Determining the Effect of the Presence of Food Deserts on the Number of Crashes in Census Tracts in Arkansas

Lizbeth Juarez-Bartolo
University of Arkansas-Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/etd

Part of the Civil Engineering Commons, Transportation Engineering Commons, and the Urban Studies and Planning Commons

Citation
Juarez-Bartolo, L. (2023). Determining the Effect of the Presence of Food Deserts on the Number of Crashes in Census Tracts in Arkansas. Graduate Theses and Dissertations Retrieved from https://scholarworks.uark.edu/etd/5172

This Thesis is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu.
Determining the Effect of the Presence of Food Deserts on the Number of Crashes in Census Tracts in Arkansas

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Civil Engineering

by

Lizbeth Juarez-Bartolo
University of Arkansas
Bachelor of Science in Civil Engineering, 2021

December 2023
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Sarah Hernandez, Ph.D.
Thesis Director

Suman Mitra, Ph.D.
Committee Member

Lekshmi Sasidharan, Ph.D.
Committee Member
Abstract

Food deserts, areas with limited access to grocery stores which sell fresh produce, dairy, and meats, are a growing concern in many communities. The presence of a food desert is known to cause great impacts on the physical health of the people residing within them, but there is little to no research on a food desert’s influence on the traffic safety of the location in which it encompasses. Based on previous studies into food deserts, studies suggest that food inaccessibility is mostly felt by rural residents. Since rural residents must travel farther distances on more hazardous rural roadways to access food retailers, it was hypothesized that there could be an increased number of crashes in areas with food deserts. Because there is little to no research into the intersection of food deserts and traffic safety, to begin the first steps into this area of research this study attempts to determine if the relationship between the presence of food deserts and number of expected vehicular crashes is statistically significant.

To determine to what degree the presence of food deserts affect the number of crashes in a tract if at all, a linear regression was implemented that was examining the effect the attributes of a tract’s built environment, which included population distribution, roadway functional class distribution, daily vehicle miles traveled, and food desert attributes, had on the expected number of crashes. While the study determined that there is a statistically significant relationship between the presence of food deserts and crash incidences, contrary to the initial hypothesis, the study found an inverse relationship between the presence of food deserts and the number of expected vehicular crashes. Specifically, tracts identified as food deserts decrease the overall number of crash incidences in the tract. This unexpected outcome was further supported by the parallel finding that a decrease in the number of grocery stores correlates with a reduction in crash incidences.
These findings suggest that there could be a complex interplay between food deserts, travel patterns, and land use that requires further research to deepen the understanding between crash incidences and food desert. Future studies should examine the socioeconomic identities, travel patterns, and driving behaviors of individuals living in food deserts to understand aspects of grocery trips exhibited by those residents. Additionally, investigating the land use composition of food deserts and its impact on traffic safety could provide valuable insights. This study takes the first steps towards bridging the research gap in identifying the impacts the presence of a food desert has on traffic safety.
Table of Contents

Chapter 1. Introduction ......................................................................................................................... 1

Chapter 2. Background .......................................................................................................................... 4

2.1. Food Deserts: Definitions and Broad Research Areas ................................................................. 4

2.1.1 Definitions .................................................................................................................................. 4

2.1.2 Impacts ...................................................................................................................................... 8

2.2. Rural Road Safety: Issues and Relations to Food Deserts ............................................................ 9

2.3. Modeling Effect of the Built Environment on the Expected Number of Crashes ................. 11

2.3.1 Variables Used for Modeling ....................................................................................................... 11

2.3.1 Modeling Approaches .................................................................................................................. 20

Chapter 3. Methodology ....................................................................................................................... 27

3.2. Data Acquired for Analysis ............................................................................................................ 27

3.2.1. Rural Designation ....................................................................................................................... 29

3.2.2. Food Desert Designation ............................................................................................................ 31

3.2.3. Characteristics of the Food Desert ............................................................................................ 34

3.2.4. Racial and Ethnic Composition ................................................................................................. 35

3.2.5. Roadway Types .......................................................................................................................... 38

3.2.6. Daily Vehicle Miles Traveled (DVMT) ....................................................................................... 40

3.2.7 Number of Crashes ...................................................................................................................... 42

3.3. Data Preparation and Modeling ................................................................................................... 44
Chapter 4. Results .......................................................................................................................... 49

4.1. Food Accessibility Mapping ............................................................................................... 49

4.2. Linear Regression Model .................................................................................................... 52

Chapter 5. Discussion .................................................................................................................. 58

5.1. Racial and Ethnic Influence ............................................................................................... 58

5.2. Roadway Type Influence .................................................................................................... 59

5.3. Daily Vehicle Miles Traveled Influence ............................................................................ 60

5.4. Food Desert Presence’s and Characteristics’ Influences .................................................. 60

Chapter 6. Conclusion .................................................................................................................. 66

References .................................................................................................................................... 71
Chapter 1. Introduction

While the term “food desert”, areas having limited access to food retailers, was first coined in Scotland in the early 1990s, it is speculated that in the United States of America (USA) the rise of food deserts happened decades before the term was popularized and researched (Burke 2010; Cummins and Macintyre 2002). The increase of food deserts appearing in the USA can be traced back to the 1970s and 1980s when the USA experienced a great migration of high-income households from the city to the suburbs. This migration led to the inner city being mostly occupied by low-income households that could no longer financially support the grocery stores in the area leading to their closures (Burke 2010). The migration of the high-income households also led to the median income of the cities to decrease which in turn deterred companies from opening new grocery stores in the city cores. This economic segregation is speculated to have started the trend of food deserts in cities. Since financial gain is the main driver for any business, while there may be significant benefits to opening grocery stores in food deserts, the possible financial loss is not worth the risk for most businesses. This systematic cycle coupled with the increase in low-income households has brought food deserts and their effects to the public to the forefront of many fields of study.

In 2009, the United States Department of Agriculture (USDA) began mapping food accessibility to identify which census tracts are impacted by food deserts. This mapping concluded that every state contained a food desert (Karpyn et al. 2019). Since the initial mapping, “low access” which is defined as the physical distance, straight-line distance from a grocery store to the center of a census tract, to a food retailer, has decreased overall with a reported 15% of the population of the USA living within a food desert (Rhone et al. 2017).
Research evolving since the 1970’s examines metrics to designate an area as a food desert, and their effect on its inhabitants. While it is widely debated what defines a food desert, either geographic or social characteristics of an area, most researchers are in agreement that food deserts are the lack of access to healthy, fresh vegetables, fruits, milk, and meats by people which leads to an increased risk for health issues such as obesity, diabetes, or other weight-related health issues (Dutko, Ploeg, and Farrigan 2012). Most of the research into food deserts has been conducted in the fields of medicine, geospatial information, sociology, and/or economics which has caused a lack of investigation of the implications caused by the transportation network system; if transportation aspects are included in the research conducted it is usually represented by if inhabitants of the food desert own vehicles or have access to public transit to travel to grocery stores (Burke 2010; Chenarides et al. 2021; Winthrop Rockefeller Institute 2022; Dutko, Ploeg, and Farrigan 2012). Past research does not use the transportation network to calculate accessibility; physical accessibility is calculated utilizing straight line distance from residents and grocery stores. There is a gap in the literature between transportation and food desert research because of the method used to define food deserts and then study how food desert characteristics along with other built environment characteristics of a tract may affect the number of crashes in an area.

Although physical access to grocery stores has overall improved in the USA, in Arkansas, grocery store openings have stagnated. This has led to fewer than two grocery stores or produce vendors for every 10,000 people in the state leading to many Arkansans having low access to grocery stores and healthy foods (Winthrop Rockefeller Institute 2022). Since many Arkansans lack accessibility to grocery stores with affordable and healthy foods, there is a strong reliance on local convenience stores, like Family Dollar, for food and other living essentials due
to these stores’ proximity and affordability (Winthrop Rockefeller Institute 2022; Chenarides et al. 2021).

Further exacerbating the challenge of food deserts in Arkansas was the closure of a food distribution warehouse in West Memphis, Arkansas. This closure would affect the ability of many Arkansans, especially those in rural areas, to access food as the warehouse served approximately 400 Family Dollar, a dollar store, across the state (Frederick Price 2022). Since rural residents would have less access to food due to the closure, these residents would have to drive longer distances on rural roads to access food. Research shows that longer exposure (longer time and distances) on rural roads is linked to the increased likelihood of being in a fatal crash (Raymond et al. 2022). However, there is a gap in the literature connecting food deserts to traffic exposure, more specifically, for the context of rural roads.

Thus, this goal of this thesis is to determine the degree to which the presence of food deserts affects the number of crashes. This goal is met through two main objectives: 1) measure the prevalence of food deserts at the census tract level in Arkansas by developing a food accessibility map specific to Arkansas and 2) determine the degree to which different built environment characteristics may impact the number of crashes by applying regression-based modeling techniques.
Chapter 2. Background

Although there has been extensive research into food deserts and how the built environment, like schools or pedestrian facilities, can affect crash frequency in areas separately, there has been no research combining those distinct areas of research. In order to develop a study that would meld both areas of research together to determine the degree to which a food desert’s presence affects the number of crashes, the following topics were reviewed in the literature: definitions of food deserts, attributes of the built environment that affect crash frequency, and methods to model the impact of a built environment on crash frequency.

2.1. Food Deserts: Definitions and Broad Research Areas

This section discusses current landscape of food desert research regarding the definitions used to identify them and their effects on the residents that live within them.

2.1.1 Definitions

In the early 1990s when the term food desert was first used in Scotland, the term meant “region in which access to food retailers that stock fresh, affordable, and healthy options are lacking or non-existent”. As time has progressed, so did the number of definitions for food deserts (Widener 2018). Given the increase in qualitative and quantitative definitions for the term ‘food desert’, there is no consensus among researchers about how to identify it. Most researchers agree that a food desert is an area in which there is little access to fresh, affordable, and healthy foods, such as fruits, vegetables, milk, and meats. However, there is no consensus on what creates a food desert: physical proximity or societal barriers. Food deserts have been identified according to the following definitions:

1. “Urban areas with 10 or fewer stores and no stores with more than 20 employees”

   (Walker, Keane, and Burke 2010),
2. “Poor urban areas where residents cannot buy affordable, healthy food” (Walker, Keane, and Burke 2010),

3. “Low-income tracts in which a substantial number or proportion of the population has low access to supermarkets or large grocery stores. Low-income tracts are characterized by either a poverty rate equal to or greater than 20 percent or a median family income that is 80 percent or less of the metropolitan area’s median family income (for areas in metropolitan areas) or the statewide median family income (for tracts in non-metropolitan areas). Low access is characterized by at least 500 people and/or 33% of the tract population residing more than 1 mile from a supermarket or large grocery store in urban areas, and more than 10 miles in rural areas.” (Dutko, Ploeg, and Farrigan 2012)

4. “A tract with at least 500 people, or 33 percent of the population, living more than 1 mile (for urban areas) or 10 miles (for rural areas) from the nearest supermarket, supercenter, or large grocery store.”

