RURALITY AND ROBUSTNESS: RURAL COMMUNITIES AND HOUSING INSECURITY RISK

By

Steve Garcia

A.A.S., B.A.

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Data Science and Engineering

(in Spatial Information Science and Engineering)

The Graduate School

The University of Maine

May, 2024

Advisory Committee:

TBD, Advisor

TBD

TBD

RURALITY AND ROBUSTNESS: RURAL COMMUNITIES AND HOUSING INSECURITY RISK

By Steve Garcia

Thesis Advisor: TBD

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Data Science and Engineering
(in Spatial Information Science and Engineering)
May, 2024

This paper explores rural housing insecurity through Swope and Hernandez's (2019) 4 C's of housing insecurity in rural areas. Rural census tracts are defined by the United States Department of Agriculture (USDA) Rural Urban Code Continuum (RUCA) codes seven through ten to avoid an overly restrictive definition. Little attention has been paid to rural areas in the conversation on housing, to facilitate further discussion this exploratory study uses unsupervised machine learning to group census tracts into risk levels across 7 sectors of ACS data based on housing insecurity factors found in the literature. Multinomial logistic regression is used to determine variation between states based on how well state risk levels can be predicted with the national dataset. Additionally, spatial autocorrelation is used to analyze how spatially clustered the risk levels and housing insecurity risk variables. The results indicate that many rural census tracts have a medium risk of housing insecurity, and the risk levels are hard to predict. The spatial autocorrelation results show that the housing insecurity variables are not as highly spatially clustered as expected.

RURALITY AND ROBUSTNESS: RURAL COMMUNITIES AND HOUSING INSECURITY RISK

By Steve Garcia

Thesis Advisor: TBD

A Lay Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Data Science and Engineering
(in Spatial Information Science and Engineering)
May, 2024

Keywords: housing insecurity, homelessness, data mining

This study delves into the issue of housing insecurity in rural areas using Swope and Hernandez's framework, which identifies four key aspects of housing insecurity. To ensure a comprehensive view, rural areas are defined based on specific codes by the USDA, allowing for a broader scope in the analysis. Despite the limited attention given to rural housing concerns, this research employs unsupervised machine learning techniques to categorize census tracts into risk levels across various data sectors related to housing insecurity.

By utilizing data from the American Community Survey (ACS) and employing multinomial logistic regression, the study assesses the differences among states in predicting housing insecurity risk levels using a national dataset. Furthermore, spatial autocorrelation analysis is utilized to gauge the degree of clustering among risk levels and housing insecurity variables.

The findings highlight that a considerable number of rural census tracts face moderate housing insecurity risks, with challenges in accurately predicting these risk levels across states. Surprisingly, the spatial analysis reveals that housing insecurity variables are not as spatially clustered as previously anticipated. This study sheds light on the complexity of

housing insecurity in rural areas, offering insights into the unpredictability of risk levels and the spatial distribution of housing insecurity factors.

DEDICATION

Dedicated to Steve and Linda Mauldin, who taught me to find love in the emptiness of the absurd

ACKNOWLEDGEMENTS

I would like to thank the following people for helping with this research project: Thank you to my advisor, Dr. Matt Dube for providing guidance and feedback throughout this project, To my wife Katie for encouraging me to pursue my dreams, to Dr. Jeremy Mhire for teaching me how to think, and to Dr. Dolliann Hurtig for teaching me how to live.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

Homelessness research has undergone a significant transformation in recent years. Traditionally, the focus was on categorizing and describing different segments of the homeless population (Lee et al., 2021). A more contemporary approach views homelessness as a spectrum rather than a binary condition (e.g., Desmond et al., 2015; Swope and Hernandez, 2019). This paradigm shift opens opportunities for preventative and reactive services to address homelessness and housing insecurity . To harness these opportunities, there are four critical areas in that require attention. First, housing research has concentrated on urban settings resulting in an urban-centric view of social issues like poverty and homelessness. Second, measuring housing insecurity proves challenging due to its dependence on circumstances and obstacles for individuals and communities (Leifheit et al., 2022). Third, housing and homelessness in urban and rural areas necessitates a multi-disciplinary approach to properly capture the aspects that contribute to them, an approach that has been scarcely taken in the extant literature. Finally, the scarcity of identified community-level risk factors in rural areas coupled with a dearth of rural-specific data and research, limits our understanding of housing insecurity and rural homelessness (Gleason et al., 2021). Studies that do exist often focus on descriptive surveys of those using public services and providers of public services (Robertson and Nofsinger 2007). Addressing these gaps by integrating rural areas into the discourse on homelessness and housing insecurity is essential for creating a just and equitable society with effective policies for preventing and addressing homelessness(O'Regan et al., 2021).

1.1 *Rural Areas*

Rurality is often defined simply as not being urban Robertson et al., 2007). Defining rural areas in contrast to urban areas largely excludes the variation between rural areas insert definitions used by different organizations here. The lack of universally accepted definitions of rurality reduces the amount of time and resources that can be dedicated to struggling communities (Samadura and Yousey 2018). Rural areas encompass a broad spectrum, including farms, ranches, villages, small towns, and many other characteristics (Cromartie and Bucholtz, 2008). Castle (1998) identified a sparse population, interdependence with urban and global systems, and enormous diversity. insert rural population in urban areas stat At their core, rural areas are a function of "space, distance, and relative population density" (Castle et al., 2011, ?).Shoup and Houma (2010) group urban areas into three categories: rural areas dependent on nearby urban centers, "destination counties" with natural or artificial amenities that attract non-permanent residents, and production communities that revolve primarily around a single industry. This variation makes defining and understanding rurality a difficult challenge. Rural areas dominate the land mass of the United States, but with 85 percent of the population living in urban areas, they are often overlooked in the public discussion (Pendall et al. 2016) In the study of housing, rural areas are almost entirely excluded from the conversation (Gkartzios and Ziebarth, 2017). Contributing significantly to this problem is a wide variety of definitions used by governmental organizations, policy makers, and scholars (Samadura and Yousey, 2018; Cromartie and Bucholtz, 2008). Recently, The main policy objective for rural communities has been the promotion of economic development and preservation of the characteristics ascribed to rural areas (Lichter and Johnson, 2007).

Figure 1.1: Map of Rural areas here vs population density here

As Figure 1.1 demonstrates, the size of rural areas compared to the population density is vast and deconstructing the urban-centric lens of housing research necessitates a novel approach that can accommodate the differences in rural areas. The size and variation necessitates addressing rural issues differently because there can be no one size fits all policy approach to improving conditions for rural people.

1.2 *Literal Homelessness*

For decades, scholars have debated if research should focus on the reasons why people become homeless or the structural forces that create homelessness (Shlay and Rossi, 2003). Prior to significant shifts in the 21st century, research on homelessness focused on identifying and describing categories of homeless people (Lee et al., 2021). In other words, much research has focused on the binary of individuals and families being housed or unhoused and trying to assign them into umbrella categories. This neglects the wide range of individual and societal factors that occur in the phases between when an individual or family is housed and becomes unhoused. Estimating the number of people that are unhoused is notoriously difficult even in urban areas. For measuring homelessness, the most popular mechanism in the United States is the Department of Housing and Urban Development (HUD) point-in-time (PIT) count and housing inventory count. These counts are used for the distribution of federal funds for combating homelessness. as Agans et al. (2014) note that the unhoused frequently relocate and the housed may quickly become unhoused, making it difficult to accurately estimate the number of unhoused people at any given time. When it comes to addressing literal homelessness, public health experts differentiate between preventative services and reactive or emergency services (O'Regan et al., 2021). Preventive services prevent households from becoming homeless, while reactive or emergency services step in after a household becomes homeless. As homelessness is often seen as an urban problem, most intervention occurs in urban areas (Gleason et al., 2021). Federal action on homelessness began at-scale with the passage of the McKinney-Vento Homeless Assistance Act of 1987. it provided funds to support a variety of programs (Evans et al., 2019). Most of these programs offered services dependent on meeting some other requirements (source). The HEARTH Act of 2009 expnded the definitions of homelessness for supported federal programs to expand those eligibile beyond the literal homeless. These included those living in a place that is not meant for habitation, people who re expected to lose their residence within 14 days, families with children that re unstably housed, and people fleeing domestic violence (source).

add more about literal homelessness

1.3 *Housing as Health*

A house is far more than four walls, a roof, and some doors. The characteristics and location of a house can make a significant impact on one's life. In the United States, housing is often a family's greatest expenditure, greatest source of wealth, and a place of safety and gathering (Braveman et al., 2011). The federal government has long acknowledged this through legislation like the Housing Act of 1949, and social programs and development goals developed by the U.S. Department of Housing and Urban Development. Housing is often seen as one of the most fundamental determinants of health, and a lack of adequate housing can produce adverse health outcomes and acts as a foundation for "social, psychological, and cultural well-being" (D'Alessandro and Appolloni, 2020, p.17; Leifheit et al., 2022). A large body of literature has documented the relationship between children's health and homelessness (Bailey et al., 2016) Part of acknowledging housing as health is moving beyond the housed/ unhoused binary in order to better understand and intervene in households that are at risk of being unhoused. This is often referred to as housing insecurity , a broader term that encompasses a continuum that affects a larger part of the population than simply housed/ unhoused (DeLuca and Rosen, 2022).

add more about housing as health

1.4 *Theoretical Framework*

Rather than focusing on literal homelessness, this thesis approaches housing from a housing insecurity perspective. housing insecurity has a variety of definitions across government organizations, but it can be characterized as: Housing Stability, Housing

Affordability,Housing quality/ safety, and Neighborhood quality/ safety (Cox et al., 2019). To further refine these broad characteristics, this thesis follows the 4 C's approach of housing insecurity. With little infrastructure for homelessness services in rural areas, the 4 C's approach to housing insecurity proposed by Swope and Hernandez (2020) can highlight areas of critical concern for devoting resources to reactive services and identify areas where preventative services can improve or expand. The pillars of the 4 C's include:

- Conditions: quality of housing
- Cost: housing affordability
- Consistency: residential stability
- Context: neighborhood opportunity

The 4 C's of housing are an interconnected web of factors that impact health and encapsulates "this unequal distribution of housing disparities along other axes of inequality, and the historical forces shaping unequal housing opportunities" (Hernández and Swope, 2019, 1). Swope and Hernandez are not the only scholars to design a model encompassing these 4 factors. Metzger and Khare (2017) proposed a similar framework that encompasses stability, affordability, internal housing conditions, and area characteristics.

1.5 *Motivation*

There are three primary reasons behind this exploration:

- The first motivation comes from the lack of attention scholars have paid to rural areas as it pertains to housing insecurity . While the literature on rural housing insecurity is growing, there has yet to be a holistic nationwide survey of rural housing insecurity. Rural areas deserve more attention, and this thesis hopes to serve as a starting point for future research on rural housing insecurity at all levels with the ultimate goal of breaking the urban-focused lens of housing insecurity.

- The second motivation is to test the efficacy of the 4 C's model of housing insecurity for the rural United States. As there is an urban lens to housing insecurity , an adequate theoretical model must be capable of adapting to areas often left out of the conversation. One study (Gleason et al., 2021) has applied the 4 C's model to housing insecurity for the state of Maine. Most applications of the 4 C's have been to study the relationships between various conditions and housing insecurity, but no studies have applied it broadly to rural housing insecurity.

- The final motivation is to provide policy-makers and researchers with a quantitative and qualitative framework to identify rural areas of housing insecurity in their constituency and create harm reduction approaches and services that can meet the unique needs of their areas. The patchwork of local, state, and federal systems that encompass the aid programs of the United States means that there are a lot of people involved in the policy-making process with, as of yet, no real mechanism for addressing housing insecurity in their particular jurisdictions.

1.6 *Approach*

In order to improve our understanding of rural housing insecurity , this thesis investigates the risk levels of rural census tracts in the United States using the 4 C's model of housing insecurity . Using k-medoid clustering, risk factors across eight different sectors are used to assign housing insecurity risk levels to rural census tracts for the continental United States. Each state is clustered with census tracts from other states within a 15-mile boundary to encapsulate how communities exist across state lines. The cluster medians are analyzed to understand the trends in housing insecurity factors across states and the clusters are relabeled based on risk factors identified in the literature so that each cluster falls into a low, medium, or high risk level. Using these risk levels, association rules learning is used to identify common patterns between sector risk levels and identify pockets of rural census tract that are at high risk of housing insecurity . To better understand how factors of housing insecurity relate to space, Moran's I spatial autocorrelation is used to determine how spatially clustered each housing insecurity factor is. Local Moran's I is used to determine how spatially clustered each risk level is to better understand the clustering of housing insecurity risk in rural areas.Finally, a multinomial logistic regression is used to determine how well each states sector risk levels can be predicted and a national model is generated for each sectors risk levels to analyze well the risk levels created by this implementation of the 4 C's model can be predicted nationally and state by state.

Beginning to understand rural homelessness requires that several questions be answered:

- How can risk factors of be used to identify risk levels of housing insecurity while accounting for the variation in rural areas?
- When measuring housing insecurity across different dimensions, how often do the same features arise?
- Are there spatial relations between the different dimensions of housing insecurity ?
- To what extent can this model of housing insecurity be used to predict risk levels across housing insecurity factors?

1.7 *Major Results*

I have some

1.8 *Intended Audience*

This thesis is intended for an audience with a significant interest in rural housing insecurity. Such an audience can include, but is not limited to policy makers, economists, political scientists, community psychologists, rural sociology, and many others concerned with housing insecurity.

1.9 *Structure of Thesis*

The thesis is structured into six chapters. Chapter 2 offers a comprehensive theoretical foundation, focusing on the application of the 4 C's of the housing insecurity model. This chapter critically reviews pertinent literature pertaining to various facets of the model. Chapter 3 delineates the methodology employed for data processing. It provides an in-depth explanation of the chosen methodology and its execution. The ensuing Chapter 4 presents the study's findings, offering a detailed analysis of the acquired results. Chapter 5 deliberates on the implications of the results, discussing their significance and impact within the scope of the study. This chapter provides a thorough examination of the noteworthy findings. Chapter 6 serves as a synthesis, summarizing the entirety of the work and offering insightful commentary on the major findings. Additionally, it delineates potential avenues for future research and study.

CHAPTER 2
LITERATURE REVIEW

There are three primary areas where the extant literature must be analyzed. The first is the unique challenges that face rural areas in order to better inform the theoretical framework. Next, it is important to understand the distinction between homelessness and housing insecurity. Finally, it is necessary to take a multi-disciplinary look at each pillar of the 4 C's of housing insecurity. Together, these 3 aspects form the theoretical basis for exploring rural housing insecurity.

