CREATING A FLOOD VULNERABILITY INDEX FOR HOUSTON, TEXAS

by

Marshall Aubrey Wilson

A Thesis Presented to the
FACULTY OF THE USC DORNSIFE COLLEGE OF LETTERS, ARTS AND SCIENCES
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

August 2020

To my parents and my grandfather

Acknowledgements

I would like to thank my committee members, Drs. Wilson and Chiang, for informing me on the correct practices and methodologies for my research. I would especially like to thank my advisor Dr. Oda for guiding me through the thesis process and helping me construct this manuscript.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACS	American Community Survey
AHP	Analytic hierarchy process
CEDS	Cadastral-based Expert Dasymetric System
FEMA	Federal Emergency Management Agency
FVI	Flood vulnerability index
GIS	Geographic information system
GSA	General sensitivity analysis
HCFCD	Harris County Flood Control District
NFHL	National Flood Hazard Layer
NSS	National Shelter System
OAT	One-at-a-time
PCA	Principal component analysis
PPU	People per unit
RD	Relative deviation
RSD	Relative standard deviation
SFHA	Special Flood Hazard Area
SSI	Spatial Sciences Institute
USC	University of Southern California
VIF	Variance inflation factor

Abstract

Flooding and its associated risks and challenges pose a persistent problem for the city of Houston, Texas. Worsened by climate change and increased urban growth, the growing flood severity appears to have far outpaced any current or past efforts towards managing floods. It is, therefore, imperative to understand how flooding can affect Houston residents, and who is the most at risk and the most vulnerable. While much has been written about flood risk in Houston, relatively little current research exists regarding flood vulnerability, which in this case can be described as the intersection of flood risk, shelter accessibility, and certain social justice factors. This study used principal component analysis (PCA) and dasymetric mapping to assess flood vulnerability in Harris County, which encompasses Houston. The goal of the project was to create a flood vulnerability index (FVI) that could be used to identify areas of high vulnerability. The results of the analysis identified several high-vulnerability areas around various watersheds in the county. Several of these areas have histories of flooding and slow recovery. These results indicated that the index could effectively identify areas of high vulnerability. The residents living in these areas would be likely to experience greater suffering during a flood than in other areas. The FVI could be used by disaster planners and managers to distribute resources and aid during a flood efficiently.

Chapter 1 Introduction

The issue of flooding in Houston, Texas is an ongoing problem that seems to be growing increasingly more severe. Each new tropical storm or hurricane brings major logistical challenges to response and mitigation. Certain neighborhoods and communities are disproportionally underserved in terms of emergency aid and access to evacuation shelters. While flood risk in Houston has been heavily researched, the issue of flood vulnerability and the social factors which may contribute to it have not received as much attention. Although an accurate assessment of areas prone to flooding is critical for flood preparation and response, it is also imperative to understand which of those flood-prone areas contain the highest concentrations of vulnerable populations. This study implemented dasymetric mapping and statistical analysis to create a flood vulnerability index (FVI) for Harris County, and develop a methodology that could be improved, updated, and re-applied over time.

1.1. Project Overview

Harris County encompasses the city of Houston, as well as several smaller surrounding towns (Figure 1). This area has historically been heavily affected by flooding, which has worsened drastically due to climate change and urbanization. The issue of flood mitigation and response remains a major logistical challenge to this city, in which many people's homes are located on a floodplain, and are at risk of flooding during periods of heavy rainfall. Some of these people have disproportionally struggled to find shelter when they were forced to leave their homes, and received insufficient aid in the aftermath of major floods. Many people living in areas at risk of flooding have limited mobility and resources due to factors associated with their socioeconomic status. People with physical limitations would struggle to evacuate or find shelter.

Low-income workers would likely face considerable financial difficulties. Areas of Houston with these kinds of vulnerable populations are also prone to flooding, and have previously been heavily affected. The purposes of this study were to develop a methodology for identifying those areas and determine where people are going to have a greater need for shelter or aid in the event of a flood.
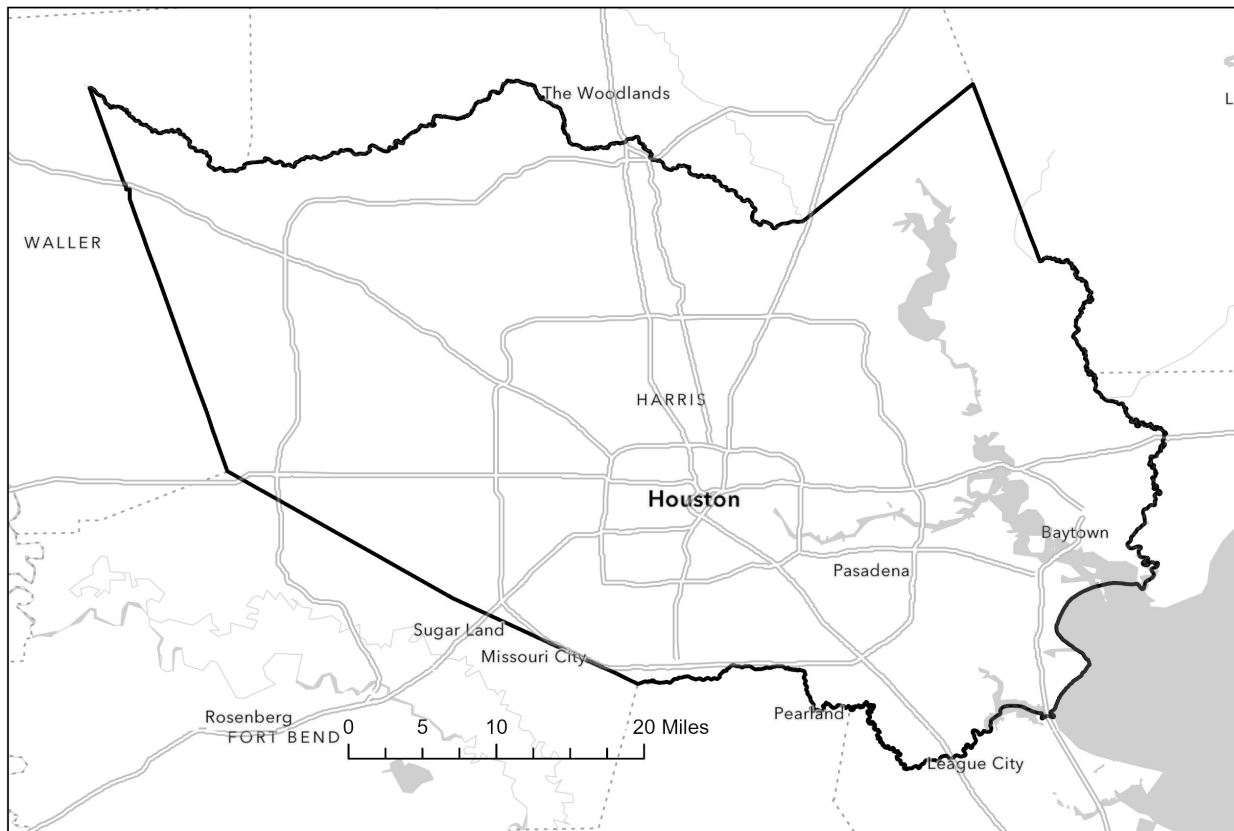
Figure 1 Harris County, Texas.

This study developed an FVI for Harris County, using spatial and statistical analysis, to identify flood-prone areas that could be in disproportionate need of aid or shelter during floods. High vulnerability tracts were identified by local neighborhood names, which were cross-referenced against reports of aid and shelter disparities in Houston. The expected results of this analysis were that the index would identify socioeconomically marginalized areas with documented histories of flooding and flood damage.

The index ranked Harris County census tracts by a weighted average of three primary factors: flood risk, shelter accessibility, and social justice. Flood risk and shelter accessibility were assessed through a dasymetric analysis, in which tract-level populations were disaggregated and redistributed among smaller parcel features. This allowed for a more precise understanding of population distribution with reference to floodplains and evacuation shelters. Principal component analysis (PCA) was used to derive the social justice factor from various American Community Survey (ACS) population estimates. This method gave statistical significance to the input variables. After a sensitivity analysis of various weighting schemes, the analytic hierarchy process was used to compare the relative importance of the three factors and assign weights to each one.

The final results of the index calculation were compared against Harris County neighborhood boundaries, to determine if high vulnerability tracts were located in previously affected areas. This analysis revealed that the highest-ranked tracts were located in neighborhoods that have historically suffered from flooding, most notably Alief (Greater Houston Flood Mitigation Consortium, 2018), Sharpstown (Hennes, 2019), and Greater Greenspoint (Rogers, 2016). Based on these results, the analysis was successful in identifying vulnerable areas of Harris County.

Although the analysis produced an FVI that correctly identified vulnerable areas in Houston, there remains the potential for improvement to both the input data and the analysis itself. For this reason, the two main stages of the analysis were developed into Python scripts, which allowed for the process to be easily repeated. The scripts modeled the analysis in a way that allowed for input workspaces and datasets to be easily changed. The landscape of flood vulnerability is constantly changing in Houston, as well as the data associated with it. New ACS

estimates are released every year, and the Federal Emergency Management Agency (FEMA) is currently re-drawing its floodplain maps for the Houston area (Despart, 2018). Python scripting allows for these updated data to be integrated into the process, producing a new and updated FVI layer.

This thesis begins with a description of the status and recent history of flooding in Houston, as well as the state of flood management in the city in Chapter 2. Previous influential studies on the subjects of flood risk, statistical and spatial analysis, and social vulnerability to natural hazards are summarized, and their influences on this study are explained. A detailed description of the analysis methodology is provided in Chapter 3, and the results of that analysis are presented in Chapter 4. Finally, the implications of those results are further discussed, and future research steps are proposed in Chapter 5.

1.2. Houston and Flooding

The nature of flood risk in Houston can be seen through an assessment of its floodplain maps. FEMA has demarcated several floodplains or zones which define varying degrees of flood risk (FEMA, 2007). The 100-year floodplain indicates areas in which there is a 1% annual chance of flooding. This is considered to be a high-risk flood zone, referred to as a special flood hazard area (SFHA). Beyond the 100-year floodplain lies the 500-year floodplain, with an expected 0.02% yearly chance of flooding. Although the degree of risk is reduced in this area, there remains the potential for flooding. With the watersheds of four major bayous passing through the city, many of its residents live at risk of flooding.

Floodwaters in the Houston area have not only exceeded FEMA's SFHA boundaries, but have done so at an alarming rate over a relatively short period of time (Blackburn, 2017). The first of these storms was Tropical Storm Allison in 2001, which passed over the city twice,

bringing flood waters that reached well beyond the 500-year floodplain and inundated about 74,000 Harris County homes. The next major flood occurred in 2012, during which the 100-year (1% annual chance) rainfall amount was exceeded within 24 hours. The 2015 Memorial Day floods brought 11 inches of rainfall in just 12 hours, which led to eight deaths and 581 water rescues. The 2016 Tax Day flood was even more severe, bringing 15 to 17 inches of rainfall in the same period, qualifying as a 500- to 1000-year flood event. The worst storm to ever impact the city was Hurricane Harvey in 2017, which brought unprecedented levels of rainfall and flooding over four days. Large areas of southeast Texas received over 40 inches of rain, with most of Harris County receiving at least 30 inches. Tropical Storm Imelda, which brought up to 43 inches of rain to some parts of southeast Texas, also impacted Houston in 2019 (Mervosh, 2019). 100- and 500-year floods have also occurred in Houston in recent decades, and life in this city is often punctuated with flooding from lesser storms and floods. However, the increasing frequency and severity of these events within the first two decades of the twentieth century is cause for concern.

1.2.1. Factors Behind Major Floods

There are several possible explanations for this noticeable surge in rainfall and flooding. Some studies have pointed to climate change as a primary culprit. The Gulf of Mexico has become increasingly warmer during this period of unprecedented flooding, providing fuel for severe weather events bringing vast quantities of rainfall (Blackburn, 2018). This pattern seems likely to continue, as indicated by a study conducted by Li et al. (2019). They used predictive modeling to demonstrate that mean annual precipitation over the Houston area would remain constant or even slightly decrease over several decades. However, it would be characterized by lengthy dry spells followed by short, intense periods of rainfall which could bring flooding.

Climate change does not appear to be delivering higher overall quantities of rainfall to the city but instead is impacting how that rainfall is distributed over time. This pattern can be seen in the weather conditions for Houston in 2017, during which the city experienced a severe drought followed shortly by Hurricane Harvey and the worst flooding in its history. Storms considered "extreme" in the past may now become increasingly more commonplace for the Houston area. This means that not only will more people likely be affected by flooding, but also that people already living in flood-prone areas will likely have to endure more severe floods than before. It is, therefore, vital to recognize this critical issue, and work towards an improved understanding of the areas of Houston which are the most at-risk, as well as its most vulnerable people.

Ever-growing urbanization can also be attributed to the increase in flood severity for the Houston area. Urbanization has been well-documented as an exacerbator of flood risk. With increases in urban growth come associated increases in impermeable surfaces (Munoz et al., 2017). Rainwater that is unable to penetrate the ground becomes runoff, which drains into local streams. The resulting increase in streamflow can cause water levels to rise quickly and overflow well into and beyond the 100-year floodplain. Through this increase in rainfall runoff, urbanization can be cited as a factor in increasing flood extents, putting people and their homes in danger of flooding, who previously would not have been considered at risk. Another connection between urban growth and severe rainfall has also been proposed by Zhang et al. (2018). A study comparing modeled hurricane simulations found that the rough urbanized ground surface of a large city such as Houston could result in greater amounts of drag on a storm, pulling it closer to the city, bringing greater quantities of rainfall. Urbanization could therefore potentially have a two-fold impact. It could be responsible for both heightened flood levels due to both increases in impermeable surfaces and wind drag, but also more frequent

occurrences of major floods as storms are drawn to the city. More research is likely required to determine which of urbanization or climate change is primarily responsible for the more frequent and severe flooding of the Houston area. Still, the current research does make it clear that this increase in flooding is a significant problem which will only get worse as both climate change and urbanization continue unabated.

1.2.2. Flood Preparation and Response

Houston's size and the number of people living in flood-prone areas present considerable challenges regrading flood preparation and evacuation. While a city-wide evacuation would be the ideal strategy for preserving human life, it is unfortunately not a practical or realistic option. Tufecki (1995) dismisses this potential evacuation strategy in favor of local evacuation options. He asserts that if an entire population of a city or county took to the few roads and highways heading away from a storm, they would put themselves at risk of creating massive traffic gridlock, exposing themselves to the elements, possibly including the storm itself. Tufecki's argument was unfortunately validated in 2005 when Hurricane Rita crossed the Gulf of Mexico in the wake of Hurricane Katrina. The response to this approaching storm was a mass evacuation, in which approximately 3.7 million people attempted to evacuate the coastal region (Baker, 2018). This evacuation resulted in massive traffic jams on every highway in the area. Hyperthermia, dehydration, and a heat-related explosion on a bus claimed the lives of over 100 evacuees, producing an evacuation death toll several times greater than that from the storm itself. There have been no large-scale evacuations from the Houston area since then. The city-wide strategy of sheltering in place has presented a whole new set of issues, however, made painfully apparent by Hurricane Harvey. Due to flooding from this storm, 30,000 people who sheltered in place rather than evacuate were forced to leave their homes (Haynie et al., 2018). This new

strategy resulted in a shelter crisis as the 230 open FEMA shelters across Texas were unable to accommodate the sudden influx of such large numbers of evacuees. Heavy rainfall could also impact the city with minimal warning. Tropical Storm Imelda made landfall within 4 hours of being classified as a tropical storm (Brown, 2019). Floods such as these would not allow enough time to effectively organize a mass evacuation effectively. While localizing evacuation is undoubtedly the best course of action for a large city such as Houston, the shelter shortage that occurred during Hurricane Harvey indicates that there is still more to be done regarding shelter and evacuation planning.

Socially vulnerable populations, who are more likely to suffer from medical problems or financial distress, are therefore more likely to require the assistance of an emergency shelter in the event of a storm. Low-income residents may not be able to afford the necessary supplies to adequately prepare for a flood. People with medical issues, especially those that affect mobility may not have access to a hospital or emergency care facility. Local evacuation shelters are critical for supporting disadvantaged people who are unable to seek help through traditional means. Karaye, Thompson, and Horney (2019) found that shelters in the Houston-Galveston area can only accommodate 36% of evacuees with housing and transportation needs. With such a significant disparity between shelter availability and potential evacuees, in the event of future storms, it is necessary to understand where shelter deficits for socially vulnerable people are highest.

1.2.3. Flood Management

Although the current response plan for major floods in Houston is in need of expansion and other improvements, the city is still making significant efforts towards mitigating floods and the damage they cause. The Harris County Flood Control District (HCFCD), which was founded

in 1937, is currently conducting a series of improvement projects on several of Houston's major watersheds (Lynn, 2017). These projects include Project Brays, a multi-phase undertaking which has been ongoing since 1994. The project was about halfway complete when the Brays Bayou overflowed during the 2015 Memorial Day flood. This situation allowed Bass et al. (2017) to compare the areas which had been improved against the areas which had not yet been improved. They found that flooding in finished areas was confined to the 100-year floodplain, while flooding exceeded the 100-year floodplain in areas where construction had either not begun or had not yet been completed. In 2018, Harris County voters passed a $2.5 billion flood bond, and now approximately 80 more improvement projects which, as of November 2019, are in various stages ranging from waiting for funding to just beginning construction (Arraj, 2019). The county adopted a "worst first" criteria for expediting individual development projects primarily based on the severity of flood risk for each project area. The HCFCD intends to have begun all 80 projects by 2022. These projects should be major undertakings that will likely significantly improve numerous flood-prone areas of Houston. However, construction will not be a quick process, and many of these projects will not be complete for several years. Considering the frequency of flooding in the Houston area, it is likely that some of the areas that are slated for improvements will be impacted by flooding before those improvements are complete. While those projects are underway, it should be a priority to ensure that effective local evacuation strategies are in place. The identification of vulnerable areas along the various at-risk watersheds could be useful for those strategies.

In addition to the numerous watersheds in Harris County which are waiting for improvements, other watersheds experience regular flooding but do not qualify for federally funded improvement projects due to the low value of the structures located in vulnerable areas.

One such neighborhood that does not meet that qualification is Greater Greenspoint, which is a low-income area that has been heavily impacted by flooding five times in the twenty-first century (Elliott, 2017). This neighborhood is intersected by the Greens and Halls Bayous, which are the source of frequent and severe flooding. Although this area is clearly in dire need of substantial mitigation, only limited work has been approved for the Greens Bayou watershed (Blackburn and Bedient, 2018). Greenspoint serves as an example of why social justice factors are an important element to consider in flood management. Residents in this area are generally low-income, and cannot afford to move or rebuild after a flood (Miller and Goodman, 2019). With 5,700 homes located on floodplains, increased flood mitigation efforts in Greenspoint (as well as similar low-income neighborhoods) would significantly reduce the number of socially and economically vulnerable people in need of aid, and allow for better distribution of resources and responders.

With increasing urbanization, climate change, and the numerous improvement projects intended to curb the side effects of those two factors, Houston's flood risk landscape will likely change significantly over time. The digital representation of that landscape will change as well, with new FEMA floodplain maps expected by 2023 (Despart, 2018). Several studies have demonstrated that FEMA's 100- and 500-year floodplain maps drastically underestimate the potential extent of flooding in Houston. One-third of the homes damaged during the 2015 Memorial Day floods were located outside the furthest extent of FEMA's floodplain map (Hunn, Dempsey, and Zaveri, 2018). The Tax Day flood just over a year later would flood numerous homes, 55% of which were located outside the 500-year floodplain. Similarly, over half the homes damaged in Hurricane Harvey were located outside the 500-yer floodplain. An earlier study of flood insurance claims from 1978 to 2008 found that 47% of all claims were located

outside of the 100-year floodplain (Highfield, Norman, and Brody, 2013). While FEMA's floodplain maps can substantially underestimate the extent of flooding, they do still indicate areas where flooding is most likely to occur, particularly in and around Houston's numerous bayou watersheds. The analysis developed for this study utilized the 100-year floodplain for its flood hazard layer, although it would benefit from more accurate data. For this reason, the dasymetric analysis model was designed as a Python script, which could be re-applied with updated floodplain data.

1.2.4. Social Justice and Flooding

People with social and economic disadvantages can have greater vulnerability to flooding than those without those disadvantages. Vulnerability in the context of flooding refers to the potential harm an individual could suffer when their home or community is affected by a flood (Balica, Wright, and van der Meulen, 2012). Vulnerable people can be described as those who are at risk of flooding and are likely to struggle to prepare, respond to, and recover from a flood. Flooding has impacted much of Houston, and has affected both rich and poor neighborhoods (Castles, 2018). However, areas with high populations of residents with limited financial means, medical problems, and restricted mobility are likely to disproportionately suffer in the event of a flood. An assessment of flood vulnerability is therefore incomplete without a social justice element, as it can indicate to planners and responders where the greatest amounts of aid should be allocated in the event of a flood.

Greater Greenspoint is not the only economically distressed Houston neighborhood to suffer from regular flooding. Still, it provides a stark example of the struggles faced by residents living in such an area. Numerous large, aging apartment buildings located near the bank of Greens Bayou flood regularly, and the impoverished residents living within them have few

options regarding preparation and evacuation (Rogers, 2016). With high costs of living elsewhere in the city, many living in these buildings have no choice other than to live through the flooding. To simply raze the damaged structures and mandate that all residents move to less flood-prone areas would fail to address the underlying socioeconomic problems which led to their settling in that neighborhood in the first place. Displaced residents could be unable to find similarly located homes, and many of those without personal vehicles may lose access to their place of employment if they were forced to move to a different part of the city. Until a solution to this cycle of poverty and flood risk is devised, the most realistic current action is to identify areas of vulnerability, so that they can receive adequate aid in the event of a flood. This study was designed to identify areas such as Greenspoint, where many residents are at risk of flooding, but lack the means to effectively evacuate and recover from it.

Social justice factors and their relation to flood risk and vulnerability have been previously examined for the Houston area, although there is not a substantial body of research on the subject. Peacock et al. (2012) performed an assessment of social justice factors in nearby Galveston and presented an analysis to inform disaster response. Their analysis used many similar criteria as this study's FVI analysis. An environmental justice index was also incorporated into Harris County's plan for implementing the HCFCD's improvement projects (Arraj, 2019). It should, however, be noted that some Harris County officials did oppose its incorporation into the criteria for determining the schedule of projects. Increased research into the subject of flood vulnerability and social justice as it relates to natural hazards could allow for methods to be refined and for the field of study to become more widely accepted.