One of the most widely used definitions of a food desert is the geographical definition developed by the USDA for their food accessibility mapping project in 2009 (Figure 1). Since this is one of the most cited definitions of a food desert, this definition was utilized to complete the spatial analysis for this study but involved a couple of changes to the procedure developed by the USDA.
To analyze food accessibility, the USDA first designated each tract in the USA as either rural or urban, as per the definition provided by the U.S. Census Bureau. Then population data was aerially allocated down to the ½ km-squared grids across the USA since these population allocations would determine the number of people or percentage of the population in a tract that did not have access to a grocery store. Finally, to assess physical distance from a grocery store, 1-mile or 10-mile buffers were created around each grocery store using straight line distance; grids within the buffer would be determined to have access to a grocery store and its population would be counted toward the tract’s population with access. If more than 500 people or 33% of the population of a tract did not have access to a grocery store, the tract would be identified as a food desert.
This study deviated from the USDA’s procedure in two ways. First, instead of utilizing $\frac{1}{2}$ km-squared grids, the analysis for this study utilized 1 km-squared grids due to computing capability constraints. Second, instead of straight-line distance buffers, this study utilized buffers that measured distance based on distance along the roadway network. Some researchers argue that using only geographical metrics to designate areas as food deserts neglects the holistic problem of a food desert. Measuring a food desert using only physical proximity, usually measured as straight-line distance, disregards important factors that affect a person’s experience living with food accessibility issues. Such factors include income, household characteristics, transportation network availability, time use, and mobility patterns (Widener 2018; Cummins and Macintyre 2002). Without considering sociodemographic factors, improvements to the metrics used to define a food desert include how physical distance is measured. Instead of merely using straight line distance, network (or routed) distance may be more representative of the physical conditions that create a food desert (Figure 2). These adaptations of the definition and method to measure a food desert create a more nuanced approach to determining a food desert (Widener 2018; Cummins and Macintyre 2002).
2.1.2 Impacts

Although there is much discourse within the research on food deserts, there is common ground on the significant impacts food deserts have on their inhabitants. Most of the research done on the adverse effects of food deserts focuses on the physiological health of the residents in a food desert. This research covers the fields of medicine, sociology, and public health. Poor access to
food retailers with healthy and affordable food options will lead to those residents having an increased exposure to energy-dense foods which are usually readily available due to the convenient presence of convenience and fast-food stores (Walker, Keane, and Burke 2010). The frequent intake of energy-dense foods, containing high amounts of fat, sugar, and sodium can lead to poor health outcomes such as diabetes, obesity, and other health-related illnesses which have now become more prevalent in children in the USA (Dutko, Ploeg, and Farrigan 2012).

In Arkansas, for example, in place of grocery stores there are corner and convenience stores that offer unhealthy processed food with low nutritional value. This leads to higher rates of diet-related diseases in communities defined as food deserts (Winthrop Rockefeller Institute 2022). While it is well understood that food deserts have a large impact on the physiological health of their residents, living in a food desert could also impact a person’s chances of being in a vehicular crash because of the longer distance those residents have to drive to access grocery stores which can be compounded with the unsafe nature of rural roads since most food deserts are in rural areas (Dutko, Ploeg, and Farrigan 2012).

2.2. Rural Road Safety: Issues and Relations to Food Deserts

Between 2000 and 2006, 6,529 census tracts in the USA (approximately 10%), were identified as being food deserts by the USDA with 2,204 (approximately 33%) of those tracts considered rural (Dutko, Ploeg, and Farrigan 2012). 23 million people (approximately 8% of the USA population) live in a census tract defined to be a food desert. Since 2009 there has been a 15% decrease in the number of food deserts measured by physical accessibility. But as the poverty rate increased there has been a net increase in the number of food deserts due to people’s inability to access affordable and healthy food because of their socioeconomic status.
Arkansas has seen a steady decline in the number of grocery and supermarket stores since 1997 according to the Arkansas Hunger Relief Alliance. Furthermore, in 2019, it was reported that every county in Arkansas had a food desert with 10 census tracts containing a population in which 60% or more of those populations had limited access to food sources (grocery or supermarket stores). Out of 72 counties in Arkansas, 52 of those are considered rural as stated by the Economic Research Service Office from the USDA. This posits that those who are most affected by food deserts in Arkansas are rural residents. As mentioned in the previous section, there are significant physiological health risks that arise when people live within food deserts but knowing that a large portion of Arkansans also reside in rural counties brings forth another potential safety risk.

Approximately 2.9 million (71%) of the 4.1 million public access roads in the USA are in rural areas (Kirk, Robert 2018). While fewer people live in rural areas, half of all fatal crashes occur on rural roads (Raymond et al. 2022). It is estimated that from 2016 – 2022, 85,022 people died in rural road crashes in the USA (Raymond et al. 2022).

The safety issues causing an increased number of deadly crashes on rural roads can be attributed to simpler roadway infrastructure, longer distances to medical services, risky driver behavior, etc. among other factors (Raymond et al. 2022). Considering that most food deserts in Arkansas are in rural areas, aside from the physiological risks of food deserts mentioned previously, there could be an additional safety risk that plagues rural residents specifically because of their residence in food deserts. Because the risk of dying on a rural road due to a vehicle crash is 62 percent higher than on an urban road for an equivalent trip length (Raymond et al. 2022), it highlights the idea that rural residents could be putting themselves in danger by
merely trying to access healthy and affordable food because of the types of roads they must drive on and how far they must go.

2.3. Modeling Effect of the Built Environment on the Expected Number of Crashes

While little to no research has been done on the effect of a food desert’s presence on the number of vehicle crashes occurring in a region using modeling tools, similar studies have examined other built environmental factors. This section discusses the various models developed to study food deserts in other contexts in terms of variables and methods used.

2.3.1 Variables Used for Modeling

Although research into how different measures of the built environmental, such as street network configurations, are present in most paper studying this topic, it is very uncommon to find features, such as the presence of school facilities, included in these studies (Merlin, Guerra, and Dumbaugh 2020). With little to no research into how the presence of specific locations, like schools or facilities, the development of what variables to include in the modeling process were derived from research that studied more common measures of the built environment and their effects on the crash risk, frequency, and counts. Common measures of the built environment range from the simplest measures, such as density, to more complex ones, like disadvantaged neighborhoods (Merlin, Guerra, and Dumbaugh 2020).

Population counts and density are some of the most common measures of the built environment within these types of studies which could be to ease of access to the data. While the most common there are mixed outcomes from several studies, some show an inverse relationship (Fischer, Sternfeld, and Melnick 2013; Dumbaugh and Rae 2009) while others show a positive effect on the dependent variable (Wier et al. 2009). (Fischer, Sternfeld, and Melnick 2013) finds that in the Los Angeles region tracts with a low population density has a larger rate of crashes.
through their investigation into how population density impacts collisions rates in a rapidly developing rural, exurban area of Los Angeles County. Similarly, (Dumbaugh and Rae 2009) find that as population density decreases so does crash risk in the study location. While those two studies both conclude that population density has an inverse effect on crash rate, (Wier et al. 2009; Guerra, Dong, and Kondo 2019) finds slightly differing results. (Wier et al. 2009) found that with an increase in overall population there was an increase in the number of crashes through their regression model of pedestrian-vehicle injury collisions based on environmental and population data in 176 San Francisco, California census tracts. Guerra, Dong, and Kondo (2019), similarly, got results in regard to populations of underprivileged populations. This study found that as the populations of black residents, residents experiencing poverty, and residents over the age of 64 increased so did the total number of crashes. Although studies have presented conflicting results as to how population or population density may affect crash frequency, the inclusion of these variables as a measure of the built environment seems to be commonly used in modeling no matter the subject matter.

Aside from population totals and density, another very common measure of built environment that is included in studies is mileage of roadways by type (Merlin, Guerra, and Dumbaugh 2020). Mileage of different roadway types was included as variables for the modeling completed for all the studies mentioned previously (Fischer, Sternfeld, and Melnick 2013; Dumbaugh and Rae 2009; Guerra, Dong, and Kondo 2019; Wier et al. 2009) as well as others (Yu and Xu 2018; Chen 2015; Dumbaugh et al. 2012). Through the inclusion of roadway types in these types of studies, it can be concluded that arterials, multi-lane streets, and roadways with high speed limits are all associated with higher crash risk and serious injuries (Yu and Xu 2018; Chen 2015; Guerra, Dong, and Kondo 2019; Dumbaugh et al. 2012). For example, Dumbaugh
and Rae (2012) finds that more crashes occur in San Antonio in districts with more kilometers of arterials even when vehicle kilometers traveled is controlled (Dumbaugh et al. 2012). Yu and Xu (2018) find similar results in their study on the influence of built environments on crashes with different levels of injury severity; this study found that high speed and high traffic facilities, like freeways and arterials, have a positive effect on the number of crashes in their study location.

Aside from roadway types, like arterials, affecting the number of crashes, the roadway type can also affect the crash severity expected. Guerra, Dong, and Kondo (2019) find that all high speed, multi-lane roadways, like highways and arterials, had a positive effect on the number of total fatal crashes; this means as the mileage of these roadways increase so does the number of total fatal crashes. The studies previously discussed all study how the built environment effect the number of vehicular crashes specifically, but studies like Chen (2015) and Wier et al. (2009) are studying the relationship between measure of built environment and pedestrian-vehicular crashes. Although Chen (2015) includes slightly different variables than the previous studies, like variables representing miles of bike lanes on arterials, the study still includes mileage of different arterials like previous studies. Chen (2015) finds that cycling on arterial roadways increases the potential severity of a crash due to its positive effect on the dependent variable.

Similarly, Wier et al. (2009) finds that with an increase in arterial mileage there is an increase in vehicle-pedestrian injury crashes per tract. Again, although the studies investigated different measures of the built environment and their effect on some type of crash, all these studies always included population and roadway mileage by type.

The previously mentioned studies all investigate how measures of the built environment effect crash frequency and crash severity; therefore, they share common variables like population and roadway configuration details. Since all the studies are investigating slightly different topics,
from density’s effect on crashes to pedestrian-involved crashes, in regard to the built environment, the studies include very different variables throughout and different quantities of variables. For example, Guerra, Dong, and Kondo (2019) tries to model relationships between neighborhood socio-demographics, urban form, roadway characteristics, traffic collisions, injuries, and fatalities on the Philadelphia region’s streets from 2010 to 2014 to answer, “Do denser neighborhoods have safer streets?” This study contained more than 30 possible variables spanning from information about ADT to public school enrollment. Conversely, Amoros (2001) only includes five independent variables (county name, year of fatal crash, kilometers of motorways, kilometers of national roads, and kilometers of county and local roads) in their study trying to compare traffic safety among several counties in France and explore whether observed differences can be explained by differences in road type distribution between counties. Based on the objective of the study, measures represented in the model as variables can range from 30 measures of the built environment to only five measures. Though the quantity of measures can vary between studies, there are common measures found throughout them all, like population and roadway characteristics, as discussed previously (Merlin, Guerra, and Dumbaugh 2020).