2.1 Challenges for Rural Areas

While there is limited research on homelessness in rural areas, previous research has documented the unique struggles of rural areas that should be addressed in a discussion on rural housing insecurity. First, previous research has identified both pockets of prosperity and pockets of deep poverty in rural areas. Concentrated poverty is "often the manifestation of an interactive and intergenerational dynamic between structural changes that restrict economic opportunities and the emergence of populations with characteristics that put members at a relatively high risk of poverty (Valasik, 2018, p#). Poverty is acknowledged more in urban areas, but poverty rates are highest in both remote rural counties and in cities (Source). Persistent poverty, typically defined as poverty levels above 20 percent, is geographically concentrated in rural regions (Crandall and Weber 2004). In 2010, the poverty rate among the rural population was higher than that of the nation overall (Lichter and Eason 2013). A cluster analysis found that of 3,017 places with about 5 percent of the nation's population, 84 percent of this population lives in rural rather than urban areas (Peters 2009). Lichter and Johnson (2007) found that 85 percent of the nearly 500 counties with poverty rates over 20 percent and the 12 counties with poverty rates over 40 percent are in nonmetro areas. The areas with persistent poverty have some similar characteristics: they have primarily agricultural or resource-based economies, reduced employment opportunities due to economic changes, or gentrification is making living costs unaffordable for many people Robertson and Nofsinger, 2007).

Looking only at poverty does not tell the full story of rural areas. There are more than 300 rural counties spread across the nation that are more "prosperous" than the rest of the nation based on measures spanning education, housing, poverty, and unemployment (Isserman et al., 2009). This highlights the need for an approach to rural areas that is relative rather than absolute. One explanation for the high rates of poverty in rural areas is their isolation. Isolation stems from limited ease of travel or access to nearby markets and population centers which can hinder economic development, meaning that greater geographic isolation is associated with both lower income and greater poverty rates (Blank 2005). Metzger and Khare (2017) highlight the tendency for Americans to segregate themselves not only based on race but on class too. A tendency for the rich and the poor to cluster around themselves could explain these findings in rural areas. This spatial inequality is critical to understanding rural poverty (Valasik, 2018). Another problem that rural areas are facing is a growing economic divide between urban and rural areas (Bjerke and Mellander, 2019). Rural communities have been hit hard by economic changes in recent decades (Pendall et al., 2016). One aspect of this is that friction is created when rural households are too distant from adequate labor markets that enable them to support their families (Sparks et al., 2013). This has created a common migration pattern where so many people move to urban areas for greater economic opportunities that rural towns are left with a smaller, older population and a less skilled labor force (Bjerke and Mellander, 2019). The effects of these population decreases span across socioeconomic factors. School consolidations, reductions in local services, closed businesses, increased infrastructure costs, poorer schools, poorer healthcare, and limited public services have all been tied to population shrinks and communities have little ability to control these processes that limit economic mobility and can perpetuate poverty (Zarecor et al., 2021; Blank 2005). There is a cyclical nature to the problems facing rural areas. For the areas affected by poverty, it becomes difficult for systemic improvements because the economic decline inherently reduces the resources available in the community for addressing the issues at hand.

2.2 Housing Insecurity

Housing insecurity is a term that stems from the shift in homelessness research from focusing on only the housed and the unhoused. The Oxford Dictionary defines housing insecurity as "the state of not having stable or adequate living arrangements, especially due to risk of eviction or because one lives in unsafe or uncomfortable conditions" (Source). DeLuca and Rosen (2022) argue that the term housing insecurity is a more dynamic concept than the traditional housed/ unhoused binary. "Housing insecurity operates through multiple mechanisms-including material hardship, stress, environmental and infectious disease exposures, social network disruption and barriers to healthcare- to produce adverse health outcomes over the life course" (Leifheit et al., 2022). (Cox et al., 2019) offer an operationalized definition of housing insecurity that aligns with the 4 C's of housing insecurity: housing stability (consistency), housing affordability (cost), housing quality/ housing safety (conditions), and neighborhood safety/ neighborhood quality (context). To understand housing insecurity in the context of the 4 C's framework, the following subsections will detail each pillar of housing insecurity under the model proposed by (Swope and Hernandez, ?). As each column forms a web rather than separate pieces, there is a significant amount of overlap between columns.

2.2.1 *Cost*

It is difficult to determine one number that determines when a household is spending too much on housing. A cost-to-income ratio is the most common way of measuring housing affordability. The threshold for housing affordability has ranged between 25 and 50 percent but the current standard is 30 percent (Kropczynski and Dyk, 2012). Housing is considered affordable if the household spends less than 30 percent of its income on housing and 50 percent or more is considered a high-cost burden (Braveman et al., 2011;Swope and Hernandez, 2020;Weicher, 2006). Inherit to any cost-to-income ratio is the understanding that housing is that there are other expenses necessary for survival (Herbert et al., 2018). Housing costs are determined by the rate of household formation and household attrition Pendall et al., 2016. Housing affordability affects individuals, families, and communities while access is largely determined by their demographic characteristics (Braveman et al., 2011;Yadavalli et al., 2020).Housing affordability is directly related to residential stability and has the potential to harm both those being forced to move, the community they are leaving, and the community they are entering (Desmond et al., 2015). Access to affordable housing affects the physical and material comfort of the population as well as the individual (Kumar, 2003). If a household cannot afford to live in their current place, they may be forced to relocate seeking more affordable housing or through eviction and foreclosure.. If too much of a household's money goes to housing, they may be forced to go without other necessities (Herbert et al., 2018). note that a body of research has found that those with high housing costs may also experience food insecurity as food is often considered a flexible expense while housing is a fixed expense (Fletcher et al., 2009;Kropczynski and Dyk, 2012). The shortage of affordable housing drives lower-income families to substandard housing in worse neighborhoods (Braveman et al., 2011). Kang (2021) characterizes housing instability as a by-product of the affordable housing shortage wherein households can be destabilized by minor financial shocks. insert housing shortage stuff These factors can create a situation where housing costs lead to residential instability, which is linked to a variety of adverse conditions, especially in children and adolescents (Desmond and Shollenberger, 2015). Part of the blanket construct of rural areas is that they are cheaper to live in. However, Kurre (2003) note that there is relatively little systematic data that supports this presumption. Rural areas face the same low per capita income and poverty problems faced by urban areas (Castle et al., 2011). Zimmerman (2008) found no consistent pattern of lower princes across all of the rural counties in ? While the dollar amount paid for housing may be lower, given the different socio-economic circumstances of rural areas, housing costs alone may not fully encapsulate the situation (Kropczynski and Dyk, 2012).

2.2.2 *Conditions*

Many scholars have identified internal housing conditions as a significant factor on health (Braveman et al., 2011;Metzger and Khare, 2017; Swope and Hernandez, 2020. In one study, decent housing was found to be a more important determinant of health than education or income (Stefan and Bittschi 2014). Previous environmental health research has identified five broad categories in which housing conditions contribute to adverse health effects: *physical conditions*, *chemical conditions*, *biological conditions*, *building and equipment conditions*, and *social conditions* (Jacobs, 2011). Adequate housing is necessarily related to public health (Matte and Jacobs, 2000). Links to an increase of disease have been tied to poverty, poor housing, and degraded environments reflecting the interconnectedness of housing insecurity issues (Rauh et al., 2008). Stefan and Bittschi (2015) found that the probability of facing a chronic disease increases when housing problems accumulate and that poor housing conditions quickly degrade subjective health. These problems are amplified in the modern world where individuals spend an estimated 90 percent of their time in doors (Palacios et al., 2020). Despite housing conditions playing such a significant role in modern life, there is not a significant sense of communal benefit and responsibility when it comes to housing (Jacobs, 2011). Rural areas face unique housing issues because one of the most common housing solutions is mobile homes.more on mobile homes Structural problems like poor construction and risks of air pollution and fire create a unique problem (MacTavish et al., 2006). An area of particular concern are marginalized populations who are more likely to be exposed to harmful housing conditions (Swope and Hernandez, 2020). Housing conditions also play a role in residential mobility as Desmond and Bell (2015) place decent housing and affordable housing as fundamentally connected and the previously mentioned rise in housing cost has not brought an increase in housing quality.

2.2.3 *Consistency*

Residential mobility is a complicated subject because, as a broad concept, it is conceived as a good thing. That one can pack up and go somewhere with more opportunity is considered a part of the American "mystique" (Molloy et al., 2011). An average of 15 percent of Americans move every year and 25 percent move over the course of two years (Bachmann and Cooper, 2014). Classic urban economic theories explain that households make trade-offs between proximity to jobs and housing prices (Hu and Wang, 2019). This puts low-income households at a disadvantage as their access to jobs may be lower than their wealthier counterparts. Consistency or residential stability plays an important role in the physical and social well-being of individuals, families, and communities. It has been linked to a variety of adverse conditions and affects the neighborhoods being entered and left. It has been identified as a more important predictor of community health than standard factors like poverty and racial composition (Desmond and Shollenberger, 2015;Desmond and Perkins, 2016, Rauh et al., 2008). An important distinction must be made between voluntary and involuntary moves (Siskar, 2019). While most moves are voluntary, there are millions of low-income households that struggle to maintain housing stability (Phinney, 2013;Kang, 2019). Housing is often the biggest expense for low-income families, often forcing them to make trade offs between necessities (Desmond and Bell, 2015). Of particular concern is forced relocation, foreclosure, eviction, and condemnation are all drivers of forced relocation (Phinney, 2013;Siskar, 2019). It is linked to an increase in residential instability and households forced to move often end up in places with greater disadvantage and are more likely to face additional moves (Desmond and Shollenberger, 2015). One issue with the study of residential mobility is the limited scope of predictors that have been linked to it (Kang, 2019). In urban areas, renters have been found to be particularly vulnerable to relocating to worse neighborhoods than the one they are exiting (Desmond and Shollenberger 2015). It is yet to be seen how this translates to rural areas, where renting is far less common than urban areas with the exception of mobile homes. insert mobile home costs

2.2.4 *Context*

Context revolves around neighborhood and community characteristics including demographics, green spaces, education, and healthcare among other things. While it is impossible to capture context in its entirety, this thesis focuses on demographics, economic diversity, housing type, and household wage/ aid factors as these have all been studied as matters related to housing insecurity that do not fall directly into the other pillars of housing insecurity. The following is an interdisciplinary review of how these factors affect housing insecurity.

2.2.4.1 *Employment*

In the United States, the labor market is the result of cumulative individual behaviors including geographical migration and educational investments (Wiener, 2020). The demand for labor is driven by firms, which must consider a wide variety of factors in deciding location (Partridge and Rickman, 2007). In recent decades, the United States labor market has entered a risk regime job market where workers hold a greater share of the risk in an employment system without the perceived promise of security and stability, which has become embedded in American social and political institutions (Lowe, 2018). It is agreed that the Fordist regime that brought unprecedented prosperity in the early 20th century came to an end in the 1970s (Stockhammer 2008). Since this shift, productivity of the average worker has increased 64.6 percent while hourly pay has only increased an average of 17.3 percent between 1979 and 2021 Economic Policy Institute). Over this same period, U.S. Housing and Urban Development data shows that the median price of a new single-family home increased from $60,600 ($232,091 adjusted for inflation) in Q1 of 1979 to $369,800 in Q1 of 2021) (U.S. Census Bureau and U.S. Department of Housing and Urban Development, 1963). The Great Recession has had a lasting impact on the housing market within the United States. As the economic recovery did not benefit all households equally, wealth inequality has grown along both racial and ethnic lines (Lochar, 2014). As wages have failed to keep up with the price of housing, the current economic system under this risk regime places those with low incomes in a precarious situation for housing affordability and residential stability. Thus, employment insecurity and income inequality are two pressing issues the United States is facing that have serious impacts on communities. "housing insecurity has risen in relative lockstep with employment insecurity" (Desmond and Gershenson, 2016, 48).

Rural communities have been hit hard by economic change, driven by the transition from a production to a consumption based economy (Pendall et al., 2016). what (Bjerke and Mellander, 2019) identified an increasing economic divide between urban and rural areas where over several decades rural areas have lost out. During this shift, employment became increasingly scarce for agricultural workers (Kropczynski and Dyk, 2012). Today, manufacturing is responsible for 21 percent of rural non-agricultural earnings (Low, 2017). Economic development is therefore a fundamental issue to rural areas. While manufacturing has grown, the majority of counties that experienced manufacturing employment growth between 2001 and 2015 had low levels of growth in terms of total employment (Low, 2017). Sherrieb et al. (2010) identify three key elements connected of economic development: the level of economic resources, the level of equality in resource distribution, level of diversity in economic resources. Economic development alongside demographic change in rural areas has been linked to the quality and condition of local housing infrastructure (Barcus, 2011). Thus, how policies shape economic development has a direct affect on the overall housing insecurity risk of rural communities. Demonstrating the interconnectedness of communities, regional economic development in one area can encourage economic stability of its neighboring regions as well (Chen, 2018). Deller and

Watson (2016) highlight the importance of economic diversity, a vital aspect of economic development, finding that more diverse economies enhance economic stability. (Blank, 2005) note that rural areas often have more limited job opportunities and are more likely to rely on one industry rather than having a diversified economy. As an insulator against economic instability, employment diversity in rural areas is a key factor that policy-makers and scholars should consider as part of a holistic approach to housing insecurity. This may be difficult to achieve for rural areas based on natural amenities, where one industry acts as the lifeblood of the community.

2.2.4.2 *Housing, race, and poverty*

Housing is affected by a variety of social, political, and economic factors. "The ability of residents to access affordable housing, whether renting or buying, is in large part determined by their demographic characteristics, such as income, race, age, and educational attainment" (Yadavalli et al., 2020, 115). While unpredictable events may narrow the disparities, "As a rule, a household's vulnerability to displacement should be shaped in predictable fashion by those characteristics that define its members' position in the [social] stratification system" (Lee and Evans, 2020, 5). Although the federal government took a direct interest in promoting home ownership in 1933, racial discrimination in the housing market was not outlawed until 1968 but enforcement of the law remained difficult until the Fair Housing Act of 1988 (Sharp & Hall, 2014). For example, the practice of redlining made it difficult for Black Americans to receive mortgages under federal aid programs and creating racial segregation that can still be seen today. At the county level, the probability of living in affordable housing decreases as the white population decreases (Brooks, 2022). In addition to racial segregation, income segregation must be considered for a holistic discussion of housing insecurity. Lichter and Brown (2011) found that 40.5 percent of high-poverty places are in high-poverty counties for non-metro areas and the poor and non-poor are becoming increasingly segregated, with higher concentrated poverty among minorities. High concentration of poverty may exacerbate housing condition issues due to a lack of revenue to maintain the necessary services at the household and local government levels. Minorities are also at a disadvantage in income segregation with poor whites being less segregated from their non-poor counterparts (Lichter et al., 2021). As a home is often a household's greatest source of wealth, the disadvantages minorities have in terms of housing are compounded as social and economic inequality are reproduced as these disparities continue (Krivo and Kaufman, 2004).

