

University of Redlands

InSPIRe @ Redlands

MS GIS Program Major Individual Projects

Theses, Dissertations, and Honors Projects

11-2014

Developing Quantitative Indicators to Measure Built Environment Using GIS: a Case Study in Cities of Loma Linda and Redlands in California

Jose J. Ruiz-Valcarcel
University of Redlands

Follow this and additional works at: https://inspire.redlands.edu/gis_gradproj



Part of the [Geographic Information Sciences Commons](#)

Recommended Citation

Ruiz-Valcarcel, J. J. (2014). *Developing Quantitative Indicators to Measure Built Environment Using GIS: a Case Study in Cities of Loma Linda and Redlands in California* (Master's thesis, University of Redlands). Retrieved from https://inspire.redlands.edu/gis_gradproj/232



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This material may be protected by copyright law (Title 17 U.S. Code).

This Thesis is brought to you for free and open access by the Theses, Dissertations, and Honors Projects at InSPIRe @ Redlands. It has been accepted for inclusion in MS GIS Program Major Individual Projects by an authorized administrator of InSPIRe @ Redlands. For more information, please contact inspire@redlands.edu.

University of Redlands

**Developing Quantitative Indicators to Measure Built Environment
Using GIS: a Case Study in Cities of Loma Linda and Redlands in
California**

A Major Individual Project submitted in partial satisfaction of the requirements
for the degree of Master of Science in Geographic Information Systems

by

José Javier Ruiz Valcárcel

Fang Ren, Ph.D., Committee Chair

Mark Kumler, Ph.D.

November 2014

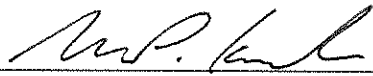
Developing Quantitative Indicators to Measure Built Environment Using GIS: a Case
Study in Cities of Loma Linda and Redlands in California

Copyright © 2014

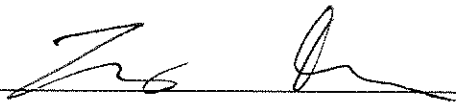
by

José Javier Ruiz Valcárcel

The report of José Javier Ruiz Valcárcel is approved.



Mark Kumler, Ph.D.



Fang Ren, Ph.D., Committee Chair

November 2014

Acknowledgements

I would like to thank my advisor Dr. Fang Ren, the professors Dr. Mark Kumler and Dr. Douglas Flewelling, Dr. Ruijin Ma, Dr. Russell Weaver, and my client Dr. Samuel Soret for their assistance throughout this project. Also thanks to my sister and parents who gave me the tools and energies to continue my studies. Lastly I would like to thank Debra Riley and Cohort 24 who helped and supported me during this year. I will never forget the team spirit of the group, the multicultural environment, and the potluck dinners at the pool. It was a hard year full of sacrifices but with a lot learned.

“Los jóvenes tienen el deber de defender su Patria con las armas del Conocimiento”

Pedro Albizu Campos

Abstract

Developing Quantitative Indicators to Measure Built Environment Using GIS: a Case Study in Cities of Loma Linda and Redlands in California

by

José Javier Ruiz Valcárcel

Adventist communities are the subject of much epidemiological research due to their homogeneous lifestyle characteristics. Little is known about how Loma Linda's built environment influences its Adventist community making it the city with the highest life expectancy in the U.S. One approach is to measure the potential of a city to promote walking, but there is a lack of tools that measure multiple dimensions of this built environment indicator. We developed a built environment toolbox taking into account different factors: access, density, diversity, and quality. A pilot study was conducted to compare the built environment of the city of Loma Linda with the city of Redlands in California. The toolset used network analysis functions in order to test the indexes. The indicators used were access to recreational areas, population density, land use diversity, and quality of sidewalks. Data were obtained from subjects of the Adventists Health Study 2. The city of Loma Linda presented better access to recreational areas, higher population density, and more diverse land use in the participants' built environment. These findings should inform urban planners in order to consider walkable indicators that can promote a healthy community.

Table of Contents

Chapter 1 – Introduction	1
1.1 Client	2
1.2 Problem Statement	2
1.3 Proposed Solution	3
1.3.1 Goals and Objectives.....	3
1.3.2 Scope	3
1.3.3 Methods	4
1.4 Audience.....	4
1.5 Overview of the Rest of this Report.....	5
Chapter 2 – Background and Literature Review	7
2.1 The Adventist Health Study.....	7
2.2 Physical Activity, a Response to Obesity and Chronic Diseases.....	8
2.3 A Framework Evaluating Built Environment	10
2.4 Summary.....	11
Chapter 3 – Systems Analysis and Design.....	13
3.1 Problem Statement	13
3.2 Requirements Analysis	13
3.2.1 Functional Requirements.....	14
3.2.2 Non-Functional Requirements	15
3.3 System Design	15
3.4 Project Plan.....	17
3.5 Summary.....	18

Chapter 4 – Database Design.....	19
4.1 Conceptual Data Model	19
4.2 Logical Data Model	22
4.3 Data Sources and Descriptions	25
4.4 Data Scrubbing and Loading	28
4.5 Summary	31
Chapter 5 – Implementation	33
5.1 Overview of Built Environment Toolbox	33
5.2 Walking Zone Generator.....	36
5.3 Access to Recreational Areas	38
5.4 Population Density.....	40
5.5 Diversity of the Land Use	42
5.6 Quality of the Sidewalks	44
5.7 Summary	46
Chapter 6 – Results and Analysis.....	47
6.1 Mapping Individual Variables.....	47
6.2 Comparing Built Environments.....	57
6.2.1 Access to Recreational Areas	59
6.2.2 Population Density.....	63
6.2.3 Diversity of the Land Use	65
6.2.4 Quality of the Sidewalk.....	66
6.3 Summary	68
Chapter 7 – Conclusions and Future Work.....	70

Works Cited	73
Appendix A. Walking Zone Generator Tool	77
Appendix B. Access Tool.....	80
Appendix C. Density Tool	85
Appendix D. Land Use Mix Tool.....	89
Appendix E. Sidewalk Quality Tool	92
Appendix F. Classify NDVI Tool.....	95

Table of Figures

Figure 3-1: Components of the System Design	16
Figure 3-2: Project Plan	17
Figure 4-1: Access to Recreational Areas and Attractiveness of the Sidewalks.....	20
Figure 4-2: Conceptual Model Density and Diversity Factors	21
Figure 4-3: Built Environment Indicators.....	22
Figure 4-4: Logical Model for the Input Data in the Built Environment Geodatabase	23
Figure 4-5: Logical Model for the Output Data in the Built Environment Geodatabase	25
Figure 4-6: Attribute Table of a Classified NDVI.	30
Figure 5-1: Environment Toolset Workflow	34
Figure 5-2: Built Environment Toolbox Workflow	35
Figure 5-3: Walking Zone Generator Tool.....	36
Figure 5-4: Walking Zone Generator Model	37
Figure 5-5: Access to Destinations Model.....	39
Figure 5-6: Accessibility Summary Table	40
Figure 5-7: Population Density of a Walking Zone	41
Figure 5-8: Population Density Tool.....	42
Figure 5-9: Walking Zone with Land Use Categories.....	42
Figure 5-10: Diversity of the Land Use Tool.....	43
Figure 5-11: Classify NDVI Tool Model	44

Figure 5-12: Classified NDVI	45
Figure 5-13: Sidewalk Quality Model	45
Figure 6-1: Adventist Sample Distribution	48
Figure 6-2: Land Use for the City of Loma Linda.....	49
Figure 6-3: Land Use for the City of Redlands	50
Figure 6-4: Cities of Loma Linda and Redlands Dissolved by Land Use Categories	51
Figure 6-5: Distribution of Recreational Areas in the City of Loma Linda and Redlands	52
Figure 6-6: Multispectral Image of the City of Loma Linda and Redlands.....	53
Figure 6-7: NDVI in the Cities of Loma Linda and Redlands	54
Figure 6-8: Classified NDVI Using a Threshold of 0.1	55
Figure 6-9: 1-Mile Walking Zones for Each AHS-2 Participant	56
Figure 6-10: Distribution of Population from the U.S. Census at Block Level	57
Figure 6-11: Workflow to Shift Participants' Locations	58
Figure 6-12: AHS-2 Participants with the Nearest Recreational Area	60
Figure 6-13: Access to Recreational Areas within One-Mile Walking Zone	61
Figure 6-14: Distance to the Nearest Recreational Area in AHS-2 Participant with No Access within One-mile Walking Zone	61
Figure 6-15: Nearest Recreational Area from a Participant's Location	63
Figure 6-16: Population Density per Squared Mile and Population per Walking Zone between The City of Loma Linda and Redlands	64
Figure 6-17: Diversity of the Land Use per Walking Zone	66

Figure 6-18: Vegetation Pixels of the City of Loma Linda and Redlands using 0.1 as
the Threshold..... 67

Figure 6-19: Percent of Vegetation pixels covered..... 68

List of Tables

Table 3-1: Functional and Non-Functional Requirements	14
Table 4-1: Project data.....	26
Table 4-2: Preprocessing tasks.....	28
Table 6-1: Percent of vegetation pixels per city.....	55
Table 6-2: Access to recreational areas for the cities of Loma Linda and Redlands	59
Table 6-3: Population Density of the cities of Loma Linda and Redlands.....	64

List of Acronyms and Definitions

AHS-1	Adventist Health Study 1
AHS-2	Adventist Health Study 2
AMS	Adventist Mortality Study
BE	Built Environment
BMI	Body Mass Index
CDC	Centers for Disease Control and Prevention
GIS	Geographic Information Systems
IDW	Inverse Distance Weighted
NDVI	Normalized Difference Vegetation Index
NHGIS	National Historical Geographic Information System
PAG	Physical Activity Guidelines
SPSS	Statistical Package for the Social Sciences

Chapter 1 – Introduction

The Seventh-Day Adventist community is regarded as one of the healthiest communities in the country. Researchers have looked at various aspects of this phenomenon. However, little has been studied about how the layout of the cities, so-called built environment, contributes to the higher life expectancy of Adventist communities. Loma Linda, California, for example, is a city designated as a Blue Zone where people have a longer life expectancy and where more people reach the age of 100 than anywhere else (Buettner, 2013). Compared to the Seventh-Day Adventists in other cities, the Seventh-Day Adventists in Loma Linda have longer life spans, which could be explained by the local built environment. In order to test this hypothesis, local built environment needs to be quantified.

According to Brownson, Hoehner, Day, Forsyth, & Sallis (2009), built environment can be measured from different aspects such as street patterns, landscaping, land use, population density, and accessibility to food stores and recreational, cultural, government, and administrative places. The purpose of the project was to make a script toolset in GIS that can calculate these geographic indicators for the researchers who are not GIS experts. With this automated process, the assessment of the built environment of any cities and counties can be conducted and the output numerical indices for built environment can be used for public health studies. The findings of this project will also have policy implications for urban planners to promote healthy communities by taking initiatives that promote physical activities.

1.1 Client

The client of this project is Dr. Samuel Soret, Associate Dean for Public Health Practice and Executive Director of the Center for Community Resilience at Loma Linda University in California. The population the client is researching is Seventh-Day Adventists in North America. Dr. Soret is currently working on the Adventist Health Study-2 (AHS-2). The AHS-2 is a longitudinal health research study of 96,000 Adventists in the United States and Canada from 2002 until the present. The goal of that study is to evaluate the effects of food intakes on the risks of colorectal, breast, and prostate cancers. Dr. Gary Fraser, Associate Dean for Research at the School of Public Health at Loma Linda University is the principal investigator of the AHS-2 study.

The client is interested in exploring how the built environment (physical form of communities) of a city affects people's physical activities, lifestyle choices, and consequently health outcomes. The client was responsible for providing data from the AHS-2 study, providing indicator definitions, journals and articles on the research topic.

1.2 Problem Statement

Earlier studies showed that Adventists live longer and have less disease as compared with non-Adventists counterparts (Loma Linda University, 2013). To explain why Adventists in Loma Linda have an even longer life span as compared to other Adventists communities, the client would like to examine the impact of built environment on life expectancy. The problem addressed in this project was how to quantify built environment indicators, such as sidewalk quality and accessibility to recreations, for community assessments and research purposes.

1.3 Proposed Solution

The proposed solution to the client's problem was to develop a toolbox in ArcGIS that can be used with minimal user input to calculate predefined geographic indicators of the built environment. To test the toolbox, the local built environment of Adventists from the cities of Loma Linda and Redlands were compared.

1.3.1 Goals and Objectives

The goal for this project was to measure built environment using a framework drawn from Brownson et al. (2009) and Maghelal et al. (2008). One of the objectives was to automate the calculation process for predefined indicators of built environment. In doing so, the client can easily repeat the process to measure built environment for any other future cities that they are interested in. The second objective was to compare local built environment of Adventists from different cities as a case study.

1.3.2 Scope

The project focused on two deliverables. The first deliverable was a set of tools scripted in Python language that can be used to calculate built environment indicators in ArcGIS. The built environment indicators defined in this project, included access to recreational areas, population density of a neighborhood, the diversity of the land use, and the quality of the walking environment.

The second deliverable was to conduct an empirical study for the cities of Loma Linda and Redlands as an example to show how the tools can be used to compare different geographic contexts of the Adventist communities.

1.3.3 Methods

The project methodology follows a waterfall schema. This approach was used to build the toolset and compare the two cities. The following steps were included to implement this methodology:

1. Define built environment indicators.
2. Acquire current parcel level and zoning data of the cities Redlands and Loma Linda.
3. Download required data from U.S. Census and Esri.
4. Prepare data in ArcGIS Desktop, including creating network datasets and geo-processing the data.
5. Develop the tools and the toolbox using Model Builder and Python.
6. Display the results on a map.
7. Write a report detailing the findings and recommendations.

1.4 Audience

This project aims to reach a variety government and academic professionals within the GIS and the public health community. The project's audience includes professionals, from urban planners to researchers, public health professionals, and policy makers. Community activists too may find this project interesting, as they could use it to effect changes in their neighborhoods. In addition, the GIS component from government offices includes built environment indicators that can be used as measures to be considered for the assessment of quality of life in cities and counties.

1.5 Overview of the Rest of this Report

The remaining of this report includes six chapters. Chapter 2 highlights information from the Adventist population, the Adventist Health Study, the Brownson and Maghelal, Capp frameworks and its applications, along with literature review on the subject. The rest of the chapters will explain in detail the GIS model use, the database design and implementation of the pilot project. Chapter 3 is devoted to the overall project plan and core requirements. Chapter 4 outlines the database design, followed by implementation in Chapter 5. Chapter 6 shows the results and analysis of the project, and Chapter 7 lists the conclusions and considerations for future work.

Chapter 2 – Background and Literature Review

Many indicators of built environments, also known as the physical forms of communities, have been developed using objective and subjective approaches. This portion of the report reviews various models presented in different studies and how they come together to support a better understanding of built environments. This chapter also reviews the contributions of a large-scale Adventist study to the understanding of how the physical form of communities affects physical activity and health. The chapter is divided in the following sections: the Adventist Health Study, physical activity as a response to obesity and chronic diseases, a framework evaluating built environment, and a summary.

2.1 The Adventist Health Study

The U.S. Department of Health and Human Services declared that the use of alcohol or tobacco is associated with cancers, cardiovascular disease, and other diseases (U.S. Department of Health and Human Services, 2007). The Seventh-Day Adventist church is a worldwide religious congregation that promotes healthy lifestyles. In North America, church members are expected to be non-smokers and non-alcohol users (Butler et al., 2008). The absence of these highly associated risk exposures in the Adventist population makes them ideal participants for epidemiological and chronic-disease studies. Due to the opportunities that this population offers, researchers from Loma Linda University are carrying out a health study titled Adventist Health Study-2 (AHS-2). The AHS-2 is a longitudinal health research study of 96,000 Adventists in the United States and Canada from 2002 until the present. The goal of this study is to evaluate in detail the effects of food intake on risks of colorectal, breast, and prostate cancers.

The Adventist Health Study 2 follows prospective studies of the California Adventist population — The Adventist Mortality Study (AMS) and the Adventist Health Study 1 (AHS-1). These two studies compared mortality rates of smokers with non-smokers. The AMS not only documented that Adventist had lower mortality rates from cancers, cardiovascular disease, and other chronic diseases as compared with non-smoking Californians and the United States white population, but also that Adventist lifestyles improve longevity in the population and provide protection against certain types of cancers and chronic diseases (Beeson, 1989). The AHS-1 focused on food intake and cases of cancers and heart disease. It reported detailed descriptions of what types of eating habits bring protection against chronic conditions, but it did not address questions about the impact of other components of the Adventist lifestyle, such as physical activity, the use of prescription drugs, and measures of religiosity (Beeson, 1989). To measure in details the effects of an Adventist lifestyle as a quality of life, new studies have sought to differentiate between eating habits and exercise patterns in the longevity of Adventists (Loma Linda University, 2013). The AHS-2 adds a physical activity questionnaire that combines eating habits as a lifestyle complement.

2.2 Physical Activity, a Response to Obesity and Chronic Diseases

The U.S. Department of Health and Human Services states that an unhealthy diet and a lack of physical activity can contribute to obesity and risks of certain type of cancers, diabetes, and cardiovascular disease, among other chronic diseases (U.S. Department of Health and Human services, 2008; Centers for Disease Control and Prevention, 2008). The National Center for Health Statistics' estimates during 2007-2008 included the following results: about 34.2% of U.S. adults 20 years old and over are overweight,

33.8% are obese, and 5.7% are extremely obese (Carroll & Surveys, 2010). As a consequence, a high-priority objective of the program Healthy People 2020 is to reduce the proportion of adults who are obese (U.S. Department of Health and Human Services, 2014). One of the recommendations for substantial health benefits made by the Physical Activity Guidelines (PAG) is to do moderate-intensity physical activity, which includes walking briskly (3 miles per hour) 150 minutes a week.

There are a number of factors that affect a person's level of physical activity, including socio-demographic, psychological, emotional, behavioral, social, cultural, and physical environment factors (Trost, Owen, Bauman, Sallis, & Brown, 2002). Trost and colleagues examined recent studies of physical activity and found that researchers are now including physical environmental indicators in their analysis. They concluded that physical environmental indicators remain a high priority for future research to better understand how physical environment could affect a person's level of physical activity.

The U.S. Department of Health and Human Services determined that physical environment factors positively associated with physical activity include the following: access to facilities, enjoyable scenery, presence of sidewalks, safe neighborhoods, and access to public transportation (Trost et al., 2002). The Centers for Disease Control and Prevention (CDC) is aware that the current U.S. transportation infrastructure focuses on motor vehicles. Nonetheless, the center is encouraging communities to change their urban design and land use to consider transportation policies that improve access to places for physical activities and improve the quality of life. The agency also emphasizes the need to create roadways designed for safe, attractive, and comfortable access for pedestrians and bicyclists of all ages and abilities, in order to prevent certain chronic diseases,

improve environmental health, and stimulate access to all people (Centers for Disease Control and Prevention, 2011).

2.3 A Framework Evaluating Built Environment

A built environment (BE) is defined as natural or constructed features that define the layout of a neighborhood. The built environments, or the physical forms of communities, include urban features such as street patterns, accessibility to food stores, recreational and cultural destinations, government and administrative places, landscaping, land use, and population density, among others (Brownson, Hoehner, Day, Forsyth, & Sallis, 2009).

A number of recent studies have researched and identified BE indicators, especially the ones associated with walking and physical activity. A variety of measurements were developed in different studies, including objective measures developed using GIS, subjective measures derived from surveys, and measures developed using audits (Maghelal & Capp, 2008). The measures derived from surveys included scales and auto-reports assessments related to levels of usage, satisfaction, and perception of suitability of walking. The audits are objective in-person data gathering where an inspector evaluates areas suitable to walk using observational methods. The measures using GIS are virtual audits that provide standardized methods that can be replicable and can overcome methodological problems, increasing reliability. From a literature review, Maghelal and Capp (2008) identified BE indicators and agreed on the need to develop indicators relevant to research on physical activity and public health. The authors also encouraged the need to develop objective indicators due to their reliability and replicative nature.

Even though pure perception of walkable areas is considered a less reliable measure because neighborhoods vary greatly (Maghelal & Capp, 2008), the presence of

preferences for walking may affect the decision to walk (Carlson, Aytur, Gardner, & Rogers, 2012). Carlson et al. (2012) adapted a framework proposed by Glass & McAtee (2006) and defined BE as risk regulators that capture the opportunities and constraints for physical activity. Further, the framework includes the perception of BE as a pathway that influences walking decisions and a person's level of physical activity.

In response to a lack of objective and standardized measures, Maghelal and Capp (2008) identified a comprehensive list of BE indicators associated with walking. They categorized the indicators into four dimensions: design, density, diversity, and quality. These dimensions gather objective measures that could be calculated using GIS, but the quality dimension includes comfort measures that are associated with perception of walkable areas and lower BMI (Tilt, Unfried, & Roca, 2007). Brownson and colleagues identified the most frequently assessed measures of the built environment that can be measured using GIS: population density, land-use mix, access to recreational facilities, street patterns, sidewalk coverage, and crime. (Brownson et al., 2009). The three research articles converge on the multidimensionality and indicators of the built environment framework and agreed that there was a need to improve the BE indicators for a better understanding on physical activity.

2.4 Summary

The benefit of observing the Adventist population for research purposes is due to its lifestyle characteristics. For decades, this population has contributed to the understanding of the benefits of certain eating habits, but questions have not been addressed about the effects of physical activity on health. These questions arise at a time when obesity is a concern and a priority for the U.S. Department of Health and Human Services. The

natural and constructed features of a neighborhood could help enhance walking and physical activity.

The chapter reviews the role of built environment in influencing physical activity, and how the BE could help prevent obesity. Although attempts to measure built environment indicators using GIS have been made, the scholars (Brownson et al., 2009; Carlson et al., 2012; Maghelal & Capp, 2008) convey the need to consider broader and multidimensional measures. This is why an adapted conceptual model of built environment was chosen. The adapted model incorporates the multidimensionality proposed by Maghelal and Capp (2008) into the framework proposed by Carlson et al. (2012). The model also incorporates many of the indicators proposed by Brownson et al. (2009). The following chapter details the client's requirement and presents the system and design overview.

Chapter 3 – Systems Analysis and Design

This chapter focuses on addressing the client's problem by analyzing the system requirements. A system design was then developed based on the requirements. In addition, a project plan was laid out and risk assessment was conducted. The chapter is divided in the following sections: problem statement, requirements analysis, system design, project plan, and a summary.