While there have been many studies investigating how simple measures of the built environment may affect crash incidences, as discussed previously, studies investigating how the presence of specific facilities, like schools or hospitals, are uncommon (Merlin, Guerra, and Dumbaugh 2020). The studies found that investigate this uncommon topic were Clark and Cushing (1999) and Clifton and Kreamer-Fults (2013). The study by Clifton et al. (2007) examines pedestrian–vehicular crashes in the vicinity of public schools, the severity of injuries sustained, and their relationship to the physical and social attributes near the schools (Clifton and Kreamer-Fults 2007a). Since the study specifically wanted to analyze the relationship between
the number of crashes occurring within the areas near public schools, the study first created
buffers around the schools to only capture data on crashes that fell within the ¼ mile straight line
buffer from the years 2000 to 2002. The study wanted to determine the relationship between the
presence of schools and its characteristics to the number of crashes in the area therefore the
researchers included seven variables representing presence of a school facility and its
characteristics; the variables on school characteristics were:

1. School Type: a categorical variable that had categories for elementary schools,
middle schools, and high schools.
2. Road Functional Class: this was a categorical variable that stated if the roadway
adjacent to the main entrance was either a primary separated (interstates, state
roads, and highways with medians), primary unseparated (interstates, state roads,
and highways with no medians),), or local unseparated roadway (city streets and
rural roads with no medians).
3. Enrollment: this was a continuous variable that represented the average annual
number of student enrolled in the school.
4. Driveway: an indicator variable that was a one if a driveway for drop-off existed
and zero if not.
5. Parking Lot: an indicator variable for the presence of off-street parking; one for
present and zero for not present.
6. Recreation Facilities: an indicator variable for the presence of playground area
and equipment; one for present and zero for not present.
7. Set Back: an indicator variable for the presence of a setback; one for present and
zero for not present.
While the study’s focus was the degree of effect from the previous variables on the expected number of crashes, the study also included several other variables that described the population and land use around the schools. Notably, these variables included: the population of non-white residents, population of children under five years old, population of children between the ages of 5-15, population density, percentage of land use types, and roadway density (miles of roadway per block group). Clifton and Kreamer-Fults (2013) utilized a linear regression approach to model the relationship that was being investigated in this study. The study found that from the school characteristics, the presence of recreation facilities and driveways were statistically significant, with the presence of recreation facilities having a positive effect of 7.06 and the presence of driveways having a negative effect of -7.66 on the number of crashes within the school buffer (Clifton and Kreamer-Fults 2007a). Furthermore, in regard to the area characteristics, the variables that were of statistical significance were commercial access, population of non-white residents, population density, and the total area of commercial property in a census block; these variables had a positive influence of the number of expected pedestrian-vehicular crashes in the buffer zones (Clifton and Kreamer-Fults 2007a) which was in agreement with the studies discussed previously.

Guerra, Dong, and Kondo (2019) found that as the populations of black residents so did the total number of crashes which is similar to the results from the study conducted by Clifton and Kreamer-Fults (2007) since their study found that as the population of non-white residents increased the number of expected pedestrian-vehicular crashes. Clifton and Kreaner-Fults (2007) also concluded that as population density increases so does the number of expected pedestrian-vehicular crashes which is consistent with previous studies (Fischer, Sternfeld, and Melnick 2013; Dumbaugh and Rae 2009). This study also found that commercial access and the total area
of commercial property in a census block had a positive effect on the dependent variable meaning as access or total area increased the number of expected ped-vehicular crashes would increase. This result aligns with previous studies that included land use in the model like Fischer, Sternfeld, and Melnick (2013), Dumbaugh and Rae (2009), and Yu and Xu (2018) which all found a positive relationship between the increase in total commercial area or percentage of commercial area and its impact in the number of expected crashes or number of expected pedestrian-vehicular crashes specifically. Since no other study was found that investigated the degree to which a school facility’s presence and attributes had an effect on crash incidences in an area, there is no way to compare the findings from the Clifton and Kreaner-Fults (2007) study about the school attributes, like the driveway and recreational facility presence, to any other study to see if this study concurs or disagree with other findings.

Another study that investigated how the presence of a physical facility affected the possibly number of crashes was a study by Clark and Cushing (1999). This study attempted to find the degree to which population density and number of hospitals in a county affected the number of expected fatal crashes. While the Clifton and Kreaner-Fults (2007) study included over 20 variables, the Clark and Cushing (1999) study only included three variables. The variables included:

1. Number of Fatal Crashes: this was a continuous variables representing the number of crashes per county.

2. Population Density: this was a continuous variable representing the number of residents in a county.
3. Mean Inter-hospital Distance: a continuous variable estimated by the square root of the reciprocal of the hospital density (Clark and Cushing 1999).

The study developed several linear regression models to investigate the relationship between each variable listed above and the crash rate. Through the creation of the linear regression models, Clark and Cushing (1999) concluded an inverse relationship between population density in a county and the expected number of fatal crashes which concurs with studies referenced above like Guerra, Dong, and Kondo (2019). The other linear regression models developed for this study showed a positive relationship between the mean inter-hospital distance in a county and the crash. Due to some overdispersion in the distance to a hospital model, the researchers added an indicator variable for Southern State Counties since there was a cluster of points in their data that corresponded to 13 contiguous southern states (Florida, Georgia, South Carolina, North Carolina, Alabama, Mississippi, Louisiana, Tennessee, Arkansas, Oklahoma, Texas, New Mexico, and Arizona); one representing it as a southern state and zero representing it is not (Clark and Cushing 1999). Even with the new indicator variable, the linear regression model still found the relationship between the distance of hospitals to have a positive relationship with the crash rate in county.

While it is rare for studies to investigate how the presence of certain facilities, like schools or hospitals, affect crash incidences, this the main objective of the study being discussed. There are not many studies outlining how to conduct this types of research, but through an investigation into how studies investigate the relationship between different measure of built environment and crash incidences an idea of what type of variables to include can be formed. Following from the literature, to examine the degree to which food deserts affect the number of
expected crashes this study will focus on environmental attributes rather than attributes of the driver(s) involved in the crash. This study is geared toward understanding how measures of built environment and presence of a food desert affect the number of expected crashes in a tract and thus socioeconomic or behavioral data is not considered in the model aside from racial distribution within the tract. The model developed in this study considers measures of the built environment such as: kilometers of roadway by functional class and designation, number of grocery stores, the presence of food deserts, density of grocery stores, population distribution by race, daily vehicle miles traveled (DVMT) and number of crashes per census tract. The variables utilized to represent area characteristics, like roadway distribution, population distribution by race, and DVMT, that are included in this study similar to other studies investigating the relationship common measures of the built environment and crash incidences (Merlin, Guerra, and Dumbaugh 2020; Dumbaugh and Rae 2009; Dumbaugh et al. 2012; Fischer, Sternfeld, and Melnick 2013; Wier et al. 2009; Guerra, Dong, and Kondo 2019; Chen 2015; Yu and Xu 2018; Amoros, Martin, and Laumon 2003; Noland and Oh 2004). The variables chosen to represent attributes of the food desert were guided by the study from Clifton and Kreaner-Fults (2007). Clifton and Kreaner-Fults (2007) included variables describing physical attributes of schools that could cause a difference in the crash incidences occurring within the buffers, but also make the school unique compared to each other. For this study on food deserts, the variables chosen to represent attributes of food deserts were the number of grocery stores in a tract and the density of grocery stores. These variables could help identify how “bad” a food desert actually is in terms of food accessibility.

Specifically, variables considered in the model are rural/urban designation, miles of specific roadway types, daily vehicle miles traveled (DVMT), population size, housing density,
racial distribution, and food desert-specific characteristics. This study presents a first step
towards understanding the relationship between the built environment, food deserts, and crash
occurrence. Therefore, any assumptions and omissions of sociodemographic variables will be
left for future work. The models will be developed at the census tract level, following from prior
studies (Fischer, Sternfeld, and Melnick 2013; Wier et al. 2009; Yu and Xu 2018). By following
the norm of other research, it will make this study more comparable to past studies since the
spatial analysis procedure conducted for mapping food accessibility can be directly compared to
other methods, such as the USDA method.

2.3.1 Modeling Approaches

Of the studies discussed previously that aim to investigate how measures of the built
environment impact crash incidences, there is almost an even split of what model type is utilized;
six of the studies utilize a negative binomial regression (Dumbaugh and Rae 2009; Dumbaugh et
al. 2012; Guerra, Dong, and Kondo 2019; Chen 2015; Yu and Xu 2018; Noland and Oh
2004) while five of the studies utilize a linear regression (Clifton and Kreamer-Fults 2007a; Clark
Overdispersion is commonly seen in crash data counts therefore this is why the negative
binomial regression model has started to become popular among researchers (Abdulhafedh
2017). Alongside the negative binomial regression’s rise in popularity is the zero-inflated
regression model. Zero-inflated regression is widely used to account for the overdispersion that
is caused by excessive zeroes in traffic data (or crash data) counts (Abdulhafedh 2017). While
the zero-inflated regression model is rising in popularity, none of the studies discussed have used
it. This could be due to the fact that the studies use the number of crashes occurring in a study
location for the dependent variable; the zero-inflated crash is usually utilized when modeling utilizing a specific severity type of crash (Abdulhafedh 2017).

For the studies that utilized the negative binomial regression method, almost all the studies cited the reason for the utilization was due to using count data (Amoros, Martin, and Laumon 2003; Yu and Xu 2018; Noland and Oh 2004; Dumbaugh and Rae 2009). For example, Amoros, Martin, and Laumon (2003) state that a negative binomial regression approach was conducted because this approach allows for overdispersion which is a common phenomenon in crash data. Since the study assumed that the model utilized would account for the overdispersion, there was no testing conducted on the crash data to see if the data was in fact over-dispersed. Since the Amoros, Martin, and Laumon (2003) was comparing traffic safety among several counties in France, and explore whether observed differences can be explained by differences in road types distribution between counties, the model resulted in finding interactions between county and roadway type as statistically significant (Amoros, Martin, and Laumon 2003). Overall this study found that national and motorway roads had a inverse relationship with the number of crash incidences for the regions of France while county and local roads had a positive relationship with the dependent variable. This result is slightly different than the results found in previous studies since previous studies found that high speed multilane roadways tend to have a positive relationship with the number of crashes in an area; this could be explained by the fact that the Amoros, Martin, and Laumon (2003) study was conducted in France while the other studies were conducted in the United State of America.

Similarly, the study from Xu and Yu (2018) also opted into utilizing the negative binomial regression model due to the assumption that count data is commonly over-dispersed.
Unlike the Amoros, Martin, and Laumon (2003) study, Xu and Yu (2018) included over 20 possible variables in the model. The inclusion of large quantities of variables is common throughout the studies that utilize the negative binomial regression; Noland and Oh (2004) included 20 variables and Dumbaugh and Rae (2009) included approximately 15 variables. While Xu and Yu (2018) included over 20 variables into the model, there were only nine variables that were statistically significant. The study found that linear milage of arterials and highways/freeways had a positive relationship with the number of total crashes in a tract which is on par with the studies previously discussed. Commercial area in a tract also had a positive relationship with the dependent variables meaning as commercial area is increased so does the number of crashes in a tract. Unlike previous studies, this model included a variable for the land use type “office area” which resulted in having a positive relationship with the dependent variable. This suggests that tracts with larger amounts of commercial and office area increase the number of crashes which could be due to the increased traffic volume of those areas since they are areas of work and recreation for residents; this is further supported by this model resulting in traffic volume having a positive relationship with the expected number of crashes (Yu and Xu 2018). This model reported a Akaike Information Criterion (AIC) of 6305.01; in general models with low AIC values tend to be better fits (Yu and Xu 2018).

Since the studies discussed utilize different measures for goodness of fit, it is difficult to compare models to each other; Amoros, Martin, and Laumon (2003) only reports an overdispersion value of 1.10, Yu and Xu (2018) report AIC for the goodness of fit measure, and Noland and Oh (2004) and Dumbaugh and Rae (2009) report log likelihood as their goodness of fit measure. While the negative binomial regression models can be difficult to compare to one another, the studies utilizing a linear regression approach utilize r-squared and adjusted r-square
as the measure for goodness of fit. Clifton and Kreamer-Fults (2007), Clark and Cushing (1999), Wier et al. (2009), Fischer, Sternfeld, and Melnick (2013), and Li et al. (2013) all measure the goodness of fit of their models with either adjusted r-squared or r-squared; their scores are 0.36 - 0.55, 0.43-0.74, 0.71, 0.42, and 0.41, respectively. Although there are a number of studies that utilize the negative binomial regression for modeling the relationship between measures of the built environment and crash incidences, there are also several studies that were discussed that utilized a linear regression instead. Included in those studies that utilized a linear regression approach were the two studies, Clifton and Kreamer-Fults (2007) and Clark and Cushing (1999), that are similar to what is trying to be accomplished in the study for this paper.