Rural areas face significant consequences for the historical forces that shape housing today. When discussing rural poverty it must be noted that there is an underlying assumption that the dynamics of poverty are fundamentally different from urban areas (Valasik, 2018). Persistent problems faced by the rural poor include "physical isolation and poor public transportation, inadequate schools, and limit access to medical care and other basic public services while institutional support services are frequently limited or simply unavailable" (Lichter and Johnson, 2007, ?). Part of this is driven by the outflow from rural areas to urban areas. Rural areas have seen a reduction in population, reducing the capabilities of public services to accommodate those in need (Bjerke and Mellander, 2019). As mentioned earlier, there are a variety of reasons why households move. In rural areas, a common reason to move is due to the friction that exists when households are too far removed from labor markets that provide adequate employment and income opportunities (Sparks et al., 2013). Valasik (2018) found that from 2000 to 2012, increases in poverty were larger in rural counties than urban counties with the highest increases in exposure and the rural black population was by far the most disadvantaged over this time period. Rural areas are not as diverse as the United States overall, and many rural minorities are geographically central in regions tied to historical and economic dynamics (Housing Assistance Council, ?). Another demographic group that is significant to rural areas is Hispanics and Latinos, despite the widespread population decline of rural areas (Lichter and Johnson, 2020). African Americans and Hispanics and Latinos face similar discrimination in the housing market with the benefits of housing are dramatically smaller for these demographics (Krivo and Kaufman, 2004). Thus, the pockets of these groups in rural areas should be considered to be at a higher risk of housing insecurity due to the effects of these historical forces.

2.2.4.3 *Housing Type*

While owning a home is considered a part of the "American Dream," many households rent their housing by choice or by necessity. While the many benefits of home ownership portray it as a means to a better life, renting is not inherently bad and may provide better opportunities for households that can afford it, but there are many potentially destabilizing consequences of high-cost renting (Drew, 2014). Nationally, median rent in a poor neighborhood is $298 compared to $225 in a middle-class neighborhood or $250 in an affluent neighborhood after regular expenses are deducted despite property values typically being much higher in middle-class or affluent neighborhoods (Desmond and Wilmers, 2019). This creates a compounding factor for the previously mentioned disparities in home ownership. Increases in household wealth and secured debt were found to decrease the likelihood of homeowners becoming renters and vice versa (Anderson et al., 2021). Money paid towards a mortgage generates long-term wealth while money paid towards rent generates wealth for the property owner. Renters with high-cost housing are unable to increase household wealth through their means of housing. In addition to whether one rents or owns their home, the type of home can play a significant role in housing insecurity. Of particular concernr is "unconventional housing' which includes dwellings not considered long-term habitation including RVs/ campers, vans, and boats. These unconventional forms of housing may keep people off the streets, but they are not always a stable mode of housing. For RV and camper living, people who are undocumented or are unable to keep up with legal or maintenance costs for vehicles end up losing their housing (Wakin, 2005). Mobile homes also carry a unique set of circumstances that may put households at a greater risk of housing insecurity and are found frequently in rural areas. Mobile homes and the land they are situated on can be either owned or rented. It is common in mobile home parks for households to own their home but not the land it is on. Key issues with mobile homes include their financing: typically done through more expensive but easier obtained means than a mortgage such as personal property or chattel loans; mobile homes do not build wealth in the same way as they typically depreciate rather than appreciate; households on rented land have little control over their length of stay; they also tend to have worse construction and higher risks of air pollution and fire than traditional homes (MacTavish, 2007). Those that live in mobile homes or unconventional housing should be a priority for discussing housing in rural areas.

2.2.4.4 *Household income, aid, and Transportation*

In his first State of the Union address, President Lyndon B. Johnson asked Congress to declare an "unconditional war on poverty... not only to relieve the symptom of poverty but to cure it and, above all, to prevent it." Since then, a patchwork of programs regulated at the federal, state, and local levels have arisen. A large part of the federal government growth in the late 20th century is from the expansion of social welfare spending (Fishback, 2020). As the primary mechanism of income distribution is what Berkowitz and Palakshappa (2023) refer to as the "factor payment system" in which those who work and those who own the means of production and one's relation to this system and the labor market is closely related to one's poverty risk. To alleviate this poverty risk, social programs which utilize different mechanisms are available to those who qualify. These mechanisms can be divided into categorical and income targeted policy designs, alongside decentralization, where some receive benefits based on "demographically defined, categorical eligibility structures" and others enjoy standardized federal assistance through social insurance with some qualifying for income-based or "means-tested" programs (Bruch et al., 2023). Households must fall below certain income and asset thresholds to qualify for means tested programs (Rank and Hirshl, 2002). For housing, there is a wide variety of housing policies and programs aimed at low-income individuals. These take the shape of voucher programs by subsidizing privately held property although some recipients live in public housing (Kim et al., 2017). For rural areas, the U.S. Department of Agriculture (USDA) has a variety of programs aimed at improving living conditions in rural areas including direct or guaranteed loans for single or multi-family housing, and infrastructure programs for water, electricity, and telecommunications ((USDA Rural Development Summary of Programs, 2023). Transportation plays a large role in social and economic life. Access to everything from education to healthcare depends on the infrastructure and ability to use available means of transportation. The expense of owning enough vehicles may prove restrictive, especially for households with high housing costs. Rural areas often do not have public transportation, leading residents to depend more on automobiles. An analysis of 2009 National Household Travel Survey data found that 72 percent of households with a yearly income of $20,000 have access to a household vehicle compared to over 97 percent of households making $50,000 (Blumenberg and Pierce, 2012). Automobile ownership can be a crucial factor in avoiding residential instability (Kang, 2019). Households are twice as likely to be auto-deficient (less than 1 car per driver) than zero-vehicle households where a vehicle is not needed (Blumenberg et al., 2020). This is a harrowing statistic in rural areas without public transportation where distances may be too far to walk or ride a bicycle or it may not be safe due to lacking road infrastructure like bicycle lanes and sidewalks.

2.3 *Summary*

Throughout this chapter, the 4 C's of housing insecurity have been covered. It is important to highlight the interconnected nature of the 4 C's. There is significant overlap between each pillar of housing insecurity. Housing costs, housing type, and housing conditions are necessarily linked to the economic conditions of a household. These economic conditions are linked to the household wage/ aid factors that encapsulate their economic status. One's relation to the poverty level and education plays a significant role in housing accessibility and these factors are intrinsically linked to the context that they grew up in. For rural areas where public transportation is scarce, one's access to adequate transportation is highly linked to one's economic opportunity. Any discussion on housing insecurity must consider the historical forces affecting modern day race and poverty, and these forces relate to all aspects of life. When taken as a web, this model encompasses the wide ranging socio-economic factors that surround housing insecurity.

CHAPTER 3

ADDRESSING RURAL HOUSING INSECURITY

3.1 *Defining Rurality*

Rather than strictly defining rurality, this thesis uses the United States Department of Agriculture (USDA) Rural-Urban Continuum spectrum. The following codes are used to encapsulate rurality:

Table 3.1: RUCA Codes

Number	**Description**
7	Small town core: primary flow within an Urban Cluster of 2,500 to 9,999 (small UC)
8	Small town high commuting: primary flow 30 percent or more to a small UC
9	Small town low commuting: primary flow 10 percent to 30 percent to a small UC
10	Rural areas: primary flow to a tract outside an urban area or urban cluster

cite ruca codes

The range of RUCA codes described in Table 1.1 was chosen to be inclusive rather than exclusive, including small towns with various levels of commuting to urban clusters and areas classified as rural. We include small towns because they often serve as hubs for rural areas, serving an important role in rural areas and source has identified a significant amount of rural people that live on the edge of urban places, like small towns. Spatial autocorrelation is used to determine how often similar rates of each variable occurred across each rural census tract in each state. Finally, multinomial logistic regression is used to determine how well the risk levels of a census tract can be predicted based on the nationwide dataset. All analysis was conducted in the R statistical language.

3.2 *Applying the 4 C's*

Applying the four C's of housing insecurity necessitates a mix of quantitative and qualitative analysis. To use the model to classify areas into risk levels, it is necessary to define thresholds for each pillar based on the literature review. For housing costs, an area is at a higher risk of housing insecurity as the number of households spending more than 30 percent of their income on housing increases. A study of urban housing insecurity would place a higher emphasis on renters with high housing costs, but the extent renting versus owning affects housing insecurity in rural areas is unknown. Housing Conditions are difficult to encapsulate because they encompass a broad range of factors. An additional challenge is a lack of rural-specific housing conditions data. This thesis measures housing conditions by the lack of complete plumbing and kitchen facilities, with the assumption that if these are missing, there are likely other factors the household is struggling with as well. The risk of housing insecurity in an area therefore increases as the number of occupied and unoccupied housing lacking these fundamental needs increases. Priority is given to occupied housing as there will always be some amount of housing not fit for habitation. Consistency, or residential mobility, is difficult to encapsulate because many households move for reasons unrelated to housing insecurity. To focus on the subset of households that are at a high risk of becoming housing insecure, the scope of residential mobility is limited to those who have moved in the past year without a college degree or are below or just above the poverty line. These groups are more likely to move to more precarious situations than those making moves for economic and social reasons unrelated to housing insecurity. Context is the most difficult pillar of the four C's to capture because it encapsulates many individual, social, and political factors. Six different sets of factors are used to capture context. Due to the influence of social, political, and historical processes, demographic diversity is used to capture the effect that race has on housing insecurity risk. The previously mentioned measures of residential stability also contribute to the context of an area, encapsulating education and poverty. The type of housing individuals in an area live in is a significant factor of context because mobile homes, while being seen as a means of affordable housing, can signify a risk of housing insecurity when taken in tandem with other factors. The final measure in context is household factors. This range of household factors is designed to encapsulate different individual, social, and economic factors that contribute to housing insecurity

3.3 *Data*

Eight different sectors of 2019 ACS 5-year variables are used to capture the 4 Cs of housing insecurity using indicators of housing insecurity identified in the literature. These sectors are housing cost, housing quality, housing type, economic diversity, education mobility, poverty mobility, and household worker/aid. For demographic variables we use seven variables including an "other" variable to account for race/ ethnicity and the number of people over or under 18 by gender. The economic diversity data is the number of people employed across 13 distinct categories. It was necessary to create three compound variables: high-cost with a mortgage, high cost without a mortgage, and high-cost rent to use the standard affordability measure of 30 percent. There are four variables for housing conditions which include houses with an incomplete or insufficient kitchen or plumbing for occupied and unoccupied housing units. Two sets of variables account for residential mobility: education and poverty that include those who did not move, those who moved in and out of county and state. Due to the housing affordability and income inequality crises, those below the poverty level and those at 125 percent of the poverty level as high risk for housing instability are included in residential mobility: poverty. For education those with and without a high school diploma are included as those without a college degree may face higher barriers to well-paying and stable employment. Wage/aid data include households without income, households that receive public assistance, households that receive supplemental security income, households with investment income, households with other income, households with 3+ workers and the household Gini index. For housing type, renters and owners of mobile homes, single family homes, small and large multi-family homes, and renters and owners of unconventional housing are included.

3.4 *Data processing*

In order to ensure the integrity of the data, census tracts that lacked specific sector-related information were excluded from the analysis. These omitted tracts were assigned a risk level of zero, a measure adopted to preserve the largest possible number of census tracts for subsequent analyses. To mitigate potential biases stemming from differences in population sizes and geographic areas, a standardized approach was employed across each sector. This involved scaling all dataset components to a common base unit. Demographic and economic diversity metrics were adjusted proportionally to the population size. Meanwhile, data pertaining to household expenses and types were scaled based on the counts of homeowners and renters. The household dataset underwent normalization corresponding to the total number of households, whereas housing condition indicators were adjusted relative to the total count of occupied and unoccupied housing units. It's essential to note that all numerical values within the dataset have been uniformly represented as percentages, except for the household Gini Index, which retains its original values.

3.5 *Methods*

Supervised and unsupervised machine learning algorithms are used alongside globl and local Moran's I spatial autocorrelation, the Queen Contiguity spatial relationship algorithm to form and analyze the housing insecurity risk assignment system, and multinomil logistic regression to examine the predictive abilities of the risk assignment system.

3.5.1 *Neighbors Algorithm*

Communities often share dependencies across state lines, making it unjust to disregard neighboring communities in a state-based housing insecurity analysis. To address this, the analysis encompasses census tracts within 15 miles of a state's outermost tract. Any census tract sharing a boundary with a tract within this range is considered, ensuring a more inclusive evaluation of rural housing insecurity. This process is repeated for each state in the continental United States.