3.1 Problem Statement

Although Adventists in the United States and Canada share similar lifestyle traits, Loma Linda has a higher concentration of Adventists with a longer life expectancy. However, there was no evidence as to whether the layout of the city of Loma Linda resulted in an increase in the life expectancy of its Adventist population. This was in part because there was an absence of tools that measure built environment indicators associated with physical activity. This absence, combined with the difficulty of gathering and assembling comparable datasets, makes it difficult to understand the effects of a built environment on people's walking behavior.

3.2 Requirements Analysis

This section outlines the functional and non-functional requirements for this project. Functional requirements are the capabilities of the tools or software under development for the user. Non-functional requirements refer to particular qualities that the tools or their output must have. The primary functional requirement was to quantify built environment indicators from readily available data. The primary non-functional requirement was to use the adapted Brownson and Maghelal frameworks (Brownson et

al., 2009; Maghelal & Capp, 2008). Table 3-1 outlines the functional and non-functional requirements of this project.

Table 3-1: Functional and Non-Functional Requirements

Functional Requirements	Non-Functional Requirements
<p>The toolbox shall be able to calculate multiple dimensions of the built environment.</p> <p>The tools shall include a user-friendly interface.</p> <p>The outputs generated by the toolbox shall include the identifier of the original records.</p> <p>Compare Loma Linda built environment with nearby cities.</p>	<p>The indicators shall follow Brownson and Maghelal framework recommendations.</p> <p>Indicators shall be gathered by readily available data.</p> <p>Data are free and accessible.</p> <p>Map results should not include the exact location of the participants.</p> <p>Include the documentation of the formulas used and deliver results to the client.</p>

3.2.1 Functional Requirements

Functional requirements included the development of a toolbox that can quantify four built environment factors, including access to recreational areas, population density, land use diversity, and sidewalk quality. Since the future users of the tools are not GIS professionals, the tools should include a graphic user interface. Since the calculated built environment indicators will be used for further statistical analysis, the outputs generated from the toolbox should retain the identifier (ID) of the individual participants. Retaining this identifier ensures that the ID of the original participant will not be lost. Therefore, all

participants can be associated with their own built environment indicators, and then later be combined with physical activity and health outcome data collected by the client. The last functional requirement was to compare the built environment of the cities of Loma Linda and Redlands using the Adventist point data as a case study.

3.2.2 Non-Functional Requirements

The first non-functional requirement was to follow Brownson and Maghelal's framework of built environment (Brownson et al., 2009; Maghelal & Capp, 2008). To allow for replication of the analysis for other cities, it was also required to use publicly available datasets, which should be free and accessible from different public sources. Due to the confidentiality of the data, another non-functional requirement was to geo-mask every point data of the participants presented on the maps shared with the public. Other non-functional requirements included documentation of the different formulas used and delivering the empirical results to the client.

3.3 System Design

The principal outcome of the project was to provide a number of scripted tools to calculate built environment indicators using data that are available from public domains. The development of the project's system design was based on the functional and non-functional requirements defined earlier in the project. Figure 3.1 outlines the major components of the system design.

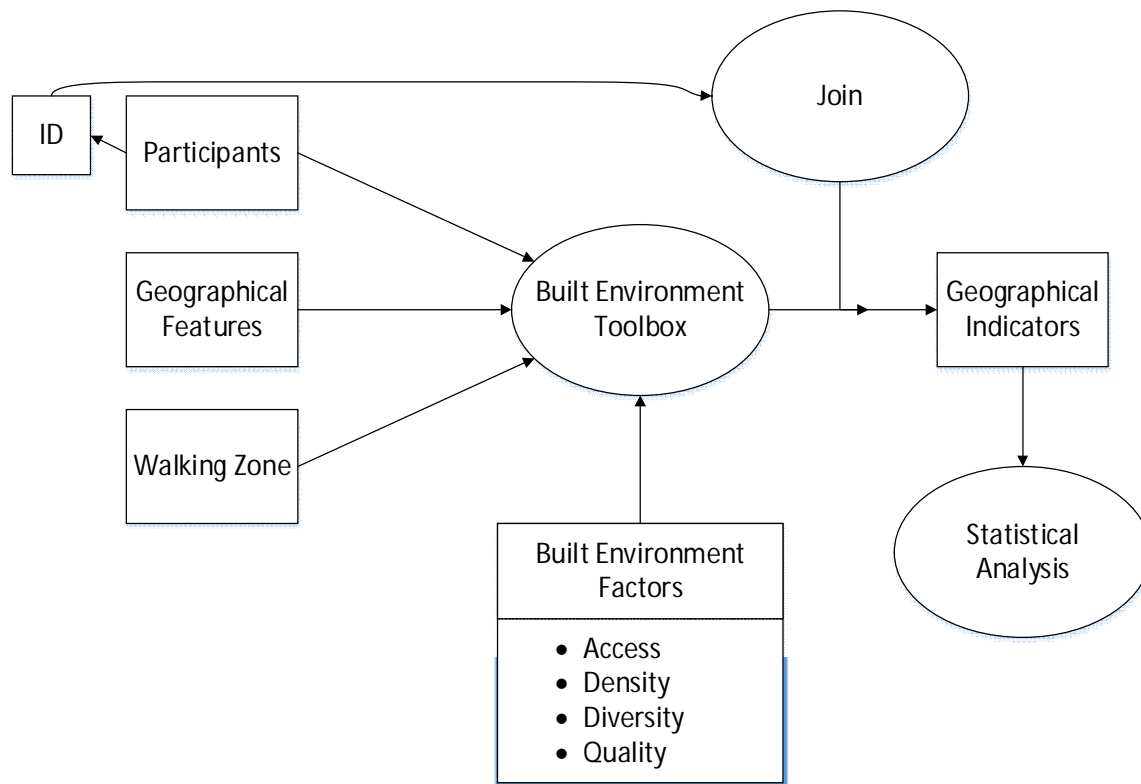


Figure 3-1: Components of the System Design

The model is composed of three major components: the input requirements, the built environment toolset, and the geographical indicator outcomes. In the input requirements, there are three major elements: the point data of participants, geographical features, and a walking zone. The participants are a sample from the Adventist Health Study 2 (AHS-2). The AHS-2 sample was provided by the client and included the point locations of the participants from the city of Loma Linda and city of Redlands at house or street level.

The geographical features considered could vary according to the interest of the user. This project used recreational areas to measure accessibility and population density to represent density. For the factor of Diversity, the project calculated land use mix for both cities. Finally, a vegetation index along the sidewalks was developed to represent Quality. In addition, the built environment toolbox includes another two tools, one for

creating a walking zone, and the other for classifying the vegetation index. The extent of the built environment in this project was defined as a one-mile walking zone in which the indicators of Density, Diversity, and Quality were measured.

The quantified geographical indicators were the outcomes. Those outcomes were further compared between Loma Linda and Redlands using statistical methods. In order to use the results for further analysis, it was necessary to join the identifier of the AHS-2 participants to the geographical indicator outcomes. After the join process, the geographical indicators were ready to be used in further analysis such as comparing people's built environments with physical activity and chronic diseases.

3.4 Project Plan

The project plan includes nine major stages, illustrated in Figure 3-2.

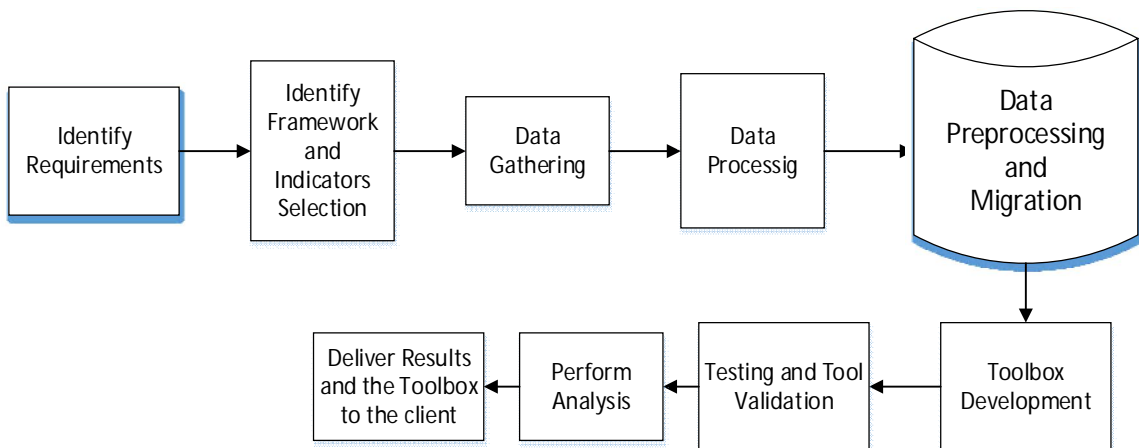


Figure 3-2: Project Plan

The first step of the project was to identify requirements. It included gathering the requirements from client in order to define specific requests, functional and non-functional requirements, and deliverables. The second step was to identify the framework

that the project used and selected the indicators used to measure the built environment. This step included a literature review and discussions with the client. Once the framework and indicators were chosen, the next step was to gather the data required to perform the analysis. The fourth step of the project was data processing to clean up the datasets. Steps two through four were repeated until all datasets were collected and cleaned. Then the data were migrated into a geodatabase for each city. The geodatabase included the point data from the AHS-2 and the data of the geographical features from the selected indicators from the two cities. The next step was the development of the toolbox. The toolbox included series of Python scripts created to calculate built environment indicators. The seventh step was to test and validate the toolsets. The purpose was to confirm reliability of the tool and decrease processing time. The eight step was to perform the analysis to compare the built environment of the city of Loma Linda with the city of Redlands. The final step was deliver the results of the analysis from the two cities and the toolbox with the pertinent documentation to the client.

3.5 Summary

This chapter provided an assessment to the project's goals. Functional and non-functional requirements were examined to satisfy the client's needs. The system design and the project plan were introduced. Also this report demonstrated how the project was completed. In addition, a risk assessment was developed in order to secure the success of the project. The next chapter discusses database design to include the conceptual and logical implementation of the project.

Chapter 4 – Database Design

This chapter introduces the conceptual and logical data models, and details the type of data used to implement the project. Several processing tools were developed for future data processing needs. The chapter is divided in the following sections: conceptual data model, logical data model, data sources and attributes, data scrubbing and loading, and a summary.

4.1 Conceptual Data Model

A conceptual data model helps visualize the objects involved in the project and their relationships. It also provides a general guidance to design a logical model. The amount of physical activity a person engages in is a combination of personal choice and conditions presented in his/her neighborhood (CDC, 2008). To increase people's physical activity, Centers for Disease Control and Prevention recommends to improve the access to parks and recreational facilities. The inclusion of sidewalks in a neighborhood represents a simple and practical way to provide access to parks and recreational areas. The presence of sidewalks can increase physical activity levels by providing an opportunity for people to safely and easily walk to their destinations. The attractiveness along the sidewalk can also influence people's decision to walk. Based on the literature review, four environment indicators were considered in this project: access to recreational areas, population density, diversity of the land use, and quality of the sidewalks.

Two conceptual models are presented in Figures 4-1 and 4-2, illustrating the impact of the four factors of the built environment on local residents. Figure 4-1 shows the conceptual model for accessibility to recreational areas and sidewalk attractiveness.

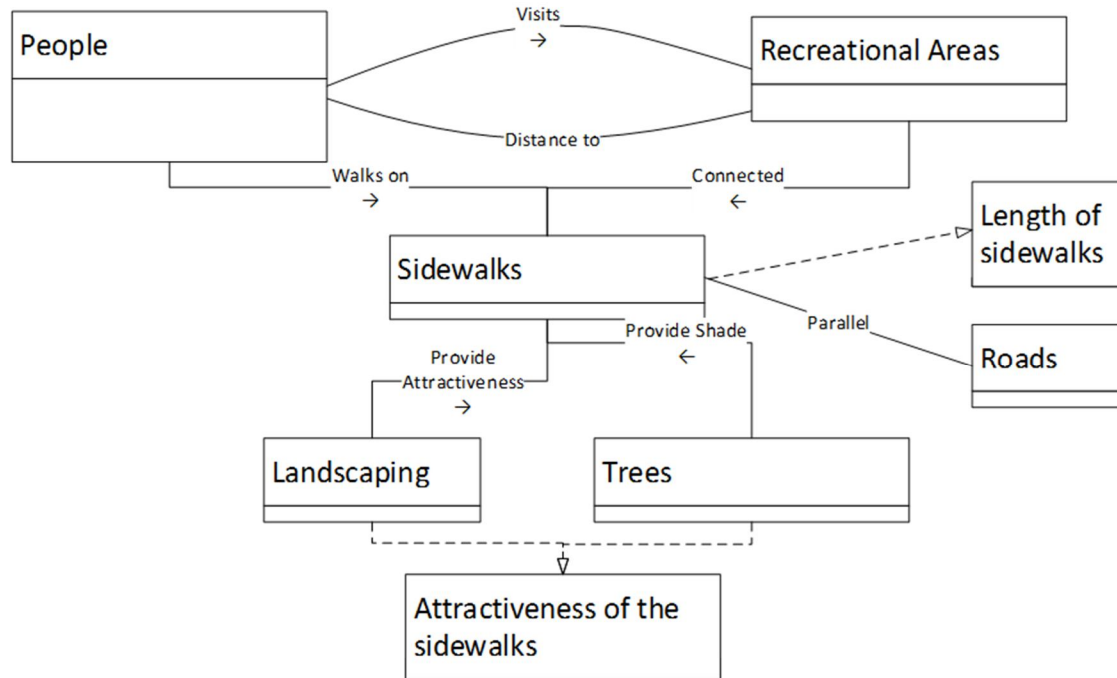


Figure 4-1: Access to Recreational Areas and Attractiveness of the Sidewalks

The inclusion of sidewalks provides connective paths between neighborhoods and recreational areas. People walk on sidewalks to visit recreational areas. An important conceptual assumption was that sidewalks align parallel to the roads. The length of the sidewalk between a neighborhood and its nearby recreational areas is an indicator of practical access to recreational areas.

The presence of vegetation is highly correlated with active and walkable communities (Tilt et al., 2007). The inclusion of trees and landscaping near sidewalks provides an enjoyable scenery that could trigger walking behavior. Landscaping and shade parallel to sidewalks are useful indicators for attractiveness and comfort to pedestrians.

It is commonly thought that an integrated community is associated with better health outcomes, less television usage, shorter commutes, and increased walkability (Leyden,

2003). Figure 4-2 identifies the impact of the population density and land use diversity on walkability.

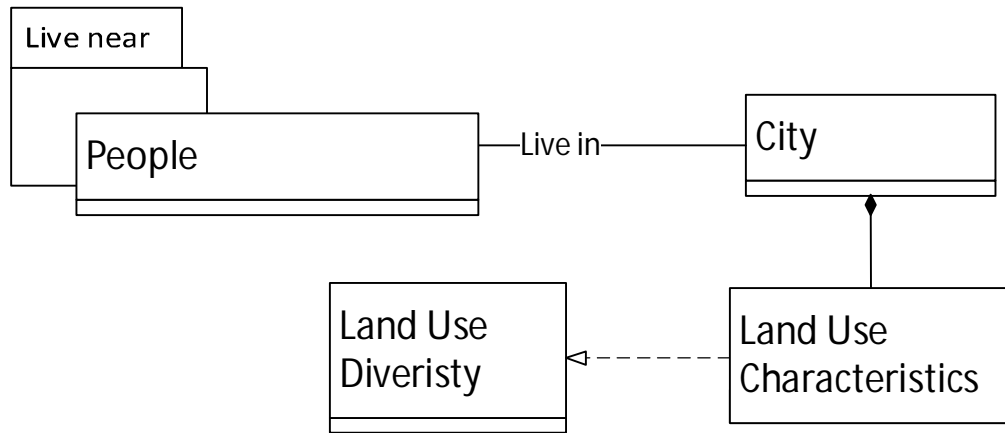


Figure 4-2: Conceptual Model Density and Diversity Factors

People live near other people to establish neighborhoods and cities. The concentration of people is often measured by population density. Every city in the U.S. has different land use characteristics. Separating residential with non-residential areas may increase a detriment of health and lifestyle (Christian et al., 2011). A good layout of a city can support physical activity by mixing residential with non-residential areas. The diversity of the land use in a city provides an opportunity for people to walk in the areas with different characteristics.

Based on the conceptual models of how built environment may affect physical activity, four indicators of the built environment were considered in this project (Figure 4-3).

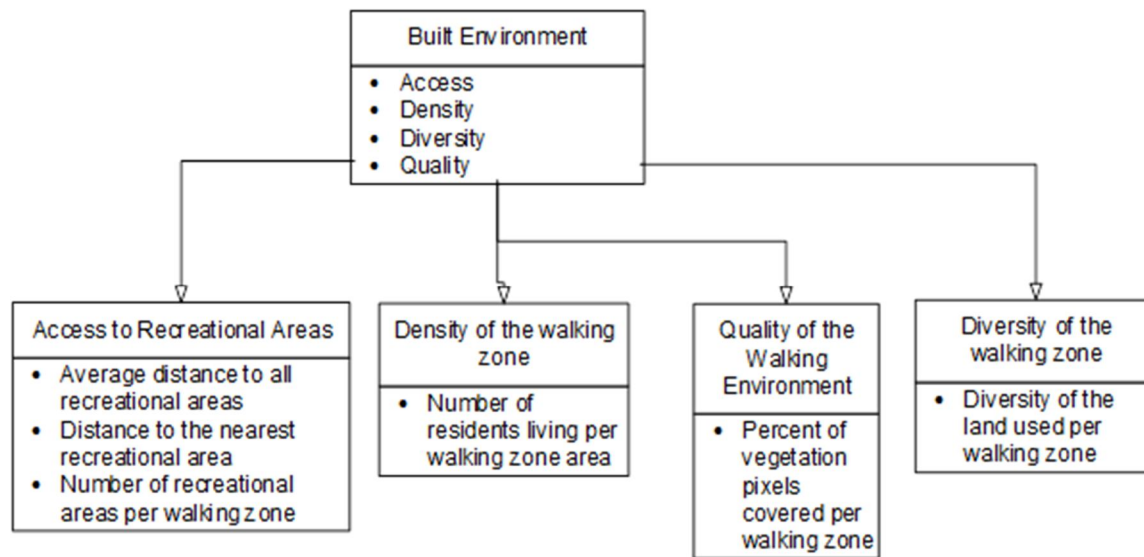


Figure 4-3: Built Environment Indicators

The access to recreational areas was measured by the average distance to all recreational areas, distance to the nearest recreational area, and number of recreational areas in the walking zone. The project defined the population density as the number of residents living per walking zone area. The attractiveness of the walking environment was measured by the percent of the sidewalks in the walking zone that are covered by vegetation. For the factor of land use diversity, the indicator was defined as diversity of the land use in the walking zone.

4.2 Logical Data Model

The logical model is presented in two diagrams for a better understanding of the model. The first part of the logical model consists of the input data required to calculate built environment indicators. The built environment indicators were compared between the cities of Loma Linda and Redlands; therefore, two geodatabases of the same structure were created for each city respectively. Figure 4-4 shows the database structure for one city.

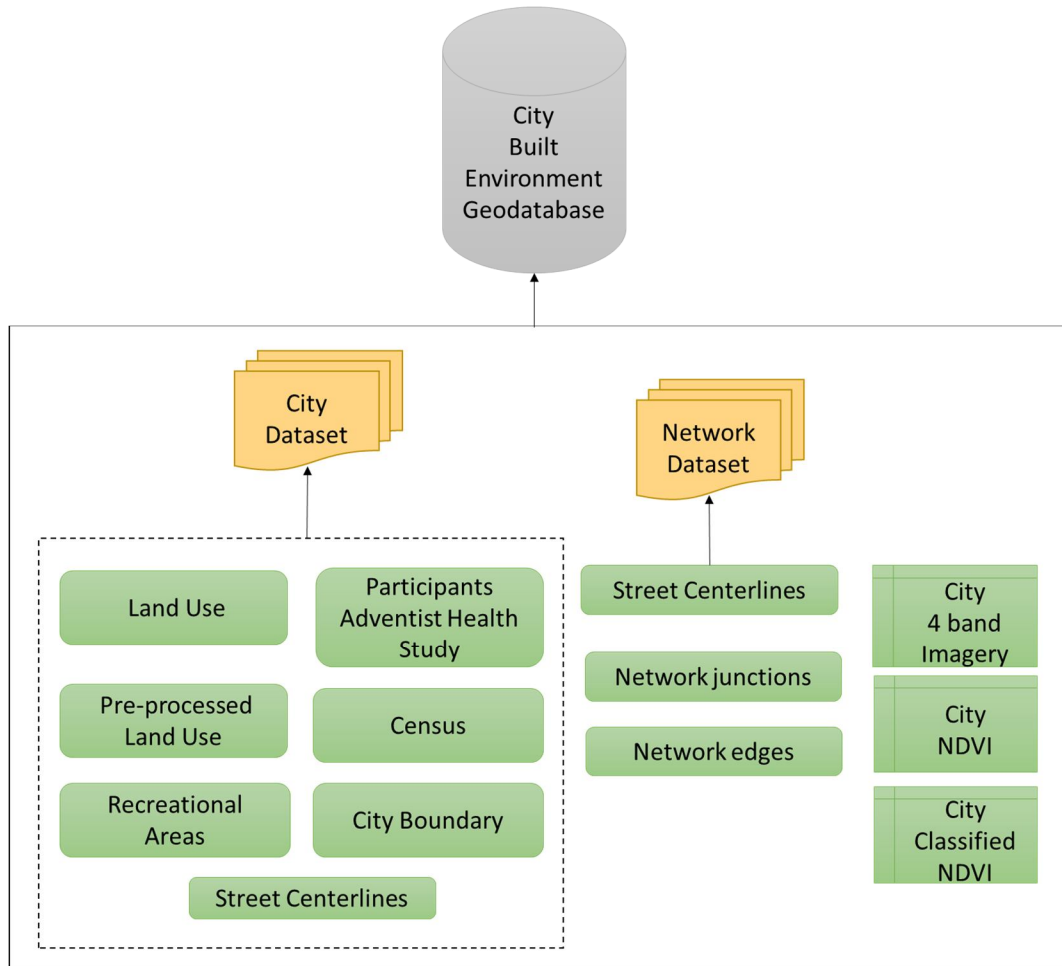


Figure 4-4: Logical Model for the Input Data in the Built Environment Geodatabase

The city built environment geodatabase includes two features datasets including the city dataset and the network dataset and imagery data. The feature classes in the city dataset include the point feature class representing participants from the Adventist Health Study 2 (AHS-2), the census information, land use and pre-processed land use, recreational areas, and street centerlines. The network dataset was developed based on the street centerlines, which has the topological rules enforced. The imagery data include three raster datasets: a 4-band imagery for the study area, a Normalized Difference Vegetation Index (NDVI) layer for the city, and a classified NDVI layer. The NDVI and the classified NDVI were the outcomes of pre-processing the 4-band imagery data. These

raster data were used to evaluate vegetation coverage along the sidewalks in the participants' built environment.