Similar to the study being performed for this paper, Clifton and Kreamer-Fults (2007) attempted addresses gaps in the knowledge by examining the physical and social conditions that are associated with pedestrian crashes near schools since this topic area had little to no research associated with it (Clifton and Kreamer-Fults 2007a). To create the first step towards understanding the relationship between the measures of the built environment, schools, and crash occurrence Clifton and Kreamer-Fults (2007) chose to utilize a linear regression model to gather a basic understanding on whether a schools’ attributes could have a statistically significant effect on total pedestrian-vehicular crashes in a school area. The study did not mention any data preparation or transformations that were done if needed. While the study contained approximately 15 variables representing attributes of the area around the schools and school features, only six variables were found to be statistically significant in effecting the number of crashes within the buffer for all ages (R² = 0.56); two variables describing school facilities and four describing area characteristics (Clifton and Kreamer-Fults 2007a). As discussed previously the area-wide attributes and their relationship with the dependent variables are consistent with
previous studies, but due to the lack of research into the relationship between schools and crash incidences there can only be speculations made about why the school attributes are significant. For example, Clifton and Kreamer-Fults (2007) speculate that the presence of recreation facilities may attract students outside of regular school hours and these children may be younger in age and thus may be more likely to be severely hurt than older children.

A second linear regression model ($R^2 = 0.36$) was developed in the same study looking at the relationship between the 15 variables and the number of crashes occurring in the buffers by age group of the school. In this second model, none of the school attributes were significant; the only variables that were significant was the population of non-white residents, population of children between the ages of 5 – 15, and population density (Clifton and Kreamer-Fults 2007b). All of which had a positive relationship with the number of pedestrian-vehicular crashes involving appropriately school aged children in the buffer zones. The relationship between numbers of school aged children living in the area and severity of crashes between this population group is intuitive. As with the previous model, population density for the area around the schools is associated with greater severity of crash (Clifton and Kreamer-Fults 2007a).

The study by Wier et al. (2009) also utilized a linear regression for their investigation into predicting changes in vehicle-pedestrian injury collisions based on changes in traffic volume (Wier et al. 2009). Wier et al. (2009) completed this research through the construction of a multivariate, area-level linear regression model of vehicle pedestrian crashes based on measures of the built environment for 176 tracts in San Francisco, California. There were 22 predictor variables included in the study that represented information about street configurations, land use, population characteristics, and commuting behavior (Wier et al. 2009).
Before modeling for the Wier et al. (2009) study began, there was data transformations and clean up that occurred to ensure the data better approximated the normality assumptions of a linear model. The researchers completed a log transformation on traffic volume and employee variables to better fit the normality assumptions. Similarly, Li et al. (2013) utilized a natural log transformation on the dependent variable, number of crashes per county, and DVMT in a county for their study on fatal crashes and countywide factors including traffic patterns, road network attributes, and socio-demographic characteristics were collected from the 58 counties in California (Li et al. 2013). Aside from the data transformation Wier et al. (2009) conducted in their study, the data clean up also included removing any outliers in the data. The data points were determined to be outliers based on an assessment conducted on the residual plots from the fitted model.

From the 22 predictor variables that were included in the model, there were only nine variables that were deemed statistically significant; they were: traffic volume, percentage of arterial streets, percentage of neighborhood commercial, percentage of residential-commercial land, land area, employee population, residential population, percentage of residents living in poverty, and percentage of residents who are 65 years of age or older (Wier et al. 2009). All the variables had a positive relationship with the number of pedestrian-vehicular crashes in a tract except for land area and the percentage of residents who are 65 years or older. With the land use and percentage of residents who are 65 years or older variables, as those variables increase in size the number of expected pedestrian-vehicular crashes are expected to decrease. Wier et al. (2009) has a similar result from other studies previously discussed in regard to the relationship between the population of older adults (65 years or older) and expected pedestrian-vehicular crashes such as the study from Noland and Oh (2004). Noland and Oh (2004) found the
relationship between the population of older adults and the number of crashes to have a coefficient of -0.037, indicating an inverse relationship. Dumbaugh and Rae (2009), though, determined differing results. Their study that the population of older adults has a positive relationship with the number of expected crashes (Dumbaugh and Rae 2009). Finally, the linear regression model completed by Wier et al. (2009) reported an adjusted $R^2$ value of 0.71 which is on the upper end of of $R^2$ values reported by the studies discussed (Clifton and Kreamer-Fults 2007a; Clark and Cushing 1999; Wier et al. 2009; Fischer, Sternfeld, and Melnick 2013; Li et al. 2013).

While the negative binomial regression model is efficient in modeling crash rates, linear regression models can be as robust and accurate if the proper data transformations and manipulation are to be done (Rakha et al. 2010). Linear regression models are typically not utilized for crash rate analysis because the crash data does not typically follow the assumptions of a linear regression model. Crash data usually does not follow the normal error structure and the constant error variance assumptions. By using data transformations on both the independent variable and dependent variables the data can then ensure the normality and homoscedasticity needed for a linear regression model as done and recommended by past studies (Wier et al. 2009; Li et al. 2013; Abdulhafedh 2017).
Chapter 3. Methodology

The main objective of this research project was to determine the degree to which the presence of food deserts affects the number of expected crashes. To gather data and perform a data-driven analysis, the research was divided into three (3) tasks (Figure 3). First, a literature review was conducted, followed by the collection of the data that would be needed for a model, and, finally, the creation of the model that would produce the degree to which the presence of food deserts would impact the frequency of crashes in census tracts.

Figure 3. Flow Chart of Tasks

3.2. Data Acquired for Analysis

To perform a data-driven analysis to see the effect of a food desert on the frequency of crashes in census tracts in Arkansas, data regarding food desert designation, food desert characteristics, tract attributes, daily vehicle miles traveled (DVMT) for each tract, and number of crashes of all
types occurring in each census tract had to be gathered. This study focused on how the physical characteristics of a tract may influence the number of crashes occurring within its boundaries therefore the data collected included very limited demographic or socioeconomic details. Racial makeup of the population in the census tract was the only demographic information collected and utilized in the model which mirrored past research conducted in this area of study (Clark and Cushing 1999; Clifton and Kreamer-Fults 2007a; Amoros, Martin, and Laumon 2003; Noland and Oh 2004).

Since the grocery store locations utilized to determine food desert designation were from 2016, all data gathered about census tract attributes, DVMT per tract, and crashes per tract were taken for the year 2016 or as close to that year as possible to keep consistency throughout the study. The independent variables used for the model that would produce the degree to which the presence of a food desert affects the number of crashes in a census tract were urban designation, tract size, tract population, housing density, racial composition, miles of roadway per roadway classification, food desert designation, grocery store density, number of grocery stores, DVMT on county roads, DVMT on city roads, and miles of rural roads while the dependent variables for the models being fitted were the number of crashes for the year 2016; all this data was gathered from different sources (Table 1). The following sections discuss why and how the data was collected.
Table 1. Description of data sets utilized in the study

<table>
<thead>
<tr>
<th>Name of Data Set</th>
<th>Source</th>
<th>Data Type</th>
<th>Date of Data</th>
<th>Variables Acquired from Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arkansas Tracts</td>
<td>U.S. Census Bureau</td>
<td>GIS Shape File</td>
<td>2010</td>
<td>• Population by Race per Tract</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Total Population per Tract</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Housing Units per Tract</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Tract Size (mi&lt;sup&gt;2&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Roadway Network</td>
<td>Arkansas Department of</td>
<td>GIS Shape File</td>
<td>2020</td>
<td>• Distribution of Roadways by</td>
</tr>
<tr>
<td></td>
<td>Transportation</td>
<td></td>
<td></td>
<td>Functional Classification</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(linear kilometers)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Distribution of Rural</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Roadways per Tract</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(linear kilometers)</td>
</tr>
<tr>
<td>Daily Vehicle Miles</td>
<td>Arkansas Department of</td>
<td>Excel File</td>
<td>2018</td>
<td>• DVMT per Tract</td>
</tr>
<tr>
<td>Traveled Crashes per</td>
<td>Transportation</td>
<td></td>
<td></td>
<td>• Number of Crashes per Tract</td>
</tr>
<tr>
<td>Tract</td>
<td>Arkansas State Police</td>
<td>Excel File</td>
<td>2016</td>
<td>• Food Desert Presence</td>
</tr>
<tr>
<td>Grocery Store Locations</td>
<td></td>
<td>GIS Shape File</td>
<td>2016</td>
<td>• Number of Grocery Stores</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Grocery Store Density</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(grocery stores per 10000 people)</td>
</tr>
</tbody>
</table>

3.2.1. Rural Designation

A tract’s designation as rural needs to be taken into account to measure the degree to which the number of crashes is affected by rural characteristics not captured by other built environment variables (Amoros, Martin, and Laumon 2003; Noland and Oh 2004). Research indicates there is an increased probability of being in a fatal crash on rural roads (Raymond et al. 2022).
Arkansas, 24% of all tracts are considered rural and contain all the rural roads in Arkansas, as determined from the Arkansas Road Inventory shape file provided by the Arkansas Department of Transportation (ARDOT).

Designation of whether a tract was considered rural or urban was determined utilizing the definition from the US Census Bureau: a tract is considered urban if the population exceeds 2,500 people, and all other tracts are considered rural (Rhone et al. 2017). This is the same definition utilized by the USDA when designating tracts as urban or rural for their food desert analysis. The urban or rural designation for tracts was displayed in the model as an independent indicator variable in which one indicates the tract is rural and zero indicates the tract is urban. By using the definition, it was identified that 25% of tracts in Arkansas were rural tracts (Figure 4).

Figure 4. Urban and Rural Designation of Tracts
3.2.2. Food Desert Designation

Since the focus of this research is to determine the degree to which food deserts affect the frequency of crashes in a census tract, a spatial analysis had to be conducted on the tracts in Arkansas to identify which tracts were considered food deserts. This work adopts the low-access food desert definition created by the USDA, as previously discussed in Chapter 1, but with modification as follows.

The spatial analysis method was slightly different than the process utilized by the USDA. The USDA’s analysis of tracts being food deserts uses spatial buffers to define the percentage of the population or number of people that had access to grocery stores within a certain distance (either 1 mile in urban tracts or 10 miles in rural tracts). In the USDA approach, a straight-line distance was used. To better consider transportation network access, instead of creating buffers using a straight-line distance, a routed distance measure was used. Routed distance is defined as the traveled distance along the roadway network (represented as lines in the GIS software) that connects the food source (represented as a point in the GIS software) to the center of the tract.

The steps followed for the mapping of food deserts following the USDA approach in this study are as follows:

1. Tract-level population data from the 2010 census were allocated down to 1-kilometer-square grids for all tracts across the state of Arkansas (Figure 5).
2. From each grocery store location (defined as a point), either a 1-mile or 10-mile (depending on urban or rural tract designation) roadway network buffer was created (Figure 6).
3. All grid squares that were touched by the buffers were considered to have network access to grocery stores therefore each square’s population was counted towards having access to a food retailer (Figure 7).

![Figure 7. Example of Identifying Which Grid Squares had Access to Grocery Stores](image)

4. Since each square belonged to a specific tract, the population of each square that was within the buffers would have its population summed together with the other squares in that buffer that shared the same tract.

5. If more than 500 people or 33% of a tract’s population did not have access to a grocery store, the tract would be designated as a food desert.