The formula for queen contiguity neighbors is shown in Equation 3.1.

$$\begin{aligned}
\text{Top-left:} \quad & (x-1, y-1) \\
\text{Top:} \quad & (x, y-1) \\
\text{Top-right:} \quad & (x+1, y-1) \\
\text{Left:} \quad & (x-1, y) \\
\text{Right:} \quad & (x+1, y) \\
\text{Bottom-left:} \quad & (x-1, y+1) \\
\text{Bottom:} \quad & (x, y+1) \\
\text{Bottom-right:} \quad & (x+1, y+1)
\end{aligned} \tag{3.1}$$

3.5.2 *K-Medoids Clustering*

K-medoids clustering is a partitioning technique aimed at dividing a dataset into K distinct and non-overlapping clusters. Unlike K-means clustering, which utilizes centroids as cluster representatives, K-medoids employs actual data points within the dataset as cluster representatives. The key advantage of K-medoids lies in its robustness to outliers and noise due to its use of real data points. The objective of K-medoids clustering is to minimize the sum of dissimilarities within clusters. Each state, including neighboring census tracts, is clustered individually. The cluster medians are analyzed to determine which clusters have a high, medium, or low risk of housing insecurity based on the factors previously identified in the literature review. To bring areas of concern, census tracts are labeled as high-risk if the sum of their risk levels is 12 and medium-risk if the sum of their risk levels is 15. Out of a total of 24, this approach highlights the areas that show the most vulnerability across sectors.The formula for K-medoids clustering is shown in Equation 3.2.

$$
\begin{aligned}
&\underset{S}{\text{minimize}} \quad \sum_{i=1}^{K} \sum_{x \in C_i} d(x, m_i) \\
&\text{where:} \\
&\qquad S : \text{The set of clusters.} \\
&\qquad K : \text{The number of clusters.} \\
&\qquad i : \text{Index representing each cluster } (1 \leq i \leq K). \\
&\qquad C_i : \text{The } i\text{-th cluster containing data points.} \\
&\qquad x : \text{A data point within a specific cluster } (x \in C_i). \\
&\qquad m_i : \text{The medoid (representative) of the } i\text{-th cluster.} \\
&\qquad d(x, m_i) : \text{The dissimilarity (distance) between data point } x \text{ and medoid } m_i.
\end{aligned}
\tag{3.2}
$$

3.5.3 *Association Rules Learning*

Association Rules learning is a data mining technique used to uncover interesting relationships between variables in large datasets. It aims to discover patterns in the form of rules indicating the co-occurrence or association between items within transactions or events.

Association rule learning involves two main metrics:

Support (s): Measures the frequency or occurrence of an itemset in the dataset.

$$\text{Support}(A \to B) = \frac{\text{Transactions containing both A and B}}{\text{Total transactions}}$$

Confidence (c): Measures the conditional probability that an item B appears in a transaction given that item A is present.

$$\text{Confidence}(A \to B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}$$

Here, association rules are used to analyze the common occurrences between sector risk levels. Of primary interest are unexpected relationships where a high-risk level is associated with a low-risk level and vice versa.

3.5.4 *Moran's I*

The Global Moran's I is a statistical measure used in spatial analysis to detect spatial clustering or dispersion of similar values within a dataset. It quantifies the degree of spatial autocorrelation by assessing whether neighboring locations exhibit similar or dissimilar attribute values. Specifically, Moran's I considers both the values of the locations and the spatial relationship between them, providing a single coefficient that ranges from -1 to 1, with 0 indicating spatial randomness. This measure helps identify patterns in spatial data, highlighting if similar values tend to be close to each other or dispersed across the study area. The Moran's I values for each variable is calculated for each state and nationally in order to analyze how the housing insecurity factors cluster in space. The formula for Global Moran's I is shown in Equation 3.3.:

$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{3.3}$$

Where:

I : Moran's I statistic, representing the degree of spatial autocorrelation.

N : Total number of spatial units (e.g., census tracts, regions).

W : Total spatial weight in the dataset.

w_{ij} : Spatial weight between spatial units i and j.

x_i : Value of the variable of interest in spatial unit i.

$\bar{x}$: Mean value of the variable of interest across all spatial units.

To measure how housing insecurity risk levels cluster in space, local Moran's I is also used to indicate the spatial relationship of housing insecurity risk levels. The formula for local Moran's I is shown in Equation 3.4. Local Moran's I does not follow the same -1 to 1 structure of global Moran's I, but it retains the structure that positive values indicate stronger spatial autocorrelations and negative values indicate stronger spatial randomness.

$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{3.4}$$

Where:

I : Moran's I statistic, representing the degree of spatial autocorrelation.

N : Total number of spatial units (e.g., census tracts, regions).

W : Total spatial weight in the dataset.

w_{ij} : Spatial weight between spatial units i and j.

x_i : Value of the variable of interest in spatial unit i.

$\bar{x}$: Mean value of the variable of interest across all spatial units.

3.5.5 *Multinomial Logistic Regression*

After the clustering is performed and the clusters are analyzed, each sector is assigned a new variable containing the risk levels for each census tract. Cross split validation is used wherein for each state, a new model is trained on all states except the target state. The probability that each census tract is its actual classification is preserved for the analysis. Additionally, to better understand how the housing insecurity factors contribute to the risk levels, for each sector a national model is trained on the entire dataset so that the model can be analyzed and prediction power can be measured under the best-case scenario. The formula for multinomial logistic regression is shown in Equation 3.5.

$$\log\left(\frac{P(Y=k \mid X)}{P(Y=K \mid X)}\right) = \beta_{0k} + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \tag{3.5}$$

Where:

log : Natural logarithm function

$P(Y=k \mid X)$: Probability of the outcome being in category k given predictor variables X

$P(Y=K \mid X)$: Probability of the outcome being in the reference category K given X

β_{0k} : Intercept for category k

$\beta_1, \beta_2, \ldots, \beta_p$: Coefficients corresponding to predictor variables $X_1, X_2, \ldots, X_p$

$X_1, X_2, \ldots, X_p$: Predictor variables

k : Specific category being predicted

K : Reference category

CHAPTER 4

RESULTS

This analysis focused on a sample of 6,364 rural census tracts with a RUCA code of seven or higher. Four states and Washington D.C. were intentionally excluded from the analysis: Alaska and Hawaii were omitted due to the presence of unique factors, particularly in their rural areas, which may not have been adequately addressed in the existing literature. New Jersey and Rhode Island were excluded from our spatial analysis due to a lack of adequate data. These states are both very urban and once discrepancies in the data were removed, there were not enough observations to include in the analysis. Figure 4.1 shows how the neighbor algorithm changed the state census tract counts. This allows the risk-level assignment to reflect the intra-state nature of communities.

4.1 RUCA Distribution

Figure 4.2 shows the distribution of RUCA codes in the dataset. Small towns with a primary flow within an urban cluster with a population of 2,500 to 9,999 (33 percent) and rural areas with a primary flow to a tract outside an urban area or urban cluster (48 percent) make up the majority of the dataset. The rest are split between small towns with high levels of commuting to a small urban cluster (13 percent) and small towns with low commuting to a small urban cluster (5 percent)

insert descriptive statistics by RUCA code

4.2 *Cluster Analysis*

Here, the results of the cluster analysis are presented for each sector. The primary mechanism for analyzing the clusters is the average cluster medians for all states. The cluster averages were analyzed as well to ensure that the same trends are found in the dataset under a different descriptive statistic. All values are represented as a percentage

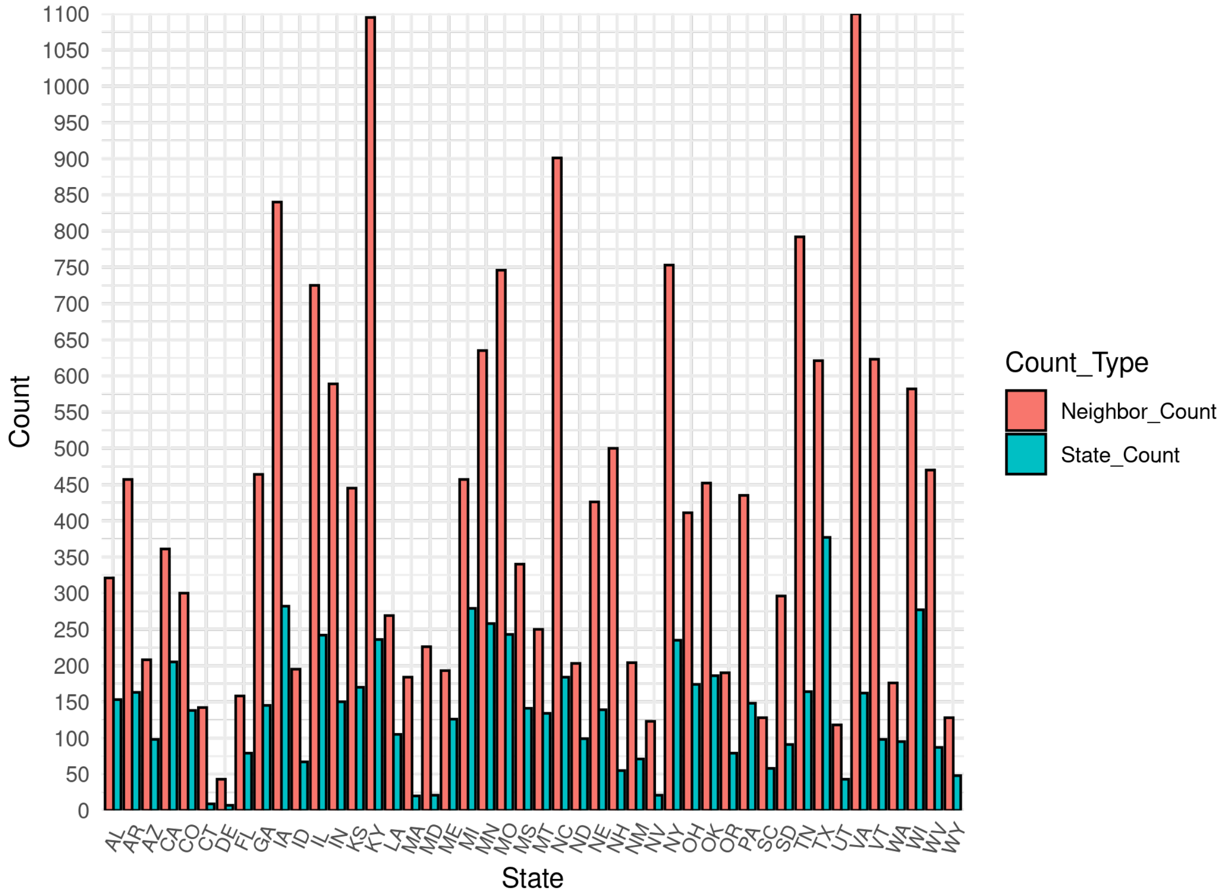

Figure 4.1: State Census Tracts vs. State Neighbors Count

corresponding to the base unit each sector is scaled to. Figure 4.2 shows the distribution of risk levels for each sector. For all sectors except housing cost and demographic diversity, there is a higher number of low-risk rather than high-risk or medium-risk level census tracts. Demographics is the only sector with notably more medium-level than low-level census tracts.

4.2.1 ***Employment Diversity***

The risk levels for employment diversity are determined based on which clusters have the highest number of maximum cluster values compared to the cluster with the lowest number of minimum cluster values. The higher the cluster medians across variables, the better the economic diversity of a cluster. Table 4.1 shows the values for this sector.

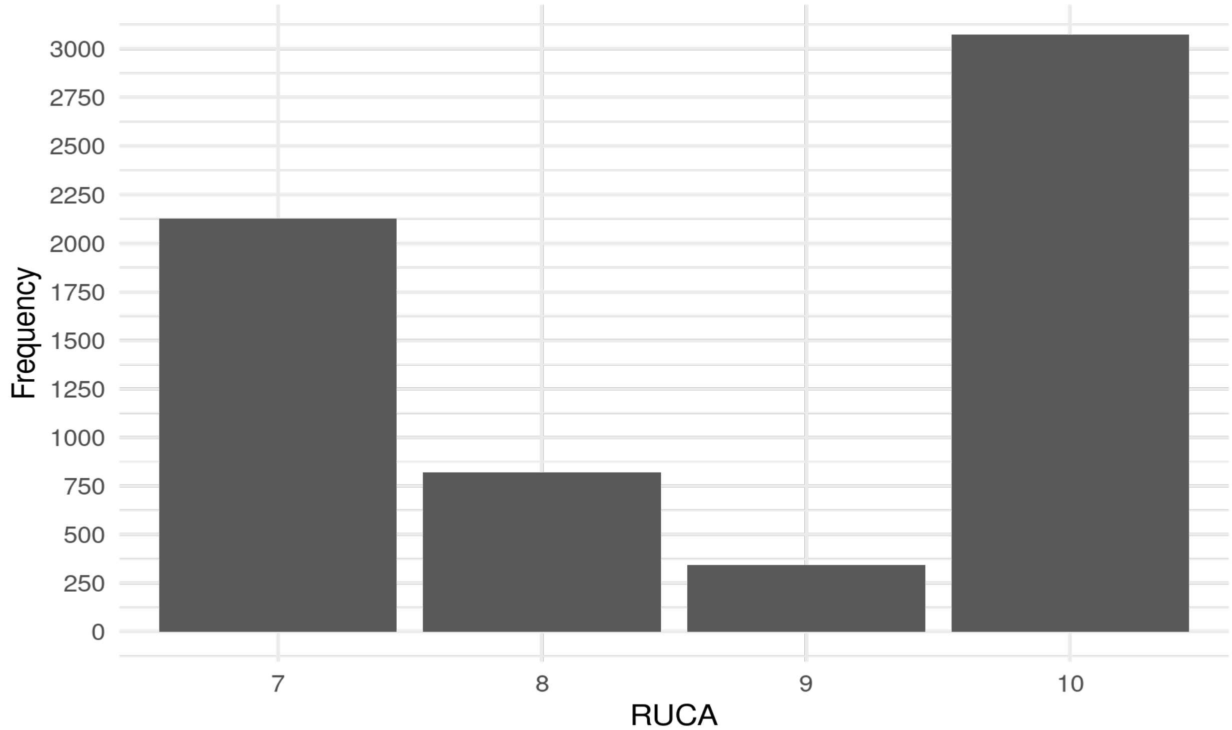

Figure 4.2: State Census Tracts vs. State Neighbors Count

Cluster 1 had the lowest cluster medians in 61 percent of variables, Cluster 2 has the highest cluster median in 53 percent of variables, and Cluster 3 has the middle value in 69 percent of cases. Based on this analysis Cluster one has the lowest level of economic diversity, cluster two has the lowest level of economic diversity, and cluster three has a medium level of economic diversity. Employment in education, health, and social work has the highest presence across each cluster followed by manufacturing.