Chapter 2 Literature Review

This project is a contribution to the growing body of knowledge regarding Houston's struggle with flooding and its work towards effective mitigation. This subject encompasses various topics of study, including flood risk, shelter accessibility, and social justice analysis. This thesis presents a methodology for determining vulnerability as the intersection of those three topics. There is a substantial quantity of literature surrounding these subjects, and a sample of this literature is described below. Much has been written on flood risk and shelter accessibility in the Houston area. Also, some studies have discussed the potential for correlation between flood risk and socioeconomic factors. However, there are relatively few works describing analyses that examine social vulnerability as an exacerbating factor for people living in at-risk areas for flooding. This literature review presents various peer-reviewed reports on subjects directly related to this project. Those subjects include flood risk delineation, dasymetric mapping, shelter accessibility, social justice factors and their relation to flood risk, flood vulnerability analysis, and the statistical challenges when conducting such an investigation. These topics are addressed in this chapter, and inform the methods and concepts which were applied in this study. This project heavily relied on existing flood risk, land use, and demographic data for its flood vulnerability analysis. The studies summarized in this chapter provide examples of the effective use of similar datasets, as well as their capabilities and limitations.

2.1. Mapping Flood Risk

There are a multitude of different strategies for demarcating flood risk using a variety of spatial analysis methods and tools. Several of these methodologies have been applied to the Houston area. Generally, these analyses use predictive modeling to determine future flood extent based on the interaction of several input datasets. The current authoritative source for flood risk

extent in the U.S. is the Federal Emergency Management Agency's (FEMA) National Flood

Hazard Layer (NFHL). This layer demarcates the floodplains using a hydrologic model.

Blessing, Sebastian, and Brody (2017) describe how that model can produce inaccurate results.

There is potential for measurement error due to limited hydrometeorological observations, as

well as inaccuracies due to changing land use over time. Houston continues to grow, and with

that growth comes increases in impermeable surfaces and rain runoff, which leads to a greater

risk of flooding (Muñoz et al., 2017). After a series of major floods in the early twenty-first

century, FEMA's floodplain maps for Houston began to come under scrutiny. They were

demonstrated to significantly underestimate the extent of flood risk for the city, as numerous

flood insurance claims over several years have been made for properties well outside of the 100-

year floodplain.

Although FEMA's dataset continues to be considered a useful indicator of potential flood

risk, analytical methods have been proposed, which could potentially improve their flood risk

maps. One such method is described by Bass and Bedient (2018), whose study combined several

predictive models to account for flooding resulting from both storm surge and rainfall. They took

into account the potential to be impacted by flooding from both sources due to the study area's

location in southeast Houston on the Texas Coast. Another study by Gori et al. (2017) addresses

the issue of changing land use by incorporating land use projections into their process for flood

risk delineation. FEMA and Harris County are redrawing the floodplain maps for the Houston

area using LiDAR and predictive flood modeling (Despart, 2018). Several floods, including

Hurricane Harvey, exposed the current maps, which were completed in 2001, as severely

underestimating potential flood extent in the Houston area. The goal of this project, which is

expected to be completed by 2023, is to gain a better understanding of flood risk extent in Harris

County. Efforts to develop a comprehensive system for identifying flood risk are ongoing, and flood risk data for Houston will continue to change over time with improvements in technology and analytical methods as well as changes to the landscape itself.

2.2. Dasymetric Mapping

Dasymetric mapping is the redistribution of spatial data to smaller, more specific spatial units for more precise analysis (Petrov, 2011). It has been utilized for mapping flood risk and vulnerability. Maantay and Maroko (2009) used a methodology referred to as the Cadastral-based Expert Dasymetric System (CEDS) for mapping flood vulnerability in New York City. Through this method, they disaggregated tract-level census data into smaller residential units. The results of the study indicated a substantial difference in calculated at-risk populations, with the dasymetric method indicating a much lower population. Maantay, Moroko, and Herrmann (2013) further describe this system in another article. This method of disaggregation utilizes tax parcel data to gain a more precise understanding of housing density within a given census tract. In urban areas where population density can vary significantly from parcel to parcel depending on structure type, this methodology allows for the most accurate possible estimation of that distribution. A simpler method, referred to as the three-class method, was utilized by Giordano and Cheever (2010) to identify communities at risk of hazardous waste exposure. This method identifies a habitable zone, and then divides that zone into three new land use classes. The three classes defined for their study were nonurban, low-density residential, and high-density residential. Populations were then redistributed across those three classes.

2.3. Shelter Accessibility

The availability of local shelter options for evacuees is critical for effective flood response. Tufekci (1995) suggests that establishing local shelter options is preferable to mass

inland evacuation in advance of an approaching hurricane. He asserts that if an entire population

of a city or county took to the few roads or highways heading away from the storm, they would

run the risk of creating a massive traffic jam, exposing them to the elements for a long period of

time, and even the storm itself if it were to change direction. The failed evacuation from

Hurricane Rita confirmed these concerns a decade later (Baker, 2018). It is therefore necessary

for any assessment of flood vulnerability to take local shelter accessibility into consideration.

As with flood risk, there are several methods for determining shelter accessibility, which

range in complexity and number of inputs. An example of one of the more complex methods is

described in a paper by Curtis (2016), whose study utilized a network analysis for determining

the closest local shelters to certain areas of the Dallas-Fort Worth Metroplex. Roadway data was

utilized to determine routes and travel times to shelters, and bridge data was used to identify

locations of potential impedances in the case of an earthquake. Travel times were compared

against each other to determine degrees of shelter accessibility. This method is among some of

the more complex methods for quantifying hurricane shelter accessibility, although it could also

be assessed by simply identifying shelter service areas, which consist of the area within a defined

radius from the shelter location (Chen et al., 2017). This service area-based methodology, which

could be performed through a buffer analysis, was applied for this study.

The association of social justice factors with shelter accessibility has also been explored

through spatial analysis. Karaye, Thompson, and Horney (2019) used spatial statistical methods

to determine shelter accessibility for people with housing and transportation needs. The study

found that Harris County, which contains Houston, had the highest shelter deficit of Texas

coastal counties. There simply are not enough established shelters in Houston to accommodate

the massive numbers of evacuees from a severe flood. The results of this analysis were made

apparent during the shelter crisis following Hurricane Harvey (Haynie et al., 2018). The lack of accessible shelter in Houston is factored into this study through the examination of proximal shelters to at-risk areas. This element of the analysis identifies areas which are especially deprived of adequate shelter options.

2.4. Social Justice and Flood Risk

Natural disasters such as major floods can expose societal inequalities in the areas they impact. Certain groups of people can struggle to evacuate or recover from a flood more than others. These people are typically economically disadvantaged and/or socially marginalized. There is a substantial body of existing research on the subject of social justice and its relation to natural hazard risk, both in terms of describing that relationship as well as quantifying it through multiple regression analysis. This section provides a sample of those works and their implications for this study.

2.4.1. Previous Studies in Houston

Prior studies have used spatial analysis to examine the spatial correlation between flood risk and social justice factors. Castles (2018) performed such a study for the Houston area, in which she utilized a variety of ACS data ranging from economic status to race to determine whether or not socially vulnerable people in Houston were concentrated in high-risk flood zones. Interestingly, although her findings indicated a concentration of marginalized populations in the inner-city areas, they did not show a direct correlation between flood risk and social vulnerability. Instead, the results indicated an indirect correlation. Another study conducted by Maldonado et al. (2014) focused on the distribution of Hispanic immigrants. They found that there is a higher likelihood for Hispanic immigrants to live on a 100-year floodplain than non-Hispanic whites. These two reports seem to contradict each other, although that is likely due to

the more comprehensive nature of Castles' analysis. Another study conducted by Chakraborty, Grineski, and Collins (2019) found that people with disabilities were disproportionally exposed to flooding during Hurricane Harvey. The results of these studies suggest that while some socially vulnerable populations may be more highly concentrated in flood-prone areas, others may not. Therefore, it cannot be assumed that all socially vulnerable people in Houston are at greater risk of flood exposure without further research regarding each specific population. Houston is intersected by many bayou floodplains, which cover much of Harris County (Figure 2). The purpose of this study, then, is to identify locations where high concentrations of socially vulnerable populations and flood risk areas intersect, indicating areas that would be at greatest need during a major flood.
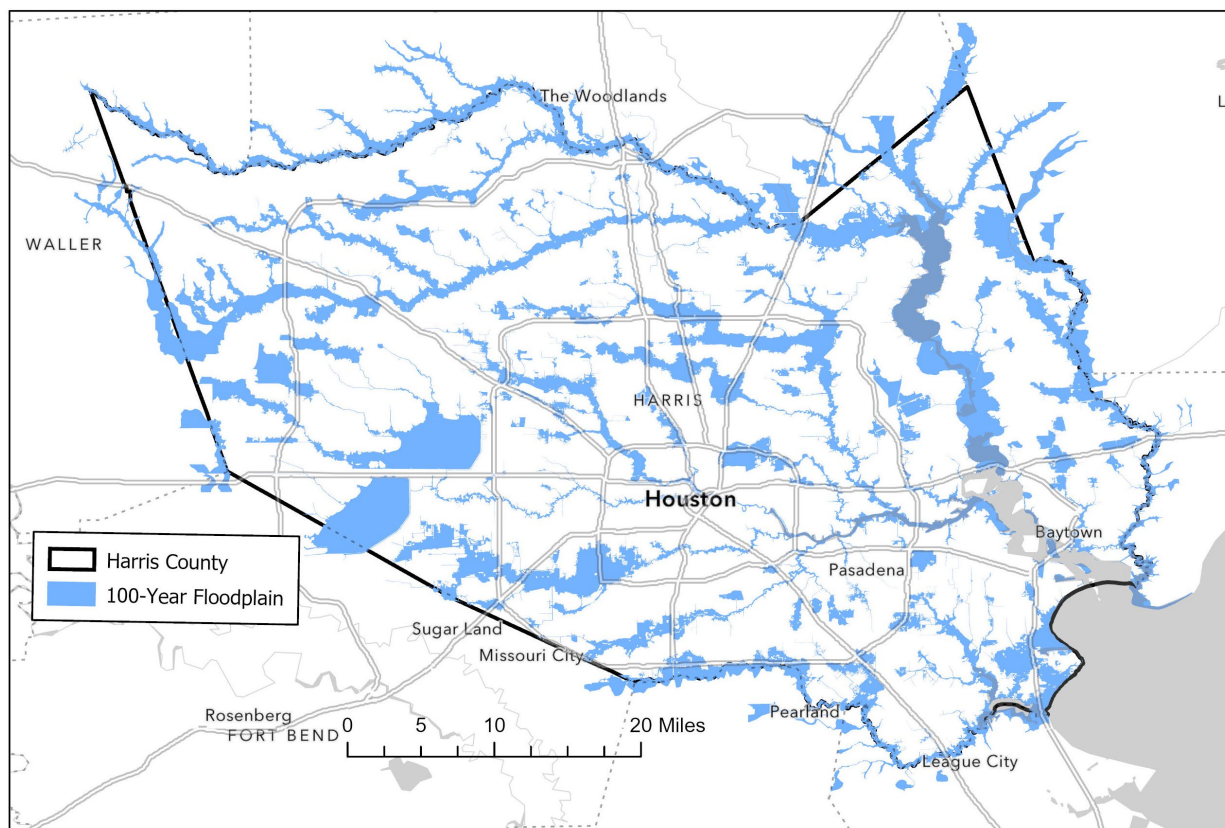

Figure 2 High-risk flood zones in Harris County, Texas.

2.4.2. Social Vulnerability Index and Multicollinearity

Multicollinearity is a potential problem for the social vulnerability index and other multiple regression models (Graham, 2003). With numerous input variables, there is the possibility for correlation between two or more variables, despite them being independent of each other. This could affect the statistical significance of the input variables, casting doubt on the analytical process. Multicollinearity among a group of variables can be assessed by examining the correlation matrix for all of the variables. In a matrix where numerous variables are highly correlated with each other, there will likely be multicollinearity within the dataset. Multicollinearity can be quantified for individual variables, by calculating the percent increase in variance caused by one variable's correlation with the other variables. This calculation produces the variance inflation factor (VIF) for each variable. Variables with a high VIF can be indicators of high multicollinearity in a linear regression model.

Cutter, Boruff, and Shirley (2003) describe a methodology for creating a social vulnerability index which has been heavily referenced by numerous subsequent studies. Their analysis used a Principal Component Analysis (PCA) to mitigate multicollinearity and reduce variance inflation. A PCA creates new, composite variables, or components from the input variables. In order, each component explains an increasingly smaller percentage of the variance within the new component dataset. Cutter, Boruff and Shirley (2003) applied this analysis to 32 independent variables used for their social vulnerability index. These variables consisted of several types of social justice factors, including personal wealth, gender, age, and ethnicity. (Table 1). PCA created 32 new components; each one was a composite score of the 32 input variables. From these components, 11 were selected for inclusion in the vulnerability index. These final components were selected through the application of the Kaiser Criterion, in which the eigenvalues of the components' correlation matrix were calculated, and components with

eigenvalues greater than 1.0 were selected for inclusion in the final index (Guillard-Gonçalves et al., 2015). This eliminated components that did not explain an acceptable amount of variance. A Varimax rotation was then applied to the values in each component, which maximized the number of very high and very low values. The variable with the highest positive or negative correlation to a component was determined to be the dominant variable for that component.

Table 1 Principal components for a social vulnerability index.

		Dimensions of Social Vulnerability		
Factor	Name	Percent Variation Explained	Dominant Variable	Correlation
1	Personal wealth	12.4	Per capita income	+0.87
2	Age	11.9	Median age	−0.90
3	Density of the built environment	11.2	No. commercial establishments/mi^2	+0.98
4	Single-sector economic dependence	8.6	% employed in extractive industries	+0.80
5	Housing stock and tenancy	7.0	% housing units that are mobile homes	−0.75
6	Race—African American	6.9	% African American	+0.80
7	Ethnicity—Hispanic	4.2	% Hispanic	+0.89
8	Ethnicity—Native American	4.1	% Native American	+0.75
9	Race—Asian	3.9	% Asian	+0.71
10	Occupation	3.2	% employed in service occupations	+0.76
11	Infrastructure dependence	2.9	% employed in transportation, communication, and public utilities	+0.77

Source: Cutter, Boruff, and Shirley (2003)

The social vulnerability element of this thesis relies on the methods described by Cutter, Boruff, and Shirley (2003). Their social vulnerability index has been repeatedly tested and applied in in areas of the United States (Sherrieb, Norris, and Galea, 2010), Canada (Oulahen et al, 2015), Brazil (Roncancio and Nardocci, 2016), and Portugal (Guillard-Gonçalves et al., 2015). Most importantly, their use of PCA to calculate index scores has been demonstrated to be an effective method for mitigating the effects of multicollinearity, which can be expected in such

an analysis. The flood vulnerability index created through this study utilizes similar inputs for identifying at-risk populations, several of which are highly correlated. The use of a PCA reduces the dimensionality of the input data and maximizes the variance among the components.

Flanagan et al. (2011) performed a study in which they used a multiple regression analysis to assess social vulnerability. They created a social vulnerability index for emergency management use in New Orleans, using variables and methods derived from Cutter, Boruff, and Shirley (2003). They assessed vulnerability at the tract level using fifteen census variables. In addition to creating the index, they demonstrated its value through comparison with data related to the impact of flooding from Hurricane Katrina. They examined mail delivery rates four years after the hurricane as an indicator of recovery in neighborhoods damaged from flooding. They found that mail delivery rates returned to or exceeded pre-Katrina rates in areas with the least social vulnerability, and were less than twenty-five percent in the Lower Ninth Ward, which contained tracts in the highest rank of the social vulnerability index. This comparison demonstrates how a social vulnerability index can be used to identify areas that are most likely to struggle to recover from a catastrophe such as a flood. The authors do, however, provide the caveat that such an index is part of a larger system including natural hazards, vulnerable infrastructure, and community resources. In order to gain a complete picture of vulnerability, those additional factors must also be assessed.

2.5. Flood Vulnerability Analysis and Criteria

Flood hazard data can be incorporated into a social vulnerability index to create an FVI. Although relatively little work has been done regarding flood vulnerability analysis in the Houston area, other such studies have been performed in other flood-prone parts of the world. Burton and Christopher (2008) developed a flood vulnerability index for the Sacramento-San

Joaquin Area in California. They examined their index in the context of potential flood risk due to levee failures. They found that sizeable clusters of vulnerable populations lived within the flood risk zone they developed using FEMA's Hazus model. These findings indicated that the areas containing those clusters would likely struggle to recover and require substantial aid in the event of a flood. The inclusion of flood risk data with the social vulnerability index afforded an additional level of specificity in assessing vulnerability to potential floods.

Flood vulnerability indices can be used for a variety of different purposes relating to flood preparation and response. Balica, Wright, and van der Meulen (2012), for example, demonstrated how their flood vulnerability analysis could be used to assess current vulnerability conditions for a range of cities, and to predict future flood mitigation needs as climate change continues to alter the hydrologic landscape of coastal cities. "Zachos et al. (2016), in their vulnerability analysis, generated an index that can be used for spontaneous disaster planning by incorporating predictive flood models as well as other ecological, economic, and social factors. An FVI can be used in a generalized context to assess vulnerability for an entire region, or it can be used to identify specific areas of vulnerability within the actual or estimated extent of a flood.

The criteria for an FVI can vary, depending on the physical characteristics of the study area and the residents living within it. Remo, Pinter, and Mahgoub (2016) developed an FVI at different spatial granularities for Illinois. In addition to concluding that vulnerability is best assessed at the block level (the smallest scale for census data), they also found that vulnerability has different characteristics in rural as opposed to urban areas. Their findings indicated that vulnerability in rural areas was more driven by losses due to flooding, while social vulnerability was the main driver in urban areas. When urban areas are assessed for flood vulnerability, the unique socioeconomic characteristics of people and communities within them must also be

considered. Balica, Wright, and van der Meulen (2012) describe social justice criteria for an FVI as factors that affect people's everyday lives, depriving them of mobility or the ability to recover, such as disability, age, or poverty. These can be relevant indicators of vulnerability in most urban areas, although there may be other factors that may be more appropriate when considered in certain areas rather than others. Oulahen et al. (2013) argue that flood vulnerability indices would be more effective if they were developed with the input of local policy workers. When developing such an index, it is critical to have an understanding of the nature of social vulnerability in the local population, and the specific challenges they may face. Indices such as that developed by Zachos et al. (2016) can be re-created in different areas with similar social and geographical makeup.

Although many social justice variables can be considered relevant to flood vulnerability, they should not all necessarily be factored into the FVI calculation. This can potentially overinflate the value of certain indicators. Balica and Wright (2011) describe a revision of a model they had previously developed four years prior. In their assessment of their model, they found that many of the variables they used were either redundant or unrelated to vulnerability. Through a process in which they factored out numerous highly correlated variables, they reduced the number of vulnerability indicators from 71 to 28. They emphasize the importance of using only the minimum number of necessary indicators. Reducing the number of variables not only improved the quality of the index data, but allowed for the index to be more flexible and easier to understand for those who wish to apply the analysis to a new study area.

2.6. ACS Data Accuracy

Population estimates compiled through programs such as the ACS can be used for a variety of purposes, including social and environmental justice analyses. The spatial component

of census data allows for it to be compared to and associated with other spatial datasets, so that the relationship between demographics, socioeconomics, and other possible correlating factors can be examined. However, not all population estimates can be assumed to represent their associated areas accurately. Each population within an ACS table also has an associated margin of error. These margins of error can vary considerably, depending on the reliability of the survey results and the size of the sampled population (Folch et al., 2016). Under a single count field within an ACS table, some counts may be analytically viable, and others may not. This inherent uncertainty in ACS estimates must be addressed in any analysis that uses them.

Spielman, Folch, and Nagle (2014) propose a method for mitigating the uncertainty in ACS data, which would be useful for the analysis described in this proposal. Potential inaccuracies in ACS estimates can be mitigated through aggregation. Although inaccurate estimates may be fed into an investigation, they can be negated to a degree by other more accurate estimates. This aggregation can be performed either by combining estimates across different spatial features, such as multiple adjacent tracts, or by combining multiple estimates for a specific geographic area. Combining multiple different ACS attributes increases sample sizes and mitigates potential error within certain variables. For multiple regression analyses such as the FVI, which considers several different social justice populations, attribute aggregation is a useful method for reducing uncertainty in the independent social justice variables.

2.7. Analytic Hierarchy Process

In multi-criteria analyses such as in this study, it is necessary to assign certain weights to the input criteria in order to return a result that reflects the anticipated effect of each criterion. The allocation involves a high degree of subjectivity, even if the relative importance of each criterion is known. A commonly used method for assigning weights is through AHP (Saaty,

1990). It derives weights through a comparison matrix, in which all criteria are compared against each other. The weights are calculated from the normalized principal eigenvector values of the matrix. AHP is a widely used methodology for analytical models for a variety of industries, including marketing, health care, energy, and numerous others (Subramanian and Ramanathan, 2012). Although the AHP is still partly subjective, it introduces a degree of statistical objectivity to the weighting process and helps justify weight assignments.

AHP has been utilized in various multi-criteria GIS models. This type of analysis is useful for the creation of indices and site suitability analyses. Wu et al. (2011) conducted an analysis using GIS and AHP to determine floor water inrush vulnerability of a coal seam in a mine in China. They present the combination of GIS and AHP as a series of three steps: process spatial data through GIS to quantify the various analysis factors, calculate factor weights through the construction and application of a comparison matrix, and then map and display the results of the combined weighted factors. Uyan (2013) conducted a GIS-based study, which analyzed several criteria, including terrain, local climate, and proximity to transmission lines to identify the best possible locations for solar farms in the Karapinar region of Turkey. Weights were assigned to the various criteria and sub-criteria of their site suitability analysis. The use of AHP allowed for multiple different variables to be compared according to their relative importance and generated weights through a logical and statistically driven process.

2.8. Sensitivity Analysis

There are two major types of sensitivity analysis: local, or one-at-a-time (OAT), and global, or general sensitivity analysis (GSA) (Feretti, Saltelli, and Tarantola, 2016). Sensitivity analysis through OAT is the simplest methodology, in which one factor is changed at a time in order to determine the models' sensitivity to the changed variable. GSA methodologies are

referred to as global because they perform an overall examination of analysis inputs and their influence on outputs (Zhou, Lin, and Lin, 2008). Although the OAT methodology is widely used, it has certain shortcomings. OAT can produce inaccurate results when applied to more complex models, as a greater number of variables increases the dimensionality of the dataset (Saltelli and Annoni, 2010). Altering one factor at a time does not account for that dimensionality, which requires a more statistically sound methodology. For this reason, GSA methods are preferred.

A simple strategy for determining sensitivity is the relative deviation (RD) method (Hamby, 1994). This method is similar to the OAT method, in that one model parameter is changed at a time. However, the RD method is different in that a much larger sample of the input distribution is used. The relative deviation for each output is calculated as the ratio of the standard deviation to the output mean. This test can indicate each factor's contribution to the variability in the model's output. Hamby (1995) compared RD against several other sensitivity analysis methods, and found it to be a reliable method for measuring a given parameter's sensitivity.