The designation of a tract would be used as an indicator variable in the final model that would be utilized to identify the degree to which the presence of food deserts influences the number of expected crashes occurring in a tract; a one (1) would indicate that the tract is a food desert while a zero (0) would indicate that the tract is not a food desert.
3.2.3. Characteristics of the Food Desert

The food desert characteristics that are being utilized for the model are the number of grocery stores in a tract and grocery store density (grocery stores per 10,000 people). These two (2) characteristics were chosen to be included in the final model because other studies also include extra variables that describe the environment being investigated. For example, in a study investigating the environmental attributes associated with pedestrian–vehicular crashes near public schools, while the public school was the main focus of the study the researchers added characteristics of the schools into the model, such as recreation facilities on the property, presence of driveways, etc. (Clifton and Kreamer-Fults 2007a).

Although food deserts are not physical places, in the way a school is, there are still specific characteristics that it can have that make a food desert subjectively “worse” or “better” than other food deserts, like the number of grocery stores in a tract or the grocery store density. These variables were added to the model as independent continuous variables. Based on the grocery store location data collected, the maximum number of grocery stores found in a tract was six (only one tract had six grocery stores) while 39% of all tracts had zero grocery stores (Table 2). Previous literature found that in Arkansas there were approximately 2 grocery stores per 10,000 Arkansans (Winthrop Rockefeller Institute 2022), conversely, the data utilized for this study showed that the maximum grocery store density for a tract was 22.55 but the average grocery store density for this data set was 3.32 (Table 2).
Table 2. Descriptive Statistics for Food Desert Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Grocery Stores</td>
<td>0</td>
<td>6</td>
<td>1.08</td>
<td>1.21</td>
</tr>
<tr>
<td>Grocery Store Density (Grocery Store per 10,000 people)</td>
<td>0</td>
<td>22.55</td>
<td>3.32</td>
<td>3.32</td>
</tr>
</tbody>
</table>

3.2.4. Racial and Ethnic Composition

The racial and ethnic composition of a tract could affect the number of fatal crashes in a county because there is evidence that Black and Hispanic Americans have higher traffic fatality rates per mile traveled than White Americans (Raifman and Choma 2022).

The racial and ethnic composition of each tract was taken from the attributes listed for each tract in the TIGER files publicly available for Arkansas tract boundaries created by the U.S. Census Bureau. The races listed in the data are: White, Black/African American, American Indian/Alaska Native, Asian, Native Hawaiian/Pacific Islander, or Other. The variable was represented as the number of people who identify as that race in each tract; if added together the population of each race in a tract would equal the total population within the tract. It is important to note that 75% of the population of Arkansas is white with the next largest racial population being Black/African Americans representing 15%. By observing the maximum number of each racial group in a tract, it is evident that the white and black populations are the majority in Arkansas (Table 3).
Table 3. Descriptive Statistics for Racial Groups per Tract

<table>
<thead>
<tr>
<th>Race</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>1</td>
<td>4498</td>
<td>313</td>
<td>482</td>
</tr>
<tr>
<td>White</td>
<td>0</td>
<td>7459</td>
<td>2507</td>
<td>1332</td>
</tr>
<tr>
<td>Black</td>
<td>0</td>
<td>4521</td>
<td>546</td>
<td>755</td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>0</td>
<td>261</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>Asian</td>
<td>0</td>
<td>3114</td>
<td>62</td>
<td>161</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>0</td>
<td>1048</td>
<td>17</td>
<td>87</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>64</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>

By mapping the populations of racial and ethnic groups in Arkansas the geographic distribution of the racial groups can be seen across Arkansas (Figure 8). From the mapping, it seems like that majority of the black population is in southeastern Arkansas while the white population is widespread, but it has a more concentrate population in northwestern Arkansas (Figure 8).
Figure 8. Distribution of Racial Populations per Tract
3.2.5. *Roadway Types*

Roadway characteristics include roadway type, number of lanes, lane width, shoulder size, etc. These characteristics are commonly used as independent variables in crash count or crash rate models applied to specific intersections or roadway segments. For the purpose of this model, a broad level of roadway characteristics is more suited (Li et al. 2013; Amoros, Martin, and Laumon 2003; Noland and Oh 2004). For this model, six roadway function classifications and rural roadways are represented in the model (Table 4).

Table 4. Details and Descriptions of Roadway Types

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Representing</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Func1</td>
<td>Interstates</td>
<td>km</td>
<td>Arterial roads providing the highest level of mobility over the longest uninterrupted distance; speed limits are usually between 55-75 MPH</td>
</tr>
<tr>
<td>Func2</td>
<td>Other Freeways &amp; Expressways</td>
<td>km</td>
<td>Expressways are designed to connect two or more roads, while freeways are designed to connect two or more cities; speed limits are usually between 50-70 MPH</td>
</tr>
<tr>
<td>Func3</td>
<td>Other Principal Arterials</td>
<td>km</td>
<td>Connect urbanized areas, cities, and industrial centers to one another; speed limits are usually between 50-70 MPH</td>
</tr>
<tr>
<td>Func4</td>
<td>Minor Arterials</td>
<td>km</td>
<td>Connect urbanized areas, cities, and industrial centers to one another; speed limits are usually between 50-70 MPH</td>
</tr>
<tr>
<td>Func5</td>
<td>Major Collectors</td>
<td>km</td>
<td>Major and minor roads that connect local roads and streets with arterials; speed limits are usually between 35-55 MPH</td>
</tr>
<tr>
<td>Func6</td>
<td>Minor Collectors</td>
<td>km</td>
<td>Major and minor roads that connect local roads and streets with arterials; speed limits are usually between 35-55 MPH</td>
</tr>
<tr>
<td>Rural_Rd</td>
<td>Rural Roadways</td>
<td>km</td>
<td>Roads designated as being rural due to their location; these can be paved or unpaved roads</td>
</tr>
</tbody>
</table>
The independent variables used in the model are the total length of each roadway type in kilometers. By utilizing a rural road variable, the model will be able to measure the degree to which the kilometers of rural roads in a tract have any effect on the number of expected crashes. This is an important variable to include because one of the assumptions in this study is that since most of the food deserts are in rural Arkansas those residents will be driving on roadways that are majority rural. This data was obtained by the roadway inventory attributes in the roadway network shape file provided by ARDOT. From an analysis completed on the descriptive statistics of roadways in Arkansas, it is shown that the roadway type that has the most roadway in a single tract is the major collector (Table 5). Major Collector roadways contained the largest length of roadways in a single tract with one tract having 254 kilometers of major collector roadways based on the network utilized in this study (Table 5).

Table 5. Descriptive Statistics for Roadways by Functional Class or Rural Designation per Tract

<table>
<thead>
<tr>
<th>Roadway Functional Class or Rural Designation</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interstates</td>
<td>0</td>
<td>71</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Other Freeways &amp; Expressways</td>
<td>0</td>
<td>39</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Other Principal Arterials</td>
<td>0</td>
<td>89</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Minor Arterials</td>
<td>0</td>
<td>101</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Major Collectors</td>
<td>0</td>
<td>254</td>
<td>31</td>
<td>39</td>
</tr>
<tr>
<td>Minor Collectors</td>
<td>0</td>
<td>87</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Rural Roadways</td>
<td>0</td>
<td>17</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Notably, major collectors also make up the largest number of roadways in the Arkansas roadway network. Major collectors comprise 47% of all roadways in Arkansas according to the analysis completed on the roadway network created for this study (Figure 9). Since the roadway network that was created for this study combined different data sets to create a comprehensive
roadway map, the roadway map created could have not included all roadways in Arkansas. All roadway network data was collected from the Arkansas Department of Transportation therefore there could have been local and county roadways that were missing from their inventory since the Arkansas Department of Transportation does not maintenance them.

![Distribution of Roadway Types by Functional Class and Designation](image)

**Figure 9. Distribution of Roadway by Function Class and Designation in the Arkansas Roadway Network**

**3.2.6. Daily Vehicle Miles Traveled (DVMT)**

One of the main points that are being tested with this analysis is that because people living in food deserts must drive farther distances to access food there is a chance that those drivers may have a higher chance of being in vehicular crashes. Since the final model does not predict crash
rate but rather the number of crashes possible in a tract based on tract attributes, this independent variable will control for the exposure as suggested by prior studies.

Distance is considered as the measure of exposure a driver has on the roadway and is expressed as a model variable in terms of Daily Vehicle Miles Travelled (DVMT). Exposure being the amount of distance that a person spends traveling which in turn affects the likelihood of them being in a traffic-related crash (Merlin, Guerra, and Dumbaugh 2020). DVMT represents the mileage traveled by all vehicles on a road system over an average day in a year. DVMT will be measured in miles for two (2) roadway types: county road DVMT and city road DVMT. This way the model can identify if there are different effects on the number of crashes happening in a tract based on the miles traveled on a specific type of roadway. Descriptive statistics for DVMT per tract on county roads and all other roadways was completed (Table 6).

Table 6. Descriptive Statistics for DVMT Type per Tract

<table>
<thead>
<tr>
<th>DVMT Type</th>
<th>Minimum</th>
<th>Maximum (in Thousands)</th>
<th>Mean (in Thousands)</th>
<th>Standard Deviation (in Thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVMT on County Roads</td>
<td>44</td>
<td>162.6</td>
<td>11.3</td>
<td>16.1</td>
</tr>
<tr>
<td>DVMT on All Other Roadways</td>
<td>77</td>
<td>487.9</td>
<td>20.4</td>
<td>42.5</td>
</tr>
</tbody>
</table>

Based on the gathered data from the ARDOT ‘Road and Street Report’ it shows that out of all daily vehicle miles traveled by Arkansans, 65% of them occur on state highways or city roads while only 35% occur on county roadways. Since half of all fatal crashes occur on rural roadways, DVMT on county roads may be a significant variable that impacts the number of fatal crashes in a tract. The DVMT is represented by two continuous variables, county road DVMT and city road DVMT.
3.2.7 Number of Crashes

The dependent variable for the model is the number of crashes of all severity types in a tract for the year 2016 since that is the year food accessibility was mapped for in this study. This data was collected from the Arkansas State Police (ASP) crash records. By plotting where in Arkansas all the crashes occurred a map could be created to show the distribution of crashes per tract throughout Arkansas (Figure 10). There seems to be a higher number of crashes per tract in areas of Arkansas that are more populated such as central Arkansas, near Little Rock, AR, or northwest Arkansas, near Fayetteville, AR (Figure 10).

![Figure 10. Distribution of Crashes per Tract](image-url)
By including all severity types in the final count for the variable representing the number of crashes in a tract, this ensured that there were no excessive zeroes in the data since all tracts, but three, had a crash of some severity occur in 2016, therefore, the utilization of a zero-inflated model was unnecessary (Figure 11).

Figure 11. Histogram of Crash Incidences per Tract

Through a look into the descriptive statistics of this variable to understand the distribution of the number of crashes per tract, it was discovered that the average tract had approximately 86 crashes occur within their borders in 2016 (Table 7).

Table 7. Descriptive Statistics for the Number of Crashes per Tract

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum (in Thousands)</th>
<th>Mean (in Thousands)</th>
<th>Standard Deviation (in Thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Crashes per Tract</td>
<td>0</td>
<td>845</td>
<td>86</td>
<td>92</td>
</tr>
</tbody>
</table>

While an excessive representation of zeroes in the count data was avoided, there were data transformation and preparations that had to be completed on the data for each variable previously
mentioned to ensure the final model was able to provide a good and accurate fit therefore allowing the determination of to what degree food deserts’ presences and other physical attributes of a tract effect the expected number of crashes in a tract.

3.3. Data Preparation and Modeling

While much of the literature advises utilizing Zero-inflated or Negative Binomial Poisson regression models when working with count data, as is being used in this study, the models suggested did not fit the data utilized (Abdulhafedh 2017; Rakha et al. 2010). Since the crash data included all severity crashes in its counts, there were only three tracts out of all 823 that had no crashes happen in 2016, there was not an excess of zeroes in the data meaning the zero-inflated regression model would not be necessary. Due to the guidance of the literature review, there was an attempt to utilize the Negative Binomial Poisson Regression model, but the lack of overdispersion in the data in this model did not warrant the use of a negative binomial regression as suggested by past research (Abdulhafedh 2017; Li et al. 2013; Khattak et al. 2021).