Figure 4-5 shows the logical model for storing the intermediary and outcome data. To calculate the four built environment indicators, a one-mile walking zone feature class was created for each participant at household level or street level as this was how the built environment was defined in this project. Then the Build Environment toolset developed in this project would calculate the four indicators using the walking zone and other input data. All the outcomes from the tools were stored as feature tables. The feature tables were then joined to the AHS-2 point data such that each participant obtained measurements for his/her built environment.

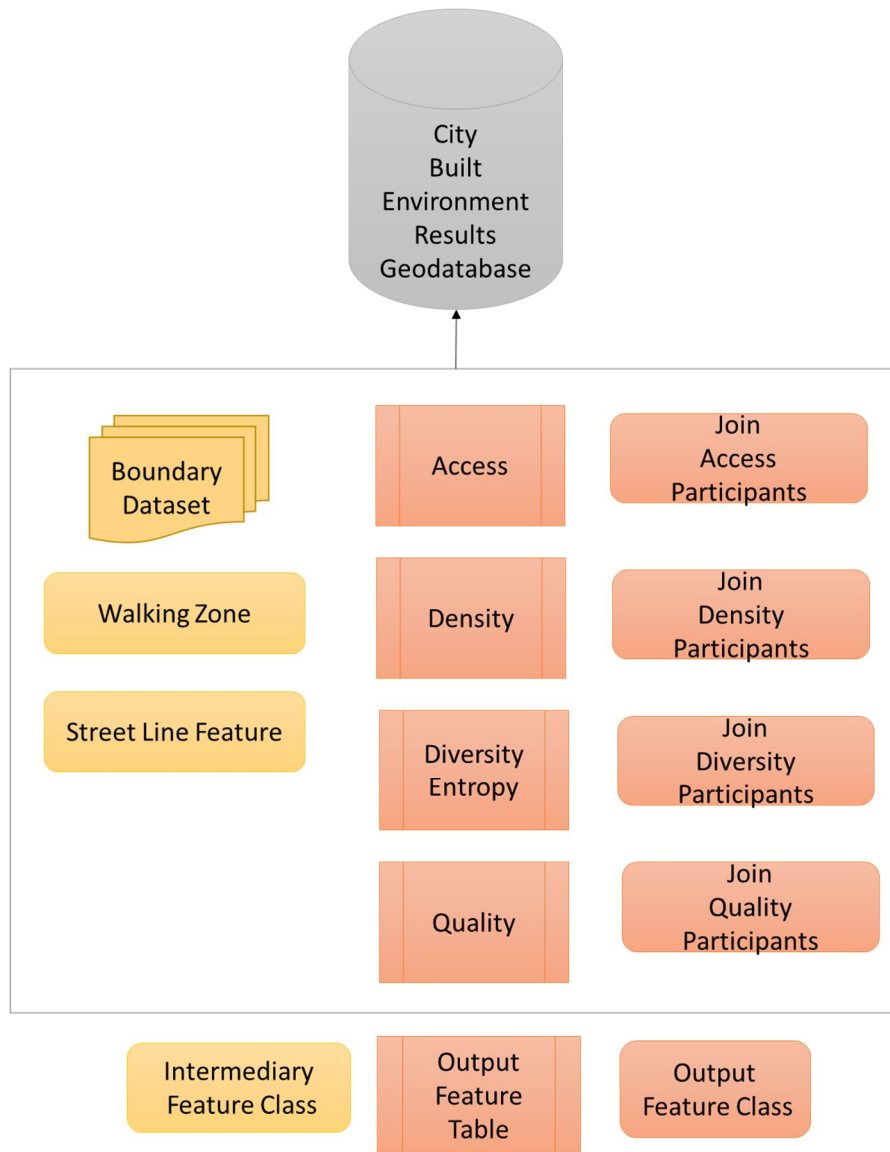


Figure 4-5: Logical Model for the Output Data in the Built Environment Geodatabase

4.3 Data Sources and Descriptions

As one of the requirements, the project gathered readily available data from private and public domains. Table 4-1 details all data used in the project. It shows the name of the data name, the use of the data, attributes used, the description of the data, and the data source.

Table 4-1: Project data

Name	Use	Attributes	Description	Source
Adventist Participants	Create the walking zones and as origins for accessibility.	<ul style="list-style-type: none"> ▪ Location ▪ Identifier 	Point data provided by the client at household or street level. It includes the location of the participants from the Adventist Health Study 2.	Client
Imagery	Measure vegetation pixels for the Quality of the Sidewalks	<ul style="list-style-type: none"> ▪ Near Infrared band ▪ Red band 	The data were from the National Agriculture Imagery Program. The Image Service in ArcGIS Online is an aerial 4-band image from 2012 at 1-meter resolution.	California Department of Fish and Wildlife
Total Population	Measure population density for the Density Factor	<ul style="list-style-type: none"> ▪ Total population 	The U.S. Census Bureau originally created the feature class but was preprocessed by the University of Minnesota. It includes the total population from California at blocks level for 2010.	National Historical Geographic Information System
Street-centerlines	Create the Network Dataset	<ul style="list-style-type: none"> ▪ Location ▪ Street types 	The street centerlines were from the U.S. and Canada Detailed Streets, 2012.	Esri
Land Use	Measure the diversity of the land uses for the Diversity factor	<ul style="list-style-type: none"> ▪ Land use categories ▪ Area 	The land use data were provided by the Planning office from the City. The two dataset were from 2014.	City of Loma Linda and Redlands

Recreational Areas	Provide destinations to measure Access	<ul style="list-style-type: none"> ▪ Location 	Recreational areas are from the U.S. Recreation Areas, 2012.	Esri
City Limits	Process data and produce maps	<ul style="list-style-type: none"> ▪ Geometry 	City limits by the U.S. Populated Place Area, 2010.	Esri

The client provided the point locations from the Adventist Health Study (AHS-2) at household or street level. The AHS-2 input data were used to create the participants walking zone and were used as origins to measure accessibility to recreational areas. The identifier field was required to be in Text or Integer format and unique. The data of recreational areas from the city of Loma Linda and Redlands were available through Esri. It was found that this dataset did not contain a complete list of the recreational sites. The recreational areas defined by Esri do not include parks from universities and schools. Other limitation, presented in both cities but, may affect the conclusions of access to recreational areas of this study, is that the dataset does not include attributes of area covered and potential entrances to the parks. Even though these limitations are presented in both cities, it presents a bias for the conclusion of the indicator of access to recreational areas because the missing recreational areas and their surroundings may be open to general public. This may present an opportunity for people to do physical activity and the potential entrances and area covered by the park could have direct implications in calculating distances and determining the indicator of access to recreational areas. For the quality of the sidewalks, the imagery data were gathered from an Image Service in ArcGIS Online. The imagery data of 1-meter of resolution contain four bands: red, green, blue, and near infrared. The near-infrared and red bands were used to calculate a vegetation index. The city limits feature classes, provided by Esri, were used for data

extraction and map production. Esri also provides street centerlines to create the network datasets for the two cities.

To calculate population density, the 2010 population data of California at the census block level were collected through a web service provided by the National Historical Geographic Information System (NHGIS), University of Minnesota (Minnesota Population Center, 2011). To measure land use mix, land use data were obtained from the City Planning Office of the two cities.

4.4 Data Scrubbing and Loading

After each dataset were obtained, the next step was to process and load them into the two different geodatabases. Several geo-processing tools were run to clean and process the data to satisfy the requirements of the Built Environment Toolbox. Table 4-2 shows the pre-processing tasks taken to ensure the quality of the data and to prepare data to be used in the tools. All the feature classes in the geodatabases were projected to NAD 1983 State Plane California V FIPS 0405 US feet Coordinate System.

Table 4-2: Preprocessing tasks

Data name	Format	Objective	Tools and Services
Adventist Participants	Point Shapefile	Clip the data to the city boundaries to assure they are completely within the city boundaries.	<ul style="list-style-type: none"> ▪ Feature class to Feature class ▪ Projection ▪ Select by Location
Imagery	Image Service	Extract the Image Service using the city boundaries to calculate and classify the NDVI.	<ul style="list-style-type: none"> ▪ Extract by Mask ▪ Projection ▪ Raster Calculator ▪ Reclassify tool
Total Population		Select the blocks from the study area to	<ul style="list-style-type: none"> ▪ Select by Location

	Shapefile and Comma Separated Values	decrease processing time.	<ul style="list-style-type: none"> ▪ Join Field Management ▪ Feature class to Feature class ▪ Projection
Street-centerlines	Feature class	Cleaned data in order to satisfy definition of potential walkable roads and create a network dataset	<ul style="list-style-type: none"> ▪ Projection ▪ Editing tools ▪ Select by Attributes ▪ Esri base maps ▪ Network dataset wizard
Land Use	Shapefile	Land Use data had to be categorized and dissolved by the categories of interest. The categories were residential, park and open spaces, and others.	<ul style="list-style-type: none"> ▪ Feature class to Feature class ▪ Projection ▪ Add Field Management ▪ Dissolve Management
Recreational Areas	Feature class	Select the recreational areas that follow within the city limits.	<ul style="list-style-type: none"> ▪ Select by Location ▪ Projection

The project objective was to compare the built environments between Loma Linda and Redlands. A Select by Location tool was performed using the city limits feature class to guarantee that the point data of participants from the AHS-2 fall within the city limits. The Imagery data were extracted using the Extract by Mask tool in ArcMap with the city limits as the feature mask. To assess vegetation coverage of sidewalks, a normalized difference vegetation index (NDVI) was calculated using the extracted imagery. The NDVI was calculated using the near infrared and red bands with the Raster Calculator tool. The formula of NDVI is as follows:

$$NDVI = \frac{NIR_{band} - R_{band}}{NIR_{band} + R_{band}}$$

where NIR_{band} is the near infrared band and R_{band} is the red band. NDVI is a continuous index that ranges from -1 to 1. To classify the pixels into non-vegetation and vegetation, the threshold of 0.1 was chosen. The pixels having NDVI values greater than or equal to 0.1 represent vegetation. This was completed with the Reclassify tool in ArcMap. The final output contains pixel value of 1 for vegetation and 0 for non-vegetation. Figure 4-6 shows an example of the attribute table of a classified NDVI, where VALUE = 1 indicating the category for vegetation pixels and VALUE = 0 representing the category of non-vegetation pixels. The field named COUNT shows the number of pixels for each category. This binary raster was then used to extract vegetation coverage along the sidewalks in the built environment, which will be explained in Chapter 5.

	OBJECTID *	VALUE	COUNT
	1	0	2303206
	2	1	5363490

Figure 4-6: Attribute Table of a Classified NDVI.

The data used to calculate population density from the NHGIS were originally gathered for the state of California and included a shapefile without attributes and a table in the format of comma-separated values. A selection of the blocks that fall in the study area was made to decrease the processing time. The table was then joined to the shapefile in order to create a feature class with the attributes of interest.

To create a network dataset that only includes streets suitable for walking, both Editing and Select by Attributes tools were used to select the street segments that could

be walkable. The highways, highway ramps, and streets with a speed limit that exceeded 50-miles per hour were excluded from the street segments. Esri base maps were used to compare and guarantee that the street segments agreed to the streets in the web service. The network dataset wizard was used to create the network dataset.

To measure the land use diversity of the walking zone, the land use data were dissolved by the categories of interest. A new field was then added in the land use feature class and the land use attributes were classified according to the client's recommendations including residential, park and open space, and other.

4.5 Summary

In this chapter, both the conceptual and logical data models were presented to show how built environment indicators would affect physical activity and how these datasets were structured to form a geodatabase. Various datasets were collected in the project including census block, imagery data, land use and parcel level data, street centerlines, and recreational areas,. These datasets were processed and cleaned using different geoprocessing tools. Finally, a geodatabase was created for either of the two cities. In the next chapter, implementation of the toolset with the feature classes in the geodatabase will be elaborated.

Chapter 5 – Implementation

This chapter describes the approaches employed to implement the Built Environment Toolbox in ArcGIS 10.2 using Python. Four built environment indicators were measured with the corresponding custom tools in the project: access to recreational areas, population density, land use diversity, and sidewalk quality. In addition, two supporting tools were developed for creating a walking zone and classifying the normalized difference vegetation index. This chapter contains following sections: measuring environmental indicators, walking zone generator, access to recreational areas, population density, diversity of the land use, quality of the sidewalks, and a summary.

5.1 Overview of Built Environment Toolbox

This project assesses the built environment characteristics that influence physical activity by developing a toolset that measures four environmental indicators considered by Brownson et al. (2009) and Maghelal and Capp (2008). Figure 5.1 shows the general workflow of the Built Environment Toolbox.

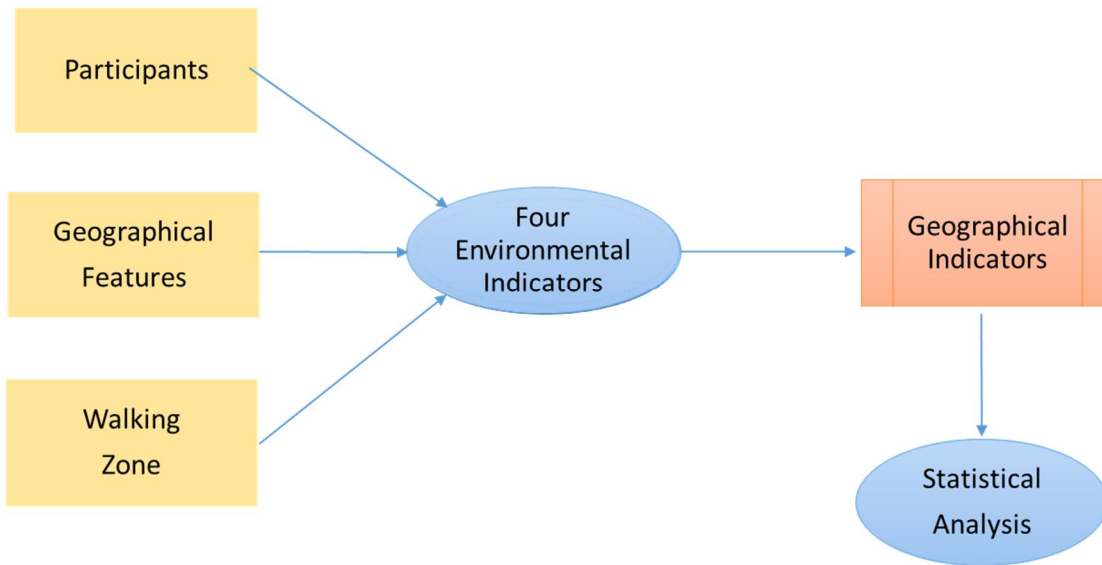


Figure 5-1: Environment Toolset Workflow

With the input data, including participants' location derived from the Adventist Health Study 2 (AHS-2), the geographical features, and a walking zone, the toolset calculates four geographical indicators and saved them in a table format for statistical analysis. Figure 5-2 details the toolset structure.

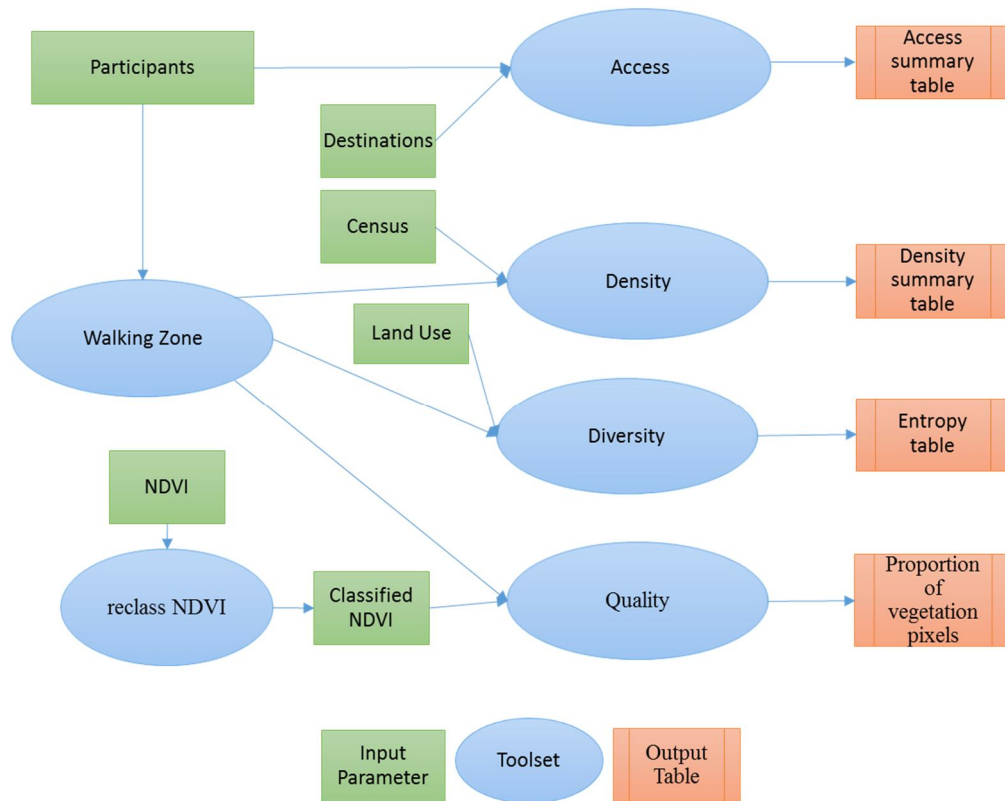


Figure 5-2: Built Environment Toolbox Workflow

The Built Environment Toolbox includes a total number of six tools: four of them were developed to measure the four built environment indicators and the other two tools were created for generating one-mile walking zones and classifying the normalized difference vegetation index (NDVI) respectively. The geographical features included in this project were: recreational destinations for measuring accessibility to recreational areas, population of Census blocks for calculating population density, land use for quantifying diversity of land use, and classified NDVI index of vegetation for evaluating sidewalk quality.

The six tools were developed in Python 2.7 using ArcPy functions, which offers more flexibility to manage the geometry of the feature classes and the attribute fields. The Integrated Development Environment used was PyScripter 2.7.

To ensure the output indicators can be joined with the original participant data, all the tools were built to handle identifiers (ID) in text, long integer, and short integer format. If the input identifier of the participants is in one of the formats specified above, the tools will produce geographical indicators populated with the same ID field name and format.

5.2 Walking Zone Generator

The outcomes of the Walking Zone Generator tool serve as the input parameters to the tools of Population Density, Diversity of the Land Use, and Sidewalk Quality. Figure 5-3 shows the input parameters and the two different outputs generated by the Walking Zone Generator tool.

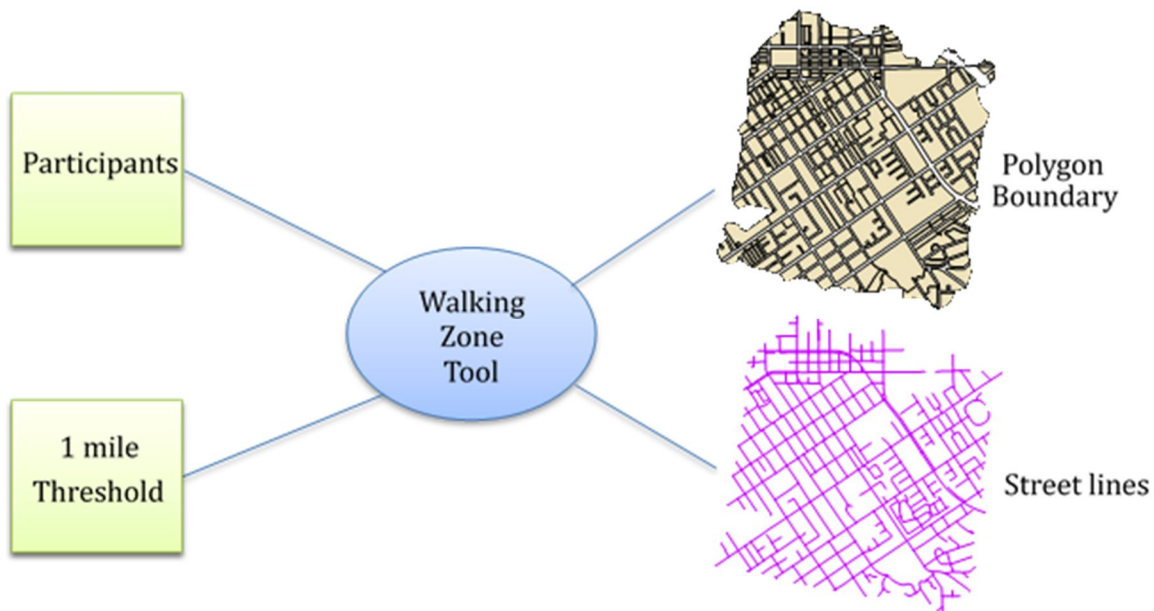


Figure 5-3: Walking Zone Generator Tool

Based on the findings of the Centers for Disease Control and Prevention (2014), a one-mile threshold was chosen to define built environment, which was also used by Müller-Riemenschneider et al. (2013). Even though one-mile network distance was used

in this project, the Walking Zone Generator tool can create walking zones using any distance thresholds. The Walking Zone Generator tool was programmed with the Service Area ArcPy functions and generates two different features classes: the boundary of the walking zone and the street lines within it. The tool requires the input data of the participants be entered as point locations with a unique identifier (ID) field. The format required for the ID can be an integer or text.

Figure 5-4 shows in detail the workflow of the Walking Zone Generator tool. Using the input network dataset, it creates a service area using ArcPy Network Analyst functions. The service area function creates six empty feature layers: Facilities, Polygons, Lines, and three different vectors Barriers.