Chapter 3 Data and Methods

This project used a combination of spatial and tabular data to identify areas of flood vulnerability in the Houston area. Census population estimates were redistributed among lot-sized parcels, and flood risk was assessed as the spatial intersection between flood hazard boundaries and populated parcels. Similarly, shelter needs were identified as populated flood hazard areas which did not have access to local shelter options. Statistical analysis through PCA was applied to various social justice population estimates for inclusion in the analysis model. The three main factors, flood risk, shelter accessibility, and social justice were combined to create an index showing flood vulnerability across Harris County. The data and methodologies used to create this index are described in this chapter.

3.1. Data

The inputs for this analysis consisted of a table with 12 columns of demographic data, and five spatial datasets (Table 2). The demographic input variables were chosen due to their use in previous vulnerability analyses (Cutter, Boruff, and Shirley, 2003; Guillard-Gonçalves et al., 2015), as well as their potential to affect disaster response and recovery. The specific nature of Houston's transportation system was also taken into consideration with the inclusion of the no-car household variable. The spatial datasets include census tract boundaries, tax and land use parcels, FEMA floodplains, and FEMA National Shelter System (NSS) locations. These datasets were incorporated into a spatial analysis, which also included the demographic data, to create the FVI.

Table 2 Input spatial and tabular datasets for the study analysis.

Dataset	Type	Description	Source
ACS Social Justice Populations	Table	Social justice population estimates were extracted from several ACS tables, and merged into a single table, which was then used as the initial input for the PCA portion of the analysis.	United States Census (data.census.gov)
Census Tracts	Polygon Feature Class	The vulnerability index was developed and displayed at the tract level, the boundaries of which were demarcated as part of the 2010 Census.	City of Houston Open GIS Data (cohgis-mycity.opendata.arcgis.com)
Flood Risk Zones	Polygon Feature Class	Flood risk areas were defined as the extent of FEMA's 100-year floodplain, which was extracted from the original floodplain dataset, and input into the analysis as a polygon feature class.	City of Houston Open GIS Data (cohgis-mycity.opendata.arcgis.com)
Land Use Parcels	Polygon Feature Class	Each parcel contained a descriptive code, which indicated the type of structure located in that parcel. The land use dataset was used to identify large residential structures.	City of Houston Open GIS Data (cohgis-mycity.opendata.arcgis.com)
FEMA Shelters	Point Feature Class	A "snapshot" of the FEMA NSS during Hurricane Harvey in 2017 is the most current spatial representation of FEMA shelters, the shelter points within were used to determine the degree of shelter need in each tract.	FEMA (gis.fema.gov)
Tax Parcels	Polygon Feature Class	This dataset contained the same parcel geometry as the land use dataset, with a different coding system. These parcels were used to identify smaller residential structures.	Harris County Appraisal District (pdata.hcad.org/GIS)

3.1.1. ACS Data

Each of the 12 social justice variables was gathered from tract-level ACS estimates, and were chosen for their potential impact in flood situations (Table 3). The reasoning for each variable has also been previously explained by Cutter, Boruff, and Shirley (2003), with the exception of No Car, the reasoning for which has been previously explained above. This study utilized a smaller number of variables compared to other studies, with particular focus given to the factors which could negatively affect one's ability to evacuate or recover from a flood.

Table 3 ACS estimates included in the social justice dataset.

Variable	Source Table	Description
Disability	Sex by Age by Disability Status (B18101)	People with disabilities could face challenges in evacuating from a flood as well as in seeking medical treatment. This impact on mobility can make them more reliant on emergency services.
Female	Sex by Age (B01001)	Women are more likely than men to struggle during natural disasters due to generally lower wages and increased likelihood of parental responsibilities.
No Car	Household Size by Vehicles Available (B08201)	Most Houston residents heavily rely on personal vehicles for transportation. The lack of a car would seriously impact one's ability to evacuate, seek medical assistance, and gather emergency supplies.
Over 64	Sex by Age (B01001)	Older people are more likely to have restricted mobility and require specialized assistance.
Part-Time Worker	Full-Time, Year-Round Work Status in the Past 12 Months by Age for the Population 16 years and Over (B23021)	Part-time workers may lack the financial capabilities to endure a prolonged natural disaster. This may be exacerbated by a lack of employment caused by the event.
Poor English Speaking	Language Spoken at Home by Ability to Speak English for the Population 5 Years and Over (B16001)	An inability to effectively communicate with disaster response personnel and other residences could impede one's ability to adequately prepare for and respond to a disaster situation.
Poverty	Poverty Status in the Past 12 Months (S1702)	Low-income residents affected by a flood would likely struggle to recover. A lack of financial means could also impact their ability to effectively prepare for a flood.
Receive Public Assistance	Public Assistance Income in the Past 12 Months for Households (B19057)	People who rely on social programs for support would likely also need additional assistance during a natural disaster. A disruption of those services could also increase their need for aid.
Renter	Tenure by Household Size (B25009)	Renters often do so out of financial necessity, as they cannot afford home ownership. If their lodging were to become uninhabitable, they could face difficulty in finding shelter or a new living space.

Single Parent	Households and Families (S1101)	Single-parent households often have limited financial means, and an increased burden due to the necessity of child care.
Under 10	Sex by Age (B01001)	Young children are typically reliant on parental support for survival. They lack the mobility, financial means, and knowledge necessary to effectively respond to a disaster.
Unemployed	Employment Status for Population 16 Years and Over	Unemployed residents are likely to be struggling financially, which would be exacerbated by the costs associated with flood evacuation and recovery.

These 12 variables, although independent of one another, could still be highly correlated with each other. Multicollinearity, in which several variables are correlated with several other variables, could inflate the variance within the social justice dataset. This variance inflation could affect the statistical significance of the social justice dataset and cast doubt on the analysis results. Multicollinearity was assessed through the calculation of the VIF for each variable. A PCA was then performed on the dataset when several high VIFs indicated a high degree of multicollinearity.

3.1.2. Parcel Data

Reliable and authoritative parcel data for the Houston area was critical for the dasymetric mapping of population data. Two different parcel datasets were available from the Harris County Appraisal District, which satisfied this requirement. Both land use and tax parcel datasets utilized the same polygon geometry, although with two different classification systems. The land use parcels were classified using four-digit, numeric land use codes, while tax parcels were classified using two-character, alphanumeric state classification codes. Although the two different codes provide similar descriptions of the parcels they classify, they do so with varying degrees of specificity. For example, a land use parcel with code 1003, improved residential, could be

coincident with a tax parcel with the state classification code 1003, single-family residential. Although both the land use and tax parcels indicate that a particular area has a residential use, the tax parcel provides a more specific description. conversely, an area classified by a tax parcel with state classification code B1: multi-family residential, could be coincident with a land use parcel coded as 4212: 4-12 story apartment structure. In this case, the land use code is more descriptive than the state classification code, and can be used to compute a more accurate estimation of the number of people living within that parcel. Through a comparison of the two different parcel datasets, land use parcels were found to better represent larger residential structures, and tax parcels were found to more accurately classify smaller structures (Figure 3). Therefore, they were both combined in the analysis for the most accurate possible assessment of population distribution within each census tract.
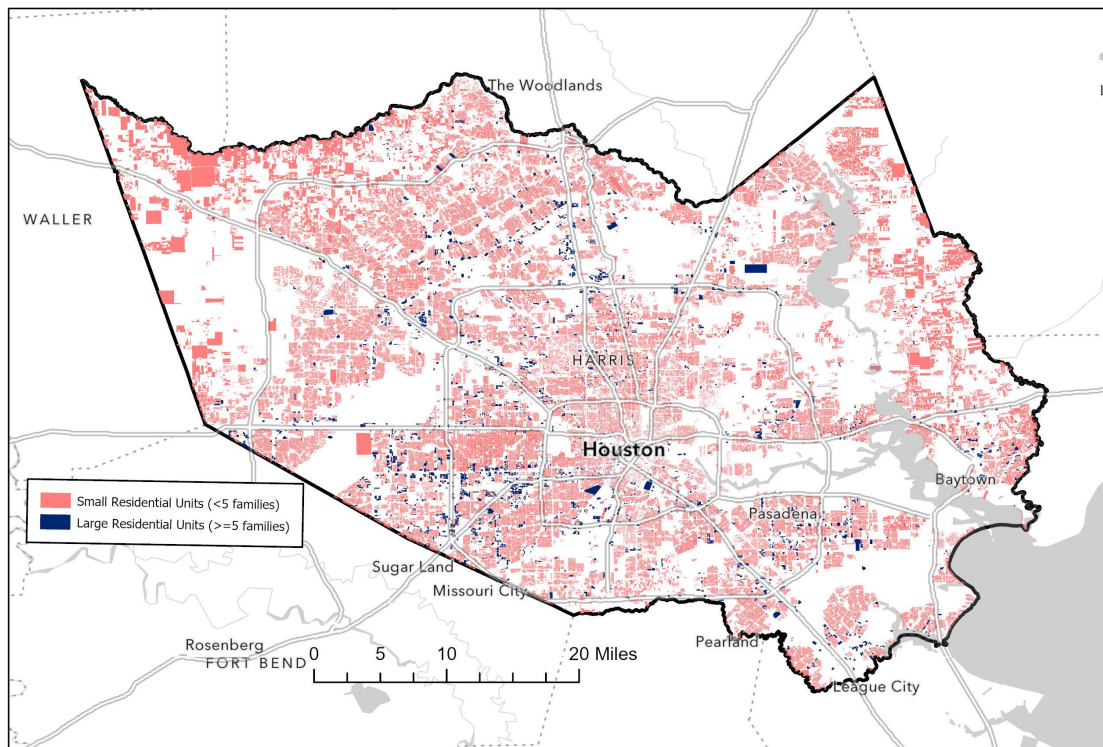

Figure 3 Distribution of parcels across Houston by residential structure size.

3.1.3. Flood Hazard Data

The extent of FEMA's 100-year floodplain was chosen as the indicator for flood hazard extent in the study area (see Figure 2). That extent was retrieved from FEMA's NFHL, which contained polygon features for all of the different floodplain types. The 100-year floodplain feature class was comprised of numerous polygon features. The whole area covered by those features was used as the SFHA layer for delineating flood risk boundaries.

Although FEMA's floodplain maps have been shown to underestimate the potential for flooding in certain areas of Houston (Bass and Bedient, 2018), they remain the current authoritative flood hazard data. This dataset was still useful for this study because while it may underestimate the potential for flooding in some areas, it does not overestimate the potential for flooding in others. All areas within the 100-year floodplain can be considered to be in danger of flooding.

3.2. Research Design

The flood vulnerability index was compiled from three main components: flood risk, social justice, and shelter accessibility. The analysis to create this index was therefore developed in three main stages, one for each major component (Figure 4). It consisted of two elements, one statistical and the other spatial. Social justice variables were assessed through a PCA, and each tract was ranked by social justice score. Flood risk and shelter needs were identified through the application of dasymetric mapping. Weights for the three main index factors were assigned through the implementation of an AHP after various weighting schemes were compared in a RD sensitivity analysis.
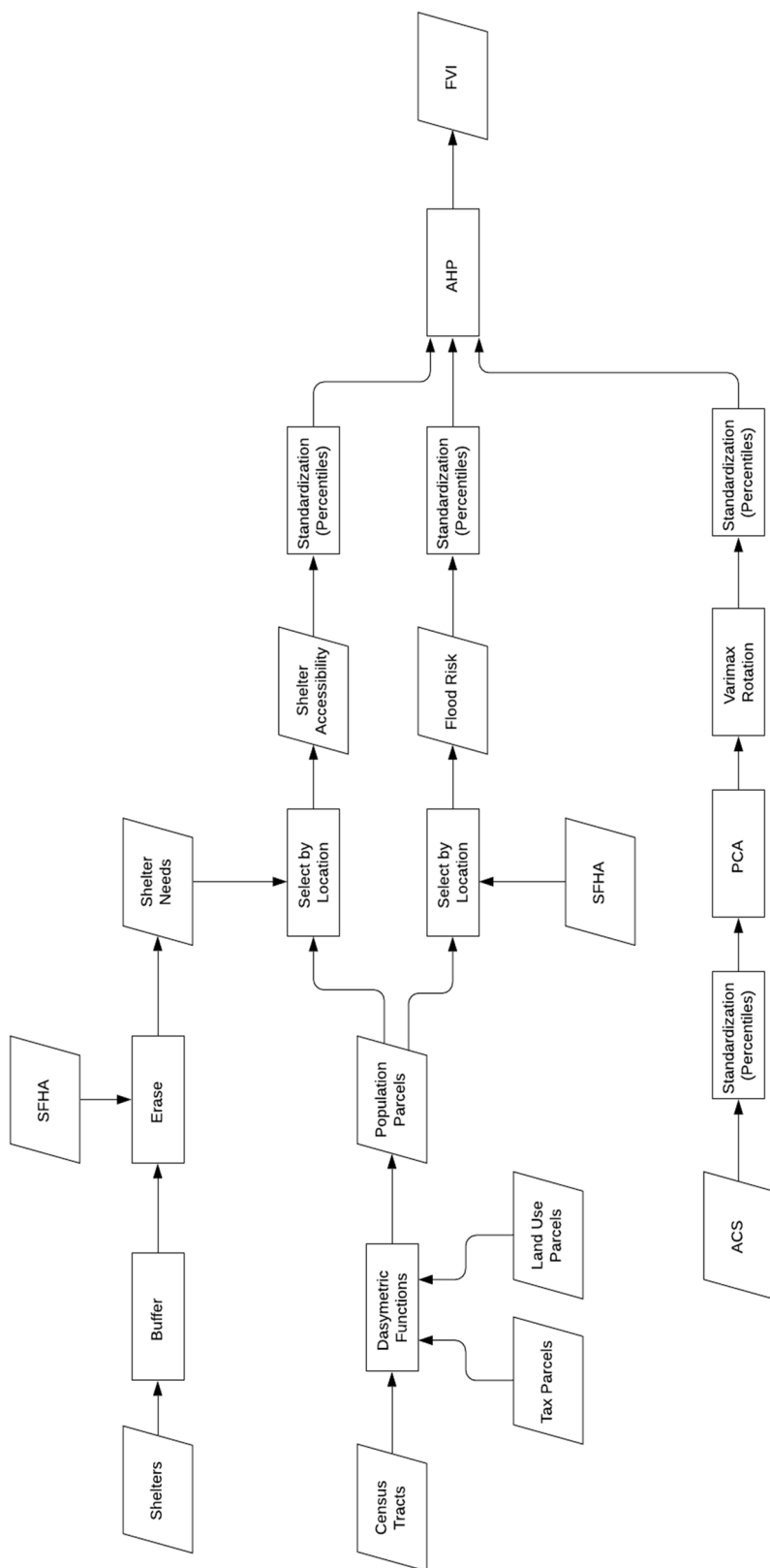
Figure 4 Study workflow.

33

3.2.1. Dasymetric Mapping

Tract-level population data was disaggregated to smaller parcel features in order to generate a more precise estimate of at-risk populations in each tract (Figure 5). This was accomplished through spatial analysis using ArcGIS, with the aid of Python scripting and the ArcPy module. The analysis utilized two different spatial datasets with identical geometry, so a degree of pre-processing was necessary to prepare the data for the analysis. Tax parcels with state classification codes were better suited for accurately identifying smaller residential structures, while land use parcels were better for identifying larger residential structures, such as apartment buildings. Parcels with certain classifications were extracted from each dataset, where the two extracted datasets covered all residential areas in Houston, with no overlapping parcels. In certain cases, multiple single-family residential parcels would occupy the same geometry, being part of the same apartment building in that parcel. Those parcels were all retained in the analysis. Additionally, land use parcels classified as correctional facilities or schools were manually identified, and all parcels containing residents (prisons, on-campus housing) were retained, and the remainder were removed.
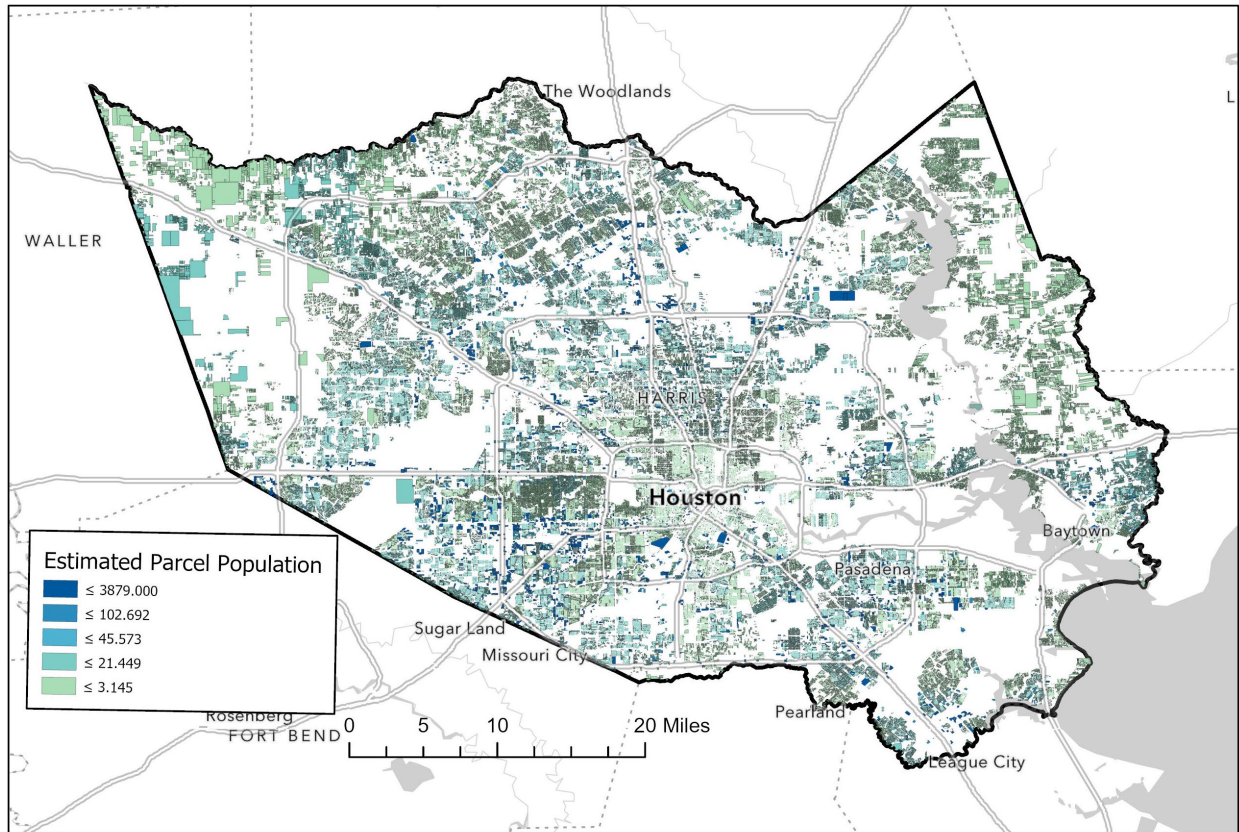
Figure 5 Estimated parcel populations in Harris County.

Specific weights were assigned to each parcel based on its land use or state classification (Table 4). The weights estimate the number of "families" living in each parcel. This is based on the smallest unit of the weighting scheme, A1, or single-family residential. The weights logically increase with two-, three-, and four-family structures, while apartment structures are weighted as double the next smallest structure. Other forms of residence, such as schools, nursing homes, and subsidized housing were weighted by comparison with structures with similar occupancy. To calculate the total number of "families" in each tract, the count of parcels for each parcel code was multiplied by its respective weight and then summed with all other weighted counts. The 2018 population estimate was divided by the family count for each tract to calculate the estimated people per residential unit (PPU). This was then multiplied by the weight of each parcel to calculate parcel populations..

Table 4 Parcel classification Codes and weights.

Code	Description	Weight
A1	Single Family Residence	1
A2	Mobile Home	1
B2	Two-Family Residence	2
B3	Three-Family Residence	3
B4	Four+ Family Residence	4
4209	4-20 Unit Apartment Structure	12
4211	Garden Apartment Structure	24
4212	Mid-Rise Apartment Structure	48
4214	High-Rise Apartment Structure	96
4221	Subsidized Housing	48
4222	Tax Credit Apartments	
4313	Dormitory	48
4316	Nursing Home	48
4613	College/University	96
4670	Jail/Prison	96

With estimated populations calculated for each residential parcel, the process for determining at-risk populations for each census tract was comparatively simple. Each tract had a unique ID that was assigned to all coincident parcels whose geometric centers crossed the tract. All parcels which intersected the SFHA layer were selected, and then all selected populations were summed for each tract ID. The populations for each tract were then ranked by percentile for the final flood risk score.

A similar strategy was applied to the shelter accessibility component of the analysis. First, 1-mile buffers were created around all shelter point features. Portions of the SFHA layer were then removed where they intersected with the shelter buffers. The same selection method that was used to calculate the total at-risk population was then used to determine the number of at-risk people without nearby shelter options, which was then also ranked by percentiles.

3.2.2. Principal Component Analysis

A PCA was performed to mitigate the problem of multicollinearity in the social justice

dataset. The degree of multicollinearity was assessed through the creation of a correlation matrix

for all 12 input variables, and the calculation of the Variance Inflation Factor (VIF) of each

variable. The VIF indicated the variables which were highly correlated with other variables, and

the correlation matrix showed the degree of correlation between individual variables. This

multicollinearity test indicated high multicollinearity among several variables, thus making a

PCA necessary.

The PCA was performed primarily through Python scripting, with the use of several

modules. The Pandas module was utilized for table reading and writing. ACS estimates could be

retrieved from the ACS API through the CensusData module. Specific fields were extracted and,

if necessary, combined to create new variables for the input social justice populations. The result

of this extraction was a Pandas DataFrame containing all of the necessary social justice variables

to be included in the PCA.

The first step of the analysis was to standardize the input data. Standardization allowed

for a more accurate assessment of the variables' relations with each other. Percentile rankings

were chosen because the goal of the index was to identify the largest concentrations of at-risk

populations. The percentile score for each value in reference to its containing field was

calculated using the SciPy module. The NumPy and SciPy modules were then used in

combination to calculate the VIF for each variable. Several VIFs with a value over 5.00 indicated

high multicollinearity, which was confirmed through the creation of the correlation matrix.

Unlike all other elements of this analysis, the correlation matrix was created in Microsoft Excel,

using the Analysis ToolPak. The PCA itself was performed using the Scikit-Learn module.

Before the PCA was performed, the data was scaled using Scikit-Learn's StandardScaler. This

standardized the input data to have a mean of zero and a standard deviation of one. The scaled data was then fit to the object created through the PCA function, transformed, then exported as 12 new components, one for each input variable.