To apply linear regression models to determine the degree to what effect the built environment affects crash rate or the number of expected crashes, the following assumptions of a linear regression must be met (Abdulhafedh 2017; Clifton and Kreamer-Fults 2007a; Noland and Oh 2004; Amoros, Martin, and Laumon 2003; Angel and Hickman 2008):

- A linear relationship between the independent variable and the dependent variable.
- The residuals of the model are independent.
- The data must have homoscedasticity.
- The residuals of the model must be normally distributed.
• There must not be multicollinearity; meaning there should not be a high degree of linear intercorrelation between independent variables (Kim 2019).

All the continuous data utilized for the model had to be transformed to meet the normal distribution assumption of a linear regression model. The raw data for all the variables had a positive skew which was determined by visualizing the data in histograms. Since all the continuous data had a positive skew, a square root transformation was applied to all continuous variables which lessened the positive exhibited by the data. Although all the continuous data was less skewed than its original form, the variable representing the number of crashes in a tract caused some overdispersion in the data set.

Through the removal of the outliers, determined to be any data points farther than 1.5 standard deviations from the mean, in the number of crashes data, the data fit a normal distribution as required by linear regression models. The removal of outliers did not have a significant effect on the size of the dataset; instead of 823 data points, which is the number of tracts in Arkansas, 788 data points were utilized in the analysis. The histograms created through the data preparations described are shown in Figure 12.
(a) Histogram of the Original Count Data for the Number of Crashes per Tract

(b) Histogram of the Count Data for the Number of Crashes per Tract After Square Root Transformation

(c) Histogram of the Count Data for the Number of Crashes per Tract After Square Root Transformation and Outliers were removed

Figure 12. Histograms for the Number of Crashes Variable thru the Data Preparation Process

To ensure that independent variables with high multicollinearity were not added to the model, an examination of the independent variables’ relationship with one another was investigated. Multicollinearity defies the assumptions of a linear regression model which could make the final model unreliable (Johnston, Jones, and Manley 2018; Kim 2019). By finding the variance inflator factor (VIF), which is utilized to detect multicollinearity, each independent
variable, the variables that exhibit VIF values higher than five were removed from the model since a value greater than five indicates multicollinearity exists (Kim 2019). Prior to fitting the model with the data, all the variables’ VIF values were calculated. Any variable with a score higher than a five was excluded from further modeling. This left 17 possible variables to be utilized in the final model (Table 8).

Table 8. Variance Inflator Factor Values for Variables Before and After Removal of Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF Value Before Removal of Variables</th>
<th>VIF Value After Removal of Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land Area</td>
<td>19.42</td>
<td>-</td>
</tr>
<tr>
<td>Total Population</td>
<td>569.38</td>
<td>-</td>
</tr>
<tr>
<td>Total Housing Units</td>
<td>201.42</td>
<td>-</td>
</tr>
<tr>
<td>Hispanic Population</td>
<td>3.17</td>
<td>2.37</td>
</tr>
<tr>
<td>White Population</td>
<td>4.02</td>
<td>2.45</td>
</tr>
<tr>
<td>Black / African American Population</td>
<td>1.79</td>
<td>1.69</td>
</tr>
<tr>
<td>American Indian or Alaskan Native Population</td>
<td>3.37</td>
<td>1.83</td>
</tr>
<tr>
<td>Asian</td>
<td>4.15</td>
<td>1.78</td>
</tr>
<tr>
<td>Native Hawaiian / Pacific Islander Population</td>
<td>2.48</td>
<td>1.80</td>
</tr>
<tr>
<td>Population Identifying as Another Race</td>
<td>3.89</td>
<td>1.45</td>
</tr>
<tr>
<td>Housing Density</td>
<td>8.59</td>
<td>-</td>
</tr>
<tr>
<td>Number of Grocery Stores</td>
<td>4.14</td>
<td>1.76</td>
</tr>
<tr>
<td>Grocery Store Density</td>
<td>3.75</td>
<td>1.69</td>
</tr>
<tr>
<td>Total Roadway Length</td>
<td>358.06</td>
<td>-</td>
</tr>
<tr>
<td>Length of Interstates</td>
<td>2.90</td>
<td>1.08</td>
</tr>
<tr>
<td>Length of Other Freeways &amp; Expressways</td>
<td>1.51</td>
<td>1.02</td>
</tr>
<tr>
<td>Length of Other Principal Arterials</td>
<td>4.14</td>
<td>1.28</td>
</tr>
<tr>
<td>Length of Minor Arterials</td>
<td>3.77</td>
<td>1.34</td>
</tr>
<tr>
<td>Length of Major Collectors</td>
<td>106.55</td>
<td>-</td>
</tr>
<tr>
<td>Length of Minor Collectors</td>
<td>14.03</td>
<td>-</td>
</tr>
<tr>
<td>Length of Rural Roadways</td>
<td>4.09</td>
<td>2.22</td>
</tr>
</tbody>
</table>
Table 8. (Cont.)

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF Value Before Removal of Variables</th>
<th>VIF Value After Removal of Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVMT on County Roads</td>
<td>2.94</td>
<td>3.20</td>
</tr>
<tr>
<td>DVMT on All Other Roadways</td>
<td>4.95</td>
<td>1.88</td>
</tr>
<tr>
<td>Food Desert Designation</td>
<td>4.85</td>
<td>1.50</td>
</tr>
<tr>
<td>Rural Designation</td>
<td>17.43</td>
<td>-</td>
</tr>
</tbody>
</table>

After all data were transformed using a square root function, outliers for number of crashes variable were removed, and independent variables were tested for multicollinearity, the data was fit using a linear regression model which was to determine the degree to which the presence of food deserts affect the number of expected crashes in a census tract.
Chapter 4. Results

The following section outlines the results from the food accessibility mapping method and details the results from the linear regression model. Combined, these results address the objective of the study: to estimate the degree to which the presence of a food desert affects the expected number of crashes.

4.1. Food Accessibility Mapping

The number of grocery stores in each tract was one of the independent continuous variables utilized in the linear regression model. 321 (39.0%) tracts did not have a grocery store within its boundaries while only 5.2% of tracts had four or more grocery stores within its boundaries (Figure 13).

Figure 13. Number of Grocery Stores per Tract
Grocery store density, represented at grocery stores per 10,000 people, was also utilized as an independent variable in the final model to add food desert characteristics into the modeling. From an analysis of grocery store density, it was determined that 505 (61.36%) tracts have a grocery store density of 3.32 or less, which is the grocery store density for the average tract (Figure 14).

![Figure 14. Grocery Store Density per Tract](image)

The mapping of food accessibility utilizing the mapping method designed for this study resulted in 598 (72.9%) tracts being identified as food deserts in Arkansas (Figure 15). From this mapping it was determined that 23.4% of food deserts are in rural tracts while 76.6% of food
deserts are in urban tracts. It is important to note that urban or rural designation was assigned by the U.S. Census Bureau’s definition stating tracts with populations lesser than 2500 people are considered rural tracts.

Figure 15. Food Desert Map Produce through Study Procedures

While the mapping of food accessibility by utilizing the roadway to measure physical distance resulted in 598 out of 823 tracts in Arkansas being identified as food deserts, the USDA only identified 386 (46.9%) tracts as being food deserts utilizing their methods and procedures
From the tracts identified as being food deserts from both mapping methods, there were 282 tracts identified as food deserts that were found in both maps (Figure 16).

Figure 16. Comparison Between the USDA Food Accessibility Map and the Food Accessibility Map from this Study

Although the USDA’s method yielded less tracts as being food deserts, it yielded a higher number of urban tracts being food deserts than the mapping method utilized in this study. The USDA’s mapping method yielded approximately 2% more urban tracts being identified as food deserts.

4.2. Linear Regression Model

Utilizing the stepwise regression method for the development of the linear regression model, the fitted model contained 11 statistically significant independent variables (p-values less than 0.05 indicated statistical significance) (Table 9). The variables represented in the model represented
common measures of the built environment like racial composition, kilometers of roadway types, and DVMT. The linear regression model also found the presence of food deserts as statistically significant enough to affect the number of crashes in a census tract (Table 9).

Table 9. Results from the Linear Regression Model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimates</th>
<th>Standard Error</th>
<th>p-Value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.419</td>
<td>0.495</td>
<td>0.000</td>
<td>30.99</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.057</td>
<td>0.011</td>
<td>0.000</td>
<td>2.39</td>
</tr>
<tr>
<td>White</td>
<td>0.044</td>
<td>0.009</td>
<td>0.000</td>
<td>2.45</td>
</tr>
<tr>
<td>Black</td>
<td>0.043</td>
<td>0.009</td>
<td>0.000</td>
<td>2.10</td>
</tr>
<tr>
<td>Other</td>
<td>0.2539</td>
<td>0.093</td>
<td>0.006</td>
<td>1.80</td>
</tr>
<tr>
<td>Number of Grocery Stores</td>
<td>0.4244</td>
<td>0.141</td>
<td>0.003</td>
<td>1.79</td>
</tr>
<tr>
<td>Interstates</td>
<td>0.4278</td>
<td>0.058</td>
<td>0.000</td>
<td>1.08</td>
</tr>
<tr>
<td>Freeways or Expressways</td>
<td>0.4425</td>
<td>0.099</td>
<td>0.000</td>
<td>1.02</td>
</tr>
<tr>
<td>Other Principal Arterials</td>
<td>0.1421</td>
<td>0.054</td>
<td>0.009</td>
<td>1.28</td>
</tr>
<tr>
<td>DVMT on County Roads</td>
<td>-0.0083</td>
<td>0.002</td>
<td>0.000</td>
<td>3.24</td>
</tr>
<tr>
<td>DVMT on All Other Roads</td>
<td>0.0043</td>
<td>0.001</td>
<td>0.001</td>
<td>1.88</td>
</tr>
<tr>
<td>Food Desert</td>
<td>-2.6446</td>
<td>0.259</td>
<td>0.000</td>
<td>1.56</td>
</tr>
</tbody>
</table>

N = 788
R² = 0.42
Adjusted R² = 0.40
F-statistic = 49.75
Prob (F-statistic) = 2.00 e -82
Skew = 0.123
Kurtosis = 3.746
Durbin-Watson = 1.709

Since the residuals need to be normally distributed to meet the assumptions of linear regression, the skew, measure of data symmetry, and kurtosis, measure of data curvature, of the residuals were tested. The calculated skew of the residuals was 0.123 meaning that the residuals of the fitted model are nearly symmetrical, and the kurtosis value was 3.746 meaning the residuals are slightly flatter than a normal distribution but could still be considered within a
normal distribution. The residuals of the fitted model have a randomly scattered pattern around zero indicating homoscedasticity in the residuals of the fitted model corroborating the results of the skew and kurtosis tests of the model (Figure 17).

Figure 17. Residual Plot for the Fitted Linear Regression Model

A Durbin-Watson test statistic tests the null hypothesis that the residuals from an ordinary least-squares regression are not autocorrelated; a value between 1 and 2 from the Durbin-Watson tests indicated homoscedasticity in the residuals for the model (Thu, May 2019). The Durbin-Watson value for this linear regression model was 1.709 signifying that the residuals are homoscedastic as required. The r-squared value for this model is 0.41 which is within the ranges achieved by prior research investigating the relationship between the built environment and crash frequency with the use of a linear regression (Li et al. 2013; Clifton and Kreamer-Fults 2007a; Amoros, Martin, and Laumon 2003; Noland and Oh 2004); the range for r-squared values in the previous studies were from 0.36 to 0.74. The r-squared value for this study ($R^2 = 0.42$) is
most similar to the r-squared value from Fischer, Sternfeld, and Melnick (2013), which had a r-squared value of 0.42 and Li et al. (2013), which had a r-squared value of 0.41.