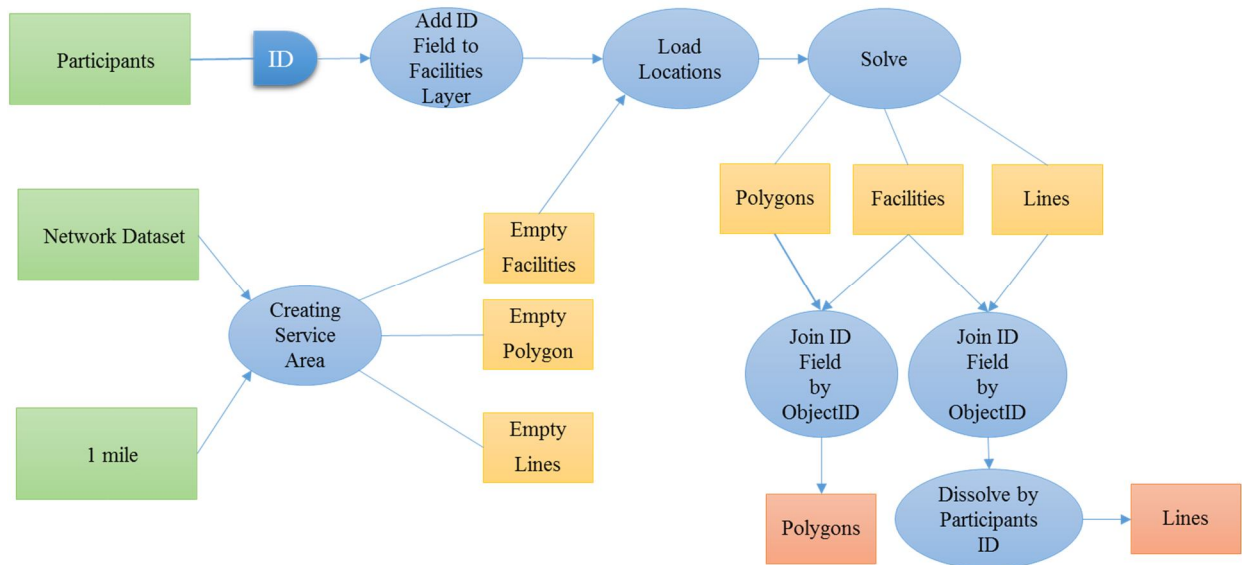


Figure 5-4: Walking Zone Generator Model

Barriers are restrictions applied to the network dataset before running the service area tool. The project does not establish any barrier criteria. Therefore, the tool only works with the Facilities, Lines, and Polygons layers. These three layers will hold the

input data of the participants (Facilities), the resulting street line features (Lines), and the service areas (Polygon).

The second step inside the tool adds a field to the Facilities layer. The field added is given the same name as the original participant identifier field from the AHS-2 dataset. The third step of the tool is to add the participants' locations to the Facilities layer as the service areas will be created for each participant (facility). After the Solve function runs, the resulting service areas are stored in the Polygons layer and the streets in the service areas are stored in the Lines layer. The identifier of Polygons and Lines (FacilityID) matches the ObjectID of the Facilities layer, so that each participant is linked to his/her own service area. The tool then uses the Dissolve function to merge all the different street segments by the participant ID. This will result in a feature class that contains a single line feature for each participant with all associated street segments.

5.3 Access to Recreational Areas

The Access tool was created to measure accessibility to recreational areas including parks, amusement parks, and golf courses. Although the tool only considers the recreational areas as the destinations for this study, it can be used on any points of interest.

The Origin-to-Destination Cost Matrix (OD Matrix) ArcPy function was applied in this tool. With the participant data from the AHS-2 study, a one-mile search distance, and the recreation destinations, the tool creates a summary table with ArcPy Statistics Summary and Search Cursors functions. The summary table includes several accessibility indicators: the average distance to all recreation destinations, distance to the nearest

recreation destination, and the count of recreation destinations within the search distance, as well as the average distance to those destinations.

Figure 5-5 shows the workflow of the Access tool. Using the network dataset, the tool creates a group of OD Cost Matrix layers that includes Origins, Destinations, and ODLines. The tool then adds an ID field to the empty Origins layer to retain the identifier of the original participant data. The next step is to load the participant point data into the Origins layer using the Load Locations ArcPy function. The following step of the tool is to load the recreation destinations into the Destinations layer.

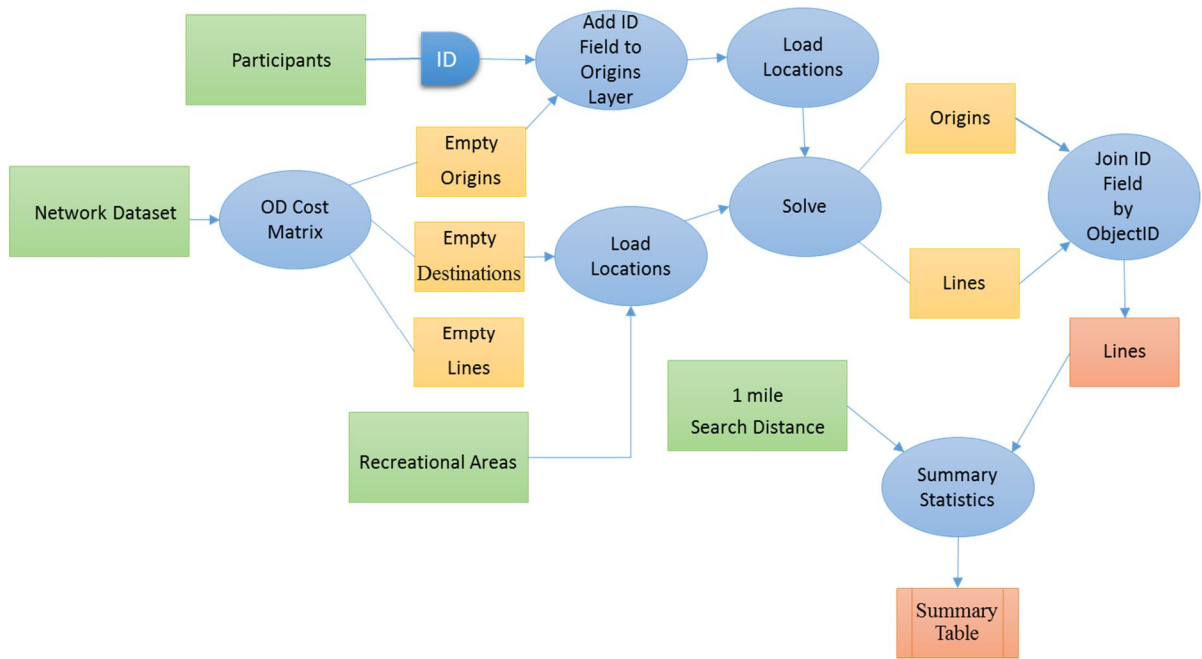


Figure 5-5: Access to Destinations Model

The Solve ArcPy function is then used to produce two different layers: Origins and ODLines, which are then joined by OriginsID and ObjectID. The resulting ODLines layer contains the distances to all destinations for each participant. Using the ODLines layer as the input, the Summary Statistics ArcPy function is used to calculate the average distance, minimum distance, sum of distances for all participants to all destinations, and

the total number of destinations. Using the Search and Update cursors, the tool can iterate through the summary table to identify the distances that are within the one-mile search threshold. For each participant, it counts how many destinations are within one mile and calculates the average distance. This will result in a table that contains various accessibility measures as shown in Figure 5-6.

AHS-2 ID	Total Destinations	Average distance to all Destinations	Distance to nearest Destination	Count of Destinations within a one mile	Average distance of to all Destinations within a one mile
----------	--------------------	--------------------------------------	---------------------------------	---	---

Figure 5-6: Accessibility Summary Table

5.4 Population Density

The Population Density tool calculates the population density for a walking zone using the input census blocks. There are different ways to calculate population density for a walking zone. One simple way is to obtain the sum of the population of all the blocks that intersects with the walking zone. A better approach would only account for the proportion of the blocks that are within the walking zone and uses the proportion to interpolate the population:

$$Density = \frac{\sum_i (P_i \cdot F_i)}{A_{WZ}}$$

where, P_i is the proportion of the i^{th} block that is within the walking zone, F_i is the population of the i^{th} block, and A_{wz} is the total area of the walking zone. Figure 5-7 illustrates only the proportion of the census block that falls in the walking zone is considered to calculate the population density.

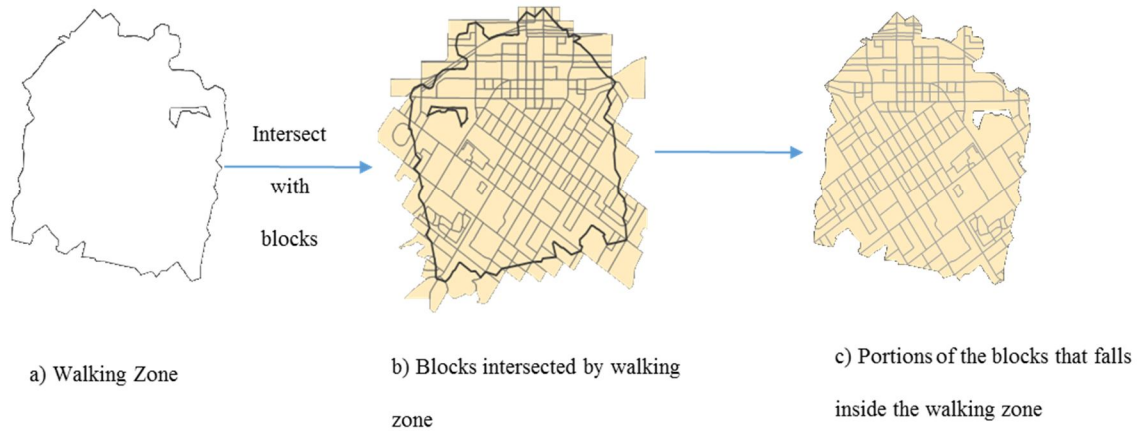


Figure 5-7: Population Density of a Walking Zone

Figure 5-8 shows how the density table is calculated. The Population Density tool iterates through each walking zone and uses the Intersect ArcPy function to intersect with block layer. This produces all the blocks that are intersected with the walking zone (Figure 5-7, c).

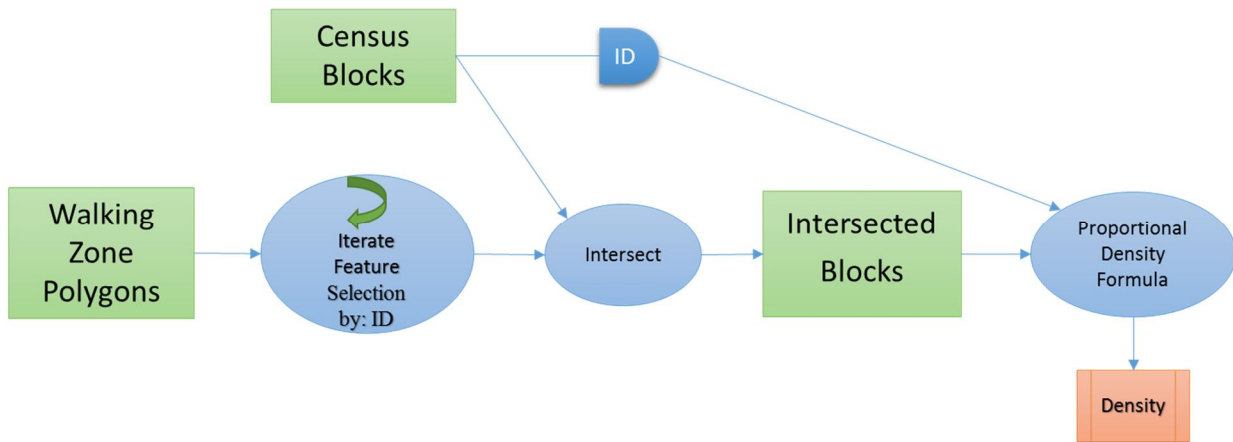


Figure 5-8: Population Density Tool

The Intersection function preserves the attributes, in specific, the identifiers from the two layers. The tool iterates through each block and calculates the proportion of the block within the walking zone. The proportion is then used to multiply the population attribute field to obtain the proportional population. The tool uses Insert cursors to save the population density values of the participants to the empty table.

5.5 Diversity of the Land Use

The Diversity tool calculates the diversity of the land use in a walking zone. Figure 5-9 shows an example of a walking zone with three different land use categories.

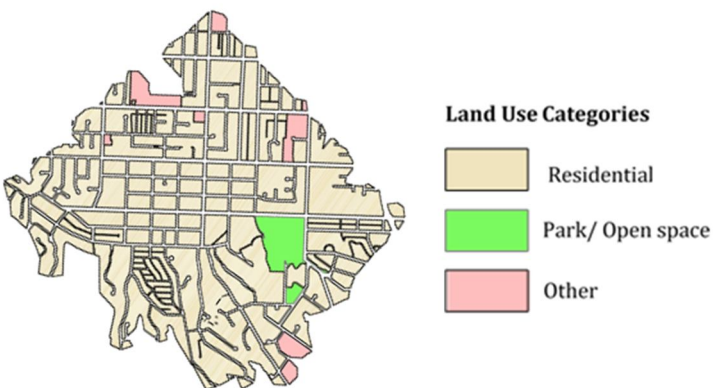


Figure 5-9: Walking Zone with Land Use Categories

The objective of this tool was to measure how evenly distributed different land uses are in a walking zone. The tool applies an entropy formula from Christian et al., (2011), originally adapted by Frank et al, (2005).

$$LUM = -1 \times \left(\frac{\sum_{i=1}^n (P_i \times \ln(P_i))}{\ln(n)} \right)$$

where P_i is the proportion of area covered by the i^{th} land use type and n is the number of land use categories of interest. The entropy values range from zero to one. A score of one indicates that there is an even distribution of the land uses. A score of zero indicates a single land use. The land uses considered in this project include residential, park or open spaces, and others land use types. As a requirement for the tool, the land use layer had to be dissolved by the land use categories of interest. Figure 5-10 shows how the Diversity of the Land Use tool calculates the diversity of land use.

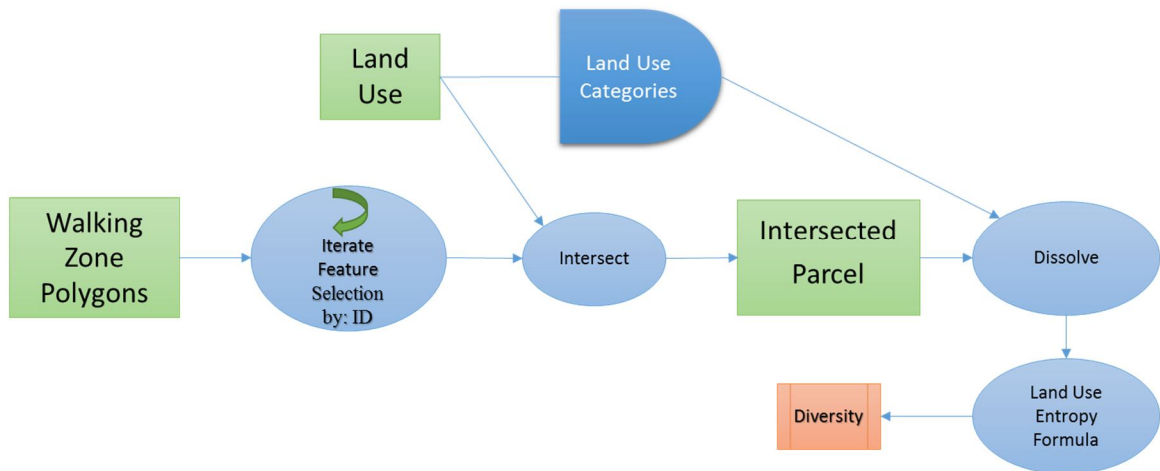


Figure 5-10: Diversity of the Land Use Tool

The tool iterates through the walking zone and uses the Intersect ArcPy function to intersect with the land use data. The parcels intersected with the walking zone are dissolved by the land use categories field using the Dissolve ArcPy function. The tool

uses Search cursors to apply the land use mix entropy formula and uses Insert cursors to populate the results into an empty table.

5.6 Quality of the Sidewalks

The Sidewalk Quality tool measures the quality of the sidewalks by calculating the proportion of vegetation pixels in the walking streets. This tool uses the street line features created by the Walking Zone Generator tool (Figure 5-3). Another input parameter for the Sidewalk Quality tool is a classified NDVI layer. The Reclassify tool was used to simplify the user experience and assure the NDVI layer to be correctly classified into vegetation and non-vegetation categories. Figure 5-11 shows the workflow of the Classify NDVI tool using an input NDVI raster and a predefined threshold as parameters. The output of this tool is a binary raster data representing vegetation and non-vegetation areas.

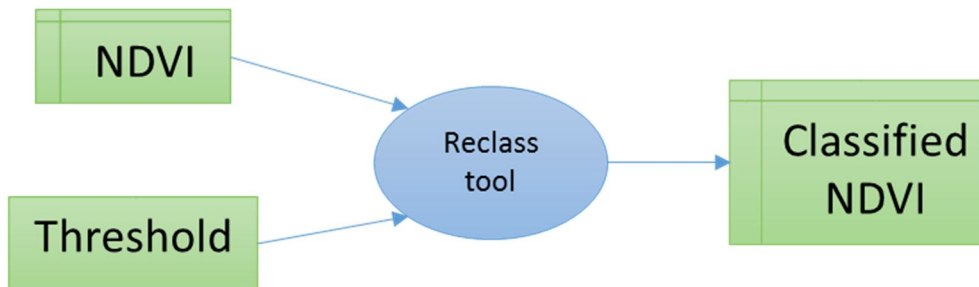


Figure 5-11: Classify NDVI Tool Model

Figure 5-12 shows an example of classified NDVI in the left figure. In order to evaluate the greenness of sidewalks, only vegetation along the sidewalks should be considered as shown in the right figure.

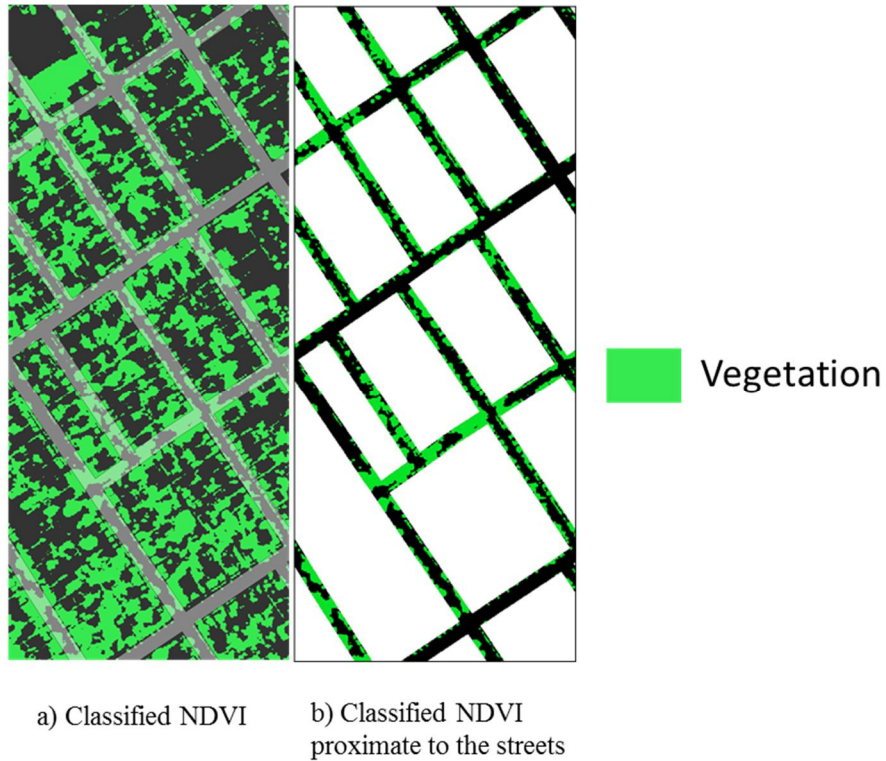


Figure 5-12: Classified NDVI

To automate this process, explained in Figure 5-13, the Sidewalk Quality tool creates a sixty-foot buffer around the street lines features and iterates using Search cursors. Although the tool can use any buffer distance and units, the city standard suggests a width of sixty feet in the local streets.

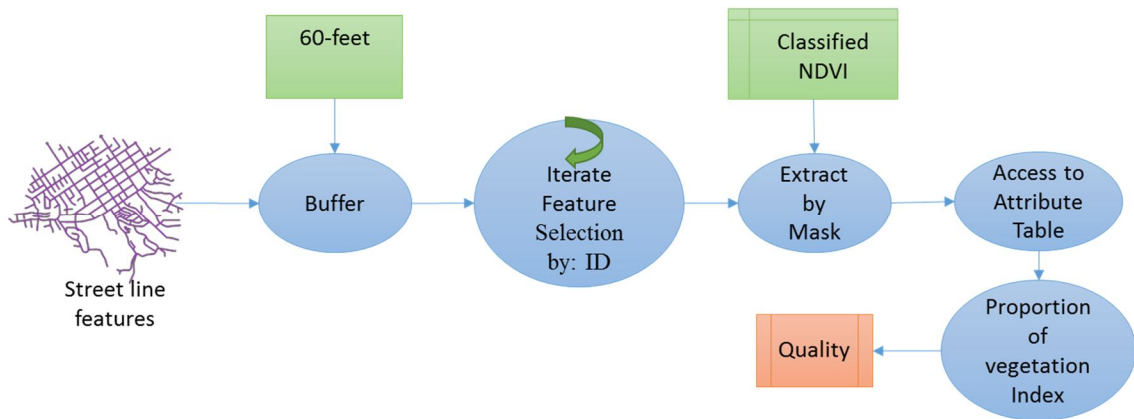


Figure 5-13: Sidewalk Quality Model

The tool then uses the Extract by Mask ArcPy function to extract the pixel values of the NDVI layer within the buffer mask. The tool then calculates the proportion of vegetation pixels that covers the walking environment.

5.7 Summary

This chapter discussed the processing tasks used to satisfy the requirements of the different tools developed. In addition, this chapter detailed the functions used for each tool to produce the different indicators. A total number of six tools were developed, four of them for measuring built environment, one for creating a walking zone, and one for classifying the Normalize Difference Vegetation Index. The following chapter shows the results of the different tools for Loma Linda and Redlands.

Chapter 6 – Results and Analysis

This chapter discusses the four quantified built environment indicators generated from the custom tools and compares these indicators between the cities of Loma Linda and Redlands. The chapter is divided in three sections: mapping individual variables, comparing built environments, and a summary.