Each new component consisted of composite scores generated from all 12 input variables, with each variable represented by specific loadings of varying weight. The amount of information accounted for by each component varied as well, in the form of explained variance. Each component explained a certain percentage of the total variance within the dataset, and only components over a certain variance threshold were kept. The Kaiser Criterion was applied to the components in order to determine which of them contained an acceptable degree of variance (Cutter, Boruff, and Shirley, 2003). The eigenvalues for each component's covariance matrix were assessed using Python and Scikit-Learn, and components with eigenvalues greater than one were retained, and the remainder discarded. This process reduced the number of components down from 12 to 3, with a combined percent variance explained of 72.84. A Varimax rotation was then applied to the remaining components with the aid of the NumPy module. The Varimax rotation was a useful step, in that it decreased the number of highly correlated variables to each component. This allowed for a more simplified approach to determining factor loadings, which was performed in Excel. The input variables and resulting components were combined in a single table, from which a correlation matrix was created. Factor loadings were then determined by examining the most highly correlated variables with each component. Although every variable had a degree of correlation with each component, the associated correlation coefficients varied from component to component, with certain variables being much more highly correlated than others. The variable with the highest (positive or negative) correlation was then determined to be the "dominant variable" for that component. These dominant variables, along with other

less but still highly correlated variables, were used to assign a name to each component and determine the sign of its associated component score. Components were given names that reflected the commonalities between the dominant variable and other heavily loaded factors. Components with generally positive factor loadings were assigned a positive sign, while those with negative factor loadings were assigned a negative sign. All input variables indicated a heightened likelihood for flood vulnerability, so negative correlation with those variables would therefore indicate a negative contribution to the final social justice score. In the case of this study, the component with the highest explained variance, the Household Characteristics component, was found to generally have a negative correlation with its most heavily loaded factors, so it was assigned a negative sign. The raw social justice scores created from the combined component scores were then ranked by percentile, so they could be standardized and incorporated into the index score with the flood risk and shelter accessibility scores.

3.2.3. Final Index Calculation and Sensitivity Analysis

After dasymetric analysis and PCA were utilized to calculate scores for the three main factors of flood risk, shelter accessibility, and social justice, specific weights were assigned to each factor, and then the weighted factors were combined for the final index score. The determination of weights was a subjective process, which took into consideration the relative importance and urgency implied by each factor in an emergency. Due to this subjectivity, the index was calculated several times with different weighting schemes, in order to determine each factor's effect. The index was first calculated with each factor weighted equally with a third of the final score. It was then calculated three more times, with one "heavy" weight at 50% and the other two at 25%, with a new heavy factor each time.

By using an OAT analysis method, the four index scores created for the sensitivity analysis were regrouped under five new classes: very low (0-10), low (10-25), medium (25-75), high (75-90), and very high (90-100). The frequency of each class was compared across the various scores, as well as the descriptive statistics, including minimum, maximum, mean, mode, and standard deviation. Noticeable differences among the indices were observed through a side-by-side comparison (Figure 6), and further statistical analysis was deemed necessary to complete the sensitivity analysis. The RD methodology for measuring sensitivity was used to evaluate the four index calculations. The relative standard deviation (RSD) was calculated as the ratio of the standard deviation to the mean of the index values. A higher RSD indicated higher variation, and therefore greater model sensitivity to the variable.
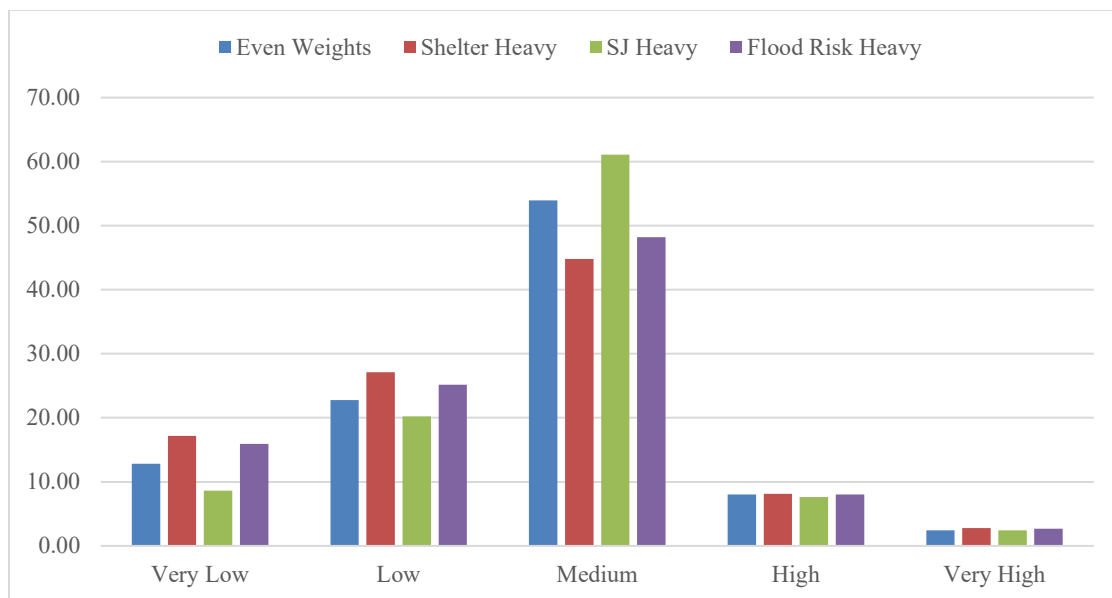
Figure 6 Distribution of index scores created through sensitivity analysis.

The final weights for the analysis were created through the application of the AHP. This was performed through an Excel template available from Business Performance Management Singapore (Goepel, 2013). Through this analysis, the three main factors were compared against each other, and a matrix was created, which showed how important each factor was in

comparison to each other factor. Shelter accessibility was ranked as generally having low

importance, as the relevant data could also be found within the flood risk layer. Flood risk was

determined to have generally high importance, as a flood vulnerability index cannot be created

without some kind of flood hazard data. Social justice was determined to be less important than

flood risk, due to the aggregate nature of the population estimates. However, it was ranked as

more important than shelter accessibility due to the uniqueness of the dataset and the

vulnerability indicators associated with it. The resulting matrix from this evaluation was then

input into the AHP template, which calculated weights based on the normalized eigenvectors of

the input. The FVI score was then computed using those weights.

Chapter 4 Results

The study accomplished its goal of creating a flood vulnerability index based on factors related to flood risk, shelter accessibility, and social justice. The analysis was developed into two scripts, which are available at https://github.com/mawilson10/Houston-FVI, along with descriptive documentation of the processes utilized in the analysis model. The scripts can also be found in Appendix A. The index layer is available in a web map at https://arcg.is/0vuDDW. The results of the PCA and dasymetric analysis are described; the results of the sensitivity analysis and the final index calculation are discussed in this chapter.

4.1. Vulnerability Index

The vulnerability analysis was performed with various spatial and tabular datasets, and scripted entirely in Python, in order to provide a concise record of the exact analysis. Scripting the entire process allowed for continuous updates to be made to the process, which could then simply be re-run, as discoveries were made about the behavior of the data and the relationships between the different datasets. Although the initial goal was to develop a single script which could accept all of the input feature classes and tables, the process was instead broken into two separate scripts, one spatially focused (dasymetric mapping), and the other statistically focused (PCA). This structure is reflective of how the study was developed, with one major phase dedicated to performing the social justice PCA and another phase dedicated to flood risk and shelter analysis. The two scripts could also be re-purposed to new models for different purposes, so they were separated in order to make them more flexible. Both scripts were shared publicly on GitHub.

4.1.1. *Principal Component Analysis*

Prior to the PCA implementation, the input social justice variables were tested for multicollinearity, in order to determine if the PCA would be necessary. This was accomplished through the creation of a correlation matrix, and the calculation of each variable's respective VIF. While a VIF of 1.00 would indicate no multicollinearity, three variables, under 10 years of age, part-time workers, and single parents were found to have VIFs over 5.00, and the female variable had a VIF over 10.00 (Table 5). These relatively high VIFs indicated that those variables might have been highly correlated with several other variables in the dataset. A correlation matrix was then used to confirm the existence of multicollinearity (Table 6). Each of the variables with high VIFs also shared correlation coefficients over 6.00 with multiple other variables. It was therefore highly likely that multicollinearity was present in the dataset, and that certain variables might explain others. A PCA was determined to be necessary to increase the variance within the dataset.

Table 5 VIFs for each social justice variable.

Variable	VIF
Female	10.312
Part-Time Worker	6.132
Single Parent	5.364
Under 10	5.124
Poverty	3.923
Disabled	3.106
Renter	2.938
No Car	2.660
Unemployed	2.578
Poor English	1.592
Public Assistance	1.439
Over 64	1.121

Table 6 Correlation matrix created from social justice variables.

	Female	No Car	Under 10	Over 64	Disability	Poor English	Poverty	Renter	Part-Time	Unem-ployed	Public Assistanc	Single Parent
Female	1.0000	0.2083	0.8512	0.0618	0.7093	0.4435	0.4583	0.4566	0.8961	0.6593	0.4294	0.7279
No Car	0.2083	1.0000	0.2692	-0.1727	0.3085	0.0122	0.6320	0.6779	0.2629	0.3413	0.2300	0.4944
Under 10	0.8512	0.2692	1.0000	0.0257	0.6093	0.2878	0.6104	0.4232	0.7567	0.6260	0.4203	0.7591
Over 64	0.0618	-0.1727	0.0257	1.0000	0.1142	0.0860	-0.1887	-0.1318	0.0310	0.0098	0.0142	-0.0349
Disability	0.7093	0.3085	0.6093	0.1142	1.0000	0.1231	0.4541	0.2859	0.6544	0.6604	0.4659	0.6709
Poor English	0.4435	0.0122	0.2878	0.0860	0.1231	1.0000	0.0298	0.3144	0.4772	0.2108	0.1452	0.2031
Poverty	0.4583	0.6320	0.6104	-0.1887	0.4541	0.0298	1.0000	0.5832	0.4433	0.5597	0.3895	0.7744
Renter	0.4566	0.6779	0.4232	-0.1318	0.2859	0.3144	0.5832	1.0000	0.5087	0.4092	0.3213	0.5810
Part-Time Worker	0.8961	0.2629	0.7567	0.0310	0.6544	0.4772	0.4433	0.5087	1.0000	0.6800	0.4081	0.6765
Unemployed	0.6593	0.3413	0.6260	0.0098	0.6604	0.2108	0.5597	0.4092	0.6800	1.0000	0.4344	0.7067
Public Assistance	0.4294	0.2300	0.4203	0.0142	0.4659	0.1452	0.3895	0.3213	0.4081	0.4344	1.0000	0.5188
Single Parent	0.7279	0.4944	0.7591	-0.0349	0.6709	0.2031	0.7744	0.5810	0.6765	0.7067	0.5188	1.0000

After the input data was scaled, the PCA was performed. The analysis was conducted according to the methodology outlined by Cutter, Boruff, and Shirley (2003). It initially produced 12 components with decreasing percentages of variance explained. The Kaiser Criterion was applied to the results, in which components with eigenvalues less than 1.00 were retained. This reduced the number of components to three. A correlation matrix including both the components and the input variables was then used to determine the components' factor loadings and sign. Each component was assigned a name based on its most heavily loaded input variable, with consideration given to other highly correlated variables. Although no input variables were included which would negatively affect vulnerability, the family characteristics component was found to have a high negative correlation with several variables. The component score sign was therefore reversed. The three component scores were then added together. The combined percent variance explained for all three components was 72.84% (Table 7). The combined score was then standardized by percentile score and incorporated into the vulnerability model.

Table 7 PCA components and dominant variables.

Component	Percent Variance Explained	Dominant Variable	Correlation
Household Characteristics	50.23%	Single Parent	-0.906
Mobility	13.61%	No Car	0.555
Communication	9.00%	Ability to Speak English	0.663

4.1.2. Dasymetric Analysis

Dasymetric mapping was used to disaggregate tract-level populations and assign estimated populations to individual residential parcels. This strategy was similar to CEDS, developed by Maantay and Maroko (2009). At-risk populations were determined for each tract by calculating the sum of populations for parcels intersecting the 100-year floodplain. Populations without access to a shelter were identified as at-risk residents located further than a mile from the nearest shelter. 567 of Harris County's 786 census tracts were found to contain at least some at-risk population, and 463 tracts were found to contain people in need of accessible shelter (Figures 7 and 8). The shelter accessibility score was determined as the number of at-risk people outside of 1-mile shelter buffers. The population was limited to only those in flood risk areas so high scores would not be assigned to tracts with low potential for flooding and therefore low need for shelter. As with the PCA score, the population estimates for both flood risk and shelter accessibility were each standardized by percentile score for inclusion in the final index score.
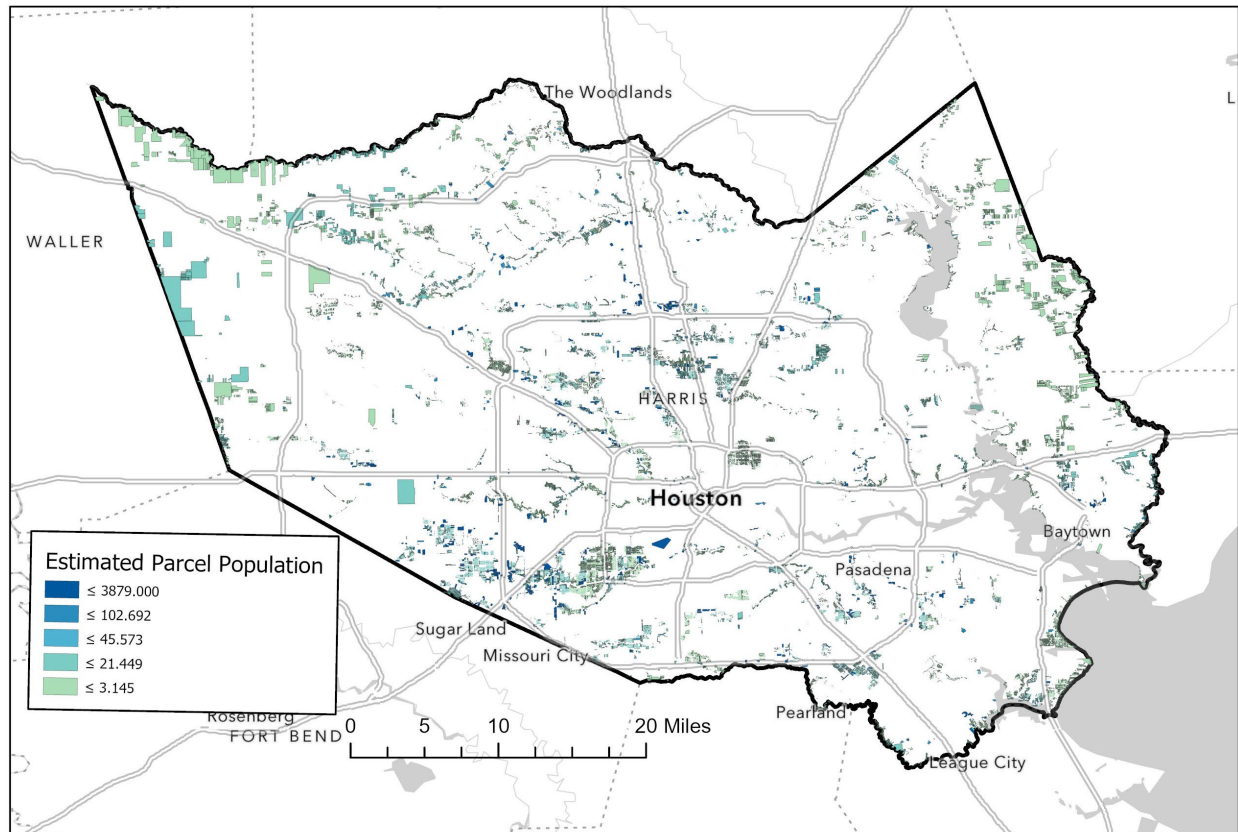
Figure 7 At-risk populations in Harris County.
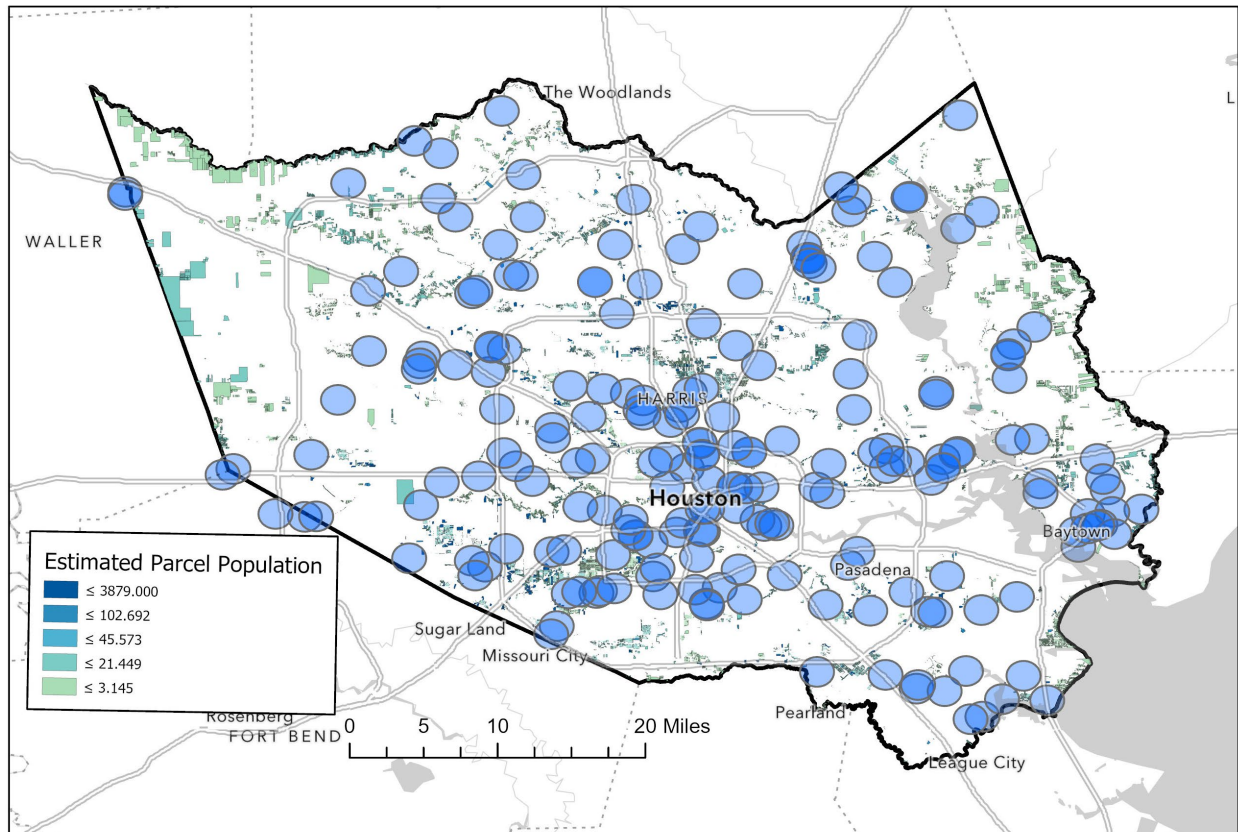
Figure 8 Areas of shelter need in Harris County.

4.2. Factor Score Results

The spatial and statistical analysis produced three significantly different sets of scores for flood risk, shelter accessibility, and social justice. A weighted average of the three scores was used to compute the final vulnerability analysis. The factor scores were mapped and compared, to determine the differences and similarities between their respective distributions across Harris County. The results of that comparative analysis are described in this section.

The flood risk analysis revealed numerous high-risk tracts, that were confined to several creek and bayou floodplains intersecting Harris County (Figure 9). However, high-risk populations were not evenly distributed across all of the Harris County watersheds. Even within individual watersheds, high-risk tracts often occurred in dispersed concentrations. Highly at-risk

tracts were identified in suburban areas south and southeast of Houston, located along Clear

Creek and at the confluence of Clear Creek and Armand Bayou. High risk was indicated in

numerous tracts surrounding Brays Bayou in southwest Houston. High-risk tracts were identified

at several different areas along the Greens, Halls, and White Oak Bayous in northern Houston.

High degrees of risk were also identified in several tracts in suburban areas north and northwest

of Houston, located along portions of Cypress Creek. Low-risk tracts were also found to be

widely distributed across the county, particularly in the downtown area of Houston.
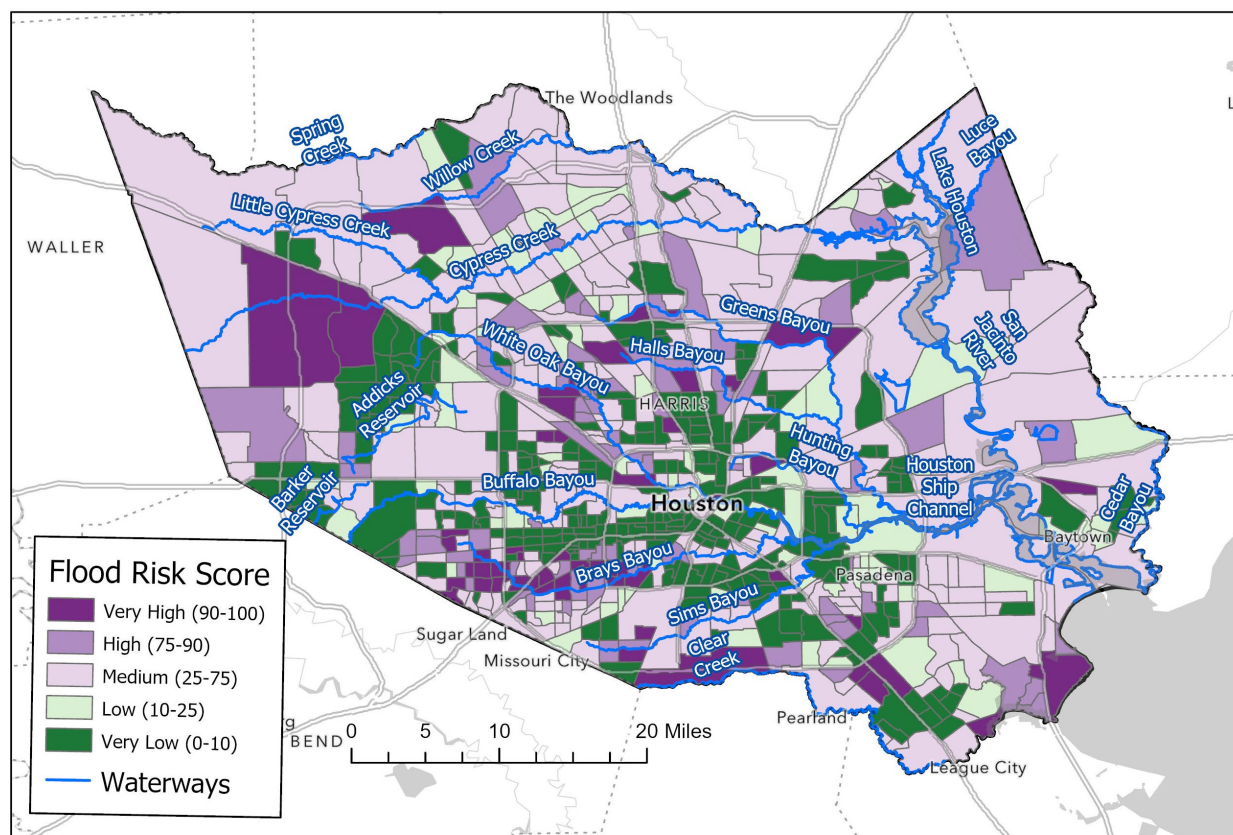
Figure 9 Harris County tracts, ranked by at-risk populations.