While the fitted linear regression found many demographic and physical attributes of tracts that impact to a degree the number of expected crashes, the main purpose of this study was to identify if the presence of a food desert was statistically significant enough to impact the number of expected crashes, and if so, to what degree. From the results of the model, it is shown that the presence of a food desert in a tract does have a statistically significant effect (p-value < 0.05) on the number of crashes expected in a tract. A tract being identified as a food desert has a negative effect on the number of expected crashes in a tract; a tract being identified as a food desert reduces the number of expected crashes by approximately seven crashes. Although two food desert characteristics were included in the model (number of grocery stores and grocery store density), only the number of grocery stores in a tract was statistically significant. The number of grocery stores has a positive effect on the number of expected crashes therefore for every 10 grocery stores in a tract approximately two crashes can be added to the overall count of expected crashes in a tract.

Aside from the food desert attribute variables, the model also found positive relationships between the population variables and the kilometers of roadway by function class variables. From the seven racial population variables used in the modeling, only four were found to be statistically significant; the populations were Hispanic, White, Black, and anyone identifying as any other race. All the racial populations found to be significant had a positive relationship with the number of expected crashes in a tract, so as the populations increased so would the expected number of crashes. If every population increased by 10,000 people, the following would be the increase in the expected number of crashes: approximately 20 crashes for the increase in the
White population, approximately 19 crashes for the increase in the White population, approximately 33 crashes for the increase in the Hispanic population, and approximately 645 crashes for the increase in the other category. The Hispanic and Other populations would have the greatest impact on the number of expected crashes due to the coefficients assigned to those variables.

Furthermore, the other common measures of built environment that was included in this study was the length of roadway types by classification. The study included six functional classes and a variable for the kilometer or rural roadways in a tract. While seven variables were added to the model to represent the streetway configuration of a tract, only three function classes were found to be significant. The three significant variables represented kilometers of interstates, freeways and expressways, and other principal arterials; all of which are highspeed, multilane roadways. As per previous studies, these variables had a positive relationship with the number of expected crashes which means as the kilometers of each roadway increases so would the number of expected crashes. If every roadway functional class increased by 10,000 kilometers, the number of crashes would increase by the following amounts: approximately 1800 crashes for an increase in the kilometers of interstate, approximately 1950 crashes for an increase in the kilometers of freeways or expressways, and approximately 200 crashes for an increase in the kilometers of arterials. It is important to note though that the kilometers of roadways by functional class was determined utilizing the roadways network built using two different roadway files that were provided from ARDOT, therefore, there is a possibility that not all existing roadways in Arkansas were present in the data. ARDOT’s roadways files only contain roadways that they maintenance, so local roads, which are maintenance by cities or counties, are not as represented in the data files.
While most of the demographic, physical attributes, and DVMT variables of a tract had a positive effect on the overall number of expected crashes, the variable representing DVMT on county roads, which are all rural roads, had a negative effect on the dependent variable. This means that for every 100,000 daily vehicle miles traveled on county roads the overall number of expected crashes in a tract will be reduced by approximately seven. Conversely, for every 100,000 daily vehicle miles traveled on city or highways roadways the expected number of crashes will increase by approximately two crashes since the “other roadways DVMT” variable has a positive effect on the dependent variable. Further discussion on the results from the mapping of food accessibility and the linear regression is presented in the next chapter.
Chapter 5. Discussion

The following section will discuss the implications of the results gathered from the linear regression model. Since this study was a “proof of concept” study to identify if there was any significant statistical relationship between the presence of food deserts on the number of crashes, and if so to what degree, the results from this study have created new avenues for potential research therefore that is also discussed in this section.

5.1. Racial and Ethnic Influence

While the population of seven races and one ethnic group, Hispanic, were added to the linear regression model, only three racial groups, White, Black, and Other Race populations, and the Hispanic population were found to be statistically significant. Both the racial populations and the Hispanic population had a positive effect on the expected number of crashes; as the populations of these racial and ethnic groups increase so will the expected number of crashes in a tract. This finding aligns with studies that have found that populations of people of color (POC) are disproportionately involved in crashes no matter the travel mode utilized (Glassbrenner et al. 2022; Raifman and Choma 2022). Similarly, in studies that investigate how racial population distributions, as a measure of the built environment, affect crash tendencies it is posited that populations of non-white racial populations have a positive with the number of crashes in the study location (Guerra, Dong, and Kondo 2019; Clifton and Kreamer-Fults 2007a). Though it is important to note that this model did not contain information about the people involved in crashes simply the racial composition of the overall tract since this study was focused on discovering what physical attributes of a tract influenced the expected number of crashes; racial composition was the only socio-economic attribute of an overall tract that was included in the study.
5.2. Roadway Type Influence

The model testing included continuous data on six roadway functional classes: the kilometers of interstates, other freeways/expressways, other principal arterials, minor arterials, major collectors, minor collectors, and there was also a variable that represented the total kilometers of rural roadways in a tract. From the results, only three functional classes, interstates, other freeways/expressways, and other principal arterials, had a statistically significant positive effect on the number of expected crashes; as the number of total kilometers of each of the aforementioned roadway classes increased so would the expected number of crashes. The positive relationship exhibited by the variables in this discussion are on par with previous studies that utilized roadway distribution by functional class in the modeling between measures of the built environment and crash incidences (Guerra, Dong, and Kondo 2019; Dumbaugh and Rae 2009; Dumbaugh et al. 2012; Yu and Xu 2018; Merlin, Guerra, and Dumbaugh 2020; Li et al. 2013; Høye and Hesjevoll 2020; Martensen and Dupont 2013). The studies found that roadways that were highspeed and had multiple lanes, like interstates or arterials, had a positive relationship with crash incidences. Furthermore, Noland and Oh (2004) report that as average number of lanes in a study location increases so do the number of crashes. Previous studies suggest that interstates and arterial roadways have more of a statistically significant effect on the number of crashes occurring in an area and the crash severity due to interstates and arterials correlation with higher traffic volume. Moreover, traffic volume may be associated with different traffic densities on different roadway types, and traffic density has been found to be associated with crash numbers (Lord, Manar, and Vizioli 2005). Since rural roads have less volume, this could explain why the kilometers of rural road in a tract was not statistically significant enough to impact the expected number of crashes.
5.3. Daily Vehicle Miles Traveled Influence

Daily vehicle miles traveled was included as a variable into the model to account for a driver’s exposure. DVMT was represented through DVMT on county roads, which are mostly rural roads, and DVMT on all other roadways. While DVMT on all other roads (state highways and city roads) had a positive effect on the expected number of crashes, DVMT on county roads had a negative effect which is unusual when compared to other studies (Li et al. 2013; Tasic and Porter 2016; Dumbaugh and Rae 2009; Dumbaugh et al. 2012). Usually it is expected as DVMT increases so does the expected number of crashes since an increased DVMT signifies more driving which in turn creates more opportunity for conflict and more crashes (Tasic and Porter 2016). In this study the DVMT on county roads had a negative effect on the dependent variable, signifying that as DVMT increases the expected number of crashes decreases. This result is in direct opposition of what previous research has theorized; since rural roads are inherently more dangerous due to design and environmental attributes there is an increased chance of being in a vehicular crash if DVMT increases (Raymond et al. 2022). While it is known that county roads are rural roads it, needs to be noted that the DVMT on state highways or city streets could also be rural roads due to their location but there was no information detailing whether the state highways or city streets were rural unlike for county roads. Although a negative effect by DVMT on county roadways was determined by the model, the DVMT on all other roadways had a positive relationship on the number of expected crashes, and this is concurrent with past studies that utilize DVMT as a measure of exposure.

5.4. Food Desert Presence’s and Characteristics’ Influences

While determining if simple measures of the built environment, like kilometers of roadway, of a tract had an effect on the expected of number of crashes in a tract was insightful, the main
objective of this study was to determine if the presence of a food desert in a tract had a statistically significant relationship with the number of expected crashes in the tract since little to no research had been done to investigate this relationship. From the results produced by this study, there is evidence pointing that the presence of food desert in a tract does have an effect on the number of crashes in a tract, but not in the way hypothesized. Initial hypotheses for this study theorized that since food deserts were mostly in rural tracts in Arkansas and rural tracts contain all rural roads, which literature states are more dangerous compared to their urban counterparts, that the presence of a food desert, if statistically significant at all, would have a positive effect on the number of expected crashes in a tract. Since there is little literature on the relationship between food desert presence and expected crashes, there can only be speculative discussion around why there could be a negative effect. One reason a food desert’s presence has a negative effect on the number of expected crashes in a tract could be because the inhabitants of the tract are travelling outside of the tract they reside in to access food retailers. By travelling outside of their home tract to access food retailer or other services, residents are causing a decrease in the tract DVMT which would decrease the number of expected crashes as suggested by the model results and other studies (Tasic and Porter 2016; Li et al. 2013), as long as the DVMT is being accrued on state highways and city roadways; this seems possible since rural roadways, as categorized by ARDOT, only comprise 6% of roadways in Arkansas as per the roadway map developed for this study. This speculation is further supported by the positive effect the number of grocery stores in a tract has on the number of expected crashes.

The results for the model created in this study also suggest that as the number of grocery stores in a tract increases so do the number of expected crashes due to the positive relationship between those two variables. This could help explain why the presence of a food desert has an
inverse relationship with the number of expected crashes in a tract. Since 39% of tracts have no
grocery store within its borders, those residents could be driving more on city streets and
highways to get to a tract with food retailers this in turn could increase DVMT in tracts with
more grocery stores while decreasing DVMT within tracts with food deserts since residents are
travelling less within the tract. As reported by the model, DVMT and the number of expected
crashes has a positive relationship; with an increase 1000 DVMT per tract the number of
expected crashes increases by approximately 18 crashes. This relationship between DVMT and
expected number of crashes is consistent with prior studies (Li et al. 2013; Dumbaugh and Rae
2009; Dumbaugh et al. 2012).

Another measure of the built environment that could help explain the current results
could be the land use makeup of tracts, but this variable was not included in this model due to
lack of accessible data for each tract. From the previously discussed studies, it can be ascertained
that commercial land use has a statistically significant effect on the number of expected crashes
in a study location. Wier et al. (2009) found that both the percentage of neighborhood
commercial and the percentage of residential-neighborhood commercial has a positive
relationship with the expected number of pedestrian-vehicle crashes; a 15% increase in the
percentage of neighborhood commercial land use would increase the number of expected crashes
by 0.41% similarly, a 15% increase in percentage of residential-neighborhood commercial land
use would increase the number of expected crashes by 0.29% (Wier et al. 2009). Likewise,
Dumbaugh and Rae (2009) find that an increase in the total land designated as retail and
commercial use near arterial roadways increases the total number of crash incidences in the study
location. Studies from Xu and Yu (2018) and Chen (2015) also found similar results.
If the amount of commercial land use in a tract can increase the number of expected crashes with an increase in the total commercial land present, this could further explain why this study found an inverse relationship between the presence of food deserts and expected number of crashes in tracts, but a positive relationship regarding the number of grocery stores. In tracts with food deserts, 39% of those tracts do not have a grocery store. While not having a grocery store in a tract means there is no commercial land use, this could mean there is less commercial land use present than in tracts with 3 or more grocery stores. This could help explain why tracts with food deserts has an inverse relationship with the number of expected crashes. Since there are less grocery stores in these tracts there may be less commercial land use which previous research has concluded means there would be less crashes occurring due to the positive relationship these variables have with each other. In tracts with more grocery stores, it can be inferred that there would be more commercial land use which would increase the number of expected crashes (Dumbaugh and Rae 2009; Dumbaugh et al. 2012; Wier et al. 2009; Yu and Xu 2018; Chen 2015). Furthermore, this could explain the positive relationship between the number of grocery stores and the number of crashes in a tract.