6.1 Mapping Individual Variables

Among the four categories of built environment indicators, accessibility to recreational areas was measured with four variables (distance to the nearest recreational area, average distance to all recreational areas, and the count and average distance of recreational areas within a search distance, while others were measured with a single variable. Five variables are presented in this chapter and they are distance to the nearest recreational area, access to recreational areas within one-mile walking zone, population density, land use diversity, and percent of vegetation pixels per walking environment. Before comparing the built environments between the two cities, it is necessary to explore the distribution of inputting variables such as participants' home locations, land use, vegetation index, population density, and recreational areas.

The participant data were obtained from the Adventist Health Study 2 (AHS-2). In a sample of 1,985 Adventists, 73% of participants are from the city of Loma Linda and 27% from the city of Redlands. Figure 6.1 shows the distribution of the participants' home locations in the study area. The terrain of both cities rises in the south and the I-10 freeway runs through Redlands. As such, the Adventists from Loma Linda are concentrated in the center and northern part of the city. In contrast, the Adventists from the city of Redlands seem to be more dispersed.

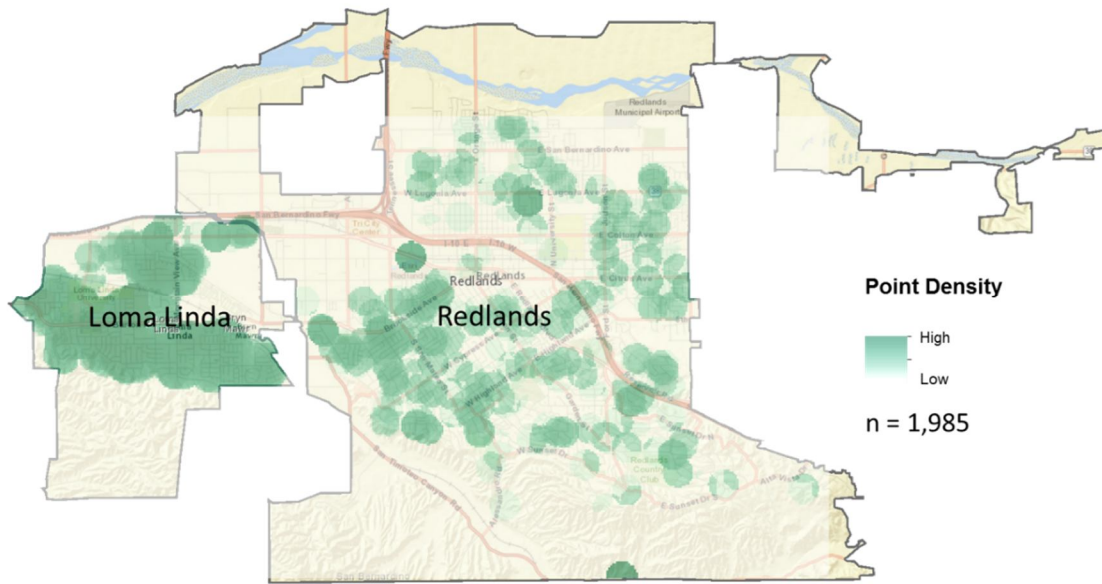


Figure 6-1: Adventist Sample Distribution

Figure 6-2 shows the land use data gathered from Loma Linda Planning Office. There are 24 categories of land uses in Loma Linda, including various commercial, residential, and open spaces.

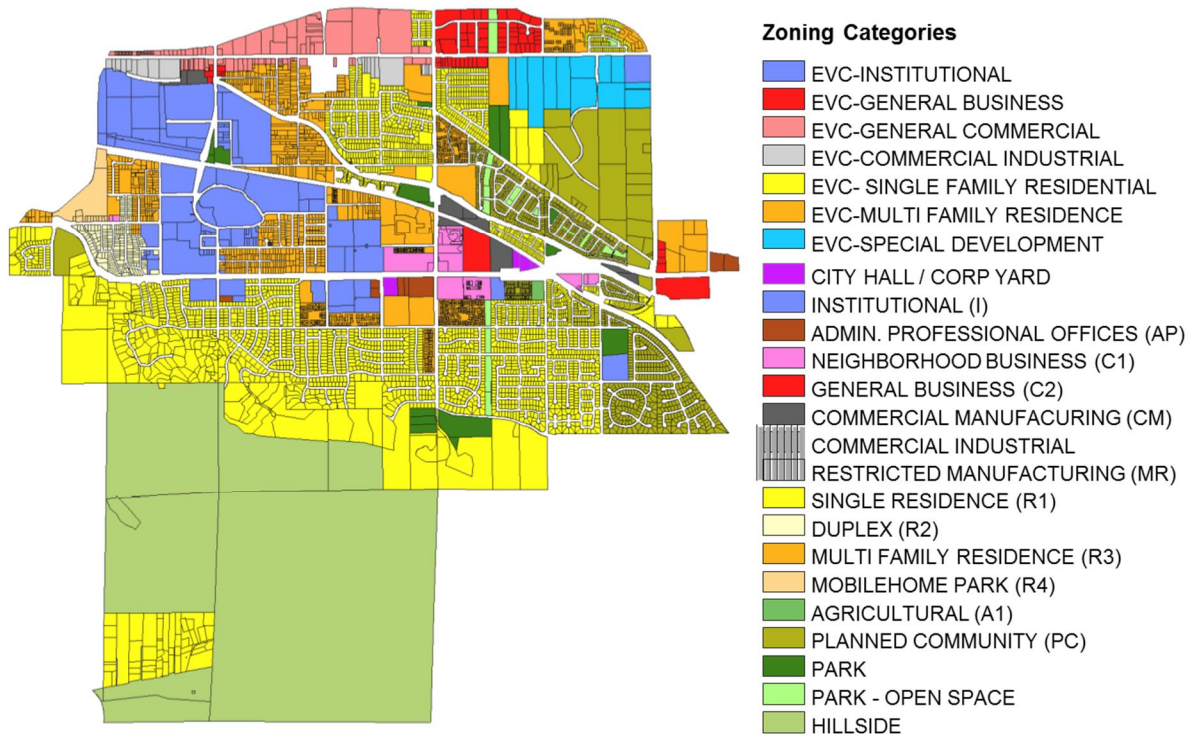


Figure 6-2: Land Use for the City of Loma Linda

Figure 6-3 shows the zoning map obtained from Redlands Planning Office, which includes 13 types of land uses.

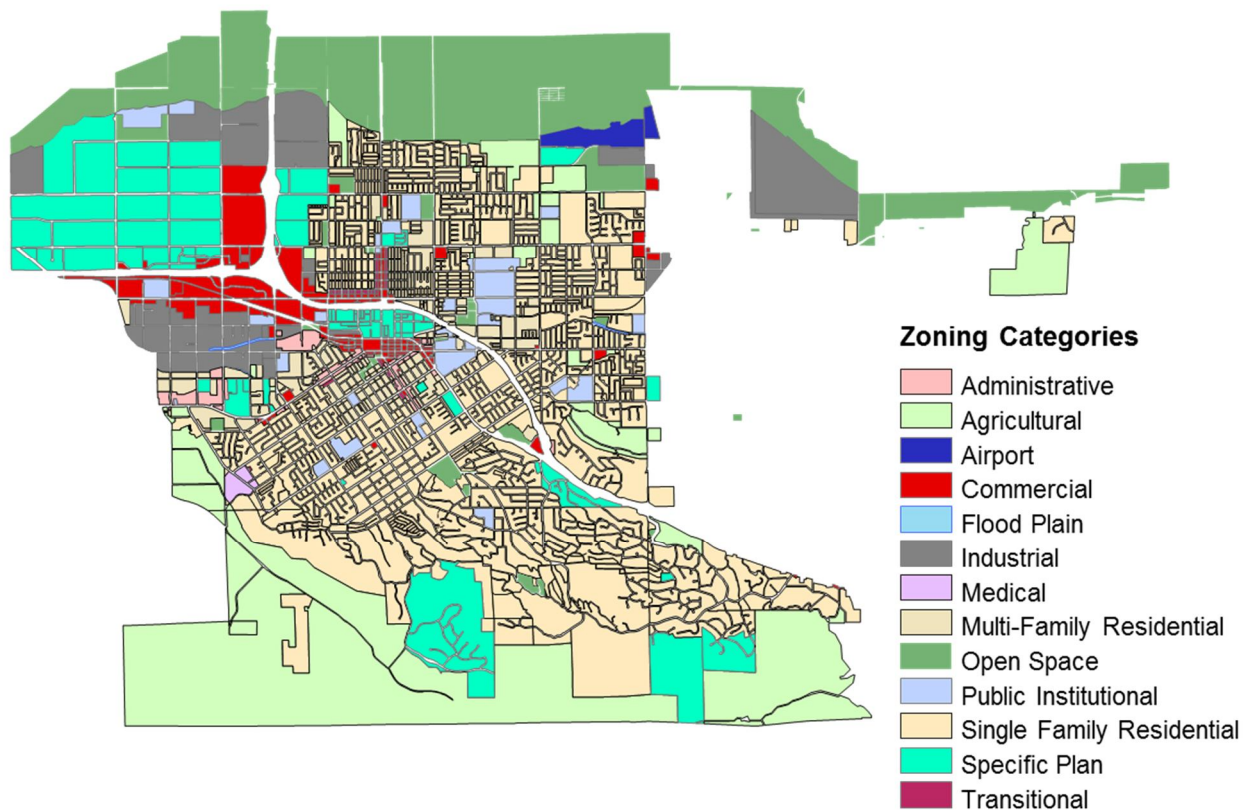


Figure 6-3: Land Use for the City of Redlands

The land uses observed in Figures 6-2 and 6-3 were combined into three categories representing residential areas, park and open spaces, and other areas to calculate diversity of land use. According to client’s request, the land use of open spaces includes parks, golf courses, and amusement parks.

Figure 6-4 shows the distribution of the three different land use categories throughout the study area. The parks and open spaces in Redlands are more dispersed as compared to Loma Linda.

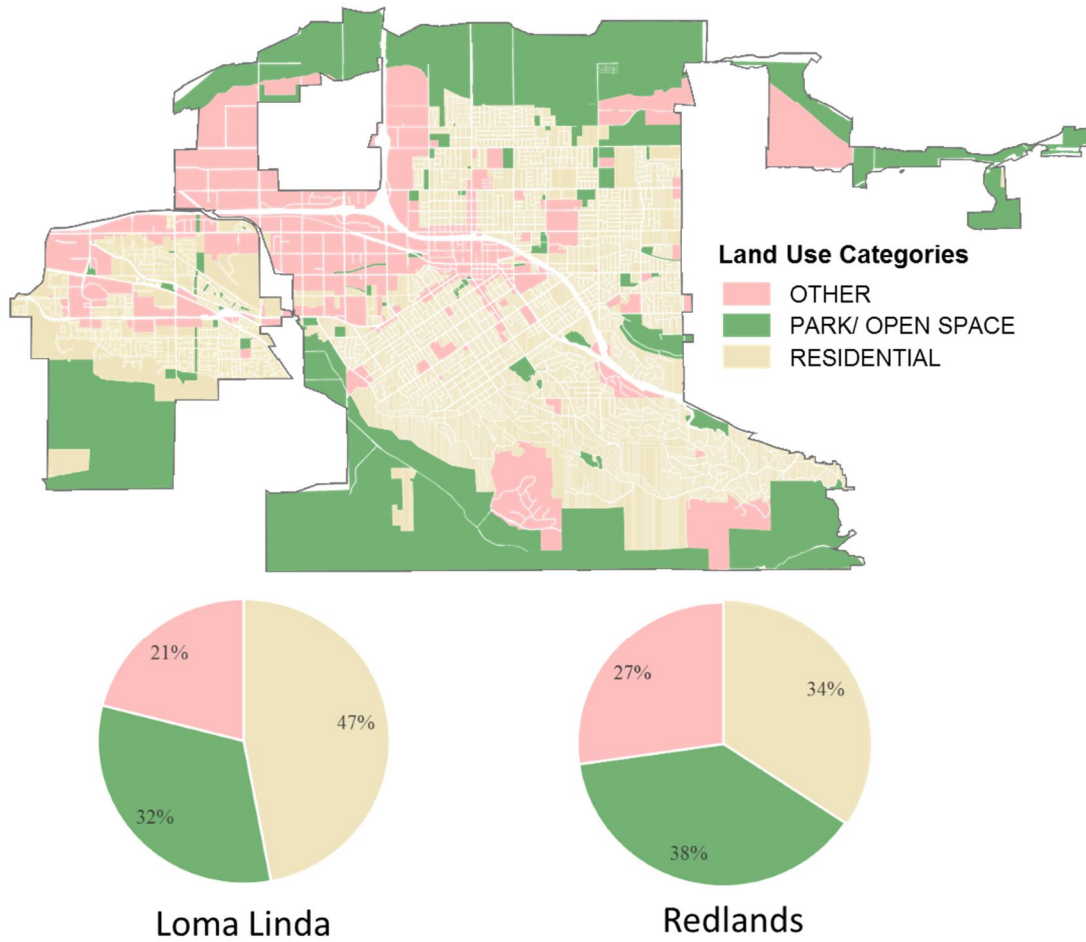


Figure 6-4: Cities of Loma Linda and Redlands Dissolved by Land Use Categories

To evaluate accessibility to recreational opportunities, the locations of recreational sites were mapped as shown in Figure 6-5.

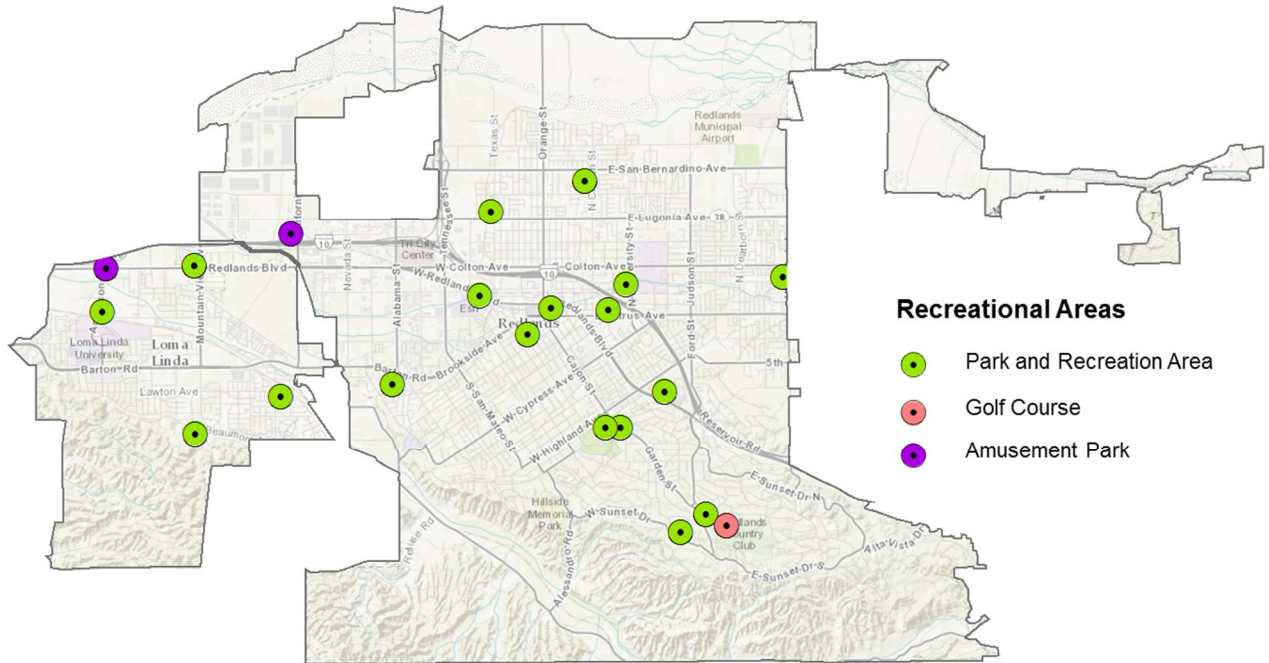


Figure 6-5: Distribution of Recreational Areas in the City of Loma Linda and Redlands

To measure the percent of vegetation pixels that covered a walking environment, a multispectral image was gathered using an Image Service in ArcGIS Online. Figure 6-6 shows an image in false color of the cities.

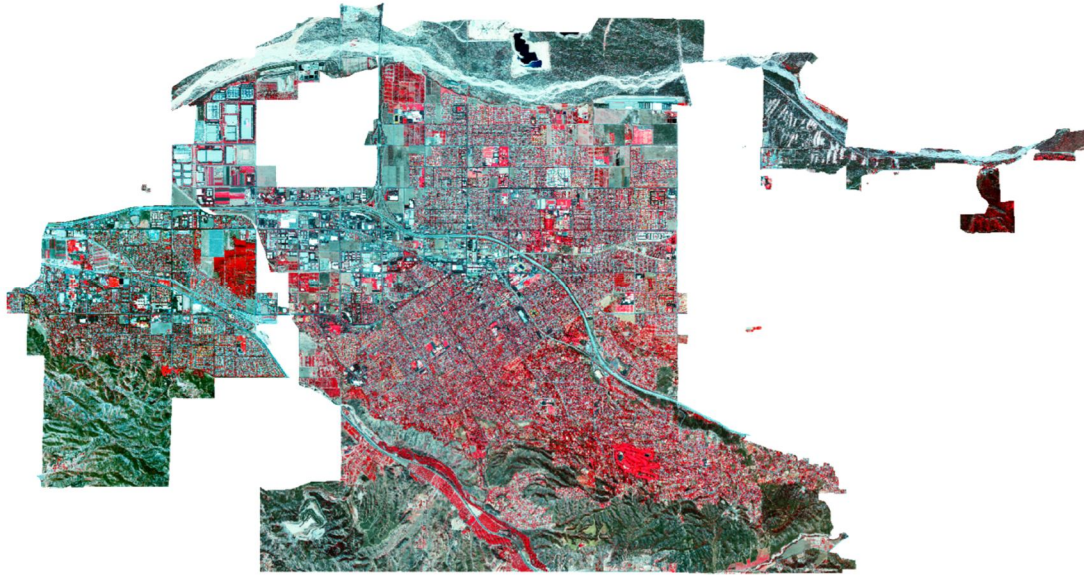


Figure 6-6: Multispectral Image of the City of Loma Linda and Redlands

The multispectral image contains near-infrared, red, and green bands. This image allows detection of vegetation in different shades of red. The normalized difference vegetation index (NDVI) accentuates the vegetation areas, given the high reflectance of the near-infrared band. The mathematical model was explained in Section 4-4. Figure 6-7 shows the NDVI of the cities.

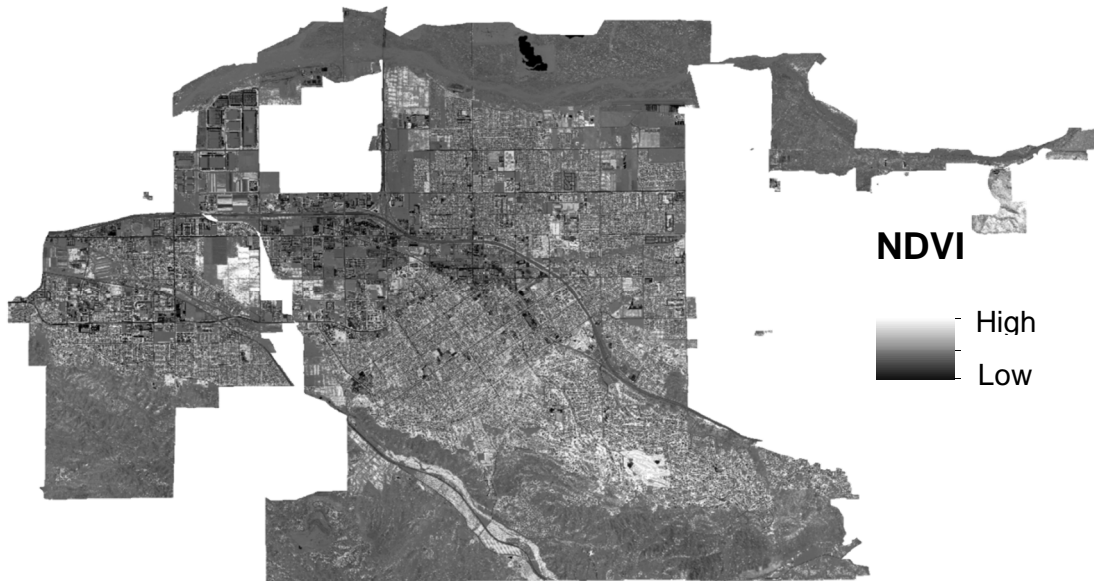


Figure 6-7: NDVI in the Cities of Loma Linda and Redlands

The brighter areas are more likely to be covered by vegetation. Figure 6-8 shows a classified NDVI using a threshold of 0.1 with values under 0.1 indicating areas that correspond to rocks, water, and others non-vegetated areas, and others representing shrubs, grassland, and other vegetated-areas (National Aeronautics and Space Administration, 2000).

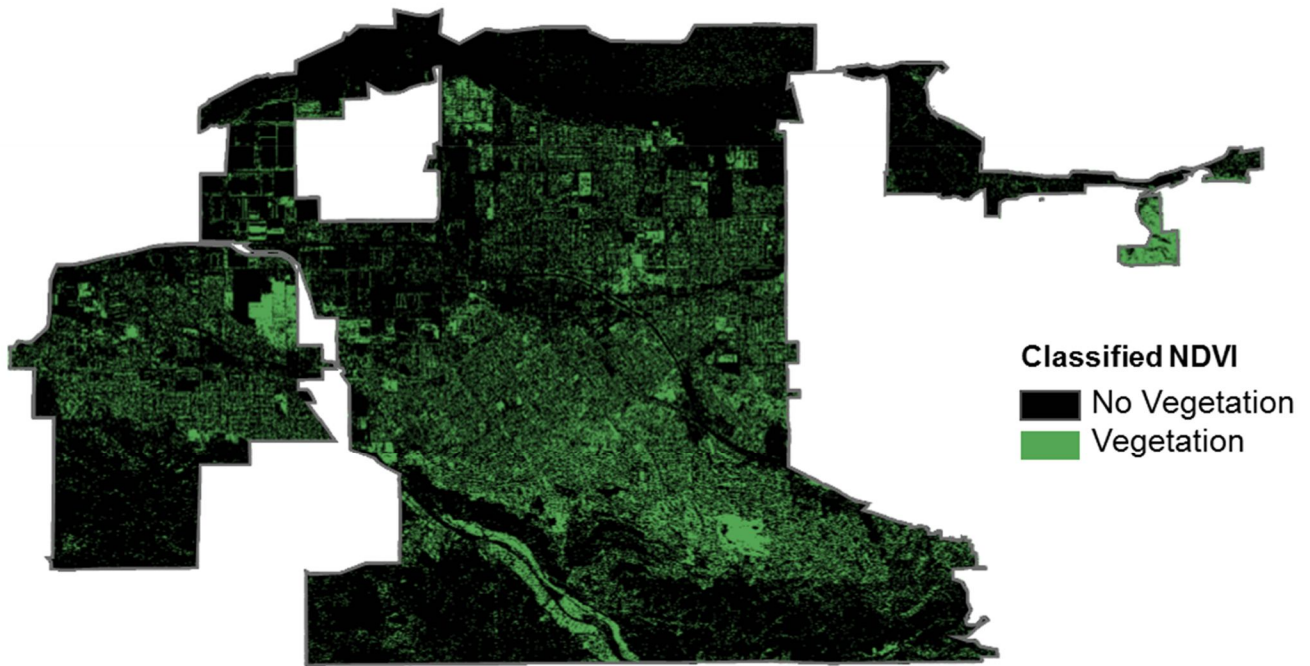


Figure 6-8: Classified NDVI Using a Threshold of 0.1

This image was used as the input of the Sidewalk Quality tool to measure the percent of vegetation pixels in a walking zone. Table 6-1 shows the percent of the city areas that are covered by vegetation for Loma Linda and Redlands. It appears that vegetation coverage is comparable between these two cities (81% vs. 79%).