The shelter accessibility analysis produced similar results to the flood risk analysis

(Figure 10), as it utilized the 100-year floodplain in addition to shelter service areas for

identifying populous areas without local shelter options. Although similar, a paired t-test

comparing the two sets of scores indicated a significant difference between them. High shelter inaccessibility was identified in many of the areas with high flood risk. However, several high-risk tracts were also found to have low shelter scores, indicating shelter availability for the at-risk populations within them. These low-risk tracts with high shelter accessibility were in the highest frequency along a portion of Brays Bayou in southwest Houston. This area also contained the highest frequency of tracts with high levels of both flood risk and shelter need. Several other high-risk, low-accessibility tracts were also dispersed along White Oak Bayou in northwest Houston. Other high accessibility tracts were sparsely distributed along Sims and Armand Bayous, as well as Clear Creek in the southern Houston area.
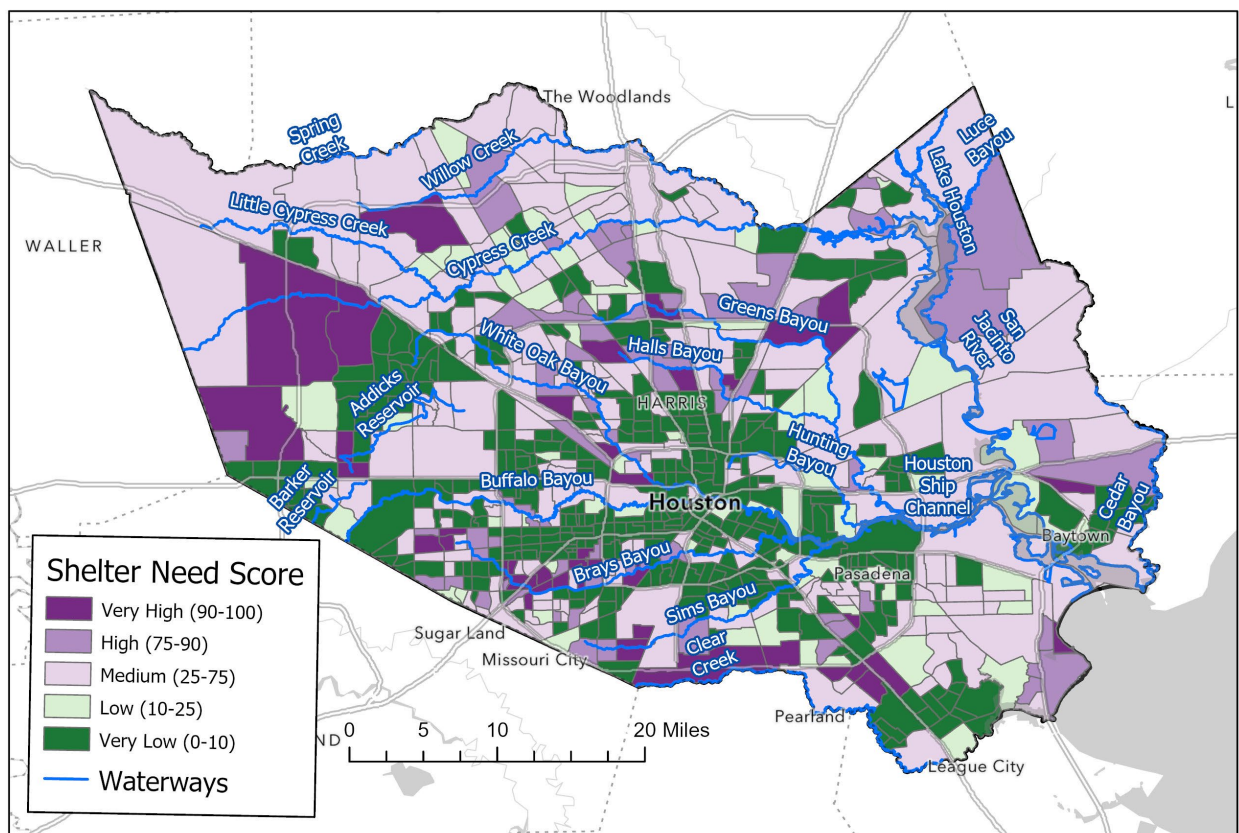
Figure 10 Harris County tracts, ranked by shelter need.

Tracts with high social justice scores were not confined to the areas surrounding Harris County waterways in the same manner as most high-risk and many low-accessibility tracts.

Generally, most of the high-scoring tracts covered large contiguous areas in smaller cities and towns surrounding Houston, as well as portions of northwest and southwest Houston. Of the 197 tracts with social justice scores classified as "high" (over 75), 38 tracts (19.2%) were found to have no flood risk. Although the remaining tracts had some degree of risk, many of those were classified as less than high-risk. However, large clusters of high-risk, highly socially vulnerable tracts were identified in areas surrounding portions of Brays Bayou, Clear Creek, and Cypress Creek.
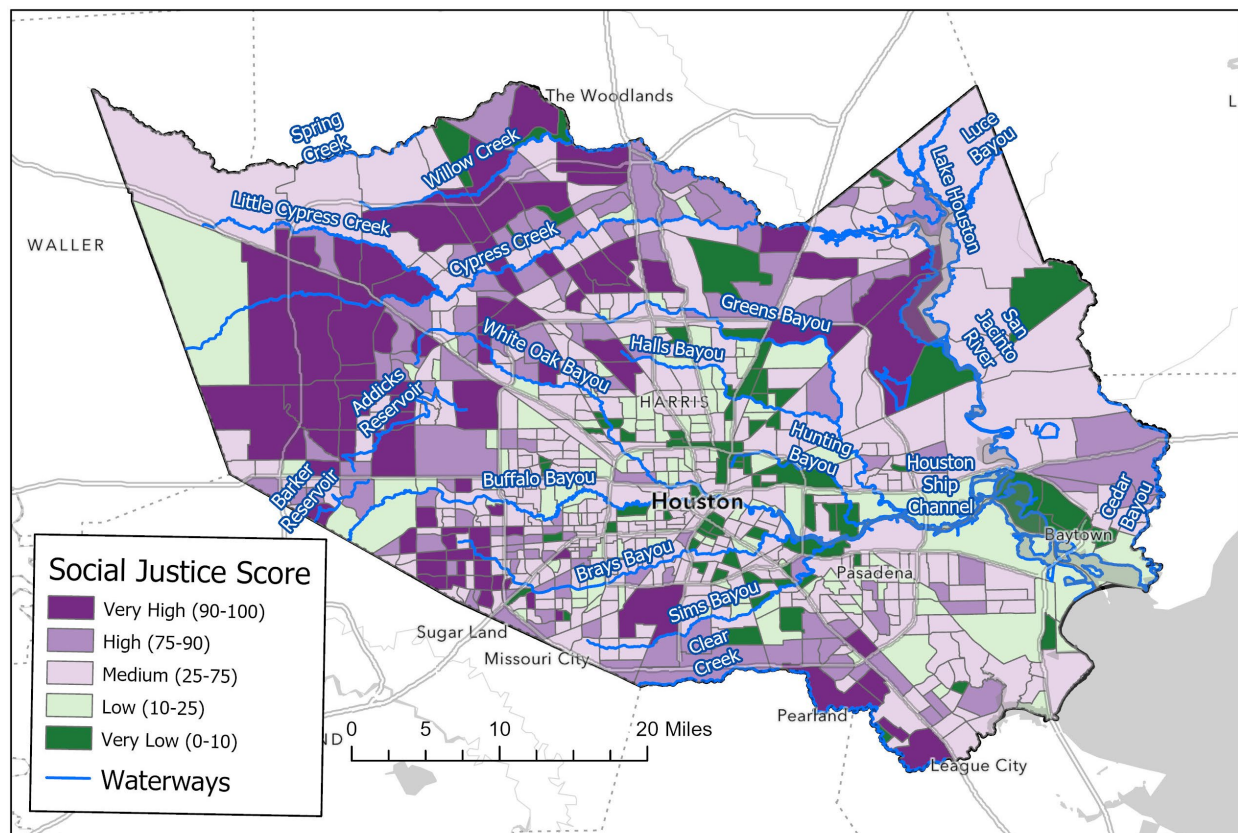
Figure 11 Harris County tracts, ranked by social justice score.

The three factors had particularly high scores in several areas of Harris County. These areas included portions of the Brays Bayou, Greens Bayou, Clear Creek, and Cypress Creek watersheds. These areas were therefore expected to be the highest-ranked areas in the final index. Tracts with low shelter need scores were expected to reduce the degree of vulnerability in certain

areas, particularly around Brays Bayou in southwest Houston. Several NSS shelters were located in or near high-risk regions of the bayou's watershed. Although several tracts with high social vulnerability scores also contained high flood risk and shelter-need scores, social justice populations were also high in tracts with little to no flood risk. The low degrees of shelter-need and flood risk in those tracts were expected to offset those high social vulnerability scores in the final index.

4.3. FVI Calculation and Sensitivity Analysis

Before the final FVI score was computed, a sensitivity analysis was performed to assess the stability of the index when subjected to different weighting systems. The FVI score was calculated as a weighted average of the three main input variables. The results of the sensitivity analysis were used to inform weight selection for each variable. Multiple index scores were calculated with varying weights and inputs to test the effect of certain variables on the final score. Four different weighting schemes were used. The first index score was calculated simply as the mean of the three scores. The remaining three index scores had one factor weighted at 50%, as the "heavy" factor, and the other two at 25% (Table 7). The RD sensitivity analysis was applied to the index scores created through those calculations (Table 8). The shelter-heavy index was found to have the highest RSD, and the flood-risk-heavy index had a similarly high RSD. The RSDs of both indices were significantly higher than that for the social justice-heavy index. This indicated that the model was particularly sensitive to those two variables.

Table 8 Factor weights for each sensitivity analysis weighting scheme.

Scheme Name	Weights		
	Flood Risk	Shelter Accessibility	Social Justice
Even Weights	33.333	33.333	33.333
Flood Risk-Heavy	50	25	25
Shelter Need-Heavy	25	50	25
Social Justice-Heavy	25	25	50

Table 9 Descriptive statistics for FV scores generated from the four weighting schemes.

Weighting Scheme	Standard Dev.	Mean	Median	Maximum	Minimum	RSD
Even Weights	24.985	35.574	34.261	97.99	0.085	70.2398
Flood Risk-Heavy	26.309	37.961	33.583	98.449	0.064	69.30534
Shelter Need-Heavy	26.329	36.307	29.409	98.493	0.064	72.5177
Social Justice-Heavy	23.669	41.444	40.081	97.03	0.127	57.1108

The RD analysis returned similar RSD values for even, flood risk-heavy, and shelter need-heavy index scores. Descriptive statistics across the board were generally similar for the three scores. The social justice-heavy index score was found to have the highest mean, and the lowest standard deviation, resulting in the lowest RSD. Although this would appear to indicate that the model was more sensitive to the social justice input data, the results of the RD analysis imply that it was more sensitive to the flood risk and shelter need inputs. The level of sensitivity was assessed through the RSDs, which indicated greater variation when more weight was

assigned to either the flood risk or shelter need scores. This is likely since the two datasets are related, and that shelter need areas are contained within flood risk areas. Heavily weighting one dataset without significantly reducing the contribution of the other could disproportionally amplify their combined effect. Therefore, the analysis could be interpreted as indicating model sensitivity to both variables. This discovery led to an assessment of the correlation between the two variables. Although a comparison matrix revealed that the two variables were highly correlated, a paired t-test returned a p-value well below zero, indicating that there was a significant difference between the two datasets.

Although the sensitivity analysis did produce varying results, the distribution of the index scores was relatively similar across the different index calculations (see Figure 6). All four scores followed the same pattern in which there were more tracts designated as very low or low vulnerability than high or very high, with the majority of tracts classified as medium vulnerability. This indicated that although the different score calculations varied, altering the weights of one factor would not drastically alter the vulnerability score. No scores changed more than one level of vulnerability when compared against different weighting schemes. Although sensitivity was identified within the analysis, it did not appear to drastically alter the nature of the output data.

Due to the apparent impact of weight variation on the final score, an AHP was implemented to calculate weights through a statistical process. Through this methodology, the high sensitivity to flood risk and shelter accessibility was addressed. In the comparison matrix, social justice was determined to be three times more important than shelter needs, and half as important as flood risk (Figure 12). Flood risk was determined to be three times more important than shelter need. Shelters are an important element of flood planning and response, but they

cannot completely mitigate the effects of flooding. Shelters in highly populated areas would likely not have the capacity to support every local resident if all of their homes flooded. Therefore, shelter accessibility's effect was deemed less significant than the other two factors. The flood risk factor was critical to the FVI, in that it demarcated the areas where flooding could occur. Without the flood risk input, the model would not be capable of accurately identifying areas of vulnerability. Social justice was also considered to be particularly important, but not as important as flood risk. Some tracts with high social justice scores did not contain any at-risk residents. Although these areas could still be affected by other problems, such as power outages or wind damage, they are not in danger of taking on significant flood damage. The social justice score was therefore given a lower importance than flood risk so as not to overstate the vulnerability of people not living at risk of flooding. Some tracts with high social justice scores could be in no danger of flooding. This was taken into consideration when determining the factor's relative importance. With the assigned importance values, the AHP returned weight values of 33.3, 14.0, and 52.8 for social justice, shelter accessibility, and flood risk, respectively (Figure 13). The flood risk weight was reduced to 52.7 so the combined weights added up to 100. The final flood vulnerability index score was calculated using these weights. Summary statistics were calculated for the final scores, which revealed that the FVI scores had a larger mean and smaller standard deviation than all previous index calculations, except for the social-justice heavy calculation (Table 9). Although the index had been calculated with a heavy weight assigned to the flood risk factor, its RSD was significantly lower than those of the previous non-social justice-heavy index scores.

Figure 12 AHP comparison matrix created in Excel worksheet developed by Goepel (2013).

Figure 13 Results of AHP analysis.

Table 10 FVI descriptive statistics.

Standard Dev.	25.324
Mean	39.845
Median	36.471
Maximum	97.958
Minimum	0.085
RSD	63.556

The FVI layer was examined using Houston waterways as a reference. Certain patterns that were observed in the examination of each of the three index factors were also present in the final index layer (Figure 14). Tracts in south and southwest Harris County, around Clear Creek and Brays Bayou, were found to be highly vulnerable. Tracts near Greens and Halls Bayous in

north Houston were also found to contain exceptionally high levels of vulnerability, as were tracts coincident with Cypress creek in northwest Harris County. These highly vulnerable tracts were consistently identified as the highest-scoring tracts for each of the three FVI factors. In instances where one variable score was significantly different than the other two variables, the index results do not appear to significantly skew towards the outlying variable (see Figure 6). This is especially apparent in tracts with high social vulnerability scores but comparatively low shelter accessibility and flood risk scores. These tracts were not highly ranked in the final index, as it was likely that only a fraction of the potentially vulnerable social justice populations were in danger of flooding. In several tracts near Brays Bayou, low shelter need scores and moderate social justice scores offset high flood risk scores, producing a moderate index score. Social justice was found to have a significant impact on several tracts as well, in which a high social justice score combined with moderate or low flood risk and shelter accessibility scores produced a high index score. These tracts were not focused in any particular watershed, but were sparsely scattered across the county.
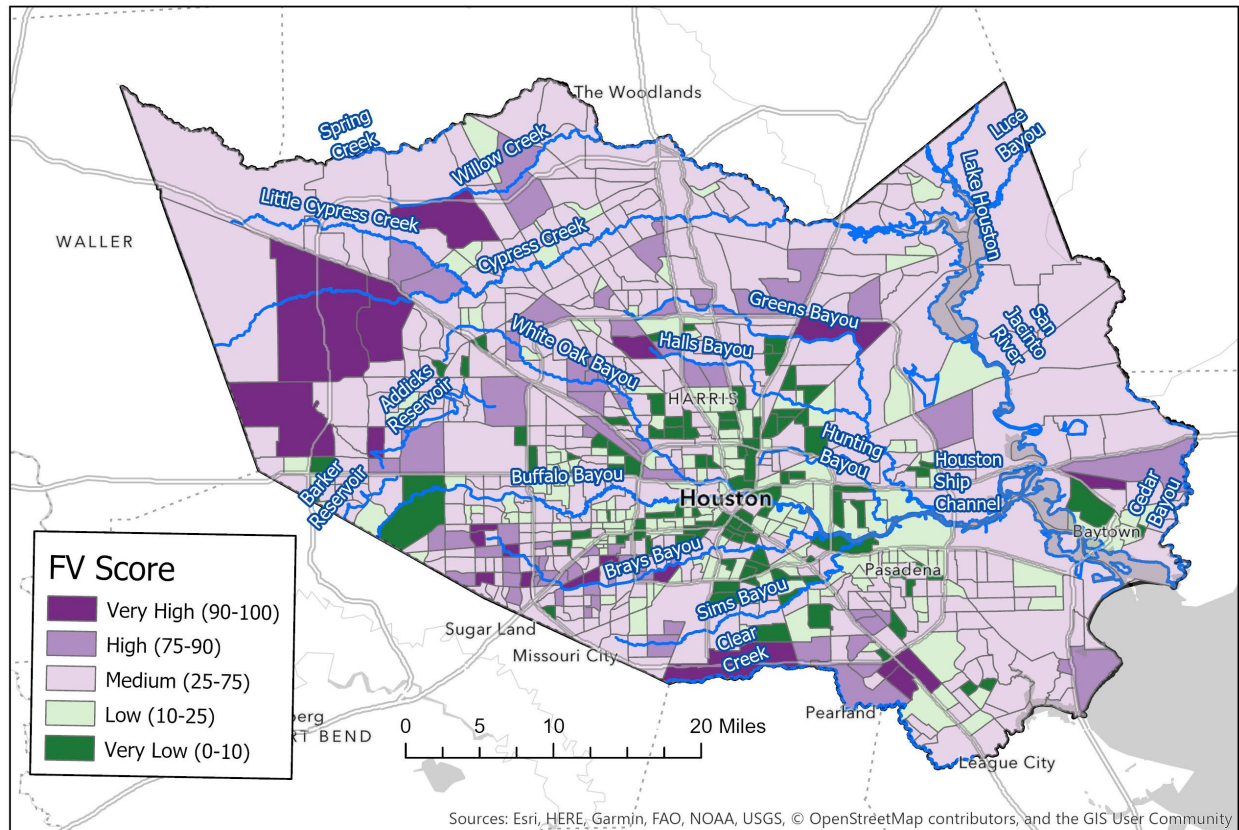
Figure 14 Final FVI layer with Harris County waterways.

The tracts with the highest FV scores were compared against coincident neighborhood boundaries in the City of Houston's Super Neighborhood layer (Figure 15). Super neighborhoods are divisions within Houston that represent distinctive communities, each with their own unique identity. These regions are often bound by major physical features such as roadways or waterways, and are comprised of interrelated commercial and residential areas (Zhang et al., 2015). The tracts with the highest vulnerability scores were found to lie within several super neighborhoods that were heavily impacted by storms from 2015 to 2019. Among the high-vulnerability neighborhoods around Braes Bayou are Alief and Sharpstown. These are two low-income neighborhoods that have been severely impacted by flooding. Alief, Sharpstown, and South Belt/Ellington, which also suffered considerable flooding during Hurricane Harvey, each

contains three tracts classified as having "very high" vulnerability. Alief and Sharpstown also each include eight tracts ranked as "high". No other neighborhoods contained as many tracts ranked as high or very high. Greater Greenspoint, which was among the neighborhoods with the highest mean vulnerability score, did not contain any tracts in the very high range. However, it includes five tracts ranked as high. In total, 22 of the 88 super neighborhoods contained at least one "very high" tract, and 41 contained at least one "high" tract. The full table ranking all of the super neighborhoods by average FV score can be found in Appendix B.

Figure 15 Houston Super Neighborhoods.
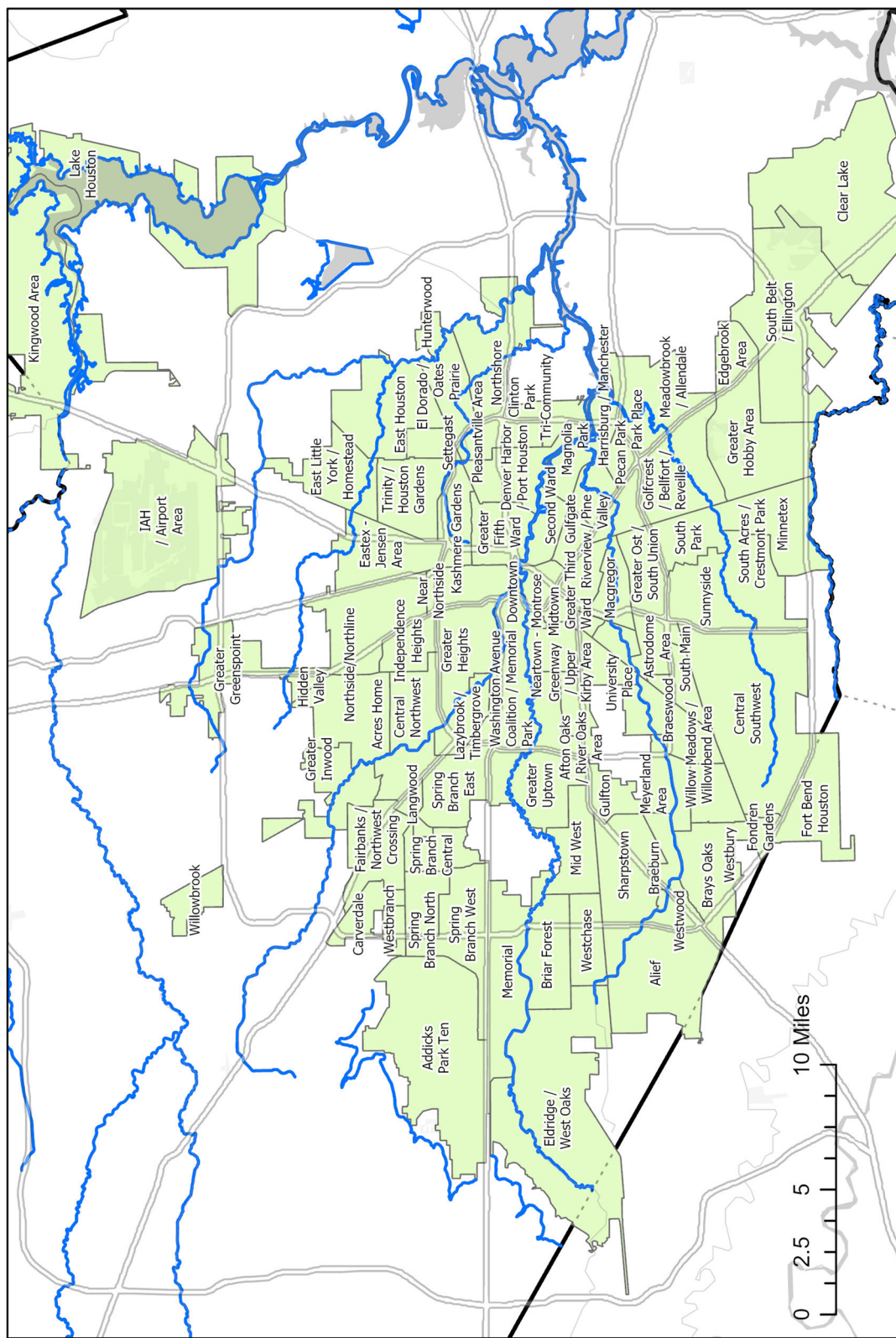
The super neighborhood layer only covers certain areas of Houston, and does not include other areas of Harris County that have been suffered from similar flooding. Further examination of the layer in relation to waterways and roads revealed additional highly vulnerable and historically impacted areas, including Baytown near the Bay of Galveston in Eastern Harris County, and the community of Cypress, intersected by Cypress Creek in northwest Harris County. This examination of the FVI in the context of super neighborhoods and other community boundaries was used to verify the accuracy of the vulnerability model.