4.2.2 *Demographics*

Due to the historical forces affecting minorities in both rural and urban areas, the risk levels for demographics are based on which clusters have the highest minority populations and the lowest white populations. Table 4.2 shows the values for this sector. This sector was decided based on the median and average highest, lowest, and medium value counts as clusters two and three had almost the same cluster median counts. Cluster three is the middle value for 90 percent of cluster median variables. Cluster three has the highest

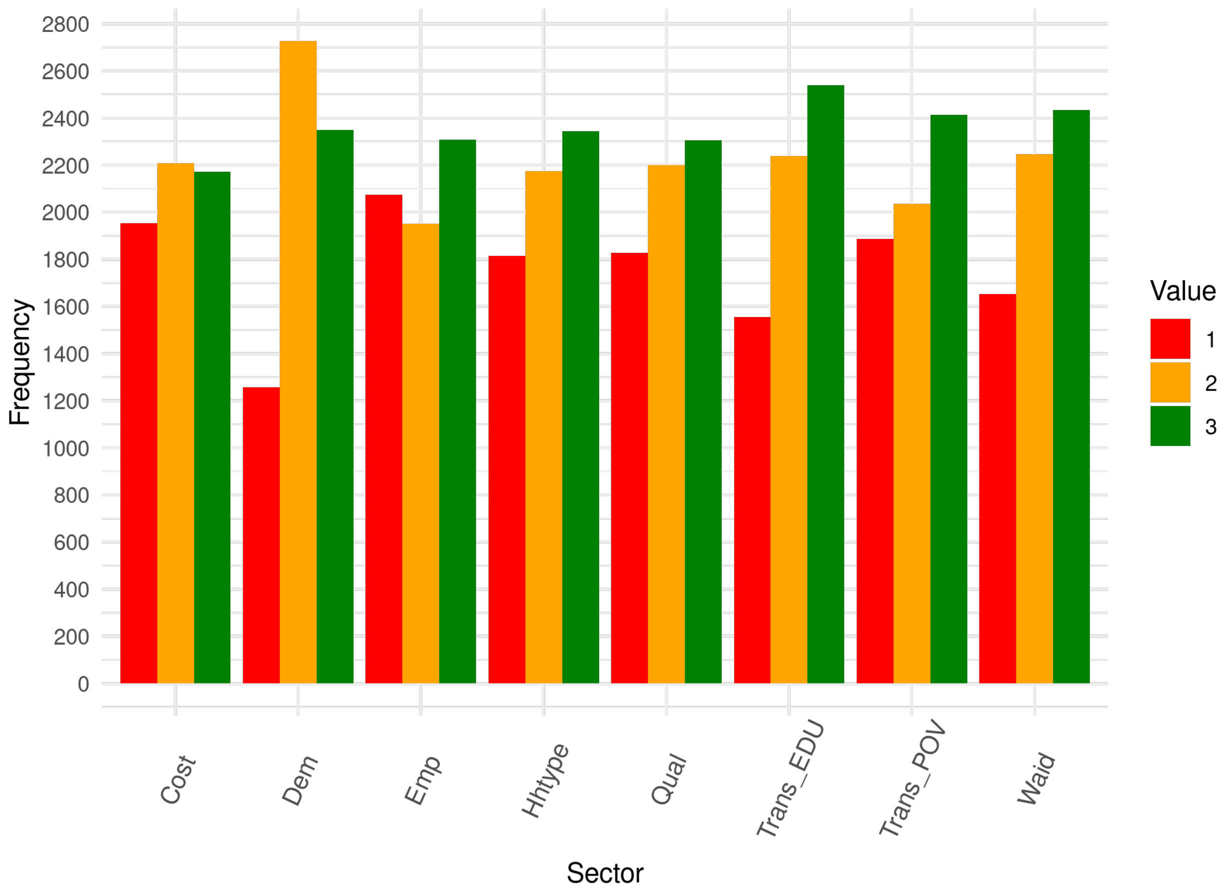

Figure 4.3: Cluster Distribution by Sector

number of highest values across means and medians with 55 percent of variables. Cluster two is the lowest for 50 percent of variables. Cluster three alsos has the largest African American and Hispanic and Latino cluster medians. Based on this analysis, cluster one has a medium risk of housing insecurity, cluster two has a low risk of housing insecurity and cluster three has the highest risk of housing insecurity. One notable observation is that the cluster medians for the male/ female over 18 years old variables are roughly three times higher than the male/ female under 18 variables. This reflects the aging population of rural areas noted in the literature.

Table 4.1: Median Values for Employment Diversity Clusters

Variable	Cluster 1	Cluster 2	Cluster 3
ag_for_fish_hunt_mining	2.54	2.24	1.87
arts_rec_food	2.86	3.12	3.10
construction	3.16	3.06	3.07
edu_health_social	9.38	9.73	9.46
fin_re_insur	1.47	1.60	1.57
information	0.35	0.41	0.37
manufacturing	4.52	5.44	4.75
othersvcs	1.76	1.88	1.91
prof_sci_mgmt_waste	2.17	2.25	2.28
public_admin	1.98	1.88	1.92
retail_trade	4.48	4.78	4.59
trans_warehouse_util	2.17	2.03	2.12
wholesale_trade	0.76	0.86	0.70

Table 4.2: Median Values for Demographic Diversity Clusters

Variable	Cluster 1	Cluster 2	Cluster 3
am_in_ala_nat	0.21	0.28	0.18
asian	0.21	0.22	0.15
black	0.72	0.72	0.85
female_o18	39.25	40.15	38.90
female_u18	10.77	9.79	10.98
haw_pac	0.00	0.00	0.00
hisp_lat	2.92	2.63	3.08
male_o18	38.26	39.38	37.99
male_u18	11.44	10.27	11.70
other	0.32	0.30	0.41
white	94.35	93.67	92.52

4.2.3 *Housing Cost*

For housing cost, risk levels are determined by which clusters have the highest cluster medians. Cluster 1 has the highest value for mortgage high cost. Cluster 2 has the lowewest mortgage and rent high cost cluster medians. cluster 3 has the highest no mortgage and rent high cost cluster medians. Table 4.3 shows the values for this sector. Cluster one becomes the medium risk level, cluster two becomes the low risk level, and cluster three becomes the high risk level.

Table 4.3: Median Values for Housing Cost Clusters

Variable	Cluster 1	Cluster 2	Cluster 3
mortgage_high_cost	5.22	4.35	4.93
no_mortgage_high_cost	2.16	2.18	2.89
rent_high_cost	15.69	14.18	16.79

4.2.4 *Housing Quality*

For housing quality, risk levels are determined by which clusters have the highest values, with preference given to occupied housing as housing conditions in unoccupied housing is of lesser concern than occupied housing. Table 4.4 shows the values for this sector. Cluster one has the highest values for unoccupied housing with incomplete kitchens and plumbing. Cluster three has the medium value for each variable.Cluster three has the highest values for occupied housing with incomplete kitchens and plumbing. Cluster one becomes the lowest risk level, cluster two becomes the medium risk level, and cluster three becomes the highest risk level.

Table 4.4: Median Values for Housing Quality Clusters

Variable	Cluster 1	Cluster 2	Cluster3
all_incomplete_kitchen	25.85	25.76	19.75
all_incomplete_plumb	24.00	22.73	17.28
occ_incomplete_kitchen	0.46	0.52	0.64
occ_incomplete_plumb	0.00	0.11	0.34

4.2.5 *Residential Mobility: Education*

For residential mobility: education, the risk levels are determined by the variables for those who moved with less than a high school education and those in the same house with less than a high school education, and the clusters where more people moved overall will be the highest risk levels. Table 4.5 shows the values for this sector. Cluster one has the medium value for 71 percent of variables including each less than high school education variable. Cluster two has the lowest values for each variable. Cluster 3 has the highest values for 71 percent of variables, including each of the less than high school education variables. Cluster one becomes the lowest risk level because it has medium levels of residential mobility but the highest level of residential stability with a high school education. Cluster two becomes the medium risk level, and cluster three becomes the highest risk level.

Table 4.5: Median Values for Residentail Mobility: Education Clusters

Variable	Cluster 1	Cluster 2	Cluster 3
moved_diff_county_hs	0.51	0.44	0.55
moved_diff_county_less_than_hs	0.13	0.10	0.18
moved_diff_state_hs	0.18	0.10	0.18
moved_diff_state_less_than_hs	0.00	0.00	0.00
moved_in_county_hs	0.95	0.84	1.36
moved_in_county_less_than_hs	0.30	0.24	0.50
same_house_hs	23.99	22.60	22.97
same_house_less_than_hs	7.71	6.97	7.82

4.2.6 *Residential Mobility: Poverty*

The residential mobility: poverty sector follows the criteria of residential mobility: education closely with those the variables for those who moved that are below the poverty level are the highest priority. Table 4.6 shows the values for this sector. Cluster one has the lowest values for every variable. Cluster two has the medium value for 57 percent of variables. Cluster three has the highest values for 57 percent of variables including three of the below the poverty level variables. Cluster one becomes the lowest risk level, cluster two becomes the medium risk level, and cluster three becomes the highest risk level. One noteable observation from the cluster medians is that cluster three has the highest value of those living in the same house below and above the poverty measure, indicating high levels of poverty for census tracts in this cluster.

4.2.7 *Wage and Household Factors*

For household wage/ aid, the clusters with the highest number of maximum cluster medians determines the risk levels with particular attention given to households with no

Table 4.6: Median Values for Residential Mobility: Poverty Clusters

Variable	Cluster 1	Cluster 2	Cluster 3
moved_diff_county_p1	0.30	0.41	0.48
moved_diff_county_p2	0.04	0.12	0.07
moved_diff_state_p1	0.05	0.10	0.08
moved_diff_state_p2	0.00	0.00	0.00
moved_in_county_p1	0.74	1.00	1.06
moved_in_county_p2	0.30	0.43	0.40
same_house_p1	9.86	10.80	12.14
same_house_p2	7.79	8.55	9.04

wage and households with three or more workers. Table 4.7 shows the values for this sector. Cluster one has the lowest cluster medians for 89 percent of variables. Cluster two has the medium value for 55 percent of variables. Cluster three has the highest values for 55 percent of variables and the middle value for the other variables. Notablable high values for cluster three include the Gini index, households with no vehicle, households with at least one worker and no vehicle, and households receiving supplemental security income.

Table 4.7: Median Values for Household Wage/ Aid Clusters

	cluster_1	cluster_2	cluster_3
gini_index	42.69	42.91	44.07
hh_3plus_worker	1.85	1.67	1.81
hh_no_investment_income	32.64	33.04	33.29
hh_no_other_income	36.28	36.81	36.79
hh_no_vehicle	1.96	2.12	2.34
hh_no_wage	13.38	14.11	14.01
hh_public_assistance	4.84	5.54	5.40
hh_ssi	2.17	2.39	2.51
hh_worker_no_vehicle	1.28	1.45	1.59

4.2.8 ***Housing Type***

For the housing type sector, owner single unit is considered the safest housing while renters and owners of unconventional housing and mobile homes are considered high risk. This sector required a combination of means and medians for the analysis because, for several variables, all cluster medians are zero. Table 4.8 shows the values for this sector.

Cluster three has the highest owner mobile and the medium value for renter mobile. Cluster three has the highest renter and owner unconventional cluster averages. Cluster one has the highest owner single and the lowest renter mobile home. Cluster one becomes the low risk level, cluster two becomes the medium risk level, and cluster three becomes the high risk level.

Table 4.8: Median Values for Housing Type Clusters

Variable	Cluster 1	Cluster 2	Cluster 3
owner_2to4	0.00	0.00	0.00
owner_5plus	0.00	0.00	0.00
owner_mobile	8.29	10.06	10.41
owner_single	90.75	88.67	88.42
owner_unconvent	0.00	0.00	0.00
renter_2to4	8.29	10.57	10.59
renter_5plus	5.78	8.50	7.70
renter_mobile	9.25	13.04	10.79
renter_single	68.16	55.67	60.95
renter_unconvent	0.00	0.00	0.00

4.2.9 *Rurality and Risk Levels*

To better understand housing insecurity risk in rural areas, it is important to look at the risk levels as they relate to the scale of rurality used. Table 4.9 shows the percentage of each RUCA code that has a high-risk level for each sector. While RUCA code 10 makes up 50 percent of all census tracts, 43 to 50 percent of all high-risk level census tracts have this RUCA code. RUCA code 7 is next with a range between 32 and 39 percent of high-risk level census tracts while making up only 33 percent of the dataset. At the opposite end of the spectrum, RUCA code 9 makes up 5 percent of the dataset and only 5 percent of high-risk census tracts. Figure 4.4 shows the risk level of each census tract across each sector. Each census tract is assigned a color red (high-risk), yellow (medium-risk), and green (low-risk) for each sector. These colors are then saturated based on the probability for each sector. The colors are then blended so that the map reflects how well the state fits into its national train-split model and the overall risk level of the census tract. Many census tracts fall somewhere between green and yellow, with pockets of light shades of red visible.

Table 4.9: High-Risk Census Tract RUCA Breakdown

	sector	RUCA	Pct
1	Qual_Cluster	10	0.5
2	Emp_Cluster	10	0.49
3	Dem_Cluster	10	0.49
4	Cost_Cluster	10	0.49
5	Waid_Cluster	10	0.49
6	Trans_POV_Cluster	10	0.48
7	Trans_EDU_Cluster	10	0.46
8	Hhtype_Cluster	10	0.43
9	Trans_EDU_Cluster	7	0.39
10	Hhtype_Cluster	7	0.37
11	Waid_Cluster	7	0.35
12	Emp_Cluster	7	0.34
13	Trans_POV_Cluster	7	0.34
14	Cost_Cluster	7	0.33
15	Qual_Cluster	7	0.33
16	Dem_Cluster	7	0.32
17	Hhtype_Cluster	8	0.15
18	Dem_Cluster	8	0.14
19	Trans_POV_Cluster	8	0.13
20	Cost_Cluster	8	0.13
21	Qual_Cluster	8	0.13
22	Emp_Cluster	8	0.12
23	Waid_Cluster	8	0.11
24	Trans_EDU_Cluster	8	0.1
25	Emp_Cluster	9	0.05
26	Dem_Cluster	9	0.05
27	Trans_POV_Cluster	9	0.05
28	Cost_Cluster	9	0.05
29	Qual_Cluster	9	0.05
30	Hhtype_Cluster	9	0.05
31	Trans_EDU_Cluster	9	0.04
32	Waid_Cluster	9	0.04

To highlight areas of particular concern, a threshold of 12 out of 24 was used where 8 equals a high-risk across all sectors and 24 equals a low-risk across all sectors to highlight areas that face the highest risk based on the cluster analysis. This results in 115 census tracts as high risk and 661 at a medium risk based on the sum of their risk level variables. Figure 4.5 highlights the high-risk areas in red, and the medium-risk levels in yellow. The