Table 6-1: Percent of vegetation pixels per city

Pixels Classification	Loma Linda N=19,446,669	Redlands N=92,818,057
Vegetation	19 %	21 %
Non Vegetation	81 %	79 %

The walking zone is a supporting feature class for measuring population density, land use diversity, and sidewalk quality. Figure 6-9 shows the one-mile walking zones for all AHS-2 participants. Each walking zone has different geometries and areas, and some overlap each other.

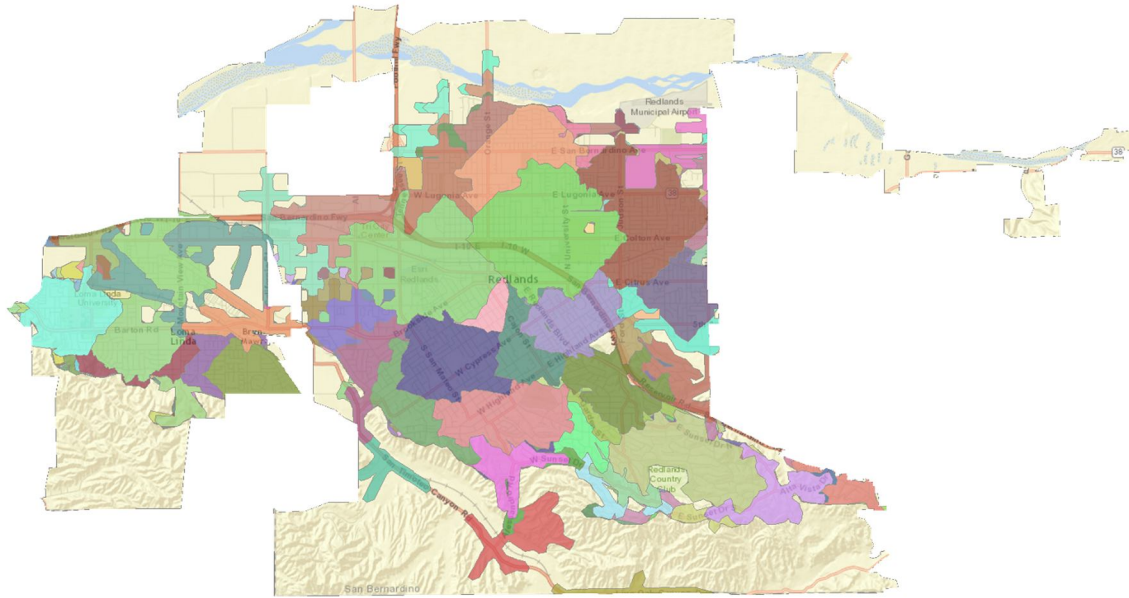


Figure 6-9: 1-Mile Walking Zones for Each AHS-2 Participant

To calculate population density, the U.S. Census 2010 data were used at the block level. Figure 6-10 show the distribution of the population by blocks. The majority of the blocks have a population of less than three hundred people.

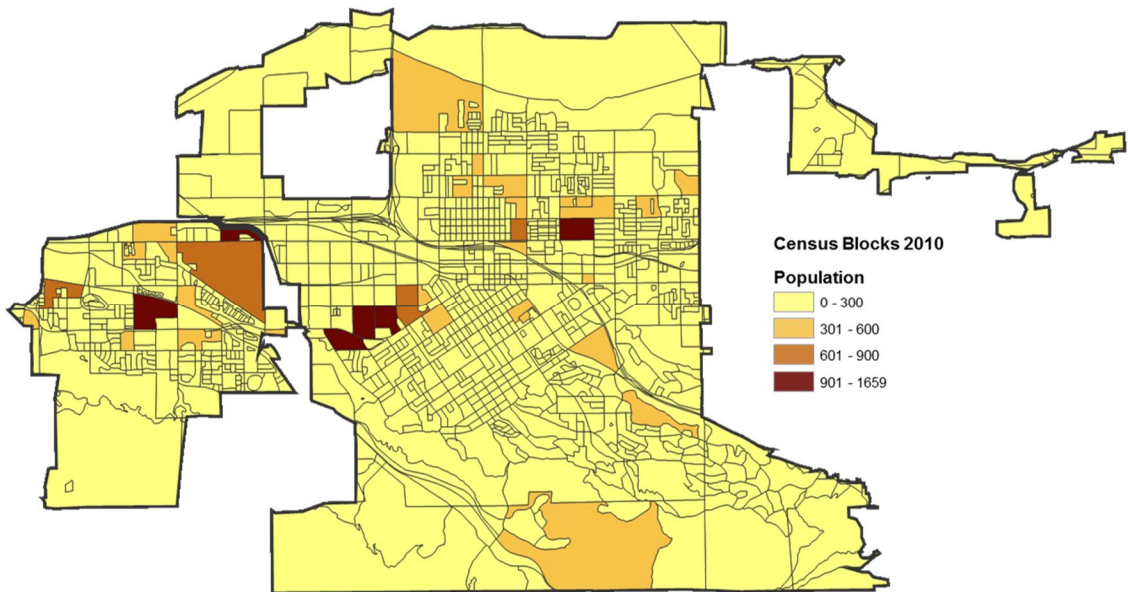


Figure 6-10: Distribution of Population from the U.S. Census at Block Level

6.2 Comparing Built Environments

The project was required to compare the built environment indicators between the two cities. The maps shown in this section were geo-masked to protect the identity of the AHS-2 participants.

The geo-masking is a process of hiding the true locations of the participants to protect their privacy. The geo-masking was accomplished by randomly shifting the original coordinates within a predefined range. Figure 6-11 shows the workflow to shift the locations of the participants using Excel.

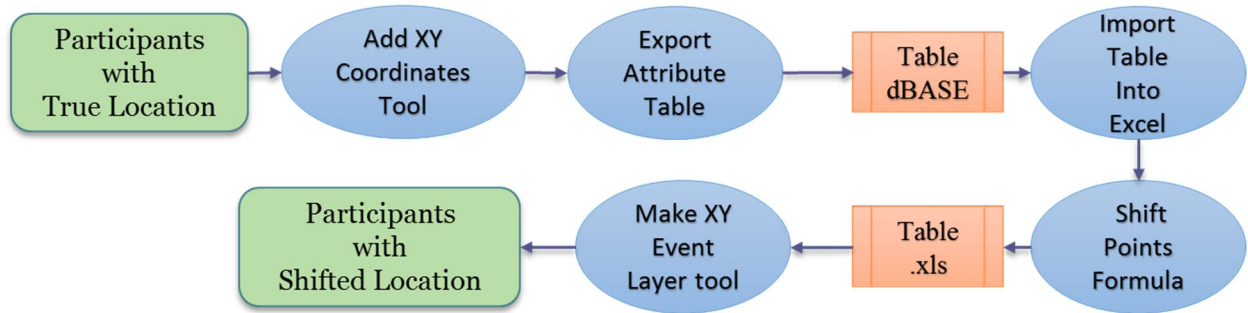


Figure 6-11: Workflow to Shift Participants' Locations

The coordinates of home locations were added to the participants' attribute table using the Add XY Coordinates tool and exported into a dBase table format to be used in Excel. The shifting procedure was made using the formula:

$$S_x = C_x + R_x$$

$$S_y = C_y + R_y$$

where S_x and S_y are the new shifted coordinates, C_x and C_y are the original coordinates, and R_x and R_y are the random values added to the true location. The random values R_x and R_y were calculated using the formula:

$$R_x = \text{IF}(\text{RAND}() \geq 0.5, 1, -1) \times (\text{RANDBETWEEN}(n, m))$$

$$R_y = \text{IF}(\text{RAND}() \geq 0.5, 1, -1) \times (\text{RANDBETWEEN}(n, m))$$

where n and m are integers that are the predefined range. The random values R_x and R_y allow a pair of coordinate to vary from $(C_x - m, C_y - n)$ to $(C_x + n, C_y + m)$. The table with the shifted locations was imported into ArcMap to create a feature layer with the Make XY Event Layer tool to present the results.

Several statistical analyses were run to compare the results between the two cities. SPSS was used to run descriptive statistics and independent samples t -test. For categorical data, the Z test for two population proportions was used.

6.2.1 Access to Recreational Areas

To evaluate accessibility to recreational areas, distance to the nearest recreational area and the proportion of participants who had access to recreational areas within one mile were calculated. Table 6-2 compares access to recreational areas between Loma Linda and Redlands.

Table 6-2: Access to recreational areas for the cities of Loma Linda and Redlands

Test on Difference	Loma Linda	Redlands
Distance to the nearest recreational area in miles ^{a*} Mean (Standard Deviation)	0.92 (0.32)	1.20 (1.20)
Access to recreational areas at one mile walking zone ^{b*}	48.4%	63.1%

^a independent sample t-test

^b z-test for two proportions

* p-value < 0.05

The *t*-test shows that on average Loma Linda participants travel significantly shorter to their nearest recreational sites. However, the percentage of Redlands participants who have access to recreational areas within one-mile walking zone is found to be significantly higher by the test on proportions. Figure 6-12 presents the distribution of the AHS-2 participants, the recreational areas, and distance surface to these destinations.

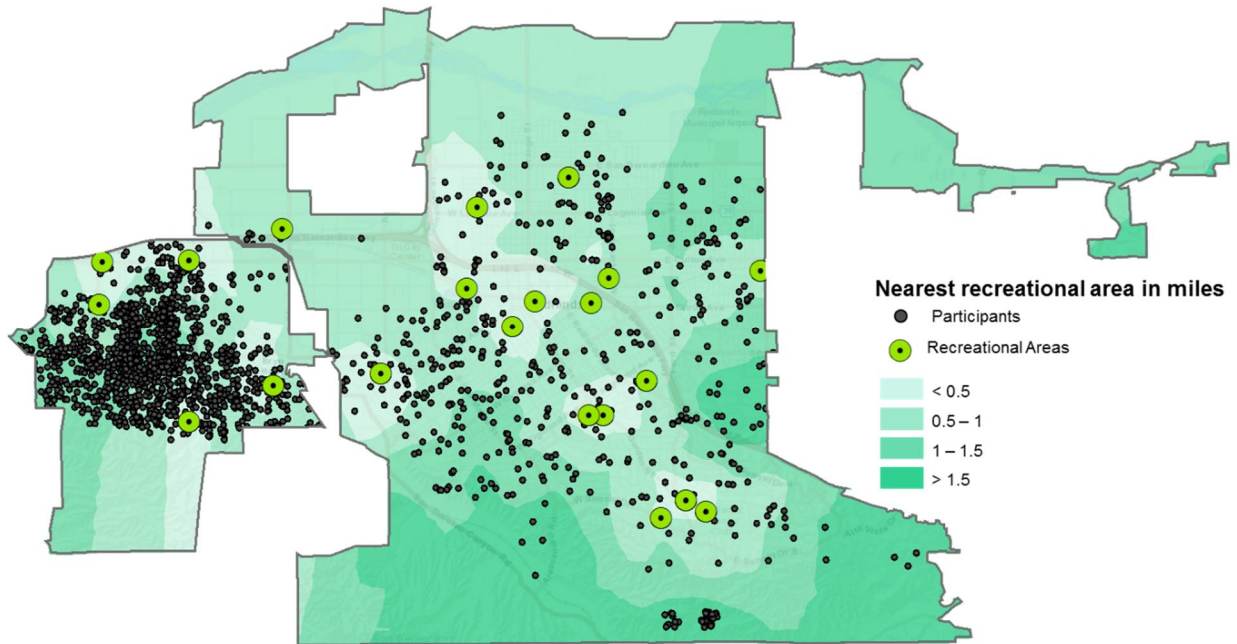


Figure 6-12: AHS-2 Participants with the Nearest Recreational Area

The map was extrapolated using the Inverse Distance Weighted (IDW) tool in the Geostatistical Analysis toolsets. The brightest areas represent shorter distance or better accessibility to the recreational destinations, while the participants located in the darker green will not have access to any recreational area within one mile of walking distance.

Figure 6-13 depicts the distribution of participants in relation to the recreational areas. The purple points in the map represent the participants with access to at least one recreational area and the orange colored points represent the participants without access to a recreational area within their one-mile walking zone. The western part of the city of Loma Linda and the southern area of the city of Redlands contain a large percentage of participants that have no practical access to recreational destinations within a walking distance.

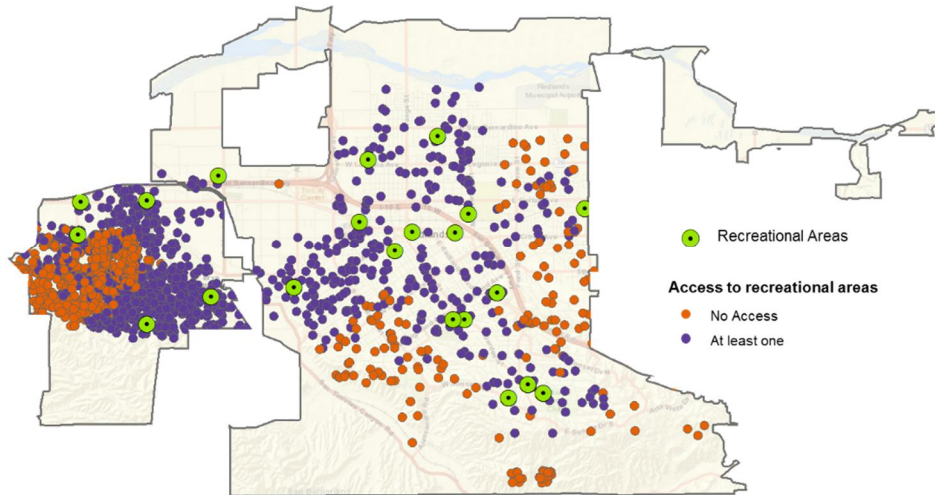


Figure 6-13: Access to Recreational Areas within One-Mile Walking Zone

The participants with no access to recreational areas within their one-mile walking zone were also further analyzed. Figure 6-14 shows a map of the distances to the nearest recreational area for people with no access in one-mile walking zone.

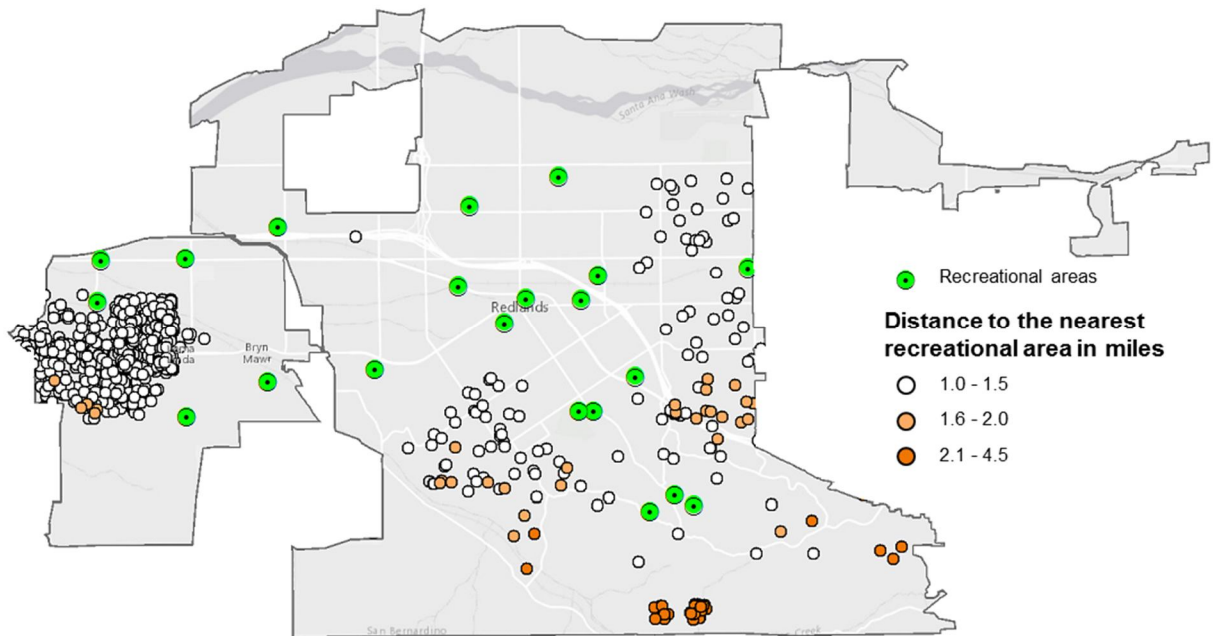


Figure 6-14: Distance to the Nearest Recreational Area in AHS-2 Participant with No Access within One-mile Walking Zone

The map illustrates that most of the AHS-2 participants in the city of Loma Linda with no access to recreational areas within their one-mile walking zone are still in the range of 1.5 miles to their nearest recreational sites, while some of the Redlands participants need to travel up to 4.5 miles to reach the nearest recreational sites.

A *t*-test shows that the Loma Linda participants who have no access to the recreational areas in one-mile walking distance still have significantly shorter distance to their nearest recreational destinations than the Redlands participants (1.19 miles vs. 2.26 miles with $p < 0.05$). This explains why the *t*-test and *z*-test yield the opposite results in Table 6-2. It also implies that the AHS-2 participants in the southern part of the city of Redlands are underserved in terms of accessibility to recreational areas.

6.2.1.1 Access Tool Limitations

The Access to Recreational Areas tool is able to calculate four variables to measure distances to recreational areas for each participants. Since the built environment was defined as one-mile walking zone, this threshold was used to compute, for example, the number of destinations within one mile of walking. This approach was found to be imperfect. For example, Figure 6-15 illustrates a case where the nearest recreational area from a participant's location is only 1.013 miles away and this recreational site is not counted as accessible to the participant. This is not reasonable given all the recreational areas, such as parks, are not a single point, but an area. Further, the entrance location of a recreational site and barriers like fences, bushes, and curbs also affect the real distance that a participant need to walk to the site.



Figure 6-15: Nearest Recreational Area from a Participant’s Location

Therefore, using a single point to represent an area may be sufficient for visualization but not accurate for the purpose of analysis. The data used in this project also do not include access points to the recreational opportunities as well as other possible shortcuts like unpaved roads and lawns along curbs. Given the limitations of the data used in the project, the result may present a bias affecting the empirical findings.

6.2.2 Population Density

Population density was calculated by population per square mile within the one-mile walking zone. Table 6-3 compares the population density between the cities of Loma Linda and Redlands.

Table 6-3: Population Density of the cities of Loma Linda and Redlands

Test on difference	Loma Linda mean	Redlands mean
Population Density per Sq. mile ^{a *}	4,951	3,986

* p-value < 0.05

^a Independent samples t-test

On average, local population density of Loma Linda participants are significantly higher than that of Redlands participants, indicating a generally walking-friendly environment in Loma Linda. However, population densities vary greatly in the study area as shown in Figure 6-16. The darkest areas represent the most populated zones. Compare to the Loma Linda, the city of Redlands has more diverse urban clusters ranging from 75 to 7,700 people per square mile.

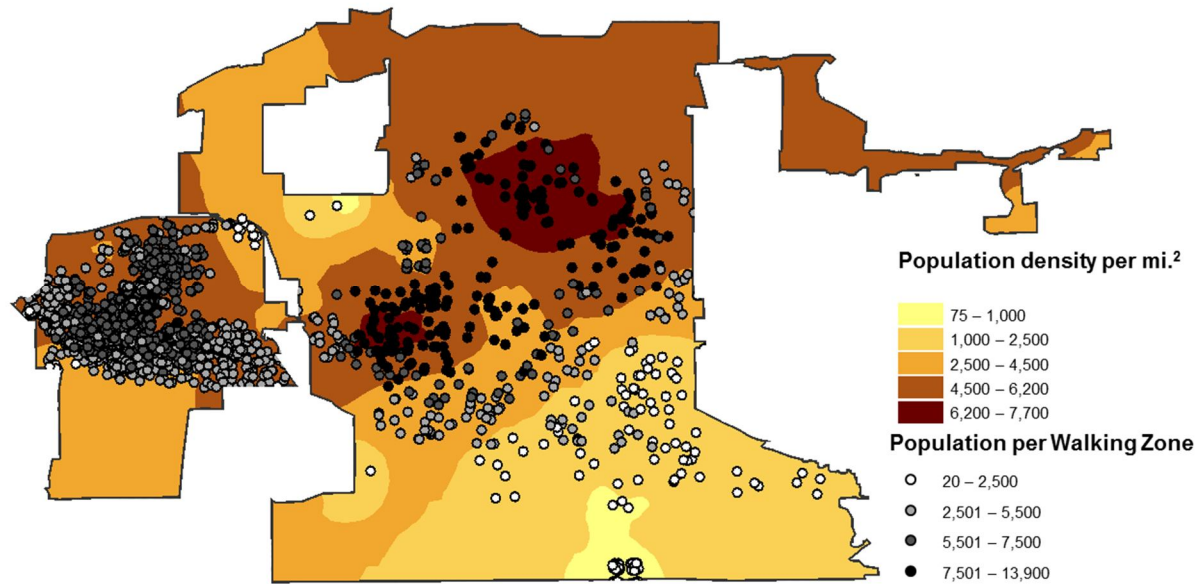


Figure 6-16: Population Density per Squared Mile and Population per Walking Zone between The City of Loma Linda and Redlands

The map also shows that the AHS-2 participants from Loma Linda are clustered in more dense neighborhoods, while Redlands participants are scattered throughout highly variable urban environments.