Due to there being a gap in the research in regard to the topic of this study, further studies need to be completed to gather a strong understanding of what the results in this study mean. First, to test the speculation that commercial land use could be a factor into why the presence of a food deserts has an inverse relationship with the number of crashes in a tract but the number of grocery stores has the opposite effect, a new model needs to be created that includes land use. This would require finding or building a data set that contains the distribution of different land use types per tract. The land use types represented could be commercial, office, industrial, land use mix, and residential since those are the types of land uses most frequently found in previous
studies (Chen 2015; Wier et al. 2009; Yu and Xu 2018; Dumbaugh et al. 2012). If the commercial land use variable resulted in having a statistically significant positive effect on the number of crashes in a tract, this would support the speculations stated in this study.

Another avenue that needs to be explored to understand the results of this study more thoroughly is what do the trips of the residents in each tract look like. In this discussion, it is speculated that do to the presence of a food desert in a tract, residents need to leave their home tract to reach a food retailer. By leaving the tract with a food desert to go to one with a larger number of grocery stores, the DVMT of the tract with more grocery stores increases while the home tract less DVMT. This speculation though rests on the assumption that residents of tracts with food deserts are driving out of the tract to get to a grocery store. Through a different study conducted by the author on Arkansas drivers’ driving behaviors and trips, it was concluded that residents from rural, suburban, and urban areas all go to grocery stores at the same frequency. Residents from all levels of urbanicities go to grocery stores at least once or twice a week, but rural residents are driving farther distances to access grocery stores. The 2023 Arkansas Driver Awareness Survey gathered that 32% of respondents who identify as living in rural areas of Arkansas state that they must travel more than 21 minutes (or 20 miles) to get to a grocery store. In comparison, 5% and 1% of respondents who identified as living in suburban and urban, respectively, state they must travel more than 21 minutes (or 20 miles) to get to a grocery store.

Knowing this preliminary information on trips involving grocery stores as destinations, shows that rural residents are driving farther distances to access a grocery store, but the question lies in where the grocery stores are that these residents are trying to access. Do the long trips to the grocery stores that rural residents are taking happen within the tract they reside in, or are the residents traveling outside of their home tract to access a grocery store? If these long trips are
occurring within the same tracts they live in, then the speculation proposed as to why food deserts and the expected number of crashes has an inverted relationship is not plausible. To gather information on travel patterns of people trying to access a grocery store, a separate study would need to be conducted to gather detailed data on Arkansas residents’ travel patterns in regard to grocery store trips. Through this research, it could be determined if residents living in tracts with food deserts do drive to tracts with more grocery stores or are they staying within their home tracts to attend a grocery store if one is available.

Due to the nature of this study being a first step toward bridging the gap in the research between food deserts and their effect on the number of crashes occurring in an area, only speculations can be made as to why certain results were what they were. While more common measures of the built environment, like population size or street configuration, have many studies that have investigated their relationships with crash incidences, the same is not true for food deserts and its characteristics. This study achieved its objective of 1) determining whether there was a statistically significant relationship between the presence of a food desert and the number of expected crashes in a tract, and if there was 2) to what degree that effect was. Through the utilization of a linear regression model, it was determined that the presence of a food desert does have a statistically significant effect on the number of crashes in a tract, but the relationship is an inverse effect. Conversely, the number of grocery stores in a food desert, an attribute of a food desert, had a positive effect on the number of expected crashes. Since this study was only to understand if the relationship was statistically significant and if so, to what degree, there needs to be further research completed in this area to understand why these relationships are and concur or disprove the speculations made in this study.
Chapter 6. Conclusion

As the prevalence of food deserts has increased throughout society, the research into metrics used to identify them and their effects on residents has advanced. Although there are many studies dedicated to food deserts, the inclusion of transportation metrics to identify food deserts is very limited. Previous studies state that lack of transportation can be an inhibitor to food accessibility and is a common aspect presented by areas considered food deserts but is rarely used as the main metric to identify food deserts. Most past studies incorporate transportation as a measure of the population that owns a vehicle, if a study location has public transportation, and/or transportation cost (Walker, Keane, and Burke 2010; Dutko, Ploeg, and Farrigan 2012; Weatherspoon, Ploeg, and Dutko 2012). While previous studies have concluded that different transportation aspects can affect food access, it is uncommon to see it utilized in the mapping of food access in order to identify food deserts.

This study attempted to bridge this gap in food desert mapping methods by incorporating an aspect of the transportation network into mapping food access. While the USDA’s food access mapping methods are the most widely utilized in food desert research, this study changed a key aspect of the USDA’s method to identify food deserts. Instead of utilizing a straight-line distance buffer to identify what portion of the population would be without access to a grocery store, this study utilized buffers that were built utilizing the roadway networks of Arkansas; this was meant to capture a more nuanced look into what portions of the population would have access to grocery stores based on the built infrastructure. Through this mapping method, 598 (72.6%) tracts identified as being food deserts due to 33% of the tract population or 500 people in the tract not having access to a grocery store. The study mapping method identified more tracts as food deserts than the USDA for the year 2016; the USDA’s mapping method identified
386 (46.9%) tracts as food deserts. Both methods had 294 tracts in common that were identified as food deserts. Although the mapping method utilized for this study identified more tracts as being food deserts, the USDA’s mapping methods identified more urban tracts as being food deserts than this study’s mapping method. By utilizing roadway network buffers instead of straight-line distance buffers, the mapping method utilized in this study created a food desert map that used the availability of transportation infrastructure (i.e. roadway networks) as a metric to identify food deserts.

Although the mapping of food access in Arkansas was an important part of this study, the main purpose of this study was not to develop a new mapping method. The purpose of this study to identify if there was a statistically significant relationship between the presence of food deserts in a tract and the expected number of crashes in that tract. To determine if there was a significant relationship, a linear regression model was utilized to determine the relationship between different measures of the built environment of a tract, including food desert presence and attributes, and the expected number of crashes. This study focused on how built environmental attributes of tracts which do or do not contain food deserts affect the number of expected crashes in those tracts not how socioeconomic or behavioral attributes of people living in tracts affect the number of expected crashes, therefore, only independent variables included in the modeling efforts represented tract attributes such as population, street configuration, DVMT, and food desert attributes.

The resulting model concluded that of the 17 predictive variables included in the fitting of the model only 11 variables has a statistically significant effect on the expected number of crashes in a tract. As consistent with previous studies that investigate the effects of common measures of the built environment on crash incidences, all the population variables that were
found to be statistically significant, which included the Hispanic, White, Black, and Other populations, had a positive relationship with the expected number of crashes in tract (Wier et al. 2009; Guerra, Dong, and Kondo 2019; Dumbaugh and Rae 2009; Dumbaugh et al. 2012).

Similarly, the roadway functional classes that were found to be statistically significant were also consistent with previous studies that found high speed, multilane roadways, like arterials, to have a positive relationship with the expected number of crashes (Fischer, Sternfeld, and Melnick 2013; Yu and Xu 2018; Chen 2015; Guerra, Dong, and Kondo 2019; Dumbaugh and Rae 2009; Dumbaugh et al. 2012). Conversely, one of the measures for DVMT, which as DVMT on county roads, had an inverse relationship with the number of expected crashes while DVMT on all other roadways was found to have a positive relationship with the dependent variable. From previous studies, DVMT is commonly found to have a positive relationship with expected number of crashes, which was only the case for DVMT on all roadways that are not county roads variables in this study (Li et al. 2013; Dumbaugh and Rae 2009; Dumbaugh et al. 2012). While most of the results for the common measures of the built environment were as expected, these measures were not the main focus of this study.

Since this study was constructed to be the first step into identifying if food deserts could affect the number of crash incidences in a tract, the main objective of the model was to determine whether the presence of a food desert had a statistically significant relationship with the number of crashes occurring in a tract. From the results of the fitted linear regression model, it can conclude that there is a statistically significant relationship between the two variables, but it disproves the original hypothesis; it hypothesized that if a tract was food desert there would be an increased number of crashes due to most tract being in rural area which contain more rural roadways, which are more dangerous than their urban counterparts.
The model found that if a food desert is present the number of expected crashes decreases due to the inverse relationship between the two variables. This result is further supported by the positive relationship had between the number of grocery stores and the number of expected crashes; with a decrease in the number of grocery stores the number of expected of crashes also decreases. Since this study is one of the first of its kind in terms of studying how the presence of food deserts can affect crash incidences, there can only be speculations made about why the model gave these results. An explanation as to why tracts with food deserts have less vehicular crashes could be that residents living in tracts with food deserts are traveling to other tracts with more grocery stores to get groceries since 39% of tracts do not any grocery stores. Although some of tracts that food desert residents are traveling to may still be food deserts, they may have more grocery stores than the resident’s home tract. By traveling into a tract with more grocery stores, the driver would also be increasing he DVMT in that tract which would increase the number of expected crashes as proposed by the model for this study. To prove or deny, this theory though a study into the travel patterns of people who do or do not live in food deserts is needed.

Furthermore, another explanation to the relationships found in this study could be linked to land use in tracts which was not explored in this study. Previous studies studying the relationship between measures of the built environment and total crashes have concluded that an increase in commercial land use increases the number of expected crashes (Chen 2015; Wier et al. 2009; Yu and Xu 2018; Dumbaugh et al. 2012). If a food desert has less commercial land use due to its decreased amount of food retailer, this could explain the inverse relationship between food desert presence and expected number of crashes. This theory could be explored by adding land use composition variables into the model utilized in this study.
Future work into the investigation into a food desert’s presence implications on traffic safety could also include the utilization of different models to determine the relationship between the presence of food deserts and crash incidences. Since this study was meant to explore and determine the significance of this relationship, a linear regression was utilized, but models that are more well-suited for count data (such as crash data) could be considered for future work. As previously mentioned, negative binomial models are widely utilized when investigating the relationship between the built environment and crash incidences, and could be utilized in future work in relation to investigating the implications the presence of food deserts have on the expected number of crashes in a location (Dumbaugh and Rae 2009; Dumbaugh et al. 2012; Guerra, Dong, and Kondo 2019; Yu and Xu 2018; Noland and Oh 2004; Amoros, Martin, and Laumon 2003). Similarly, a geographically weighted regression could also be utilized in future modeling efforts since this method can capture these spatially varying relationships in the county-level crash data between the rural and urban tracts (Li et al. 2013).

This study took the first step toward investigating the effect the presence of a food desert may have on traffic safety. The presence of food deserts in a tract does have a significant relationship with the number of expected crashes in a tract, but the relationship is inversed, therefore, if a food desert is present in a tract, the number of crashes decreases. While the result of the study was not what was initially hypothesized, by confirming there is a statistically significant relationship and discussing theories as to why the effects are what they are this study opens the door to further research in this topic. The study performed was conducted to confirm a relationship, but lacks much detail about how the socioeconomic identities, travel patterns, and driving behaviors of people living in food deserts may impacted the crash incidences in tracts therefore this area of research is filled with potential research possibilities.
References


Dutko, Paula, Michele Ver Ploeg, and Tracey Farrigan. 2012. “Characteristics and Influential Factors of Food Deserts.”