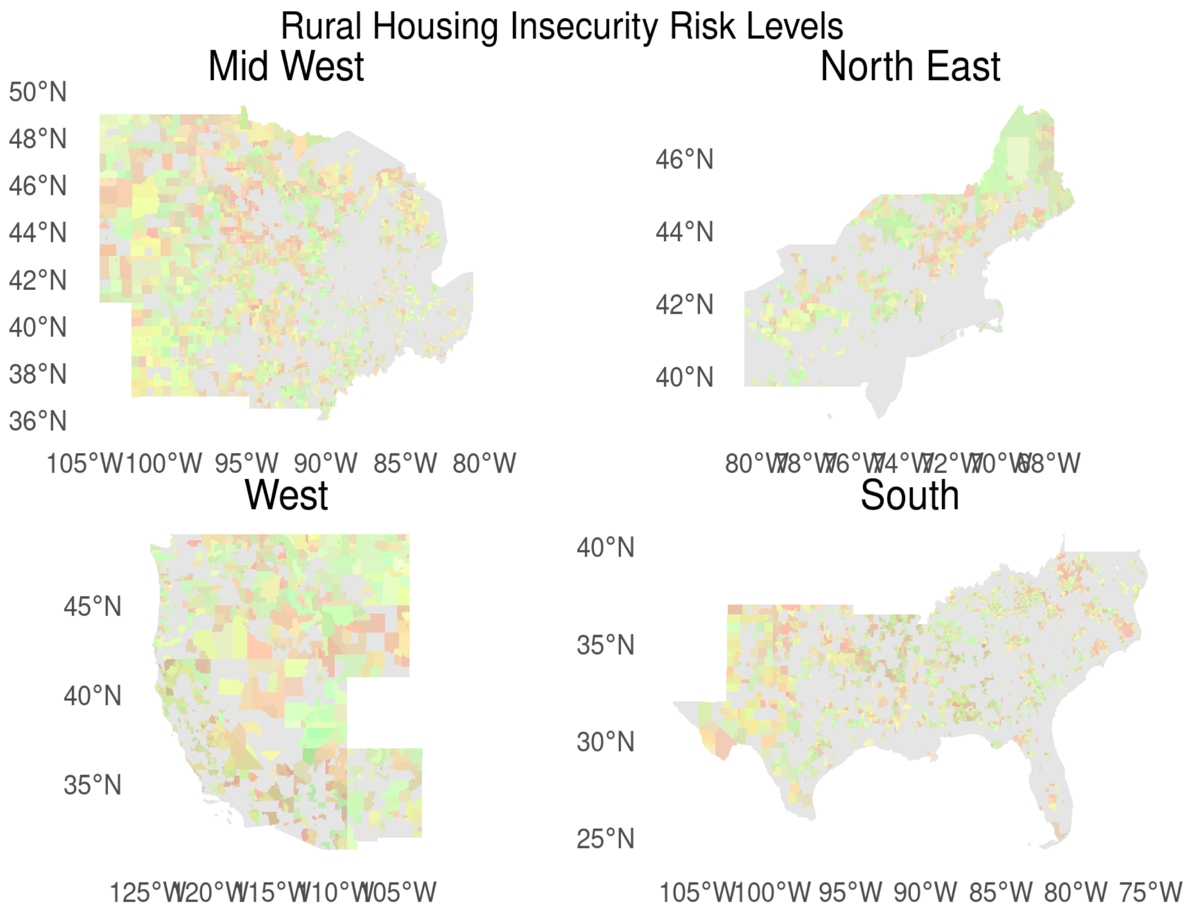

Figure 4.4: Risk Level Across Sectors

majority of the high-risk census tracts are in Minnesota (26), Wisconsin (26), Texas (24), Arizona (21), Missouri (18), Georgia (16), North Carolina (13), Montana (11), North Dakota (11), and Oklahoma (10). The other 104 high-risk census tracts are spread across 27 other states.

4.3 *Association Rules*

There are three areas of investigation for the association rules generated from the housing insecurity risk levels. First are high-risk-to-high-risk associations (1:1), second are low-risk-to-low-risk associations (3:3), third are inverse relationships: low-risk-to-high-risk associations (3:1) and high-risk-to-low-risk associations (1:3). Tables 4.10, 4.11, 4.12, and 4.13 show the average support, average confidence, coverage, and average lift for the

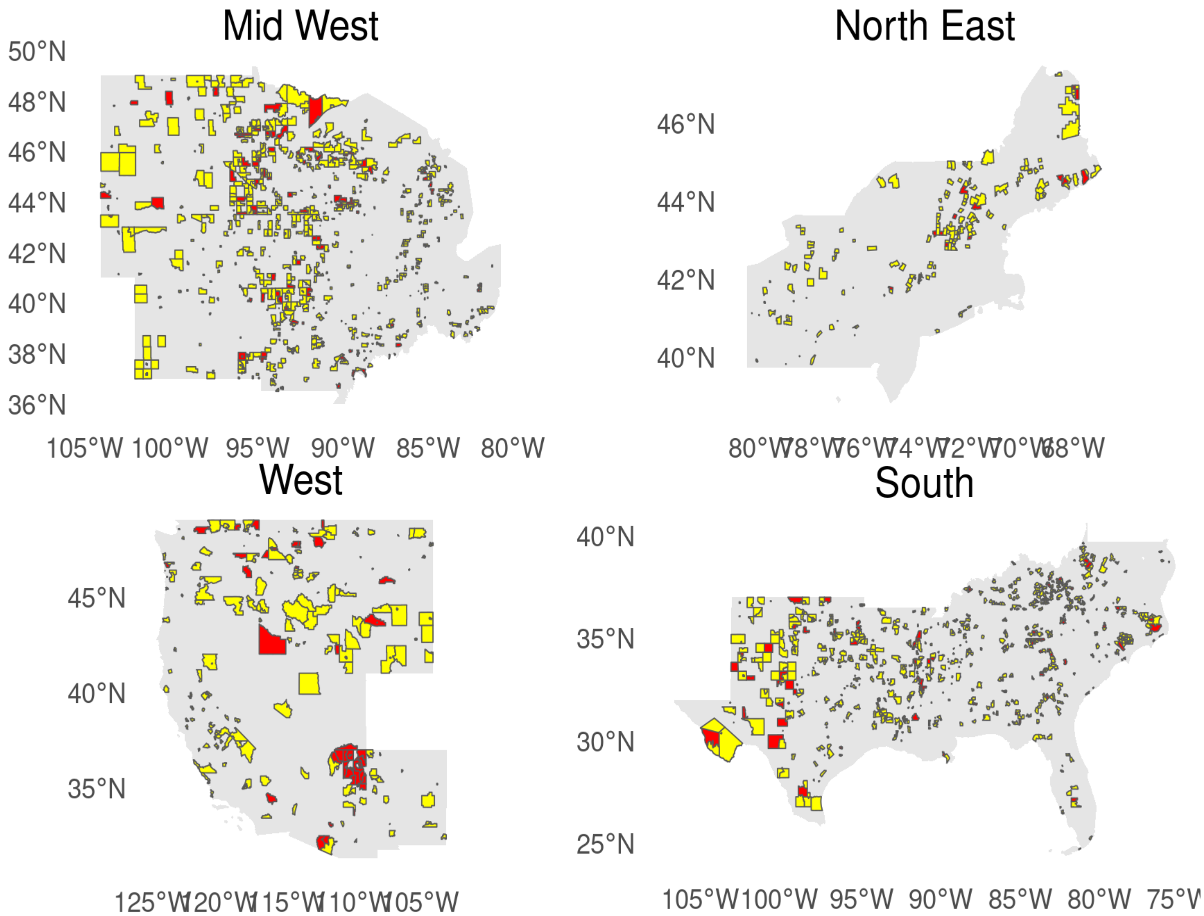

Figure 4.5: Risk Level Across Sectors

different association rules. Figure 4.6 shows the overall trends in the association rules. Support is low with confidence below 0.2 for most of the rules. There are 23 rules with support greater than 0.2 that reflect the occurrence of each risk level for each sector. The plot shows a significant amount of clustering around the 0.35 confidence and 0.1 support range. For each set of association rules, their average lift values indicate that the likelihood of finding the items together is only slightly more or slightly less than their likelihood of being found together by chance. The high-risk-to-high-risk associations have the lowest average support values of the four groups of rules, and low-risk-to-low-risk associations have the highest average support. All average confidence values range from 0.2 to 0.4, indicating that for the risk level on the left-hand side of the transaction, there is an average 20 to 40 percent probability of each other risk level being on the right-hand side of the transaction. Overall, the association rules indicate that there is little consistency in census tracts showing signs of housing insecurity risk at different levels.

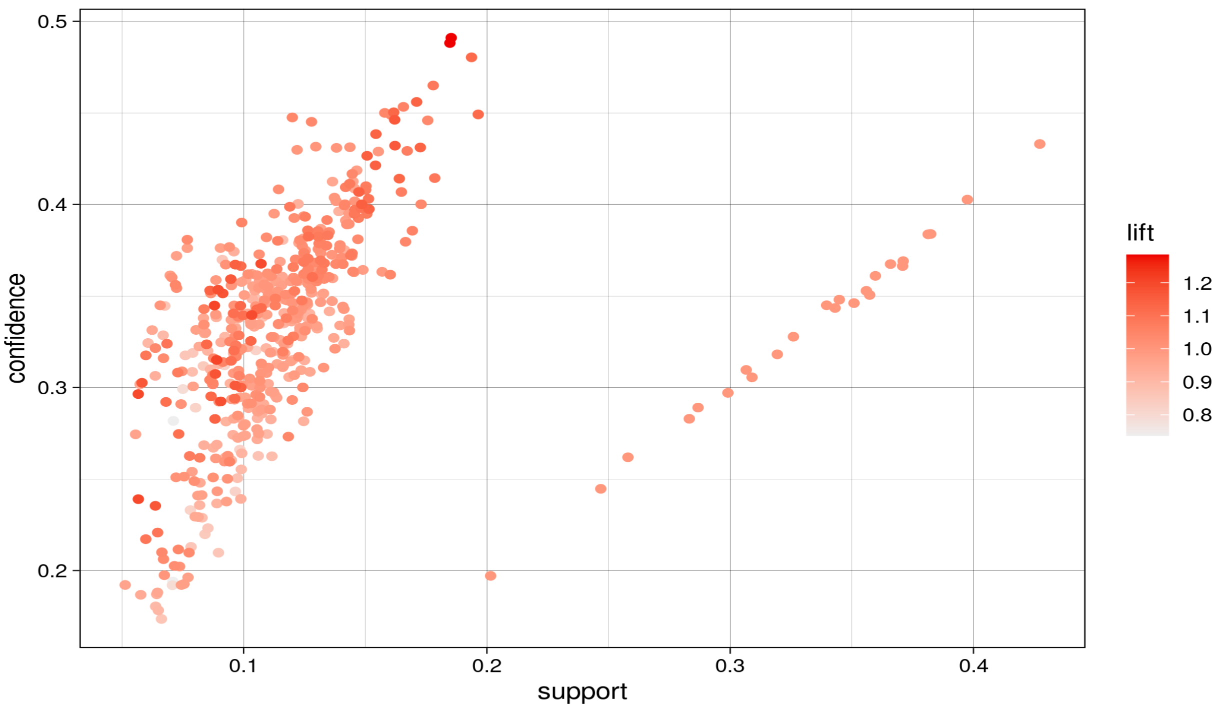

Figure 4.6: Scatter plot of Association Rules Statistics

Table 4.10: High Risk Association Average Statistics

Sector	Average Support	Average Confidence	Average Coverage	Average Lift
employment	0.09	0.27	0.33	0.99
demographics	0.06	0.31	0.2	1.1
rm: education	0.07	0.3	0.25	1.1
rm: poverty	0.09	0.3	0.3	1.1
cost	0.09	0.29	0.31	1.1
qual	0.08	0.29	0.29	1.1
housing type	0.08	0.3	0.29	1.1
household factors	0.08	0.32	0.26	1.2

4.4 *Moran's I*

While the association rules dealt exclusively with the housing insecurity risk levels, Moran's I spatial autocorrelation is used to examine how values group in space for all

Table 4.11: Low Risk Association Average Statistics

Sector	Average Support	Average Confidence	Average Coverage	Average Lift
employment	0.14	0.39	0.37	1.00
demographics	0.13	0.36	0.37	1
rm: education	0.15	0.38	0.4	1
rm: poverty	0.15	0.4	0.38	1.1
cost	0.13	0.37	0.34	1
qual	0.14	0.38	0.36	1
housing type	0.14	0.39	0.37	1
household factors	0.15	0.4	0.38	1.1

Table 4.12: Low to High Risk Association Average Statistics

Sector	Average Support	Average Confidence	Average Coverage	Average Lift
employment	0.09	0.25	0.37	0.95
demographics	0.11	0.28	0.37	0.99
rm: education	0.11	0.27	0.4	0.95
rm: poverty	0.09	0.24	0.38	0.89
cost	0.09	0.27	0.34	0.99
qual	0.1	0.27	0.36	0.97
housing type	0.1	0.26	0.37	0.96
household factors	0.1	0.25	0.38	0.91

variables. Moran's I is calculated for every state and the entire dataset. Table 4.14 shows the descriptive statistics for the significant Moran's I values Table 4.15 shows the average of statistically significant Moran's I values across sectors. Variable averages show similar trends as the sector averages. Manufacturing has the highest average Moran's I statistic at 0.43, followed by white at 0.38 and ag-for-fish-hunt-mining at 0.34. There is a weak spatial autocorrelation between the different levels of rurality at 0.29. The sector averages are low, ranging from 0.19 to 0.32. While averages are low, certain observations deserve further attention. There are 134 statistically significant Moran's I values greater than 0.5 not including the nationwide calculations. These observations are spread across 38 states, with Arizona and New Mexico accounting for 16 percent of high Moran's I statistics. These high values are explored further in the next section on outliers in calculated Moran's I values.

