences on human health and function,” ser. The Oxford handbook of environmental and conservation psychology, 2012. 73
- [142] S. Liu and X. Zhu, “An integrated GIS approach to accessibility analysis,” *Transactions in GIS*, vol. 8, no. 1, pp. 45–62, 2004. 73
- [143] W. Luo and F. Wang, “Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region,” *Environment and Planning B: Planning and Design*, vol. 30, no. 6, pp. 865–884, 2003. 73
- [144] S. Nicholls, “Measuring the accessibility and equity of public parks: A case study using GIS,” *Manag. Leis.*, vol. 6, Oct. 2001. 73
- [145] J. Pearce, K. Witten, and P. Bartie, “Neighbourhoods and health: a GIS approach to measuring community resource accessibility,” *Journal of Epidemiology & Community Health*, vol. 60, no. 5, pp. 389–395, 2006. 73
- [146] T. Yigitcanlar, N. Sipe, R. Evans, and M. Pitot, “A GIS [U+2010]based land use and public transport accessibility indexing model,” *Australian planner*, vol. 44, no. 3, pp. 30–37, 2007. 73
- [147] W. G. Hansen, “How accessibility shapes land use–Journal of the American Planning Association 25-2,” DOI: <http://dx.doi.org/10.1080/01944365908978307>, 1959. 73

- [148] A. E. Joseph and P. R. Bantock, "Measuring potential physical accessibility to general practitioners in rural areas: a method and case study," *Social science & medicine*, vol. 16, no. 1, pp. 85–90, 1982. 73
- [149] J. W. Weibull, "An axiomatic approach to the measurement of accessibility," *Regional science and urban economics*, vol. 6, no. 4, pp. 357–379, 1976. 73
- [150] R. L. Church and J. R. Marston, "Measuring accessibility for people with a disability," *Geographical Analysis*, vol. 35, no. 1, pp. 83–96, 2003. 73
- [151] A. M. El-Geneidy and D. M. Levinson, "Access to destinations: Development of accessibility measures," 2006. 73
- [152] T. Neutens, "Accessibility, equity and health care: review and research directions for transport geographers," *Journal of Transport Geography*, vol. 43, pp. 14–27, 2015. 74
- [153] D. C. Hall, "Sidewalk replacement program dallas texas," 2016. [Online]. Available: <https://dallascityhall.com/443/departments/public-works/Pages/SidewalkReplacementProgram.aspx> 76
- [154] T. T. Foundation, "Dallas Urban Forest Master Plan," 2019. [Online]. Available: <https://www.texastrees.org/projects/dallas-urban-forest-master-plan/> 76
- [155] D. Burden and T. Litman, "America needs complete streets," *ITE journal*, vol. 81, no. 4, pp. 36–43, 2011. 76
- [156] J. LaPlante and B. McCann, "'Complete streets: We can get there from here.'", *ITE journal*, vol. 78, no. 5, p. 24, 2008. 76
- [157] J. Richardson, "Equality focused evaluation of complete streets policies," Ph.D. dissertation, California State University, Northridge, 2021. 76
- [158] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018. 81