Chapter 5 Discussion and Conclusions

This study developed an FVI for Harris County, Texas, through a statistically driven multi-criteria regression analysis. The index was created as a weighted average of three primary factors: flood risk, shelter accessibility, and social justice. The goal of this study was to identify areas of Harris County, where large concentrations of socioeconomically marginalized people lived in danger of flooding. Such people can be considered vulnerable, as they are more likely to suffer and struggle to recover from a flood. The dasymetric analysis was used to increase the spatial granularity of population data for more precise identification of at-risk populations. PCA was utilized to mitigate the effect of multicollinearity among the twelve social justice populations. The final index weights were calculated through the application of an AHP. The analysis expected that the FVI would identify vulnerable tracts located in historically impacted areas, indicating the index's accuracy. Comparison of the FVI with local neighborhood boundaries indicated that the tracts with the highest FV scores were located in low-income neighborhoods that had been repeatedly and severely affected by flooding from 2015 to 2019.

This chapter reviews and interprets the results for each of the three primary factor scores, as well as the calculation of the final FVI score. The implications of these results are described, particularly regarding how the different factors contribute to the FVI and their accuracy in identifying vulnerable areas. The benefits of scripting this analysis in Python are also explained. The study and its results are placed in the greater context of relevant existing research, from which this analysis' methods were derived. The potential for expansion of this research through further examination of flood risk and vulnerability in the Houston area is also explained.

5.1. Study Findings

The FVI analysis revealed the nature of flood vulnerability across Harris County. The comparison of the three contributing factors with Houston waterways identified certain patterns of distribution that were also apparent in the index layer. The final index was computed as a weighted average of the primary factors. A sensitivity analysis was performed to assess each factor's effect on the index score. The factor weights were computed through an AHP, in which the three factors were compared based on their relative importance to flood vulnerability. The results of the sensitivity analysis were also used to inform that comparison. The implications and meanings of the FVI and sensitivity analyses are discussed in this chapter.

5.1.1. Index and Individual Factor Results

The analysis revealed similar, but varying distributions for the three primary factors across the county. Many flood risk and shelter accessibility scores were similarly ranked for the same census tract. Large social justice populations were also identified in many areas with high flood risk, although several tracts with high social vulnerability were also located outside of flood risk boundaries. The similarities between these three factors resulted in an FVI in which a single factor seldom heavily influenced the scores. However, individual factors were still capable of making meaningful contributions to the final analysis results in certain areas.

Tracts with high flood risk scores were generally clustered along certain portions of several Houston area waterways (see Figure 9). These clustered tracts were often located in areas that had previously been severely affected by flooding. This dispersed distribution of flood risk is indicative of the varying degrees of flood management and at-risk housing density throughout the county. Large, affordable, but flood-prone multi-unit housing structures such as those in Greenspoint contributed to disproportionate levels of risk in several low-income areas (Miller

and Goodman, 2019). The area surrounding Brays Bayou, which includes the neighborhoods of Alief and Sharpstown, was found to contain numerous high-risk tracts. This pattern implies that further flood mitigation improvements to the bayou watershed are necessary to reduce the potential floodplain extent and protect at-risk residents. The current ongoing improvements to Brays Bayou have been demonstrated to contain flooding within the 100-year floodplain (Bass et al., 2017). However, residents living within the 100-year floodplain are still at risk. These at-risk residents will continue to suffer from flooding until more considerable efforts in either flood management or re-housing are made.

As with the flood risk factor, shelter accessibility scores were also confined to the 100-year floodplain (see Figure 10). The difference between the two sets of scores lies in the incorporation of 1-mile radius shelter service areas to the accessibility analysis. Naturally, the similarities between the two analyses led to a correlation between them. These two scores were not excessively correlated, however, as they were found to have some significant differences. The association between shelter needs and flood risk indicates that at-risk populations in Houston lack local shelter options. This correlation will decrease if the shelter needs populations are reduced through the establishment of additional shelters in or around high-risk areas. This similarity between flood risk and shelter needs is reflective of the shelter shortage exposed by Hurricane Harvey, during which many residents in inundated regions were unable to find shelter (Haynie et al., 2019). Substantial flooding, such as that inflicted by Hurricane Harvey, can make roads impassable, and impact mid- to long-range travel. More significant numbers of shelters in areas of high risk would allow evacuees to quickly locate a local shelter with minimal exposure to the elements.

The distribution of social justice populations differed from those of flood risk and shelter accessibility. The greatest numbers of socially vulnerable people were found to be primarily located in smaller cities and suburban areas on the outskirts of Houston (see Figure 11). Although a number of these tracts were within the vicinity of the 100-year floodplain, others were in areas with no apparent flood risk. This pattern is indicative of the lower cost of living in less urbanized areas further from the city. This distribution supports previous findings that indicated that there was no strong positive correlation between social vulnerability and flood risk in Houston (Castles, 2018). This does not mean that there are not significant numbers of socially vulnerable people living at risk of flooding, however. Clusters of tracts with high social vulnerability were identified around several waterways, including Brays Bayou, White Oak Bayou, Cypress Creek, and Clear Creek. These results indicate that social vulnerability to flooding varies across Harris County and that social vulnerability alone cannot be used to identify areas of high flood vulnerability. Instead, the social vulnerability factor was used for the purpose described by Cutter, Boruff, and Shirley (2003): as an indicator of areas where the local population is most likely to suffer disproportionately in the event of a natural disaster. Highly socially vulnerable people within the 100-year floodplain were considered to be especially susceptible to flooding. In areas of high risk and shelter need, the inclusion of social justice helped identify specific tracts within those areas that would likely be disproportionately affected by flooding. Areas with high social vulnerability, that were outside of the 100-year floodplain, were also useful for the FVI analysis. Low-risk tracts are less susceptible to flooding, but the residents within them could still suffer from other indirect problems, such as loss of power or discontinuation of vital services. The inclusion of social vulnerability added an extra dimension to the FVI, resulting in a more comprehensive and descriptive index.

The final index results were reflective of the three individual factor scores. The highest-ranked tracts in the FVI were highly ranked for each of the three factors. Each factor was found to make a meaningful contribution to the FV score. The flood risk score ranked tracts by the estimated number of people living in the 100-year floodplain. This identified the most populous at-risk areas in Harris County. The shelter accessibility factor revealed populated areas that required local shelter options. Social justice was used to identify high-risk tracts where the residents were most likely to suffer disproportionally. All three of these flood vulnerability elements had noticeable effects on the final FV score.

The three factors effectively counterbalanced each other in the tracts where one factor score differed considerably from the other two. This is evident in areas of high social vulnerability, but low risk and shelter need. In this case, the social justice scores' effect was significantly mitigated by the two lower scores, but still reflected in the final score. Shelter accessibility also had a noticeable impact on the final score, particularly in high-risk areas that lay within shelter service areas. Shelter accessibility had a lower factor weight than flood risk and social justice due to uncertainties regarding capacity and availability. However, in the tracts where shelter need was significantly lower than the other two factors, it did moderately affect the FV score. These interactions between the three main factors indicated balance within the analysis, in which each factor made a meaningful contribution that was reflective of its importance during a flood.

The spatial results of the analysis revealed several large, contiguous areas of high vulnerability in Harris County. These areas were generally located in the vicinity of several major waterways in the county (see Figure 14). Areas surrounding portions of Brays Bayou, Clear Creek, and Cypress Creek contained numerous highly ranked tracts. Smaller numbers of

high-vulnerability tracts were dispersed along portions of several other watersheds. This distribution of vulnerability indicates that flooding is a widespread problem for residents across Harris County. The challenges faced by planners and responders are evident in these results, as a county-wide flood would require carefully calculated distribution of limited resources and aid across the entire county. However, by identifying areas of high vulnerability, the FVI can be used to inform that distribution of resources.

Additional analysis through the comparison of the FVI with Houston Super Neighborhood boundaries further validated the analysis results (see Appendix B). Large numbers of highly-ranked tracts were located in three neighborhoods that had previously been severely affected by flooding: Alief, Sharpstown, and Greater Greenspoint. Alief and Sharpstown were situated within the large concentration of high vulnerability tracts around Brays Bayou, and as such, contained numerous tracts ranked as high or very high vulnerability. In contrast, the high-vulnerability tracts associated with Greenspoint were part of a dispersed series of tracts along Greens Bayou. The numerous high-ranking tracts within and near to these neighborhoods indicate that this analysis can accurately identify areas of flood vulnerability.

The comparison of the FVI with super neighborhood boundaries could be useful to both flood responders and potentially at-risk residents. The general boundaries of these areas are known by many Houston residents, who could use them to gain an understanding of vulnerability within their communities. This understanding could inform their own preparation for potential floods. The history of flooding in many neighborhoods is not as well-documented as it is in especially vulnerable areas such as Alief, Sharpstown, and Greenspoint. This comparison could help to direct attention and awareness to less expected areas of vulnerability throughout Houston.

5.1.2. Sensitivity Analysis and AHP

The sensitivity analysis revealed the need for a logical weighting system for the final index calculation. It also informed the comparative analysis that was used for that weighting system. The FVI was not created using any of the four schemes used in the sensitivity analysis, but the combined results were used to inform the comparison matrix of the AHP. The sensitivity analysis indicated that the model was found to be highly sensitive to both the flood risk and shelter accessibility scores due to their correlation with each other. Based on the results of this sensitivity analysis, shelter accessibility was determined to be significantly less important than the other two factors. This ranking is consistent with shelter accessibility's role in flood management, in which it can be a mitigating factor in heavily affected areas, but does not address all of the problems associated with flooding. Flood risk, social justice, and shelter accessibility were therefore ranked in that order for the AHP weight calculation.

Although the sensitivity analysis indicated that weight changes did not drastically alter the output FV scores, an AHP was still deemed necessary to calculate the final weights due to the three factors' varying degrees of importance. The AHP was applied to add a degree of statistical significance to the subjective process of weight assignment. The resulting weights were reflective of the purpose of this analysis. Flood risk accounted for over half of the FV score, as it was the primary indicator of potential vulnerability. Social justice accounted for nearly a third of the score, due to its potential to amplify vulnerability, especially in areas of flood risk. Shelter accessibility's weight of only fourteen percent reflected its partial contribution to flood response and recovery. The varying weights still allowed each factor to make a noticeable impact on the final index score. The FVI conclusively showed the combined effects of the three factors.

5.2. Advantages of Python Scripting

The most significant benefit of Python scripting for this study was that the analysis model could easily be re-run repeatedly, as new revisions were made to the input data. This advantage was especially evident with the application of the dasymetric mapping model. An initial analysis of the land use and tax parcel datasets identified several different residential structure codes which could be included in the model. The initial application of the model revealed extremely high parcel populations in some tracts and a complete absence of residential tracts in others. Land use parcels classified as universities, prisons, and tax credit apartments had not been included in the initial list of residential codes, which caused results to be skewed in certain tracts, and for others to return null values. With a Python script already developed, the list of residential parcel codes and weights could be updated with additional codes when necessary. Weights could also be easily altered to create a balanced population density map. This flexibility was a useful advantage during the application of the PCA as well, as it allowed for new variables to be easily added to the social justice dataset as they were discovered. The ability to re-run the script allowed for assessment of the data throughout the development of the model, without minimal time lost re-applying the analysis as updates to the process were made with the discovery of new information.

Python scripting allowed multiple different types of processes to be integrated into a single workflow. Six different modules were used to perform the FVI analysis. Scripting was used to download data, manipulate tables, and perform spatial and statistical analyses. A variety of tasks, which would usually require multiple different types of software, were linked together in a script which could complete all the necessary tasks and create a spatial representation through ArcGIS. The benefit of concatenating processes through Python does have its limits,

however. The dasymetric model and the PCA script were initially intended to be combined into a single index model. That model was deemed to be impractical, as it would have combined two different types of analyses with numerous differing inputs, limiting the flexibility of the FVI analysis. The two scripts were thus developed separately, although their outputs were all combined in the final step of the analysis. The separation of the FVI model into two different elements, spatial and statistical, is reflective of the manner in which the analysis was conducted. The PCA and dasymetric scripts were completed in two entirely different stages, each requiring a different kind of analytical thinking and knowledge of very different script libraries.

Through Python, the entire analysis was able to be mapped out as a process which would always produce the same results with the same input data, but could also produce new results, with different input data. The two scripts developed through this process could be utilized for purposes other than FVI calculation. PCA can be implemented in any linear regression analysis, and the script created through this study has the flexibility to accept new input data. The dasymetric mapping model is much more specifically geared towards Houston's unique parcel classification methods, but with some modifications to the source code, the model could be reconfigured to be applied with any kind of classified parcel data, using the same processes and calculations used to create the FVI. The PCA and dasymetric analysis methods designed for this study could also be applied in other studies. Therefore, the scripts and detailed documentation were uploaded and shared on GitHub. The PCA script is a concise methodology for implementing the analysis, and can be performed on any table of values. There is likely a multitude of uses for it, which could be helpful to future researchers as they seek to understand the factors which control their world. The dasymetric mapping model script was also shared with the intention that other spatial analysts could apply the methodology and even improve it. Python

scripting allowed for a definitive record of this study's analysis to be created in a standardized, logical language. Every method and action can be repeated, reapplied, scrutinized, and modified.

5.3. Study Limitations

Although the FVI analysis was successful in identifying areas of flood vulnerability in Harris County, and its results were demonstrated to identify historically vulnerable neighborhoods accurately, there remains room for improvement in the analysis. Issues such as data quality and the sophistication of the analysis can increase the uncertainty of the analysis' results. Further improvements to both the input data and the analytical process can produce more accurate results.

An index can only be as accurate as the data used to create it. This is made especially clear by the FEMA floodplain data used to create the FVI. Areas of Houston that were heavily impacted by flooding during Hurricane Harvey were not located on FEMA's floodplain maps. Using the SFHA layer as a flood risk input, therefore, produced an index that likely underestimated vulnerability in those areas. More accurate flood risk boundary data would significantly elevate this analysis. As such data is not yet available, the current floodplain maps were used as the best possible source for flood risk extent. When FEMA completes their updated floodplain maps for the Houston area, the model should be re-applied with that updated data. The results would likely differ significantly from those described in this report, identifying areas of high vulnerability not present on the previous floodplain maps.

The shelter accessibility portion of the analysis could potentially be refined for more precise results. A buffer analysis was used, as at the time of the analysis, ArcGIS Pro did not have a tool for creating a network dataset from raw road feature data. With the entire analysis performed through ArcGIS Pro, network analysis would not have been practical for building a

cohesive, scripted model. A tool to create a network dataset was included with the release of ArcGIS Pro 2.5. The creation of a network dataset and its incorporation into the FVI calculation would further enhance the analysis and add increased precision to the results. This methodology could also increase the variation between flood risk and shelter need populations.

The dasymetric mapping element of the analysis was performed by assigning specific weights to each parcel classification, reclassifying them by estimated population density. The weights were assigned using a logical process, and overall were effective at estimating and mapping urban density, but other strategies could likely be devised which better estimate parcel populations. A sliding weighting scale, which determines parcel weight by factoring in tract population as well as the number and different types of parcels in each tract, could potentially provide more accurate parcel-level population estimates.

5.4. Further Research

There is great potential for continued work on the subject of flood vulnerability and the analytical methods associated with it. Flood risk is a critical issue in Houston and many other parts of the world, and it appears there is an ever-growing number of challenges associated with it. Increased research into spatial methods for identifying and quantifying vulnerability could allow for areas of need to be pinpointed, and for resources to be distributed accordingly.

Houston's flood vulnerability landscape will likely change significantly in the years and decades to come. Major projects are currently underway to mitigate flooding in several high-risk areas of the city (Lynn, 2017). As the city continues to grow, new challenges regarding flooding will likely arise as well. Emergency planners will need to adapt to this ongoing change. The data used to create this index could be obsolete by the time the next major flood impacts the city. For this reason, the analysis was developed as a Python script, that could be re-applied with updated

data. The script itself could also be changed to incorporate new methodologies or introduce new vulnerability factors. This flexibility allows for continuous refining and testing of the analysis.

The analysis described in this study utilized the NFHL layer to identify the areas of flood risk. Although this layer effectively identified high-risk areas for the whole of Harris County, there are other potential options for demarcating flood risk. The analysis script is equipped to accept any polygon features as flood risk areas, so the analysis could be applied with other flood risk representations. Predictive flood modelling, based on factors such as urban land use and storm data, could provide a more accurate representation of potential flooding. The analysis developed by Gori et al. (2017) delineated flood risk based on future land use projections. The results of their study could be incorporated into this study, to examine how flood vulnerability in Harris County will change over time. Other predictive models could be used to gain a more precise understanding of flood risk in specific areas. Harris County is too large of an area to be evenly affected by every flood. Bass and Bedient (2018) developed an analysis that delineated flood risk based on potential storm surge and rainfall. With appropriate storm data, their analysis could be used to locate areas of risk in the way of incoming floods. The vulnerability index could therefore be confined to the extent of potential flooding. Narrowing the focus of the analysis to regions in the path of a storm would allow for a more precise assessment of vulnerability in the most heavily affected areas.

There is much potential for continued work regarding the shelter accessibility component of this study's analysis. A network analysis, such as that developed by Curtis (2016), could be used in place of this study's buffer analysis. This would likely produce significantly differing results, as it could identify points of impedance, where floodwater could block routes to shelters. It would also generate more varied results, as Curtis' network analysis ranked areas by travel

times to shelters, while this study only classified residential parcels as either in or out of shelter service area range. Additional factors, such as at-risk populations and shelter capacity, could also be incorporated into the network analysis to gain a comprehensive understanding of shelter accessibility in Harris County.

Both the flood risk and shelter accessibility factors were analyzed through the dasymetric mapping of at-risk populations throughout Harris County. A dasymetric analysis for this study was developed using Houston's unique land use and tax parcel datasets. This methodology was based on Maantay and Moroko's (2009) CEDS analysis. They found their results differed considerably from simpler, tract-level assessments of flood vulnerability. If a similar comparison were performed for this study's flood risk and shelter accessibility factors, it would likely produce the same results. As shown through Maantay and Moroko's analysis, a tract-level assessment would likely overestimate flood risk and shelter need in certain tracts where only a fraction of the residential population lived in a floodplain. This study's results could also be compared to those of other dasymetric methodologies. The application of Giordano and Cheever's (2010) three-class method would likely produce results that would differ from both a tract-level analysis and this study's dasymetric methodology. The reclassification of land use areas as either nonurban, low-density residential, or high-density residential would produce more precise results than a non-dasymetric methodology. However, that classification method may not fully represent the full range of urban and suburban housing density of the Houston area. Residential structures in Houston range from single-family houses to high-rise apartment structures, with numerous different types and sizes of arrangements in between. Due to this variety of housing density, Giordano and Cheever's (2010) analysis could potentially overestimate population density in some areas, and underestimate it in others. A direct

comparison between their methodology and this study's parcel-based method would help to quantify the difference between the two different residential land classifications.

The social justice component of this study was based on Cutter, Boruff, and Shirley's (2003) methodology for identifying social vulnerability to natural hazards. A comparison between their methods and those of this study would reveal useful information regarding the input social justice variables. The two analyses utilized similarly themed variables, although this study used less than a third of the total variables used by Cutter, Boruff, and Shirley. Their analysis, if applied to Houston, would likely produce similar results. Howeber, it would give more weight to areas with high African American or Hispanic populations, which were not factored into this study. The most significant difference between these two studies was the incorporation of flood risk and shelter accessibility factors. While Cutter, Boruff, and Shirley's (2003) analysis quantified social vulnerability evenly across their study area, this study used flood risk and shelter accessibility data to determine where that social vulnerability was the most impactful. A comparison of social justice with the other two factors revealed noticeable differences between them. The social justice layer showed the social vulnerability landscape for the entire county, regardless of flood risk. The inclusion of flood risk and shelter accessibility data helped further direct scrutiny to vulnerable populations that were most likely to be affected by flooding. The data created through Cutter, Boruff, and Shirley's social vulnerability analysis served as a base upon which to build a more specific, multicriteria assessment of flood vulnerability. Social justice is a critical issue in any kind of hazardous event. This study provided an example of how their methodology can be applied so that future assessments of vulnerability to natural hazards can utilize it as well.

5.5. Conclusions

The goal of this study was to use spatial analysis to create an FVI for Harris County, that could be used to visualize the extent and distribution of flood vulnerability in the Houston area. The analysis sought to identify the neighborhoods in and around Houston, where the greatest numbers of people were vulnerable to flooding. Flood vulnerability was assessed as a combination of three main factors: flood risk, shelter accessibility, and social justice. The highest-ranked areas of the index were expected to be socioeconomically marginalized areas with documented histories of flooding. The results revealed various distributions of high-vulnerability tracts along several major waterways. Several of the highest-ranked tracts were located in low-income neighborhoods that had disproportionally suffered from flooding from 2015 to 2019, including Alief, Sharpstown, and Greater Greenspoint. These results demonstrated that the index analysis effectively identified areas of high vulnerability in the Houston area.

This study contributes to a growing body of research on the subject of vulnerability to flooding and other natural hazards. Several similar studies to this one have been conducted for different parts of the world, but relatively little work has been done to identify vulnerable populations in Houston. This study therefore sought to address that research gap. The results of the study indicated the usefulness of such an analysis, as they included neighborhoods that were known to suffer considerably from flooding. The results could be used to identify lesser-known areas of vulnerability outside of those neighborhoods as well. In this way, the FVI provided a comprehensive representation of flood vulnerability across all of Harris County. Such an analysis ensures that no vulnerable people will be forgotten or ignored when they are affected by a flood.