Table 4.13: High to Low Risk Association Average Statistics

Sector	Average Support	Average Confidence	Average Coverage	Average Lift
employment	0.12	0.37	0.33	0.98
demographics	0.7	0.36	0.2	0.96
rm: education	0.09	0.35	0.25	0.95
rm: poverty	0.11	0.35	0.3	0.96
cost	0.11	0.36	0.31	0.94
qual	0.1	0.36	0.19	0.95
housing type	0.1	0.36	0.29	0.96
household factors	0.09	0.33	0.26	0.89

Table 4.14: Moran's I Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
N	2,018	445.284	1,153.592	12	6,333
Morans_I	2,018	0.259	0.141	0.014	0.935
std_dev	2,018	5.451	6.023	1.646	72.172
variance	2,018	0.004	0.005	0.00001	0.083
expectation	2,018	−0.006	0.007	−0.091	−0.0002
p_value	2,018	0.005	0.011	0.000	0.050

4.4.1 Moran's I Outliers

Here, outliers based on the interquartile range method are calculated for the Moran's I statistics are analyzed to highlight areas that do not follow the overall trends. Figure 4.8 shows the distribution of Moran's I for each sector. Demographics, household wage/ aid, do not have any outliers based on the IRQ method. mobility poverty has 13 outliers. The mean of all mobility poverty observations is 0.19 while the mean for the outliers is 0.47. 69 percent of these outliers are the same house below the poverty line variable. Connecticut, Nevada, and Arizona have surprisingly high Moran's I statistics for the mobility poverty risk levels variable. The average for these three states is 0.45 compared to 0.21 for the same variable. There are 4 outliers in the residential mobility: education sector with same_house_less_than_hs in Ohio, California, and all states. The final outlier is same_house_hs in Maryland. These outliers have an average of 0.61 while all sector observations have an average of 0.23. For housing type, there are two outliers:

Table 4.15: Average Moran's I by Sector

sector	Morans_I	std_dev	variance	expectation	p_value
Demographics	0.32	5.87	0.00	-0.01	0.00
Employment	0.25	4.58	0.00	-0.01	0.01
Household Wage/ Aid	0.26	4.48	0.00	-0.01	0.00
Housing Cost	0.21	3.78	0.00	-0.01	0.01
Housing Quality	0.27	4.76	0.00	-0.01	0.00
Housing Type	0.25	4.44	0.00	-0.01	0.01
RUCA	0.31	5.25	0.00	-0.01	0.00
Transience: Education	0.23	4.08	0.00	-0.01	0.01
Transience: Poverty	0.20	3.54	0.00	-0.01	0.01

owner_single and owner_mobile, both in South Dakota. These outliers have an average Moran's I of 0.63 while the sector has an average of 0.25. For housing quality there are two outliers: occupied incomplete plumbing and occupied incomplete kitchen, both in the state of New Mexico. The sector average is 0.28 while these outliers have an average of 0.66. Housing cost has six outliers: mortgage high cost in Arizona, Maryland, Minnesota, Nevada, and New Mexico. The variable average is 0.28 while these observations have an average of 0.5. For economic diversity there are 16 outliers, 10 of these observations are for manufacturing nationally and in Virginia, Florida, Indiana, Kentucky, Mississippi, Ohio, Pennsylvania, South Dakota, and Virginia. The average Moran's I statistic for this sector was 0.47 while these outliers have an average of 0.68. five of these outliers are for the agriculture, forestry, fishing, hunting, and mining variable in New Mexico, Oklahoma, Texas, Washington, and nationally. The average Moran's I statistic for this variable is 0.36 while these outliers have an average of 0.61. Figure 4.7 shows the distribution of Moran's I for each state by region.

4.4.2 Risk Level Analysis

The following table shows the local spatial autocorrelation for each cluster across each sector. There are notable local Moran's I statistics for the low and high-risk level census tracts. The medium-risk level census tracts had very small local Moran's I statistics. The results indicate that the extremes of the scale tend to cluster around each other: high-risk

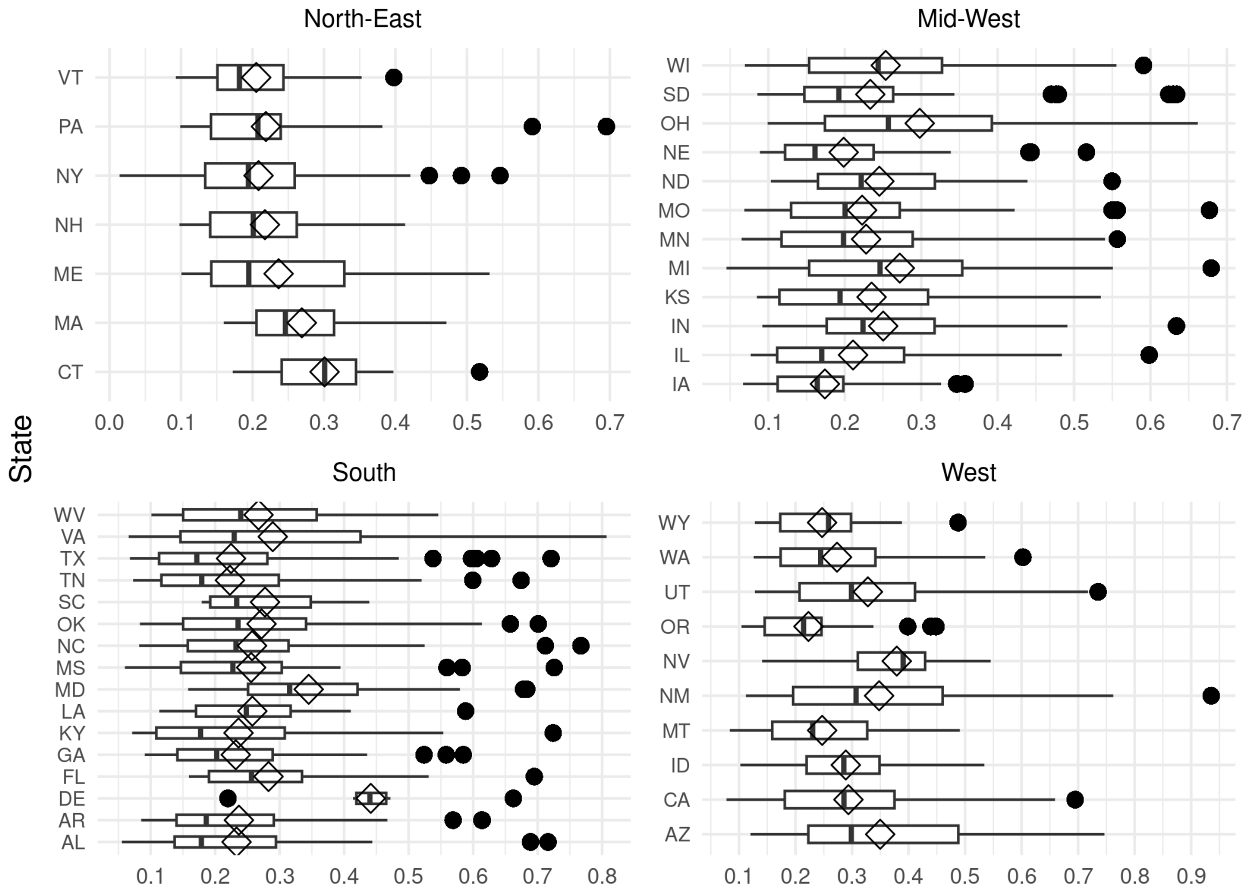

Figure 4.7: Boxplot of Moran's I by Region

census tracts are close to high-risk census tracts and low-risk census tracts are close to low-risk census tracts while there is a level of spatial randomness in the grouping of medium-risk level census tracts.

4.5 *Multinomial Logistic Regression*

The final method applied in this study is a multinomial logistic regression performed on each sector of data and tested on the data for each state. The probability that a predicted risk level is the actual risk level is used to measure how well the data for each state can be predicted based on a model trained on the other states alongside the confusion matrices for each sector's actual and predicted classification. National models using in-sample evaluation are used to measure how well a census tract's risk levels can be predicted.

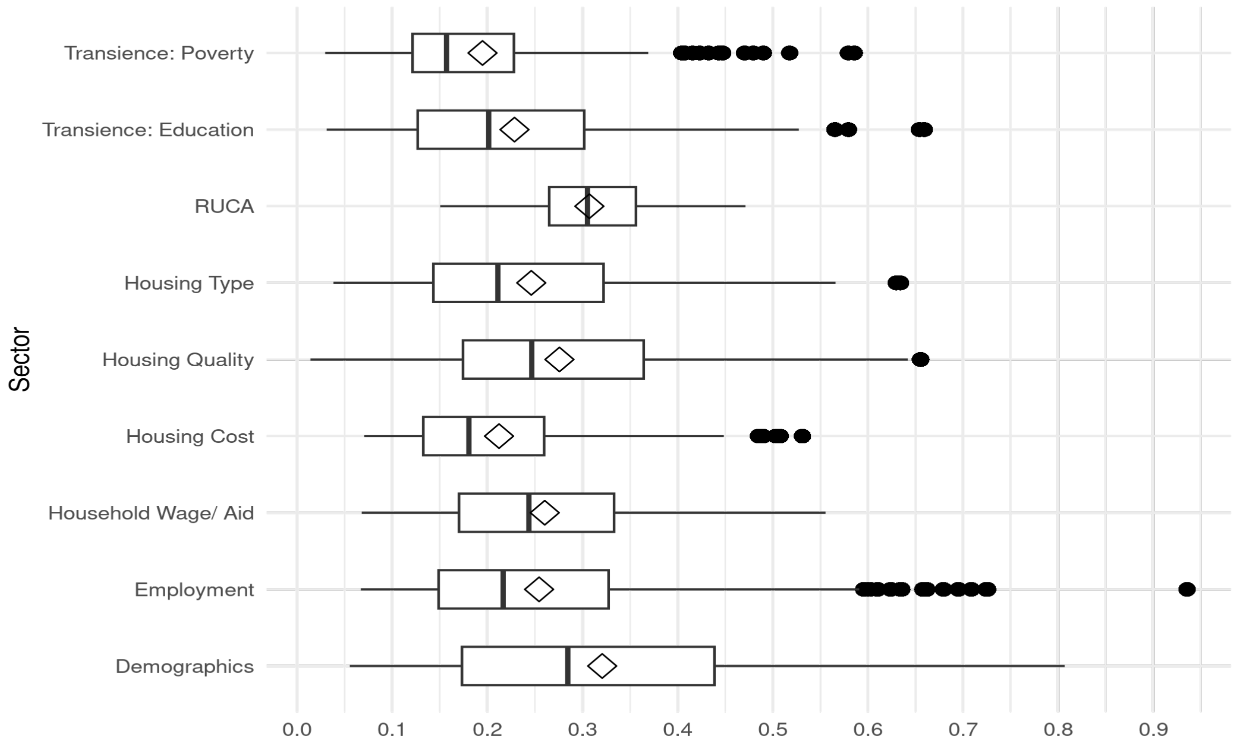

Figure 4.8: Boxplot of Moran's I by Sector

Table 4.16: Local Morans I Risk-Level Results

sector	c1	c2	c3
emp_cluster	0.83	0.00	0.89
dem_cluster	2.09	0.14	0.58
trans_edu_cluster	1.28	0.10	0.25
trans_pov_cluster	1.16	0.05	0.48
cost_cluster	0.79	-0.00	0.57
qual_cluster	1.06	0.00	0.74
hhtype_cluster	0.91	-0.00	0.68
waid_cluster	1.52	0.09	0.22

4.5.1 Probability

Average probability for all sectors was low as demonstrated by Figure 4.11: employment diversity, housing quality, residential mobility: poverty and household factors had an average probability of 34 percent; housing type and housing cost had an average of 0.35; residential mobility: poverty had an average of 36 percent; demographics had the highest average probability at 0.38. Demographics had the highest standard deviation at 14

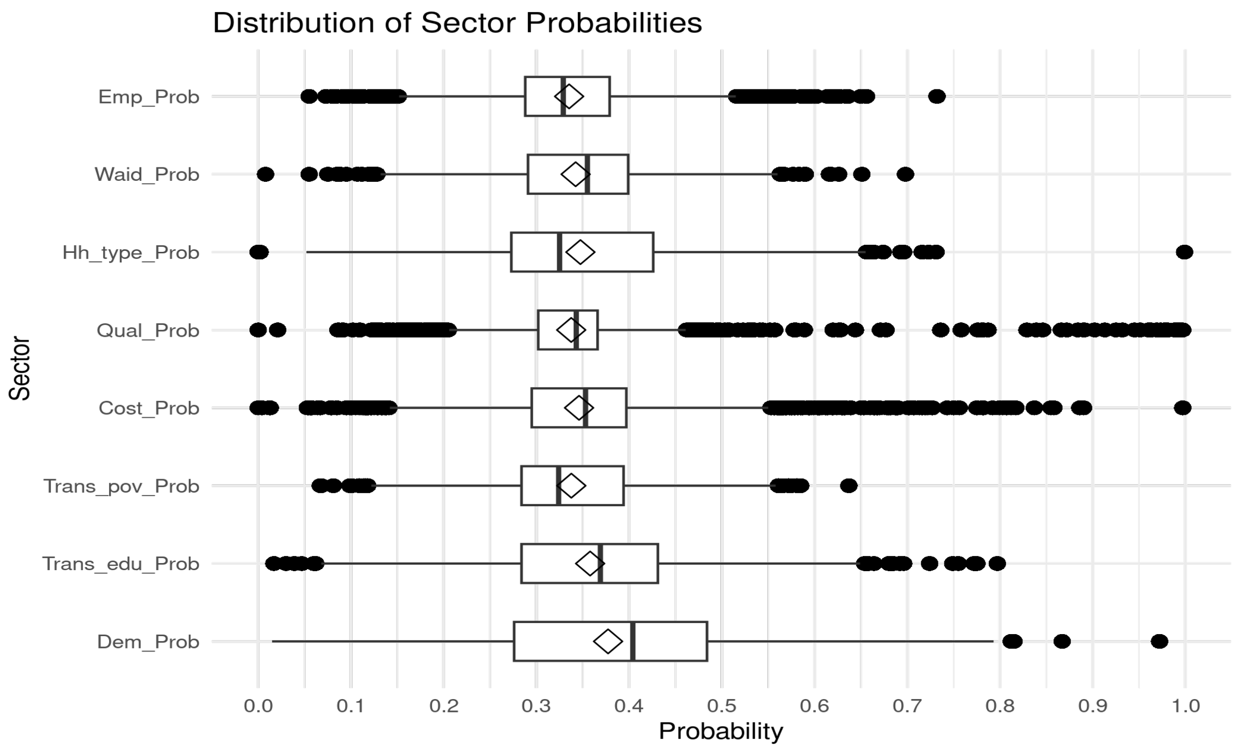

Figure 4.9: Boxplot of Moran's I by Sector

percent, indicating a high degree of variation in predictability. For each sector in each state, Utah had the best prediction results with an average of 41 percent and Minnesota was the hardest to predict at 31 percent between sectors. Average probabilities for each cluster across sectors were similarly low. Across every sector except demographics, the models predicted the highest average probabilities for low-risk level census tracts. For demographics, the models had the highest average probability for the medium-risk level census tracts. Figure 4.11 Shows the distribution of average probability for each state. With an average of 0.35, no states performed well across sectors. The test-train split models performed best on the state of Utah, which had an average accuracy of just over 40 percent at 41 percent. One last area of interest is any trends that may exist between the probabilities for each sector. Figure 4.10 Shows that there are no significant correlations between the probabilities across sectors. The following subsections explore the performance of the state models and national models for each sector.