6.2.3 Diversity of the Land Use

The diversity of the land use was calculated using an entropy formula from Christian et al., (2011), originally adopted by Frank et al., (2005). The Land Use Mix tool calculated the entropy index for each participant, ranging from 0 to 1.

On average, the participants from Loma Linda have a higher land use diversity with a mean of 0.632 in the entropy index as compared with the Redlands participants with a mean of 0.472 ($p < 0.05$). Figure 6-17 presents the map that was extrapolated from the entropy indices using the inverse distance weighted interpolation (IDW) method. The dark areas have highly diverse land uses in terms of the selected land use categories. Most of the participants from the city of Loma Linda come from more diverse land use types. The south of the city of Redlands has the lowest entropy values indicating a more uniform land use in terms of residential, parks or open spaces, and others land uses.

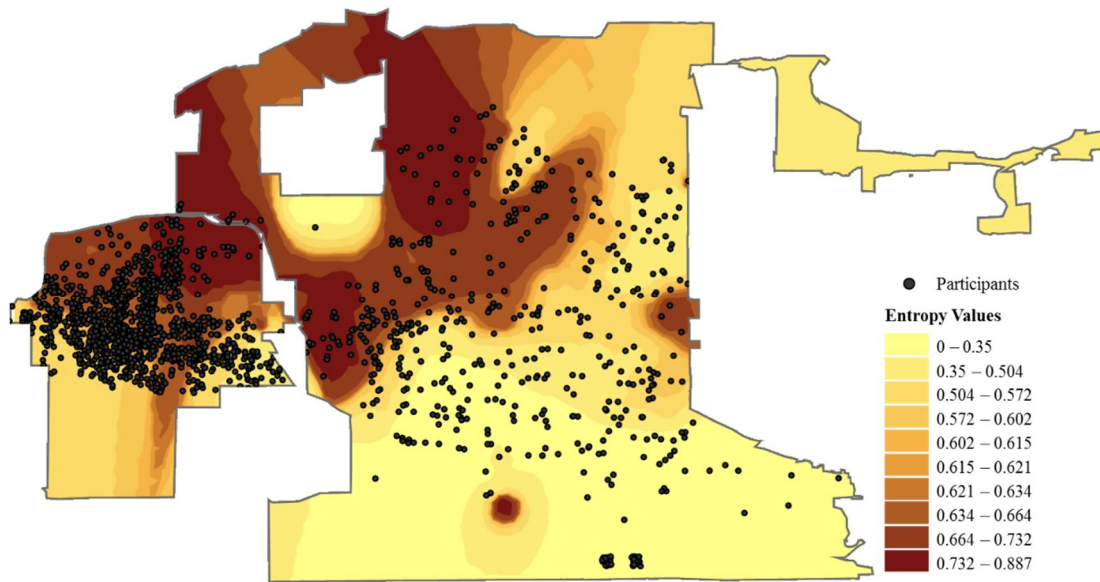


Figure 6-17: Diversity of the Land Use per Walking Zone

6.2.4 Quality of the Sidewalk

The attractiveness of the sidewalks was measured by the percentage of sidewalks that are covered by vegetation. Vegetation was derived from NDVI with a threshold of 0.1. Figure 6-18 shows the distribution of vegetations (pixels with NDVI value greater than 0.1) in both cities.

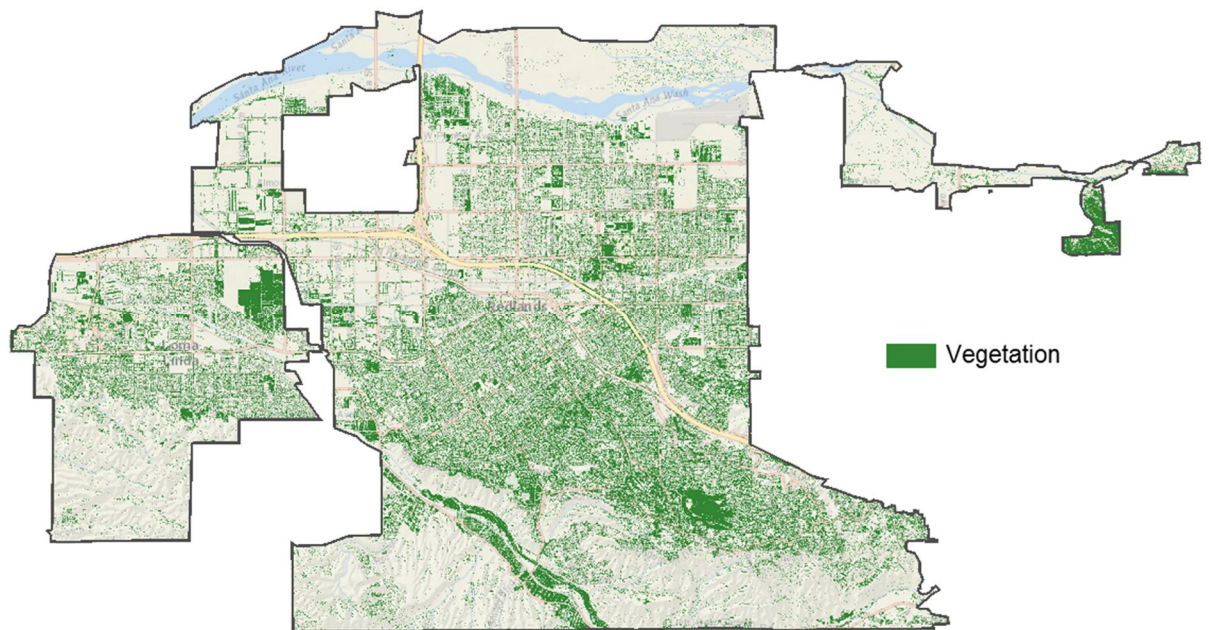


Figure 6-18: Vegetation Pixels of the City of Loma Linda and Redlands using 0.1 as the Threshold

There are some clusters of vegetations in the two cities. The clustered vegetations in the northeastern part of the city of Loma Linda and southeastern portion of the city of Redlands are a farming area and a golf course respectively. Even though those areas are highly vegetated, only the vegetations in the walking environment of the participants were considered.

On average, approximately 18 percent of the walking environment is covered by vegetations for all participants from both cities; however, according to the standard deviation of the percentage of vegetated pixels, the quality of sidewalks of Loma Linda participants appears more uniform than that of Redlands participants. Figure 6-19 shows the distribution of percent of local sidewalks covered by vegetations. The map was extrapolated from the sidewalk quality measurements of all participants using the IDW interpolation method.

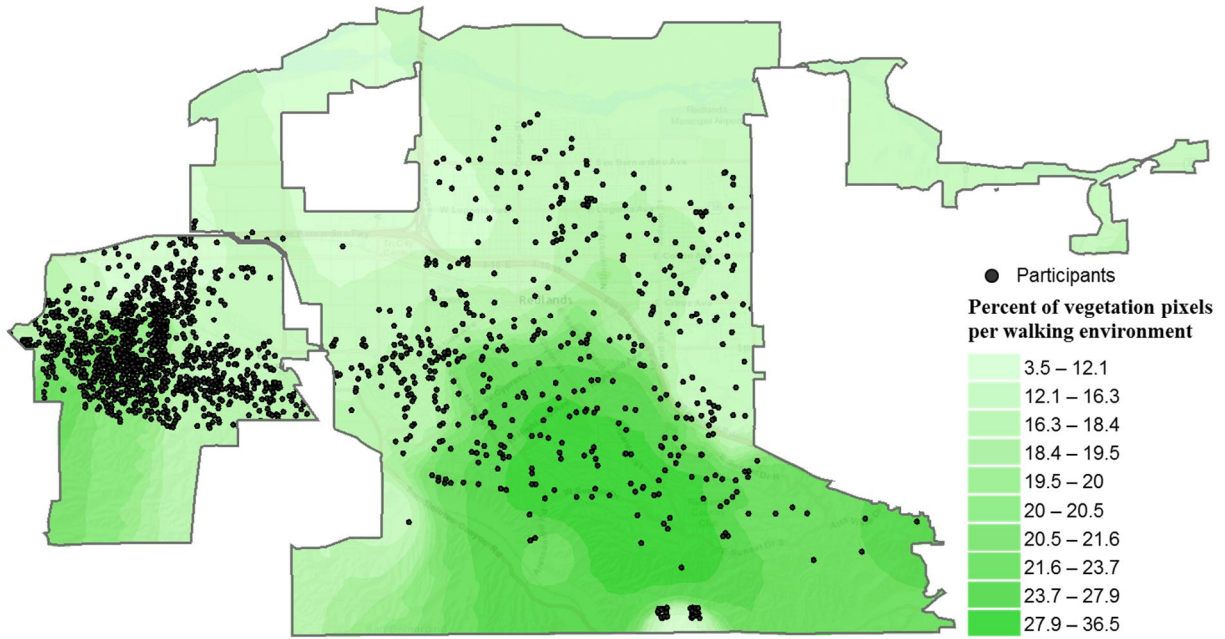


Figure 6-19: Percent of Vegetation pixels covered

The darkest green represents the areas with the highest percent of vegetated sidewalks in the local walking environments. The western part of the city of Loma Linda has a better walking environment as compared with the eastern part of the city. The center and southern part of the city of Redlands have a higher vegetation coverage as compared with the northern part. The highest vegetated sidewalks in Figure 6-19 are located in the center of the city of Redlands. More than 27.9 percent of the sidewalks are covered by vegetation.

6.3 Summary

Five built environment indicators were used to compare the cities of Loma Linda and Redlands. In general, the city of Loma Linda presents better access to recreational areas, higher population density, and more diversified land use in participants' built environment. Although participants from these two cities share similar kinds of sidewalks in their

neighborhoods, the sidewalk quality in city of Redlands is more variable than that of Loma Linda. Further, the city of Redlands shows more diversified built environments for its residents and some of the areas are underserved by recreational places.

Chapter 7 – Conclusions and Future Work

The project goal was accomplished by developing a toolset with a user-friendly interface that helps standardize the process of measuring built environment indicators for statistical analysis. The built environment indicators were defined following Brownson et al. (2009) and Maghelal et al. (2008) frameworks and the input datasets were obtained at no cost from readily available data sources. The custom toolbox allows the user to calculate four built environment indicators including access to recreational areas, population density, land use diversity, and sidewalk quality. There are six tools contained in the toolbox: one tool that creates a walking zone, four tools that create four factors of the built environment, and one additional tool for the reclassification process of the Normalized Difference Vegetation Index. With this toolbox, the client would be able to calculate built environments for other cities of their interest. In the meanwhile, the client may change parameters to meet their needs. For example, they can change definition of built environment from one-mile walking distance to a desired distance range to accommodate different urban structures.

To test this toolbox, an empirical study was conducted for the cities of Loma Linda and Redlands. Using these outputs, the client would be able to conduct further analysis to test their hypothesis of how the local environment would affect health.

During the empirical analysis, the results of access to recreational sites were found to be biased for the participants that are merely located beyond one-mile network distance. To address this issue, one possible solution could be increasing the one-mile distance threshold. Other possible solutions could be using areal data to represent recreational destinations and improving the network dataset by adding potential entrance to

recreational areas, a sidewalk dataset, and natural and built barriers like fences and bushes.

In addition, the built environment indicators are readily to be linked back to the AHS-2 individual health outcomes. Understanding the effects the built environment on physical activity and obesity will provide a better understanding of how the urban design effects public health. As described by Carlson and colleagues (2012), future researchers should not avoid subjective factors like surveys of participants. Having responses from people who use walking as a physical activity or as a method of transportation may increase the reliability of the indicators. As data related to sidewalks, curbs, fences, natural or built barriers, the elevation terrain, potential entrance to destinations, among others, are gathered in cities, further development of the tools is needed to add those measures into the network dataset for analysis. Brownson and colleagues (2009) recommend a new generation of indicators that takes into account measures that vary by age, physical abilities, and culture. All of these new developments would be fruitful endeavors for future research.

Works Cited

- Beeson, W. L., Mills, P. K., Phillips, R. L., Fraser, G. E. (1989). Chronic Disease Among Seventh-day Adventists, A Low-Risk Group. Rationale, methodology, and description of the population. *Cancer*, 64(3), 570–81.
- Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A., & Sallis, J. F. (2009). Measuring the built environment for physical activity: state of the science. *American Journal of Preventive Medicine*, 36(4 Suppl), S99–123.e12.
doi:10.1016/j.amepre.2009.01.005
- Buettner, D. (2013). Blue Zones. Retrieved from <http://www.bluezones.com/>
- Butler, T. L., Fraser, G. E., Beeson, W. L., Knutsen, S. F., Herring, R. P., Chan, J., ... Jaceldo-Siegl, K. (2008). Cohort profile: The Adventist Health Study-2 (AHS-2). *International Journal of Epidemiology*, 37(2), 260–5. doi:10.1093/ije/dym165
- Carlson, C., Aytur, S., Gardner, K., & Rogers, S. (2012). Complexity in built environment, health, and destination walking: a neighborhood-scale analysis. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 89(2), 270–84. doi:10.1007/s11524-011-9652-8
- Carroll, M. D., & Surveys, E. (2010). Prevalence of Overweight, Obesity, and Extreme Obesity Among Adults: United States, Trends 1960–1962 Through 2007–2008, 1994(June), 1–6.
- Centers for Disease Control and Prevention. (2008). *Preventing Chronic Diseases: Investing Wisely in Health*. Retrieved from www.cdc.gov/nccdphp
- Centers for Disease Control and Prevention. (2011). *Transportation Recommendations*. Retrieved from <http://www.cdc.gov/transportation/recommendation.htm>

- Centers for Disease Control and Prevention. (2014). *Vital Signs*. Retrieved from <http://www.cdc.gov/vitalsigns/>
- Christian, H. E., Bull, F. C., Middleton, N. J., Knuiman, M. W., Divitini, M. L., Hooper, P., ... Giles-Corti, B. (2011). How important is the land use mix measure in understanding walking behaviour? Results from the RESIDE study. *The International Journal of Behavioral Nutrition and Physical Activity*, 8(1), 55. doi:10.1186/1479-5868-8-55
- Glass, T. A., & McAtee, M. J. (2006). Behavioral science at the crossroads in public health: extending horizons, envisioning the future. *Social Science & Medicine* (1982), 62(7), 1650–71. doi:10.1016/j.socscimed.2005.08.044
- Leyden, K. M. (2003). Social capital and the built environment: the importance of walkable neighborhoods. *American Journal of Public Health*, 93(9), 1546–51. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1448008&tool=pmcentrez&rendertype=abstract>
- Loma Linda University. (2013). *Loma Linda University, School of Public Health*. Retrieved from Adventist Health Studies: <http://www.llu.edu/public-health/health/index.page?>
- Maghelal, P. K., & Capp, C. J. (2011). Walkability: A Review of Existing Pedestrian Indices. *Journal of the Urban and Regional Information Systems Association*, 23(2),5-20. Retrieved from <http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=4&sid=10d90a44-ef70-4cdf-9459-3d71b5e55958%40sessionmgr198&hid=124>

- Minnesota Population Center. *National Historical Geographic Information System: Version 2.0*. Minneapolis, MN: University of Minnesota 2011. Retrieved from: <http://www.nhgis.org/>
- Müller-Riemenschneider, F., Pereira, G., Villanueva, K., Christian, H., Knuiman, M., Giles-Corti, B., & Bull, F. C. (2013). Neighborhood walkability and cardiometabolic risk factors in Australian adults: an observational study. *BMC Public Health, 13*(1), 755. doi:10.1186/1471-2458-13-755
- National Aeronautics and Space Administration. (2000, August 30). *Earth Observatory*. (J. Weier, & D. Herring, Editors) Retrieved from Measuring Vegetation: <http://earthobservatory.nasa.gov/Features/MeasuringVegetation>
- Tilt, J. H., Unfried, T. M., & Roca, B. (2007). Using objective and subjective measures of neighborhood greenness and accessible destinations for understanding walking trips and BMI in Seattle, Washington. *American Journal of Health Promotion: AJHP, 21*(4 Suppl), 371–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17465183>
- Trost, S. G., Owen, N., Bauman, A. E., Sallis, J. F., & Brown, W. (2002). Correlates of adults' participation in physical activity: review and update. *Medicine and Science in Sports and Exercise, 34*(12), 1996–2001. doi:10.1249/01.MSS.0000038974.76900.92
- U.S. Department of Health and Human Services. (2007, January). *National Institute on Alcohol Abuse and Alcoholism*. Retrieved from Alcohol Alert: <http://pubs.niaaa.nih.gov/publications/AA71/AA71.htm>

U.S. Department of Health and Human Services. (2008). *Office of Disease Prevention and Health Promotion*. Retrieved from Physical Activity Guidelines for Americans: <http://www.health.gov/paguidelines/>

U.S. Department of Health and Human Services. (2014, May). *Office of Disease Prevention and Health Promotion*. Retrieved from Healthy People 2020: <https://www.healthypeople.gov/2020/Leading-Health-Indicators/>

Appendix A. Walking Zone Generator Tool

The purpose of this Toolset was to create a Polygon feature and line feature using Network Analyst Service Area tools. The polygons and line features are the parameters for others tools (Density, Diversity, and Quality Tool). Every Polygon and Line Feature has the Identifier from the Original Point Data of the Participants

The tool:

- 1-Creates a service area using the Service Area Parameters.
- 2-Determine the format of the Identifier ID of the point data
- 3-Add a field to the Facilities Service Area sublayer with the same format of the point data ID field
- 4-Add the point data into the Facilities layer
- 5-Run the Service Area tool
- 6-Save the Polygon sublayer and the Line sublayer
- 7-Join the ID field of Facilities to the Polygons and Lines
- 8-Dissolved the Line feature by ID

The Parameters are:

Parameters	Format
Input Street Network Dataset	Network Dataset Layer
Input Boundary Threshold	Double
Input point data of Participants	Point Feature Layer
Input Participants ID	Field
Output Polygons	Feature Class
Output Dissolved Lines	Feature Class

```
# Start
```

```
# Importing ArcPy Functions library
```

```
import arcpy
```

```
#Over Write
```

```
arcpy.env.overwriteOutput=True
```

```
# Check out any necessary licenses
```

```
arcpy.CheckOutExtension("Network")
```

```
#Input Parameters
```

```
Street_Network_Dataset = arcpy.GetParameterAsText(0)# Input Street Network Dataset
```

```
Zone_Boundary=arcpy.GetParameterAsText(1) # Input Boundary Threshold
```

```

in_table = arcpy.GetParameterAsText(2) # Point data of participants
Input_ID= arcpy.GetParameterAsText(3) # Participant Identifier Field
out_polygons=arcpy.GetParameterAsText(4) # Output Polygons Feature Class
out_Lines_diss=arcpy.GetParameterAsText(5) # Output dissolved Line Feature Class

# Tool Message
arcpy.AddMessage("Making Service Area...")

# Process: Make Service Area Layer
arcpy.MakeServiceAreaLayer_na(Street_Network_Dataset, "Service_Area", "Length",
"TRAVEL_FROM", Zone_Boundary, "SIMPLE_POLYS", "NO_MERGE", "DISKS",
"TRUE_LINES_WITH_MEASURES", "OVERLAP", "NO_SPLIT", "",
"", "ALLOW_UTURNS", "", "TRIM_POLYS", "100 Meters",
"LINE_SOURCE_FIELDS", "NO_HIERARCHY", "")

#Get the service area layer called "Walking Zone"
saLayer = arcpy.mapping.Layer("Service_Area")

#Get the list of fields from the Participants feature layer in the table of contents
fields = arcpy.ListFields(in_table)

# Get the Facilities sublayer called "Facilities"
facilitiesSubLayerName = arcpy.na.GetNAClassNames(saLayer)["Facilities"]

# Get the format of the "Input_ID" Participants ID Field
CopyFields=arcpy.ListFields(in_table,Input_ID)
Type_fields=CopyFields[0].type
if Type_fields=="String":
    arcpy.na.AddFieldToAnalysisLayer(saLayer, "Facilities",str(Input_ID),"TEXT")
else:
    arcpy.na.AddFieldToAnalysisLayer(saLayer, "Facilities",str(Input_ID),"LONG")

#Get the network analysis class names from the service area layer
naClasses = arcpy.na.GetNAClassNames(saLayer)
#Get the polygons sublayer from the service area layer
faSublayer=arcpy.mapping.ListLayers(saLayer,naClasses["Facilities"])[0]
polygonsSublayer = arcpy.mapping.ListLayers(saLayer,naClasses["SAPolygons"])[0]
LinesSublayer = arcpy.mapping.ListLayers(saLayer,naClasses["SALines"])[0]

# Get the List of Fields that are used in the Facilities Sublayer
# used in the Add Location functions
fieldMappings = arcpy.na.NAClassFieldMappings(saLayer,
facilitiesSubLayerName,False, fields)

```

```

# Tool Message
arcpy.AddMessage("Adding Locations...")
# Load the point data of Participant points into the service area facilities sublayer using
field mappings
arcpy.na.AddLocations(saLayer, facilitiesSubLayerName, in_table, fieldMappings, "")

# Solve the Service Area
arcpy.na.Solve(saLayer)

# Tool Message
arcpy.AddMessage("Joining...")

# Add the ID of The Participants into the Polygons and Lines Service Area sublayers
# Key Fields where the join will be based on: polygon and Lines sublayers are =
"FacilityID"
# Facilities sublayer = "ObjectID"
arcpy.management.JoinField(polygonsSublayer, "FacilityID", faSublayer, "ObjectID",
[str(Input_ID)])
arcpy.management.JoinField(LinesSublayer, "FacilityID", faSublayer, "ObjectID",
[str(Input_ID)])

# Export the Polygons sublayer to a Polygon feature class on disk.
arcpy.management.CopyFeatures(polygonsSublayer, out_polygons)

# Creates in_memory to save the Lines Sublayer into features lines
# Variables converted in local memory (in_memory)
out_Lines = "in_memory\\out_Lines"
arcpy.management.CopyFeatures(LinesSublayer, out_Lines)

# Tool Message
arcpy.AddMessage("Dissolving street lines...")
# Dissolving the in_memory out_Lines
arcpy.Dissolve_management(out_Lines, out_Lines_diss, Input_ID)

# Tool Message
arcpy.AddMessage("Walking Zone Generator Tool Completed")

#End

```

Appendix B. Access Tool

The purpose of this Tool was to measure the distance of point of participants to any point destination. The tool will create a Summary Table, given point of participants, a numeric search Distance, and destinations points.