The methods and strategies described in this thesis are not limited solely to the assessment of flood vulnerability in Houston. They can be applied to other types of vulnerability

studies as well. Various kinds of hazards, such as earthquakes, fires, and chemical spills, threaten different parts of the world. Processes such as PCA and dasymetric mapping can also be used to assess vulnerability to any of these hazards. Simple modifications to the scripts created for this analysis would allow them to be incorporated into new and different vulnerability analyses. Future researchers of vulnerability to natural hazards are encouraged to continue to build on this study and those that came before it, to promote equality and assistance to vulnerable people around the world.

References

Arraj, Shawn. "Harris County Adopts 'Worst-First' Guidelines for Remaining Flood Bond Projects." *Community Impact Newspaper*. August 28, 2019. https://communityimpact.com/houston/city-county/2019/08/28/harris-county-adopts-worst-first-guidelines-for-remaining-flood-bond-projects. Accessed November 17, 2019.

Baker, Karen. 2018. "Reflections on Lessons Learned: An Analysis of the Adverse outcomes Observed During the Hurricane Rita Evacuation." Disaster Medicine and Public Health Preparedness 12, Issue 1: 115-120.

Balica, Stefania and Nigel G. Wright. 2011. "Reducing the Complexity of the Flood Vulnerability Index." *Environmental Hazards* 9, Issue 4: 321-339.

Balica, Stefania, Nigel G. Wright, and F. van der Muelen. 2012. "A Flood Vulnerability Index for Coastal Cities and its Use in Assessing Climate Change." *Natural Hazards* 64: Issue 1, 73-105.

Bass, Benjamin and Philip Bedient. 2018. "Surrogate Modeling of Joint Flood Risk Across Coastal Watersheds." *Journal of Hydrology* 558: 159-173.

Bass, Benjamin, Andrew Juan, Avantika Gori, Zheng Fang, Philip Bedient. 2017. "Memorial Day Flood Impacts for Changing Watershed Conditions in Houston." *Natutal Hazards Review* 18, Issue 3.

Blackburn, Jim D. 2017. "Living with Houston Flooding." *Baker Institute for Public Policy*: 1-32.

Blackburn, Jim D. 2018. "Houston at the Crossroads: Resilience and Sustainability in the 21st Century." *Baker Institute for Public Policy*: 1-23.

Blackburn , Jim D. and Philip B. Bedient. 2018. "Houston a Year after Harvey: Where we are and Where We Need to Be." *Baker Institute for Public Policy and SSPEED*: 1-55.

Blessing, Russel, Antonia Sebastian, Samuel D. Brody. 2017. "Flood Risk Delineation in the United States: How Much Loss Are We Capturing?" *Natural Hazards Review* 18, Issue 3.

Brown, Daniel. 2019. "Tropical Depression Imelda Discussion Number 2." National Hurricane Center. https://www.nhc.noaa.gov/archive/2019/al11/al112019.discus.003.shtml?. Accessed November 17, 2019.

Burton, Christopher, and Susan L. Cutter. 2008. "Levee failures and social vulnerability in the Sacramento-San Joaquin Delta area, California." *Natural Hazards Review* 9, no.

Chen, Wei, Guofang Zhai, Chenjing Fan, Wenbo Jin, Ying Xie. 2017. "A Planning Framework Based on System Theory and GIS for Urban Emergency Shelter System: A Case of Guangzhou, China." *Human and Ecological Risk Assessment: An International Journal* 23, Issue 3: 441-456.

Castles, Katherine Lacey. 2018. "Examining the vulnerability of Communities and Residents in the Houston Metropolitan Statistical Area with Special Attention to Hurricane Harvey" UT Electronic Theses and Dissertations. https://repositories.lib.utexas.edu/handle/2152/65817. Accessed September 29, 2019.

Chakraborty, Jayajit, SaraE. Grineski, and Timothy W. Collins. 2019. "Hurricane Harvey and People with Disabilities: Disproportionate Exposure to Flooding in Houston, Texas." *Social Science and Medicine* 226: 176-181.

Curtis, Crystal Eden. "Preparing for Earthquakes in Dallas-Fort Worth: Applying HAZUS and Network Analysis to Assess Shelter Accessibility." Master's Thesis, University of Southern California, 2016.

Cutter, Susan L., Bryan J. Boruff, and W. Lynn Shirley. 2003. "Social Vulnerability to Environmental Hazards." *Social Science Quarterly* 84, Issue 2: 242-261.

Despart, Zach. "Harris County to Begin Work on New Floodplain Maps." *The Houston Chronicle*, September 25, 2018. https://www.chron.com/politics/houston/article/Harris-County-to-begin-work-on-new-floodplain-maps-13256550.php#item-85307-tbla-2. Accessed March 23, 2020.

Elliott, Rebecca. 2017. "Greenspoint-Area Residents Again Face Devastation." *Houston Chronicle*. August 27.

Federal Emergency Management Agency, 2007. "Managing Floodplain Development Through the NFIP". https://www.fema.gov/media-library/assets/documents/6029. Accessed September 29, 2019.

Feretti, Federico, Andrea Saltelli, and Stefano Tarantola. 2016. "Trends in Sensitivity Analysis Practice in the Last Decade." *Science of the Total Environment* 568: 666-670.

Flanagan, Barry E., Edward W. Gregory, Elaine J. Hallisey, Janet L. Heitgerd, Brian Lewis. 2011. "A Social Vulnerability Index for Disaster Management." *Journal of Homeland Security and Emergency Management* 8, Issue 1: 1-22.

Folch, David, Daniel Aribas-Bel, Julia, Koschinsky, Seth Spielman. 2016. "Spatial Variation in the Quality of American Community Survey Estimates." *Demography* 53, Issue 5: 1535-1554.

Goepel, Klaus D. 2013. "Implementing the Analytic Hierarchy Process as a Standard Method for Multi-Criteria Decision Making in Corporate Enterprises – A new AHP Excel Template with Multiple Inputs." *Proceedings of the International Symposium on the Analytic Hierarchy Process, Kuala Lumpur*.

Giordano, Alberto. 2010. "Using Dasymetric Mapping to Identify Communities at Risk from Hazardous Waste Generation in San Antonio, Texas." *Urban Geography* 31, vol. 5: 623-647.

Gori, A., A. Juan, R. Blessing, S. Brody, P.B. Bedient. 2017. "Characterizing Urbanization Impacts on Floodplain Through Integrated Land Use, Hydrologic, and Hydraulic Modeling: Applications to a Watershed in Northwest Houston, TX." *Journal of Hydrology* 568: 82-95.

Graham, Michael H. 2003. "Confronting Multicollinearity in Ecological Multiple Regression." *Ecology* 84, Issue 11: 2809-2815.

Greater Houston Flood Mitigation Consortium. 2018. *Strategies for Flood Mitigation in Greater Houston, Edition 1*.

Guillard-Gonçalves, Clémence, Susan Cutter, Christopher Emrich, and José Luís Zêzere. 2015. "Application of Social Vulnerability Index (SoVI) and delineation of natural risk zones in Greater Lisbon, Portugal.

Hamby, David M. 1994. "A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models." *Environment Monitoring and Assessment* 32, Issue 2: 135-154

Hamby, David M. 1995. "A Comparison of Sensitivity Analysis Techniques." *Health Physics* 68, Issue 2: 195-204.

Harris County Flood Control District. 2019. "Z-10 County-Wide Floodplain Mapping Update." https://www.hcfcd.org/projects-studies/countywide-or-multi-watershed/z-10-county-wide-floodplain-mapping-update/. Accessed September 29, 2019.

Haynie, Aisha, Sherry Jin, Leann Liu, Sherrill Pirsamadi, Benjamin Hornstein, April Beeks, Sarah Milligan, Erika Olsen, Elya Franciscus, Natasha Wahab, Ana Zangene, Delisabel Lopez, Lyndsey Hassmann, Deborah Bujnowski, Martina Salgado, Norma Arcos, Amanda Nguyen, Vishaldeep Sekhon, Richard Williams, Valeria Y. Brannon, Jennifer Kiger, Brian Reed, Mac McClendon, Les Becke, and Umair Shah (2018). "Public Health Surveillance in a Large Evacuation Shelter Post Hurricane Harvey." *Online Journal of Public Health Economics* 10, No. 1. http://www.firstmonday.dk/ojs/index.php/ojphi/article/view/8955. Accessed September 29, 2019.

Hennes, Rebecca. "City Data Shows the Houston Area Neighborhoods, Streets that Flooded the Most in 2019-19." *The Houston Chronicle*, November 8, 2019. https://www.houstonchronicle.com/neighborhood/article/Houston-streets-flood-most-data-Harvey-Imelda-14818397.php. Accessed March 26, 2020.

Highfield, Wesley E., Sarah A. Norman, Samuel D. Brody. 2013. "Examining the 100-Year Floodplain as a Metric of Risk, Loss, and Household Adjustment." *Risk Analysis* 33, Issue 2: 186-191.

Hunn, David, Matt Dempsey, and Mihir Zaveri. 2013. "Harvey's Floods: Most Homes Damaged by Harvey were Outside Flood Plain, Data Show." *Houston Chronicle*. March 30.

Karaye, Ibraheem M., Courtney Thompson, Jennifer A. Horney. 2019. "Evacuation Shelter Deficits for Socially Vulnerable Texas Residents During Hurricane Harvey." *Health Services Research and Managerial Epidemiology* 6: 1-7.

Li, Zhiying., Xiao. Li, Yue Wang, and Stephen Quiring. 2019. "Impact of Climate Change on Precipitation Patterns in Houston, Texas, USA." Anthropocene 25: 1-14.

Lynn, Kevin A. 2017. "Who Defines 'Whole': an Urban Political Ecology of Flood Control and Community Relocation in Houston, Texas." *Journal of Political Ecology* 24, No. 1: 951-987.

Maantay, Juliana and Andrew Maroko. 2009. "Mapping Urban Risk: Flood Hazards, Race, & Environmental Justice in New York." *Applied Geography* 29, Issue 1: 111-124.

Maantay, Juliana, Andrew Maroko, and Christopher Herrmann. 2013. "Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymetric System." *Cartography and Geographic Information Science*34, Issue 2: 77-102.

Maldonado, Alejandra, Timothy W. Collins, Sara E. Grineski, Jayajit Chakraborty. "Exposure to Flood hazards in Miami and Houston: Are Hispanic Immigrants at Greater Risk than Other Social Groups?" *Environmental Research and Public Health* 13, Issue 8 (2016): 775-795.

Mervosh, Sarah. "'I Can't Do This': Imelda Left Texas With at Least 5 Deaths and Historic Rainfall." *The New York Times*. September 20, 2019. https://www.nytimes.com/2019/09/20/us/tropical-storm-imelda-houston-texas.html. Accessed November 17, 2019.

Miller, Alexandra and Jeffrey Goodman. 2019. "Striving for Equity in Post-Disaster Housing." *Planning* 85, issue 8: 23-27.

Muñoz, Leslie A., Francisco Olivera, Matthew Giglio, and Philip Berke. 2017. "The Impact of Urbanization on the Streamflows and the 100-Year Floodplain Extent of the Sims Bayou in Houston, Texas." *International Journal of River Basin Management* 16, Issue 1: 60-69.

Peacock, Walter Gillis, Shannon Van Zandt, Dustin Henry, Himansha Grover, and Wesley Highfield. 2012. "Using Social Vulnerability Mapping to Enhance Coastal Community Resiliency in Texas" in *Lessons From Hurricane Ike* edited by Philip B. Bedient. 66-81.

Petrov, Andrey. 2011. "One Hundred Years of Dasymetric Mapping: Back to the Origin." *The Cartographic Journal* 49, Issue 3: 256-264.

Remo, Jonathan, Nicholas Pinter, and Moe Mahgoub. 2016. "Assessing Illinois's Flood Vulnerability Using Hazus-MH." *Natural Hazards* 81, Issue 1: 265-287.

Rogers, Susan. "Greenspoint, Poverty and Flooding." *The Houston Chronicle*, April 22, 2016. https://www.houstonchronicle.com/local/gray-matters/article/Greenspoint-poverty-and-flooding-7303300.php. Accessed March 23, 2020.

Saaty, Thomas L. 1990. "How to Make a Decision: The Analytic Hierarchy Process." *European Journal of Operational Research* 48, Issue 1: 9-26.

Saltelli, Andrea and Paola Annoni. 2010. "How to Avoid a Perfunctory Sensitivity Analysis." *Environment Modelling and Software* 25: 1508-1517.

Spielman, Seth, David Folch, and Nicholas Nagle. 2014. "Patterns and Causes of Uncertainty in the American Community Survey." *Applied Geography* 46: 147-157.

Subramanian, Nachiappan and Ramakrishnan Ramanathan. 2012. "A Review of applications of Analytic Hierarchy Process in Operations Management." *International Journal of Production Economics* 138, Issue 2: 215-241.

Roncancio, D.J. and Nardocci, A.C. 2016. "Social Vulnerability to Natural Hazards in São Paulo, Brazil." *Natural Hazards* 84, Issue 2: 1367-1383.

Tufekci, Suleyman. 1995. "An Integrated Emergency Management Decision Support System for Hurricane Emergencies." *Safety Science* 20: 39-48.

Uyan, Mevlut. 2013. "GIS-Based Solar Farms Site Selection Using Analytic Hierarchy Process (AHP) in Karapinar Region, Konya/Turkey." *Renewable and Sustainable Energy Reviews* 28: 11-17.

Wu, Qiang, Yuanzhang Liu, Donghai Liu, Wanfang Zhou. 2011. "Prediction of Floor Warter Inrush: the Application of GIS-Based AHP Vulnerability Index Method to Donghuantuo Coal Mine, China." *Rock Mechanics and Rock Engineering* 44, Issue 5: 591-600.

Zachos, Louis, Charles Swann, Mustafa, Altinakar, Marcus Mcgrath, Devin Thomas. 2016. "Flood Vulnerability Indices and Emergency Management Planning in Yazoo Basin, Mississippi." *International Journal of Disaster* 18: 89-99.

Zhang, Wei, Gabriele Villarini, Gabriel Vecchi, James A. Smith. 2018. "Urbanization Exacerbated the Rainfall and Flooding Caused by Hurricane Harvey in Houston." *Nature* 563: 384-388.

Zhang, Yan, Jihong Zhao, Ling Ren, and Larry Hoover. 2015. "Space-Time Clustering of Crime Events and Neighborhood Characteristics in Houston." *Criminal Justice Review* 40, Issue 3: 340-360.

Zhou, Xiaobo, Henry Lin, and Henry Lin. 2008. "Global Sensitivity Analysis." In *Encyclopedia of GIS*, edited by Shashi Shekhar and Hui Xiong. Springer, Boston, MA

Appendices

Appendix A. Python Scripts

Census data download and PCA script

```python
import censusdata as cd
import pandas as pd
from scipy import stats
from sklearn.decomposition import PCA
from sklearn import datasets
from sklearn.preprocessing import StandardScaler
from numpy import eye, asarray, dot, sum, diag
from numpy.linalg import svd


# This script performs a principal component analysis (PCA) on 12 social
# justice factors derived from the American Community Survey (ACS). The
# study area for the analysis is Harris County, Texas. social justice
# populations were selected due to their potential effects during major
# flood events, and are assessed at the tract level. All input data are
# retrieved from the ACS API, which can be accessed through the censusdata
# Python module. The final output consists of two tables: one containing
# the final component scores created through the PCA, and the other
# containing the factor loadings for each component.

out_path = r'C:\Users\MWilson\Houston\Final Thesis Data and Docs'

# Specific fields from relevant tables are retrieved from the ACS API.

harris = cd.download(
    "acs5", 2018,
    cd.censusgeo([("state", "48"), ("county", "201"), ("tract", "*")]),
                ["GEO_ID",

                # Total population:
                "B01001_001E",

                # Female population:
                "B01001_026E",

                # Under age 10:
                "B01001_003E", "B01001_004E", "B01001_027E",
                "B01001_028E",
```

```
                # Over age 64:
                "B01001_020E", "B01001_021E", "B01001_022E",
                "B01001_023E", "B01001_024E", "B01001_025E",
                "B01001_044E", "B01001_045E", "B01001_046E",
                "B01001_047E", "B01001_048E", "B01001_049E",

                # With a disability:
                "B18101_004E", "B18101_007E", "B18101_010E",
                "B18101_013E", "B18101_016E", "B18101_019E",
                "B18101_023E", "B18101_026E", "B18101_029E",
                "B18101_032E", "B18101_035E", "B18101_038E",

                # Poverty status:
                "B17020_002E",

                # Unemploymed:
                "C18120_006E",

                # Part-time workers:
                "B23027_005E", "B23027_010E", "B23027_015E",
                "B23027_020E", "B23027_030E", "B23027_035E",

                # Renters:
                "B25009_010E",

                # Recieve public assistance:
                "B19057_002E",

                # Single parent households:
                "B09005_004E", "B09005_005E",

                # Poor English speakers:
                "C16001_005E", "C16001_008E", "C16001_011E",
                "C16001_014E", "C16001_017E", "C16001_020E",
                "C16001_023E", "C16001_026E", "C16001_029E",
                "C16001_032E", "C16001_035E", "C16001_038E",

                # No vehicle available:
                "B08201_002E"])

# Social justice estimates are created from the retrieved fields.

harris["GEOID"] = harris["GEO_ID"].str.split("S", n=1, expand = True)[1]
```

```python
harris["TOTALPOP"] = harris.B01001_001E

harris["FEMALE"] = harris.B01001_026E

harris["UNDER10"] = harris.B01001_004E + harris.B01001_003E + \
                    harris.B01001_027E + harris.B01001_028E

harris["OVER64"] = harris.B01001_020E + harris.B01001_021E + \
                   harris.B01001_022E + harris.B01001_023E + \
                   harris.B01001_024E + harris.B01001_025E + \
                   harris.B01001_044E + harris.B01001_045E + \
                   harris.B01001_046E + harris.B01001_047E + \
                   harris.B01001_048E + harris.B01001_049E

harris["DISABILITY"] = harris.B18101_004E + harris.B18101_007E + \
                       harris.B18101_010E + harris.B18101_013E + \
                       harris.B18101_016E + harris.B18101_019E + \
                       harris.B18101_023E + harris.B18101_026E + \
                       harris.B18101_029E + harris.B18101_032E + \
                       harris.B18101_035E + harris.B18101_038E

harris["POVERTY"] = harris.B17020_002E

harris["UNEMP"] = harris.C18120_006E

harris["PART_TIME"] = harris.B23027_005E + harris.B23027_010E + \
                      harris.B23027_015E + harris.B23027_020E + \
                      harris.B23027_030E + harris.B23027_035E

harris["RENTER"] = harris.B25009_010E

harris["PUB_ASSIST"] = harris.B19057_002E

harris["SINGLE_PARENT"] = harris.B09005_004E

harris["POORENG"] = harris.C16001_005E + harris.C16001_008E + \
                    harris.C16001_011E + harris.C16001_014E + \
                    harris.C16001_017E + harris.C16001_020E + \
                    harris.C16001_023E + harris.C16001_026E + \
                    harris.C16001_029E + harris.C16001_032E + \
                    harris.C16001_035E + harris.C16001_038E

harris["NOCAR"] = harris.B08201_002E

c_fields = ["FEMALE", "UNDER10", "OVER64", "DISABILITY",
```

```python
            "POVERTY", "UNEMP", "PART_TIME", "RENTER", "PUB_ASSIST",
            "SINGLE_PARENT", "POORENG", "NOCAR"]


all_fields = ["GEOID"] + c_fields


SJ_pops = pd.DataFrame(harris, columns = all_fields)


SJ_pops.reset_index(inplace=True)


# Social justice variables occur on several different scales
# (population, households, population over 16), so they need
# to be standardized. In this script this is done through
# percentile rankings.


def percentileTable(in_data, fields):
    """This function Calculates the percentile score for each
    value in a numeric field. It takes 2 arguments:

        in_data - input pandas dataframe containing values to be
        ranked.
    fields - list containing names of fields within dataframe
        to be ranked.

    The ouput is a pandas dataframe containing the newly created
    percentile fields.
    """
    for field in fields:
        vals = list(in_data[field])
        arr = [i for i in vals if i != 0]
        pctile = [stats.percentileofscore(
            arr, n) if n != 0 else 0 for n in vals]
        in_data["P_" + field] = pctile

    p_fields = ["P_" + field for field in fields]

    return pd.DataFrame(in_data, columns=p_fields)

# Scikit-Learn is used to perform the PCA. The PCA_kaiser function
# first scales the input values, in order to increase the
# variance within each variable. The PCA is then performed on the
# scaled data, and components are created which are equal in number
# to the input fields. the eigenvalues of the correlation matrix of each
# component are then assessed. The Kaiser rule is then applied to the
# components, in which only those with an eigenvalue of 1.00 or greater
```

```python
# are retained.

def PCA_kaiser(in_data):
    """This function performs a PCA for an input dataset with multiple
    independent variables. The Kaiser rule is applied to the resulting
    components. The output is a pandas dataframe with all remaining
    components with and eigenvalue over 1.00."""

    component_cnt = len(in_data.columns)
    X_scaled = StandardScaler().fit_transform(in_data)
    pca = PCA(component_cnt)
    f = pca.fit(X_scaled)
    t = pca.transform(X_scaled)
    PCA_Components = pd.DataFrame(t)
    keep_components = 0
    for eigval in pca.explained_variance_:
        if eigval > 1:
            keep_components = keep_components + 1
    return pd.DataFrame(PCA_Components.iloc[:, 0:keep_components])

# The output components from the PCA are rotated in order to further
# increase the variance within the dataset.

def varimax(Phi, gamma = 1.0, q = 20, tol = 1e-6):
    """This function performs a varimax rotation for a set of PCA components.
    the input is a pandas DataFrame containing the component scores, and the
    output is a pandas DataFrame with the rotated scores"""
    p,k = Phi.shape
    R = eye(k)
    d=0
    for i in range(q):
        d_old = d
        Lambda = dot(Phi, R)
        u,s,vh = svd(
            dot(Phi.T,asarray(
                Lambda)**3 - (gamma/p) * dot(
                    Lambda, diag(diag(dot(Lambda.T,Lambda)))))))
        R = dot(u,vh)
        d = sum(s)
        if d_old!=0 and d/d_old < 1 + tol: break
    return pd.DataFrame(dot(Phi, R))

# Factor loadings are determined by assessing the degree of
# correlation (positive or negative) between each component and the
# input variables. Variables with high correlation coefficients
```

```python
# are considered to be the most heavily loaded. The dominant variable
# is that which has the highest absolute correlation.


def factorLoadings(PCA_table, in_table):
    """This function computes the factor loadings for each component
    in a PCA dataset. The function takes 2 arguments:

        PCA_table - dataframe containing PCA scores
        in_table - original input dataframe used to create the PCA scores

    The output is a dataframe containing the factor loadings for each
    component.
    """
    compare = pd.concat([PCA_table, in_table], axis=1, sort=False)
    corr = compare.corr()
    pca_cols = len(PCA_table.columns)
    return corr.iloc[pca_cols:, :pca_cols]

percentiles = percentileTable(SJ_pops, c_fields)
in_PCA = PCA_kaiser(percentiles)
PCA_rotated = varimax(in_PCA)
fl = factorLoadings(PCA_rotated, percentiles)
# The final PCA scores and factor loadings are exported to CSV files
# in the output filepath.

PCA_rotated.to_csv(out_path + '\PCA_rotated.csv')
fl.to_csv(out_path + '\FactorLoadings.csv')
```