The tool:

- 1-Creates a OD Matrix.
- 2-Determine the format of the Identifier ID of the point data
- 3-Add a field to the Origins sublayer with the same format of the point data ID field
- 4-Add the participants point data into the Origins layer
- 5-Add the Destinations point data into the Destinations layer
- 6-Run the OD-Matrix tool
- 7-Join the ID field of Participants to the Lines Table
- 8-Creates a Summary Table
- 9- Calculates the Distances given the Search Distance
- 10- Adds the Results to the Summary Table

The Parameters are:

Parameters	Format
Input Street Network Dataset	Network Dataset Layer
Input Boundary Threshold	Double
Input point data of Participants	Point Feature Layer
Input Participants ID	Field
Input Destinations	Feature Class
Input Search Distance	Double
Output Table	Table

```
# Start
```

```
# Importing ArcPy Functions library  
import arcpy
```

```
#Over Write  
arcpy.env.overwriteOutput=True
```

```
# Check out any necessary licenses  
arcpy.CheckOutExtension("Network")
```

```

#input parameters
Street_Network_ND = arcpy.GetParameterAsText(0) #Input Street Network Dataset
in_table=arcpy.GetParameterAsText(1) #Input point data of Participants
Input_ID=arcpy.GetParameterAsText(2) # Input Participants ID field
Input_business=arcpy.GetParameterAsText(3) # Input Destinations
DisThres = float(arcpy.GetParameterAsText(4)) # Input Search Distance
Stats_test = arcpy.GetParameterAsText(5) #Output Summary Table

# Tool Message
arcpy.AddMessage("Making OD Matrix...")

# Process: Make OD Cost Matrix Layer
arcpy.MakeODCostMatrixLayer_na(Street_Network_ND, "Service_Area", "Length", "",
"", "", "ALLOW_UTURNS", "", "NO_HIERARCHY", "", "NO_LINES", "")

desc = arcpy.Describe(Street_Network_ND)

# Getting the Units of the Network Dataset
attributes=desc.attributes
for attribute in attributes:
    if attribute.name == "Length":
        Typeformat=attribute.units
    else:
        continue

#Get the service area layer called "OD MATRIX" from the table of contents
saLayer = arcpy.mapping.Layer("Service_Area")

#Get the list of fields from the Participants feature layer in the table of contents
fields = arcpy.ListFields(in_table)

fieldInputID=arcpy.ListFields(in_table,str(Input_ID))
InIDType=fieldInputID[0].type

OriginsSubLayerName = arcpy.na.GetNAClassNames(saLayer)["Origins"]

if InIDType == "String":
    # Process: Add Field to Analysis Layer

```

```

    arcpy.AddFieldToAnalysisLayer_na("Service_Area", "Origins", str(Input_ID),
"TEXT", "", "", "", "", "NULLABLE")
else:
    # Process: Add Field to Analysis Layer
    arcpy.AddFieldToAnalysisLayer_na("Service_Area", "Origins", str(Input_ID),
"LONG", "", "", "", "", "NULLABLE")

#Get the network analysis class names from the service area layer
naClasses = arcpy.na.GetNAClassNames(saLayer)

OriSublayer=arcpy.mapping.ListLayers(saLayer,naClasses["Origins"])[0]
destinationSublayer = arcpy.mapping.ListLayers(saLayer,naClasses["Destinations"])[0]
linesSublayer = arcpy.mapping.ListLayers(saLayer,naClasses["ODLines"])[0]

fieldMappings = arcpy.na.NAClassFieldMappings(saLayer,
OriginsSubLayerName,False, fields)

arcpy.na.AddLocations(saLayer, OriginsSubLayerName, in_table, fieldMappings, "")

#fields_b=arcpy.ListFields(Input_business)
DestinationSubLayerName = arcpy.na.GetNAClassNames(saLayer)["Destinations"]
arcpy.na.AddLocations(saLayer, DestinationSubLayerName, Input_business, "", "")

#Solve Tool
arcpy.na.Solve(saLayer)

arcpy.management.JoinField(linesSublayer, "OriginID",OriSublayer,"ObjectID",
[str(Input_ID)])

# Process: Summary Statistics
arcpy.Statistics_analysis(linesSublayer, Stats_test, [{"Total_Length",
"SUM"},{"Total_Length", "MIN"},{"Total_Length", "MEAN"}], Input_ID)

# Adding Fields to the Summary Table
arcpy.AddMessage("\nAdding Field...")
arcpy.AddField_management(Stats_test,"Count_" + str(int(DisThres)) + "_" +
str(Typeformat),"LONG" )
arcpy.AddField_management(Stats_test,"AVRG_" + str(int(DisThres)) + "_" +
str(Typeformat),"DOUBLE" )

# Updating the Summary Table to add the Distance Threshold Information

```

```
TAVRG2=arcpy.da.UpdateCursor(Stats_test,[str(Input_ID),"Count_" +
str(int(DisThres)) + "_" +str(Typeformat),"AVRG_" + str(int(DisThres)) + "_" +
str(Typeformat)])
```

```
# If the Identifier field format is String
```

```
if InIDType == "String":
```

```
# ID selection
```

```
for row in TAVRG2:
```

```
    ID=row[0]
```

```
    expression= "" + str(Input_ID) + "' = ' + "" + str(ID) + ""
```

```
    Tcount=0
```

```
    Tsum=0
```

```
# Calculating the indicators by Identifier
```

```
    TSearch=arcpy.da.SearchCursor(linesSublayer,
    ["Total_Length"],expression)
```

```
for row2 in TSearch:
```

```
    if row2[0]<DisThres:
```

```
        Tcount+=1
```

```
        Tsum+=row2[0]
```

```
    else:
```

```
        continue
```

```
if Tcount!=0:
```

```
    row[1]=Tcount
```

```
    row[2]=Tsum/Tcount
```

```
else:
```

```
    row[1]=0
```

```
TAVRG2.updateRow(row)
```

```
del row,row2,TAVRG2,TSearch
```

```
else:
```

```
# If the Identifier field format is Integer
```

```
# ID selection
```

```
for row in TAVRG2:
```

```
    ID=row[0]
```

```
    expression= "" + str(Input_ID) + "' = ' + str(ID)
```

```
    Tcount=0
```

```
    Tsum=0
```

```
# Calculating the indicators by Identifier
```

```
    TSearch=arcpy.da.SearchCursor(linesSublayer,
    ["Total_Length"],expression)
```

```
for row2 in TSearch:
```

```
    if row2[0]<DisThres:
```



```
Tcount+=1
Tsum+=row2[0]
else:
    continue
if Tcount!=0:
    row[1]=Tcount
    row[2]=Tsum/Tcount
else:
    row[1]=0

TAVRG2.updateRow(row)

del row,row2,TAVRG2,TSearch
```

```
#Tool Message
arcpy.AddMessage("\n Access Tool Completed.")
```

```
#End
```

Appendix C. Density Tool

The purpose of the Tool was to calculate the Population Density of a Walking Zone using the Census blocks

The tool:

- 1-Iterates through the Walking Zone Feature Class
- 2-Intersect each Walking Zone with the Census Blocks
- 3-Apply the Proportional Density Formula
- 4-Save Results

The Parameters are:

Parameters	Format
Input Walking Zone	Feature Class
Input Participants ID	Field
Input Census Block	Feature Class
Input Census Block ID	Field
Input Census Block Statistic Field	Field
Output Table	Table View

#Start

```
import arcpy,os  
import math
```

#Over Write

```
arcpy.env.overwriteOutput=True
```

#Input Parameters

```
Input_poly = arcpy.GetParameterAsText(0)# Input Walking Zone  
Input_ID= arcpy.GetParameterAsText(1) # Input Walking Zone Identifier  
Input_blocks = arcpy.GetParameterAsText(2) # Input Census Blocks  
CensusID=arcpy.GetParameterAsText(3)# Input Census ID  
Input_field_stats=arcpy.GetParameterAsText(4)#Input Statistics Field  
output_Table=arcpy.GetParameterAsText(5) # Output Indicator Table
```

#Create Empty Table path

```
outDir = os.path.dirname(output_Table)  
outName = os.path.basename(output_Table)
```

#Looking the field type of the Service Area ID field

```
Pfields= arcpy.ListFields(Input_poly,Input_ID)
```

```

PTypeField=Pfields[0].type
#Create empty Table
table=arcpy.CreateTable_management(outDir, outName)

if PTypeField=="String":
    arcpy.AddField_management(table,str(Input_ID),"TEXT") # Identifier TEXT Field
else:
    arcpy.AddField_management(table,str(Input_ID),"LONG") # Identifier INTEGER
    Field

arcpy.AddField_management(table,"Total_"+ str(Input_field_stats),"DOUBLE")
arcpy.AddField_management(table,"Prop_Total_"+ str(Input_field_stats),"DOUBLE")
arcpy.AddField_management(table,"Density","DOUBLE") # Density Result

#Looking the field type of the CensusID field
fields= arcpy.ListFields(Input_blocks,CensusID)
CTypeField=fields[0].type

if PTypeField == "String":

    PolySearch=arcpy.da.SearchCursor(Input_poly,[Input_ID])
    for row in PolySearch:
        ID=row[0]
        expression= "" + str(Input_ID)+ " = " + "" + str(ID) + ""

        SFC=arcpy.MakeFeatureLayer_management(Input_poly,"SFC1",expression)

        ##create in memmory intersect
        copyTable = "in_memory\\copyTable"
        #Intersect
        arcpy.Intersect_analysis([SFC,Input_blocks],copyTable ,"ALL")

        TotalAreaWZ=0
        numerator=0
        OrigPop=0
        #Iterate through the Intersected Polygon

        InterSearch=arcpy.da.SearchCursor(copyTable,[Input_ID,"SHAPE@",
            CensusID,Input_field_stats])
        for cursor in InterSearch:
            ID=cursor[0]
            FragShape=cursor[1]
            FragArea=FragShape.area
            CID=cursor[2]

```

```

FStats=cursor[3]
if CTypeField=="String":
    expression= "" + str(CensusID)+ "' = ' + ""
                + str(CID) + ""

else:
    expression="" + str(CensusID)+ "' = ' + str(CID)
# Save The Original Area of the Census Block Polygon

CensusSearch=arcpy.da.SearchCursor(Input_blocks,["SHAPE@",Input_field_stats],expression)
for CScursor in CensusSearch:
    Pop=CScursor[1]
    OrigPop+=Pop
    OrigShape=CScursor[0]
    OrigArea=OrigShape.area
# Get the Proportional attribute of the fragmented Polygons
    numerator+=(FragArea/OrigArea)*FStats
    TotalAreaWZ+=FragArea
# Get the Density
    Dens=numerator/TotalAreaWZ
    PopTable=arcpy.da.InsertCursor(table,(str(Input_ID),"Total_" +
str(Input_field_stats),"Prop_Total_" + str(Input_field_stats),"Density"))
    PopTable.insertRow((ID,OrigPop,numerator,Dens))

del row, PolySearch

else:

PolySearch=arcpy.da.SearchCursor(Input_poly,[Input_ID])
for row in PolySearch:
    ID=row[0]
    expression= "" + str(Input_ID)+ "' = ' + str(ID)

SFC=arcpy.MakeFeatureLayer_management(Input_poly,"SFC1",expression)

##create in memory intersect
copyTable = "in_memory\copyTable"
#Intersect
arcpy.Intersect_analysis([SFC,Input_blocks],copyTable ,"ALL")

TotalAreaWZ=0
numerator=0
OrigPop=0
#Iterate through the Intersected Polygon

InterSearch=arcpy.da.SearchCursor(copyTable,[Input_ID,"SHAPE@",

```

```

CensusID,Input_field_stats])
for cursor in InterSearch:
    ID=cursor[0]
    FragShape=cursor[1]
    FragArea=FragShape.area
    CID=cursor[2]
    FStats=cursor[3]
    if CTypeField=="String":
        expression= "" + str(CensusID)+ "' = ' + ""
                    + str(CID) + ""
    else:
        expression="" + str(CensusID)+ "' = ' + str(CID)
    # Save The Original Area of the Census Block Polygon

CensusSearch=arcpy.da.SearchCursor(Input_blocks,["SHAPE@",Input_field_stats],expression)
for CScursor in CensusSearch:
    OrigPop+=CScursor[1]
    OrigShape=CScursor[0]
    OrigArea=OrigShape.area
    # Get the Proportional attribute of the fragmented Polygons
    numerator+=(FragArea/OrigArea)*FStats
    TotalAreaWZ+=FragArea
    # Get the Density
    Dens=numerator/TotalAreaWZ
    PopTable=arcpy.da.InsertCursor(table,(str(Input_ID),"Total_"+
str(Input_field_stats),"Prop_Total_"+ str(Input_field_stats),"Density"))
    PopTable.insertRow((ID,OrigPop,numerator,Dens))

del row, PolySearch

arcpy.AddMessage("\n Density Tool Completed.")

# End

```

Appendix D. Land Use Mix Tool

The purpose of this Toolset was to measure land use Diversity, given a Walking Zone and the Land Use Feature Class of a City. The tool applied the entropy formula originally used by Frank et al 2005. $LUM = -1 \times \sum(P \times \ln(P)) / \ln(N)$; where P is the proportion of the Area and N is the total of classes of interest. A requirement for the tool is that the zoning feature class has to be Dissolved by the Categories of interest

The tool:

- 1-Iterates through the Walking Zone Feature Class
- 2-Intersect each Walking Zone with the Land Use Feature Class
- 3-Dissolve by the Land Use Categories
- 4-Apply the Entropy formula
- 5-Save Results

The Parameters are:

Parameters	Format
Input Walking Zone	Feature Class
Input Participants ID	Field
Input Land Use	Feature Class
Input Land Use Category	Field
Output Table	Table View

```
# Start
```

```
# Importing ArcPy Functions library
```

```
import arcpy,os
```

```
import math
```

```
#Over Write
```

```
arcpy.env.overwriteOutput=True
```

```
#Input Parameters
```

```
Input_poly = arcpy.GetParameterAsText(0)# Input Walking Zone
```

```
Input_ID= arcpy.GetParameterAsText(1) # Input Walking Zone Identifier
```

```
Input_zoning = arcpy.GetParameterAsText(2)#Input Zoning Feature Class
```

```
Input_zoning_cat=arcpy.GetParameterAsText(3)# Input zoning category
```

```
output_lux=arcpy.GetParameterAsText(4)# Output Entropy Table
```

```

arcpy.AddMessage("\nCreating Table...")
outDir = os.path.dirname(output_lux)
outName = os.path.basename(output_lux)

#Looking the field type of the Input ID field
fields= arcpy.ListFields(Input_poly,Input_ID)
ITypeField=fields[0].type

#Create empty Table
table=arcpy.CreateTable_management(outDir, outName)

#Adding Fields to the Empty Table
if ITypeField == "String":
    # If the Field format = String, create a field of TEXT format
    arcpy.AddField_management(table,str(Input_ID),"TEXT")
else:
    # If the Field format = LONG, create a field of LONG format
    arcpy.AddField_management(table,str(Input_ID),"LONG")

arcpy.AddField_management(table,"Entropy","DOUBLE")

# calculate the denominator of the Entropy Formula
arcpy.AddMessage("\nSearch Cursor...")
TLUCount=0
Ifreq=arcpy.da.SearchCursor(Input_zoning,[Input_zoning_cat])
for counter in Ifreq:
    TLUCount+=1

Isearch=arcpy.da.SearchCursor(Input_poly,[Input_ID,"SHAPE@"])
for row in Isearch:
    ID=row[0]
    Mask=row[1]
    if ITypeField == "String":
        expression= "" + str(Input_ID)+ " = ' + "" + str(ID) + ""
    else:
        expression= "" + str(Input_ID)+ " = ' + str(ID)

SFC=arcpy.MakeFeatureLayer_management(Input_poly,"SFC1",expression)

```

```

    ##create in memmory intersect
copyTable = "in_memory\\copyTable"

#Intersect
arcpy.Intersect_analysis([SFC,Input_zoning],copyTable,"ALL")

#Create In Memmory Dissolve
output_diss = "in_memory\\output_diss"
#Dissolve
arcpy.Dissolve_management(copyTable,output_diss,[Input_ID,Input_zoning_cat])

#To calculate the Toltal Area of the fragmented Polygon
Diss=arcpy.da.SearchCursor(output_diss,["SHAPE@"])
TLUArea=0
for row2 in Diss:
    geom1=row2[0]
    TLUArea+=geom1.area
#Applying Entropy formula
ALOGA=0
Diss2=arcpy.da.SearchCursor(output_diss,["SHAPE@"])
for Eq in Diss2:
    geom=Eq[0]
    ALOGA+=float((geom.area/TLUArea))*float(math.log(geom.area/TLUArea))
Entropy_V=(-1*float((ALOGA)))/float(math.log(TLUCount))
#Saving Results
PopTable=arcpy.da.InsertCursor(table,(str(Input_ID),"Entropy"))
PopTable.insertRow((str(ID),Entropy_V))

del row, Isearch

# Tool Message
arcpy.AddMessage("Land Use Mix Tool Completed")
#End

```


Appendix E. Sidewalk Quality Tool

The purpose of this Tool was to calculate the Percent of Vegetation Pixels that are in a walking environment using a classified NDVI and the line features of a walking zone.

The line features parameters are calculated using the Walking Zone Generator. The outcome table has the Identifier from the Original Point Data of the Participants.

The tool:

- 1-Creates a buffer to the lines feature.
- 2-Determine the format of the Identifier ID of the lines feature
- 3-Creates an empty Table with the fields of interest
- 4-In a for loop using Search cursors and tokens, use the outcome of Step 1 to use it in the extract by mask function
- 5-In another for loop calculate the percent
- 6-Insert the results

The Parameters are:

Parameters	Format:
Input NDVI	Raster
Input Line features	Feature layer
Identifier	Field
Buffer Width	Double
Buffer Units	List
Output Table	Table

```
# Start
```

```
# Importing ArcPy Functions library
```

```
import arcpy,os  
from arcpy.sa import *  
import math
```

```
#Over Write
```

```
arcpy.env.overwriteOutput=True
```

```
# Check out the ArcGIS Spatial Analyst extension license
```

```
arcpy.CheckOutExtension("Spatial")
```

```
#Input Parameters
```

```
inRaster = arcpy.GetParameterAsText(0) # Input classified NDVI raster
```

```

Lstreet=arcpy.GetParameterAsText(1) # Input lines feature
ID = arcpy.GetParameterAsText(2) # Participants ID Field
WidthBuff= arcpy.GetParameterAsText(3) # Buffer width
WidthUnits= arcpy.GetParameterAsText(4) # Buffer units
output_Table= arcpy.GetParameterAsText(5) # Output table with the indicator

# Merging Path Names
outDir = os.path.dirname(output_Table)
outName = os.path.basename(output_Table)

#create in memmory frequency table
LBuffer = "in_memory\\LBuffer"

# Tool Message
arcpy.AddMessage("Adding a buffer...")
# Process: Buffer
arcpy.Buffer_analysis(Lstreet, LBuffer, str(WidthBuff) + " " + str(WidthUnits), "FULL",
"ROUND", "LIST", str(ID))

#Saving the field type of the Street line feature ID field
Pfields= arcpy.ListFields(Lstreet,ID)
PTypeField=Pfields[0].type

#Create empty Table
table=arcpy.CreateTable_management(outDir, outName,"","")

if PTypeField=="String":
    # If the Field format = String, create a field of TEXT format
    arcpy.AddField_management(table,str(ID),"TEXT") # Identifitfier
else:
    # If the Field format = Integer, create a field of LONG format
    arcpy.AddField_management(table,str(ID),"LONG") # Identifitfier
arcpy.AddField_management(table,"Value_1","LONG") # Value=1
arcpy.AddField_management(table,"Count_1","LONG") #Count when Value =1
arcpy.AddField_management(table,"SUM_Count","LONG") # Sum of the Count
arcpy.AddField_management(table,"Perc_1","DOUBLE") # Percent of the Count

# Tool Message
arcpy.AddMessage("Extracting by mask...")

#Creating a Search cursors with the outcome of the buffer
Buff_mask=arcpy.da.SearchCursor(LBuffer,[ID,"SHAPE@"])
#Iterates through each line buffer
for buff in Buff_mask:
    identifier=buff[0]

```

```

inMaskData=buff[1]

# Execute ExtractByMask using the classified NDVI and the outcome of the Buffer
function
outExtractByMask = ExtractByMask(inRaster, inMaskData)
#Saving the attribute table of the ExtractByMask outcome
out_view=arcpy.MakeTableView_management (outExtractByMask, "out_view1")
#Creating a Search cursor to iterate through the Attribute Table
MaskS=arcpy.da.SearchCursor(out_view,["Value","Count"])
SUM=0
Value1=0
Count1=0
perc=0
for cursor in MaskS:
    add=cursor[1]
    SUM+=add
    SelValue=cursor[0]
    if SelValue==1:
        Value1=SelValue
        Count1=add
    perc=Count1/SUM
# Inserting the indicators into the empty table

Mtable=arcpy.da.InsertCursor(table,(str(ID),"Count_1","SUM_Count","Value_1","Perc_
1"))
Mtable.insertRow((identifier,Count1,SUM,Value1,perc))
del Mtable

del cursor,MaskS, Buff_mask,buff

# Tool Message
arcpy.AddMessage("Sidewalk Quality Tool Completed")

# End

```

Appendix F. Classify NDVI Tool

The purpose of this Tool was to classify the NDVI for quality purposes. The tool use the NDVI and a threshold to create a Raster of two categories: 0 for Non-Vegetated pixels and 1 for Vegetated pixels.

The Parameters are:

Parameters	Format:
Input NDVI	Raster
Threshold number	Double
Output classified NDVI	Raster

```
# Start
```

```
import arcpy
from arcpy import env
from arcpy.sa import *
```

```
inNDVI=arcpy.GetParameterAsText(0) # NDVI file name
Tnum=float(arcpy.GetParameterAsText(1)) # Treshold Number
outVegFile=arcpy.GetParameterAsText(2) #Output Vegetation Slicing file name
```

```
# Classification Process
```

```
outReclassify=Reclassify(inNDVI,"Value",RemapRange ([[ -1,Tnum-
0.0001,0],[Tnum,1,1]]), "NODATA")
outReclassify.save(outVegFile) # Saving Classified Raster
```

```
# Tool Message
```

```
arcpy.AddMessage("Classify NDVI Tool Completed")
```

```
#End
```