Dasymetric analysis and final index score calculation

```python
# Flood Vulnerability Index Model

# This model utilizes tax parcel, land use, flood hazard, shelter location,
# and census data to create a flood vulnerability index for Houston, Texas.
# dasymetric mapping is utilized to disaggregate tract-level population
# estimates to the parcel level, based on weight of parcel code.
# Populated parcels are then selected where they intersect flood hazard
# and shelter need areas. The results of the dasymetric analysis are ranked
# by percentile and combined with social justice data calculated through
# principal component analysis in a different script. Factor weights were
```

```python
# determined through analytical hierarchy process (AHP). The final output of this

# model is a tract-level vulnerability index score, which is added to the
# input census tract boundary feature class as a new field.

import arcpy as ap
from scipy import stats

ap.env.overwriteOutput = True

ap.env.workspace = r'C:\Users\MWilson\Houston\Final Thesis Data and Docs\Houston\
Houston.gdb'

# The input spatial data for the model is listed below.

# Land use parcels containing 4-digit numeric land use codes are used to
# identify large residential structures.

land_use = "COH_LAND_USE"

# Tax parcel data containing 2-character alphanumeric tax codes are used
# to identify smaller residential parcels.

parcels = "Parcels"

# Final index scores are stored in the census tracts input feature class.

tracts = "Tracts"

# 100-year floodplain also known as special flood hazard area (SFHA) are
# used to demarcate areas of flood risk.

SFHA = "SFHA"

# FEMA shelter points are used to identify local shelter locations.

shelters = "Shelters"

# The population table contains American Community Survey (ACS)
# population estimates for each census tract.

pop_table = "TotalPop"

# The population field from the population table is added to the
# tract feature class table.
```

```python
pop_field = "TotalPop"

# The tract ID or GEOID is utilized for various join operations.

tract_id = "TRACT"

# Individual scores are calculated for flood risk and shelter
# accessibility, while the social justice score is provided
# from a seperate analysis. All three scores are weighted
# and combined for the final FV score.

FR_Score = "FR_Score"

SA_Score = "SA_Score"

SJ_Score = "SJ_Pctile"

FVI_Score = "FVI_SCore"

# Factor weights were determined through an AHP.

factor_weights = [52.7, 14.0, 33.3]

# A unique weight is assigned to each residential land use or state
# classification code based on structure size and estimated residential
# population density.

lu_codes = [['4209', 12], ['4211', 24], ['4212', 48], ['4213', 24],
            ['4214', 96], ['4221', 48], ['4222', 48], ['4313', 48],
            ['4316', 48], ['4319', 48], ['4670', 96], ['4613', 96]]

state_codes = [['A1', 1], ['A2', 1], ['B2', 2],
               ['B3', 3], ['B4', 4], ['E1', 1]]

# Two functions in this model utilize a field calculator codeblock function.

codeblock = """

# This function returns 0 if an input value is null, and if the value is present,

# multiplies that value by a specified weight.

def getcount(count, weight):
    if count is None:
```

```python
            return 0
        else:
            return count * weight


# This function simply returns 0 for a null value and the value if it exists.


def NoNulls(v):

    if v is None:
        return 0
    else:
        return v


"""



# The JoinField function takes an input feature class and join table, and
# creates a new field in the input table from a field in the join table, with
# nulls replaced by zeros. For this model, the join field for both inputs is
# always the tract ID/GEOID.


def JoinField(in_fc, join_table, in_field, out_field, join_field)
    """This function can be used to create a new field in a feature class
    table from a field in a joined table. Unlike the ArcPy JoinField_management
    function, this allows for a different name to be given to the output field
    This function takes 5 arguments:
    in_fc - input feature class
    join_table - table to be joined to feature class
    in_field - field in join_table to be added to in_fc
    out_field - name of new field in in_fc
    join_field - field on which the two tables will be joined. Must be the same
    name for both datasets"""
    ap.AddField_management(in_fc, out_field, "DOUBLE")
    ap.MakeFeatureLayer_management(in_fc, "layer")
    ap.AddJoin_management("layer", join_field, join_table, join_field)
    ap.CalculateField_management(
        "layer", out_field,
        "NoNulls(!{}.{}!)".format(join_table, in_field),
        "PYTHON3", codeblock)
    ap.Delete_management("layer")

# JoinField is used to create the tract population field.

JoinField(tracts, pop_table, pop_field, pop_field, tract_id)
```

```python
# Some multi-unit land use parcels overlap with numerous single-family
# parcels. In this case, the state classifications are a more accurate
# indicator of the number of people living in a parcel, and the coincident
# land use parcel must be removed.

ap.MakeFeatureLayer_management(
    land_use, "lu_lyr",
    """"LANDUSE_CD" IN (
        '4209', '4211', '4212',
        '4213', '4214', '4221',
        '4222', '4313', '4316',
        '4319', '4613', '4670')""")

ap.CopyFeatures_management("lu_lyr", "lu_res")
ap.Delete_management("lu_lyr")
ap.MakeFeatureLayer_management("lu_res", "res_lyr")
ap.MakeFeatureLayer_management(
    parcels, "parcel_lyr",
    """"StClsCode" IN ('A1', 'A2',
    'B2', 'B3', 'B4')""")

ap.SelectLayerByLocation_management(
    "res_lyr", "ARE_IDENTICAL_TO", "parcel_lyr")
ap.DeleteFeatures_management("res_lyr")
ap.Delete_management("res_lyr")
ap.Delete_management("parcel_lyr")

# The GetCodeCounts function uses summary statistics to calculate counts for
# each different parcel code in each tract. Tracts are spatially joined to
# parcels in order to assign the tract ID field to each parcel Weighted counts
# are then calculated by multiplying each count by its respective parcel code wei
ght.
# All weighted counts for each tract are then summed for the total weighted count
.

def GetCodeCounts(
        parcel_fc, tract_fc, joined_parcels,
        code_field, tract_id, sum_field, code_list):
    """This function generates tract-level weighted code counts from Houston
    parcel data. Weights are determined by estimated housing size. The function
    takes 7 arguments:

    parcel_fc - input parcel feature class
    tract_fc - census tract feature class
    joined_parcels - name of join feature class created from spatial join of
```

```
    tracts and parcels
    code_field - name of input parcel code field
    tract_id - unique identifier field for tracts (GEOID)
    sum_field - name of field containing weighted code counts
    code_list - list containing parcel codes and thier associated weights,
    entered as [[code1, weight1], [code2, weight2], etc]

    The final output is a weighted code count field in the tract feature class
    """
    ap.SpatialJoin_analysis(
        parcel_fc, tract_fc, joined_parcels,
        "JOIN_ONE_TO_ONE", "KEEP_ALL", "#", "HAVE_THEIR_CENTER_IN")

    for code in code_list:
        ap.MakeFeatureLayer_management(
            joined_parcels, "parcel_lyr",
            code_field + """ = '{}'""".format(code[0], code_field))

        ap.Statistics_analysis(
            "parcel_lyr", "stats_{}".format(code[0]),
             [[code_field, "COUNT"]],  tract_id)

        ap.Delete_management("parcel_lyr")


    ap.MakeFeatureLayer_management(tract_fc, "tract_lyr")

    for code in code_list:
        ap.AddField_management(
            "tract_lyr", "Weighted_{}".format(code[0]), "LONG")
        ap.AddJoin_management(
            "tract_lyr", tract_id,
            "stats_{}".format(code[0]), tract_id)
        ap.CalculateField_management(
            "tract_lyr", "Weighted_{}".format(code[0]),
            "getcount(!stats_{}.COUNT_{}!, {})".format(code[0],
                                                        code_field,
                                                        code[1]),
            "PYTHON3", codeblock)
        ap.RemoveJoin_management("tract_lyr")

    ap.Delete_management("tract_lyr")
    sum_codes= []
    for code in code_list:
        sum_codes.append('!Weighted_{}!'.format(code[0]))
```

```python
        weighted_codes  = str(sum_codes).replace("'", "")

        ap.AddField_management(tract_fc, sum_field, "LONG")
        ap.CalculateField_management(
            tract_fc, sum_field,
            "sum({})".format(weighted_codes), "PYTHON3")

# GetCodeCounts is applied to both land use and tax parcels, to get weighted coun
ts
# from both datasets. The two counts for each tract are then combined for the tot
al
# weighted residential parcel count.

GetCodeCounts(
    "lu_res", tracts, "lu_join", 'LANDUSE_CD',
    tract_id, "LU_WeightedSum", lu_codes)

GetCodeCounts(
    parcels, tracts, "parcel_join","StClsCode",
    tract_id, "Parcels_WeightedSum", state_codes)


ap.AddField_management(tracts, "WeightedSum_Total", "LONG")
ap.CalculateField_management(
    tracts, "WeightedSum_Total",
    '!LU_WeightedSum! + !Parcels_WeightedSum!', "PYTHON3")

# The parcels containing the tract ID field created through GetCodeCounts are
# merged to create a single residential parcel layer. A 'Res_Units' field is
# added to the feature class table, which is then populated with the assigned
# weight for a given parcel's classification code.

ap.Merge_management(["lu_join", "parcel_join"], "All_Res")
ap.AddField_management("All_Res", "Res_Units", "LONG")

ap.MakeFeatureLayer_management("All_Res", "res_lyr")

for code in lu_codes:
    ap.SelectLayerByAttribute_management(
        "res_lyr", "NEW_SELECTION",
        """"LANDUSE_CD" = '{}'""".format(code[0]))
    ap.CalculateField_management(
        "res_lyr", "Res_Units", code[1], "PYTHON3")

for code in state_codes:
```

```python
    ap.SelectLayerByAttribute_management(
        "res_lyr", "NEW_SELECTION",
        """"StClsCode" = '{}'""".format(code[0]))
    ap.CalculateField_management(
        "res_lyr", "Res_Units", code[1], "PYTHON3")


# Tract populations, weighted counts, and the Res_Units field created above
# are used to estimate each parcel's population. Tract populations are divided
# by weighted counts to get a 'people per unit' (PPU) field, which is then
# multiplied by Res_Units to calculate the parcel population.

def ParcelPopulation(
        tract_fc, parcel_fc, pop_field,
         count_field, res_units, tract_id):

    """This function estimates the residential populations for parcel features,
    based on weighted parcel code counts and tract-level populations. It takes
    6 arguments:

    tract_fc - census tract feature class with population and weighted count fiel
ds
    parcel_fc - residential parcel feature class
    pop_field - population field in tract table
    count_field - weighted code count field in tract table
    res_units - field in parcel_fc containing estimated number of residential uni
ts
    (code weight)
    tract_id - Unique identifier field for tracts (GEOID)

    The output is a parcel population field in the input parcel feature class"""

    ap.AddField_management(tract_fc, "PPU", "DOUBLE")
    ap.CalculateField_management(
        tract_fc, "PPU",
        "!{}!/!{}!".format(pop_field, count_field))
    ap.AddField_management(parcel_fc, "ParcelPop", "DOUBLE")
    ap.MakeFeatureLayer_management(parcel_fc, "parcel_lyr")
    ap.AddJoin_management("parcel_lyr", tract_id, tract_fc, tract_id)
    ap.AddField_management("parcel_lyr", "PPU", "DOUBLE")
    ap.CalculateField_management(
        "parcel_lyr", "PPU",
        "!{}.PPU!".format(tracts), "PYTHON3")
    ap.Delete_management("parcel_lyr")
    ap.CalculateField_management(
        parcel_fc, "ParcelPop",
```

```python
        "!{}!*!PPU!".format(res_units), "PYTHON3")

ParcelPopulation(
    tracts, "All_Res", "TotalPop",
    "WeightedSum_Total", "Res_Units", tract_id)

# Areas of shelter need (shelter inaccessibility) are defined as areas
# where at-risk populations are not located within one mile of a shelter.
# These areas are created as all SFHA areas which do not intersect a
# shelter buffer.

ap.Buffer_analysis(shelters, "Shelter_Buffers", "1 MILE")
ap.Erase_analysis(SFHA, "Shelter_Buffers", "Shelter_Erase")

# Flood risk and shelter accessibility are computed at the tract level
# through a selection of residential parcel within flood risk and shelter
# need polygon layers

def DasymetricSelection(
        parcel_fc, pop_field,
        tract_id, select_fc, out_table):
    """This function uses parcel-level populations to determine
    tract-level populations living within certain boundaries.
    It takes 5 arguments:

    parcel_fc - input parcel feature class
    pop_field - parcel population field
    tract_id - tract ID (GEOID) field in the parcel table
    select_fc - feature class demarcating boundaries for selection
    out_table - name of the output table with tract-level dasymetric
    population estimates"""
    ap.MakeFeatureLayer_management(parcel_fc, "res_lyr")
    ap.SelectLayerByLocation_management(
        "res_lyr", "INTERSECT", select_fc)
    ap.Statistics_analysis(
        "res_lyr", out_table,
        [[pop_field, "SUM"]], tract_id)


DasymetricSelection(
    "All_Res", "ParcelPop",
    tract_id, "Shelter_Erase", "NoShelter")

DasymetricSelection(
    "All_Res", "ParcelPop",
```

```
        tract_id, "SFHA", "FloodHazard")

# JoinField is used to add the population fields from the statistics tables
# created through the DasymetricSelectionFunction to the census tract layer.

JoinField(
    tracts, "FloodHazard",
    "SUM_ParcelPop", "AtRisk", tract_id)

JoinField(
    tracts, "NoShelter",
    "SUM_ParcelPop", "ShelterInacc", tract_id)

# The GetPercentile uses the scipy module to calculate the percentile score
# for each record of an input field, counting nulls and zeros as 0. This
# function is used to standardize the dasymetric population counts,
# which can then be combined with the social justice field, which
# is also ranked by percentiles.

def GetPercentile(in_fc, in_field, out_field):
    """This function computes percentile scores for each value in an input
    table field, and writes them to a new field in the same table. It takes
    3 arguments:

    in_fc - input feature class or table
    in_field - field to be ranked
    out_field - name of output percentile score field"""

    ap.MakeTableView_management(
        in_fc, "table_view",
        '{0} IS NOT NULL AND {0} <>0'.format(in_field))

    ta = ap.da.TableToNumPyArray("table_view", [in_field])

    array = ta[in_field]

    ap.AddField_management(in_fc, out_field, "DOUBLE")

    cursor = ap.da.UpdateCursor(in_fc, [in_field, out_field])

    for row in cursor:
        if row[0] !=0 and not row[0] is None:
            row[1] = stats.percentileofscore(array, row[0])
        else:
            row[1] = 0
```

```
        cursor.updateRow(row)

GetPercentile(tracts, "AtRisk", "FR_Score")
GetPercentile(tracts, "ShelterInacc", "SA_SCore")
#GetPercentile(tracts, SJ_Score, "SJ_Score")

score_fields = [FR_Score, SA_Score, SJ_Score]

# The finalScore function takes two input lists of equal length for the
# three factor scores and their respective weights. Each factor is weighted
# and then all three are combined for the final FVI score.

def FinalScore(in_fc, in_fields, weights, fv_score):
    """This function calculates a weighted average based on a list of input
    fields and a corresponding list of weights. It takes 4 arguments:

    in_fc - input feature class or table containing fields to be averaged
    in_fields - list of input fields
    weights - list of weights for each input field
    fv_score - name of final score field containing weighted averages"""

    weightcalcs = [weight/100 for weight in weights]
    fieldcalcs = ["!" + field + "!" for field in in_fields]
    ap.AddField_management(in_fc, fv_score, "DOUBLE")
    ap.CalculateField_management(
        in_fc, fv_score,
        "({}*{})".format(fieldcalcs[0], weightcalcs[0]) +
        "({}*{})".format(fieldcalcs[1], weightcalcs[1]) +
        "({}*{})".format(fieldcalcs[2], weightcalcs[2]), "PYTHON")


FinalScore(tracts, score_fields, factor_weights, FVI_Score)
```

Appendix B. Super Neighborhoods, Ranked by Mean FV Score

Super Neighborhood	Mean FV Score	Maximum FV Score	Minimum FV Score	Very Low	Low	Medium	High	Very High
Braeburn	75.69	94.33	46.23	0	0	4	3	2
Meyerland Area	67.58	92.79	41.99	0	0	7	6	1
Westwood	64.19	85.83	27.75	0	0	4	5	0
Alief	63.84	95.53	25.34	0	0	18	8	3
IAH/Airport Area	57.65	92.92	30.00	0	0	8	1	1
Hidden Valley	57.35	79.45	32.13	0	0	4	2	0
Brays Oaks	57.20	94.33	6.52	1	0	12	4	2
Fairbanks/Northwest Crossing	56.88	87.13	5.34	1	0	7	3	0
Greater Inwood	56.73	87.13	13.61	0	2	12	5	0
Sharpstown	56.32	94.33	8.47	1	5	8	8	3
Willowbrook	55.54	85.98	29.49	0	0	5	1	0
Braeswood	55.48	93.66	9.41	1	2	9	3	1
Fondren Gardens	55.07	71.50	26.99	0	0	4	0	0
Greater Hobby Area	54.15	95.72	0.08	2	1	5	5	1
South Belt/Ellington	54.14	95.71	8.09	2	3	8	6	3
Central Southwest	53.82	95.72	8.56	2	1	9	3	1
Edgebrook Area	53.30	86.87	8.52	1	0	4	3	0
Lake Houston	52.91	71.50	13.88	0	1	14	0	0
Eldridge/West Oaks	51.71	97.96	6.05	1	4	12	4	2

Greater Greenspoint	50.99	83.90	6.40	1	4	9	5	0
Kingwood Area	49.06	71.50	12.11	0	2	13	0	0
Museum Park	48.95	69.04	9.79	1	0	2	0	0
Lazybrook/Timbergrove	48.70	86.73	3.86	1	2	4	3	0
South Main	47.17	93.66	18.30	0	2	3	0	1
Addicks Park Ten	47.04	91.11	6.05	2	1	6	3	1
Westbury	47.01	78.60	20.45	0	2	9	2	0
Westchase	46.58	97.96	19.41	0	4	9	3	2
South Acres/Crestmont Park	45.97	95.72	8.74	1	1	5	0	1
Langwood	45.97	77.75	5.34	2	0	3	2	0
Gulfton	45.65	92.79	9.32	2	7	3	3	2
Westbranch	43.21	76.96	25.90	0	0	2	1	0
Carverdale	43.14	76.96	25.90	0	0	5	1	0
Willow Meadows/ Willowbend Area	42.68	93.66	10.08	0	4	5	1	1
Fort Bend Houston	41.95	69.51	26.99	0	0	3	0	0
El Dorado/Oates Prairie	41.30	69.02	9.02	1	1	7	0	0
Briar Forest	41.16	97.96	13.64	0	5	9	0	1
University Place	40.91	78.18	6.23	3	1	5	1	0
Clear Lake	40.39	95.71	5.63	2	5	12	0	2
Spring Branch East	40.14	86.73	3.86	4	2	6	4	0
Minnetex	38.77	95.72	3.48	2	2	3	1	1
Acres Home	38.04	81.14	6.41	3	2	9	1	0

Central Northwest	37.49	81.14	1.53	5	2	9	2	0
Hunterwood	37.20	65.74	12.28	0	1	2	0	0
Memorial	36.79	75.38	6.05	1	7	16	1	0
Medical Center Area	36.56	69.04	9.79	1	2	5	0	0
Astrodome Area	36.20	93.66	1.69	2	4	6	0	1
Washington Avenue Coalition/Memorial Park	34.19	86.73	3.86	1	6	9	2	0
Northside/ Northline	34.17	82.87	4.75	3	7	8	2	0
Sunnyside	33.52	71.83	5.85	3	3	6	0	0
Spring Branch North	32.97	76.96	2.70	3	2	9	1	0
Settegast	32.93	64.68	9.02	1	1	2	0	0
Northshore	32.84	69.02	4.19	2	4	9	0	0
Greater Heights	32.65	86.73	5.17	2	10	9	2	0
Independence Heights	31.95	72.64	7.12	3	2	6	0	0
Park Place	31.83	69.96	12.72	0	2	3	0	0
Mid West	31.20	77.69	2.03	4	9	13	1	0
Eastex - Jensen Area	30.86	82.87	0.76	6	6	8	1	0
Spring Branch Central	30.82	75.65	2.70	2	5	8	1	0
Pecan Park	30.19	69.96	0.55	2	2	6	0	0
Meadowbrook/ Allendale	29.64	69.96	0.17	1	5	5	0	0
Golfcrest/ Bellfort/Reveille	28.56	69.96	0.08	4	7	11	0	0
Greater Uptown	28.24	65.62	1.40	4	9	11	0	0
Kashmere Gardens	27.80	64.68	0.76	4	1	6	0	0

Gulfgate Riverview/Pine Valley	27.59	64.14	0.97	2	2	4	0	0
East Little York/Homestead	27.33	92.92	1.61	4	4	4	0	1
Spring Branch West	27.11	76.96	2.70	3	9	8	1	0
Macgregor	26.65	69.04	0.97	5	3	5	0	0
East Houston	26.02	47.73	5.42	2	2	3	0	0
Fourth Ward	25.10	50.75	12.20	0	3	2	0	0
Greater Fifth Ward	24.39	64.68	1.65	5	4	6	0	0
Pleasantville Area	23.50	64.68	0.85	3	2	3	0	0
Lawndale/ Wayside	23.44	40.18	0.55	1	4	5	0	0
South Park	23.16	49.94	2.25	4	3	5	0	0
Trinity/Houston Gardens	22.40	64.68	1.61	5	3	3	0	0
Harrisburg/Manchester	22.32	69.96	0.17	2	4	4	0	0
Greenway/Upper Kirby Area	21.53	68.74	3.39	7	5	3	0	0
Magnolia Park	20.86	29.53	13.01	0	6	2	0	0
Afton Oaks/River Oaks Area	19.36	55.19	1.40	4	7	3	0	0
Greater Eastwood	19.33	47.89	0.47	3	3	4	0	0
Second Ward	18.38	42.21	4.32	2	6	2	0	0
Downtown	18.28	53.25	0.47	5	3	3	0	0
Midtown	18.07	68.02	1.02	5	4	2	0	0
Near Northside	17.92	45.43	0.76	4	8	3	0	0
Denver Harbor/Port Houston	17.46	64.68	1.74	4	6	2	0	0
Neartown - Montrose	17.28	50.75	2.29	5	8	3	0	0

Greater Third Ward	16.93	68.02	0.47	8	1	3	0	0
Clinton Park Tri-Community	16.22	37.60	0.85	2	2	2	0	0
Greater Ost/ South Union	11.79	44.24	0.97	7	5	1	0	0