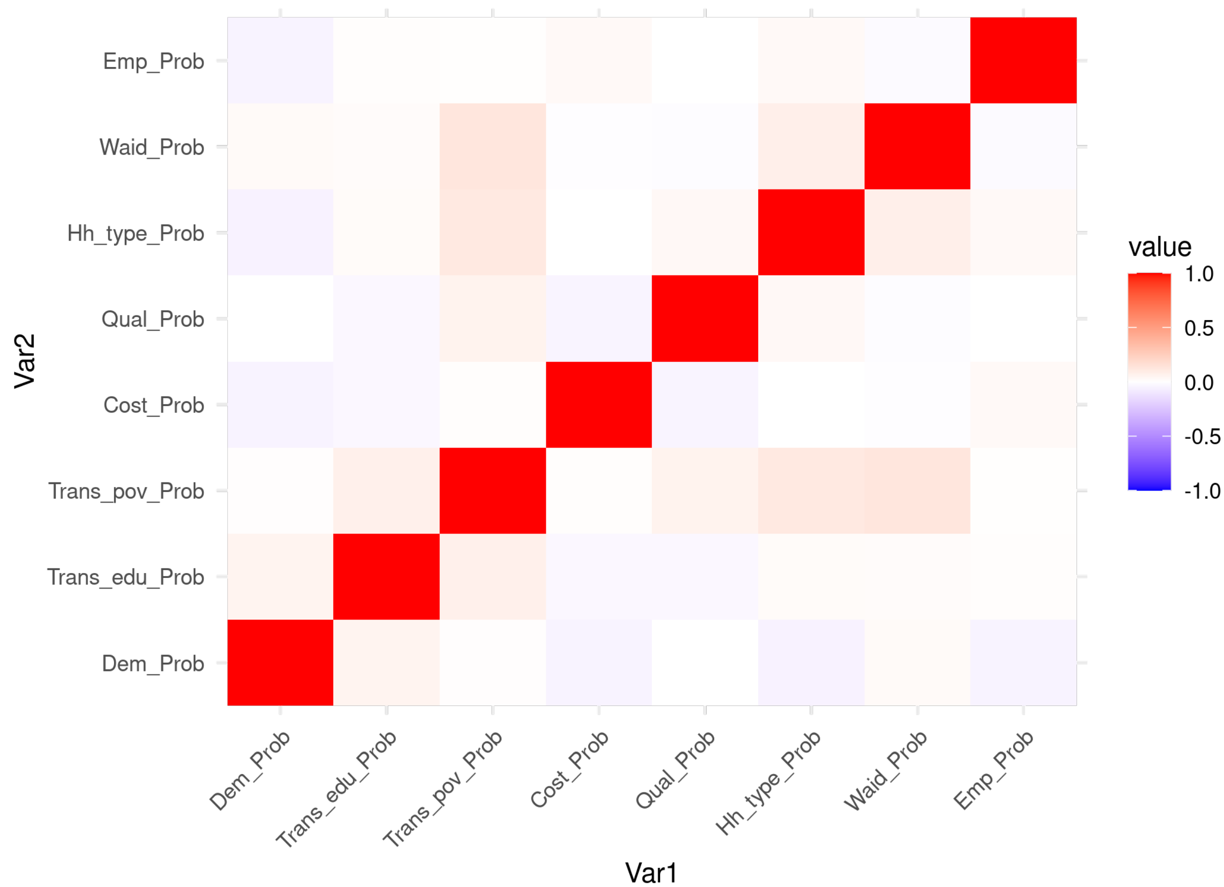

Figure 4.10: Correlation Plot of Sector Probabilities

4.5.2 Accuracy

The confusion matrices for each sector show that accuracy is low, with the models for most sectors over-classifying census tracts as low risk significantly harmed their accuracy. Presented here are also the accuracy results for national models tested using in-sample evaluation to measure accuracy under the best-case scenario.

Table 4.17 shows that the models struggled to classify the medium risk levels and high risk levels with the best performance on the low risk levels for the economic diversity sector. These models were more successful at classifying census tracts with higher levels of economic diversity. The state models correctly classified 34 percent of low-risk census tracts, 11 percent of medium-risk census tracts, and 53 percent of low-risk census tracts.

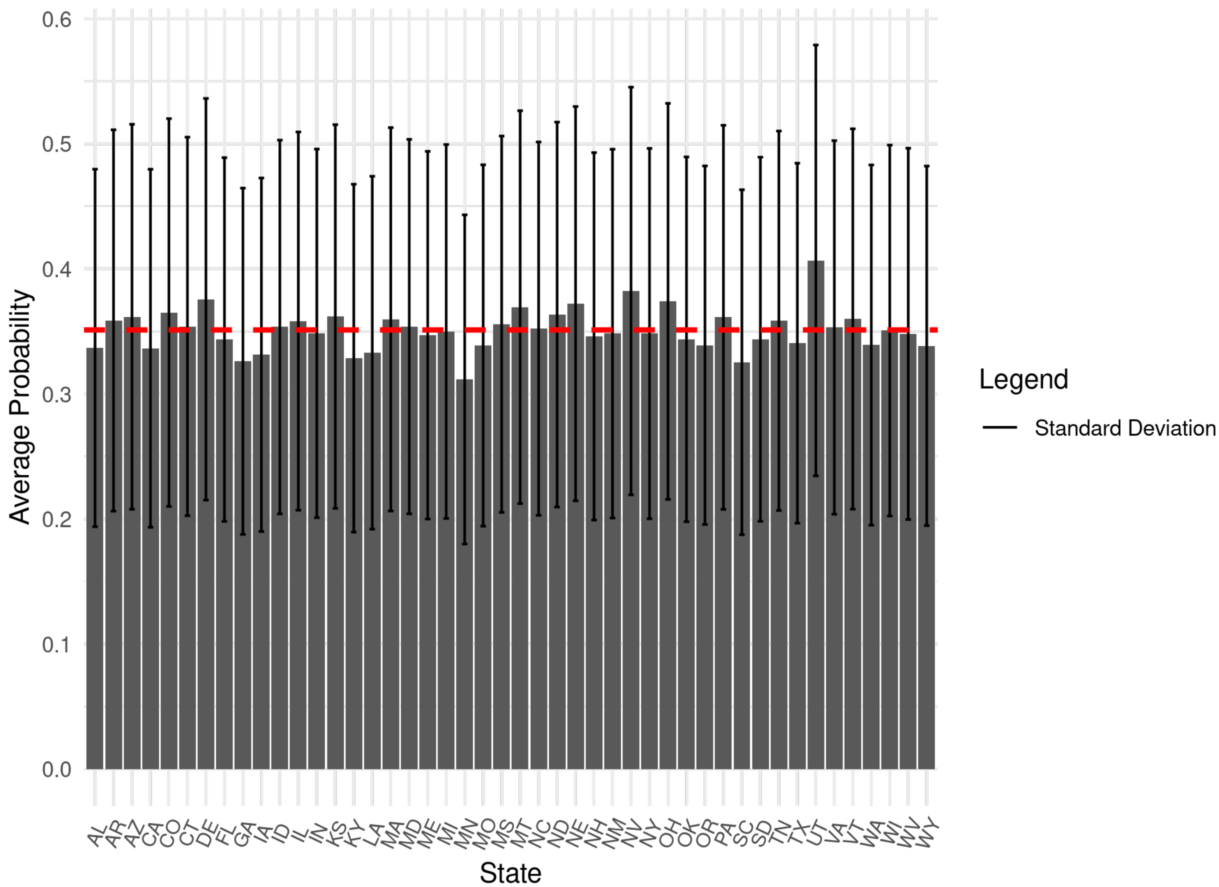

Figure 4.11: Bar Graph of Average Probabilities by State with Error Bars

Overall, the state models were 34 percent accurate and the national model was 41 percent accurate.

Table 4.17: Employment Confusion Matrix and Statistics

	High Risk	**Medium Risk**	**Low Risk**
High Risk	708	675	654
Medium Risk	386	232	431
Low Risk	987	1051	1238

The demographic diversity models were able to predict medium risk levels and low risk levels significantly better than high risk levels for the demographic diversity sector. Table 4.18 shows The models were most capable of predicting medium-risk level census tracts. The models accurately predicted three percent of high-risk level census tracts while classifying medium-risk census tracts with 72 percent accuracy and 42 percent of low-risk census tracts. The state and national models predicted 48 and 52 percent of census tracts accurately.

Table 4.18: Demographics Confusion Matrix and Statistics

	High Risk	Medium Risk	Low Risk
High Risk	48	95	76
Medium Risk	843	1975	1274
Low Risk	378	666	1007

Table 4.19 shows that the housing cost models struggled to classify the low risk level census tracts . They also struggled to differentiate between medium risk level and high risk level census tracts . The state models accurately predicted 34 percent of high-risk level census tracts, 44 percent of medium-risk level census tracts, and 43 percent of low-risk level census tracts. The state models were 45 percent accurate and the national model was 41 percent accurate.

Table 4.19: Housing Cost Confusion Matrix and Statistics

	High Risk	Medium Risk	Low Risk
High Risk	669	446	422
Medium Risk	660	962	817
Low Risk	625	803	938

Table 4.20 shows that the housing quality models significantly over-classified census tracts as low risk level . The state models correctly classified 15 percent of high-risk census tracts, 13 percent of medium-risk level census tracts, and 63 percent of low-risk level census tracts. Overall, the state models had 39 percent accuracy and the national model had 32 percent accuracy.

Table 4.21 shows that the residential mobility: education models significantly over-classified census tracts as low risk level . They successfully predicted 14 percent of low-risk census tracts, 39 percent of medium-risk level census tracts, and 60 percent of

Table 4.20: Housing Quality Confusion Matrix and Statistics

	High Risk	**Medium Risk**	**Low Risk**
High Risk	285	278	179
Medium Risk	592	296	673
Low Risk	954	1632	1458

low-risk census tracts. The state models had an accuracy of 46 percent and the national model had an accuracy of 42 percent.

Table 4.21: Residential Mobility: Education Confusion Matrix and Statistics

	High Risk	**Medium Risk**	**Low Risk**
High Risk	219	200	197
Medium Risk	427	881	810
Low Risk	917	1169	1542

Table 4.22 shows that the residential mobility: poverty models significantly over-classified census tracts as low risk level . they successfully predicted 15 percent of high-risk census tracts, 14 percent of medium-risk census tracts, and 72 percent of high-risk census tracts. The state models had an accuracy of 41 percent and the national model had an accuracy of 37 percent.

Table 4.22: Residential Mobility:Poverty Confusion Matrix and Statistics

	High Risk	**Medium Risk**	**Low Risk**
High Risk	298	394	353
Medium Risk	420	287	324
Low Risk	1182	1356	1748

Table 4.23 shows that the household factor models significantly over classified census tracts as low risk level . They successfully predicted 12 percent of high-risk level census tracts, 33 percent of medium-risk level census tracts, and 54 percent of low-risk level census tracts. The state models had an accuracy of 42 percent and the national model had an accuracy of 36 percent.

Table 4.23: Household Factors Confusion Matrix and Statistics

	High Risk	**Medium Risk**	**Low Risk**
High Risk	195	163	208
Medium Risk	492	748	897
Low Risk	972	1352	1335

Table 4.24 shows that the housing type models significantly over-classified census tracts as low risk level . they successfully predicted 6 percent of low-risk level census tracts, 33 percent of medium-risk level census tracts, and 54 percent of high-risk census tracts. The state models had an accuracy of 45 percent and the national models had an accuracy of 43 percent.

Table 4.24: Housing Type Confusion Matrix and Statistics

	High Risk	**Medium Risk**	**Low Risk**
High Risk	195	163	208
Medium Risk	492	748	897
Low Risk	972	1352	1335

CHAPTER 5

DISCUSSION AND FUTURE DIRECTIONS

Housing insecurity affects all aspects of life for the individuals that experience it and it has grave consequences for communities. This is especially true for rural areas where a conundrum of factors over recent decades has reduced the amount of community resources available for combatting housing insecurity and homelessness. These problems include insert list of rural problems from literature review here. Unfortunately, our understanding of how these factors affect housing insecurity and homelessness in rural areas is limited due to the urban-centric lens used by researchers and policy-makers. The present analyses examined variables associated with housing insecurity in a sample of rural census tracts based on RUCA designations to group census tracts into risk levels that show similar signs of housing insecurity risk. Several data mining techniques were then applied to analyze how risk levels and variables relate to each other.

5.1 *Interpretation of Findings: RQ1*

How can risk factors of be used to identify risk levels of housing insecurity while accounting for the variation in rural areas and what do the risk levels say about rural areas?

K-medoid clustering was used to cluster census tracts with similar values across 8 different sectors of variables, encapsulating different aspects of housing insecurity. To account for the variation in rural areas, a neighbors algorithm was used so that bordering census tracts that make up rural communities could be included in the clustering for each state. There are three benefits to this approach. First, clustering by each state and neighboring census tracts prevents the most vulnerable communities in one state from influencing the risk level of the most vulnerable communities in another state. Second, this state-by-state approach makes this research a tool that policy-makers and researchers can use in their states because everything is based at the state level. Finally and most significantly, a relative approach to measuring housing insecurity can capture the variation in rural areas better than an absolute approach with rigidly defined categories. Whether or not this methodology adequately allows for these differences requires further research to validate.

The cluster analysis highlights ? important observations about rural areas. First is the importance of education, health, and social work employment in rural areas (See Table 4.1). The overall average for this variable was 10 percent and the cluster medians for each risk level is 9 percent. This is significant because the variables for jobs that rural areas typically depend on like agriculture, construction, and forestry, have notably lower averages and cluster medians. The impact of this is two-fold. First, many of the jobs that fall into education, health, and social work fall into employment in the public sphere, which is greatly affected by the decreases in funding caused by several processes affecting rural areas (Source). Second, it demonstrates the decline of manufacturing and agriculture in rural areas. These industries used to dominate rural areas, but scholars have identified a significant decline in their prevalence over several decades (Source).

The demographic variable cluster results confirm previous studies on the presence of pockets of minorities in rural areas. The cluster with the lowest white percentage of the population has the highest percentage of Black, Hispanic or Latino, and Other race variables (Source). 20 percent of census tracts in this dataset fall into this category with an average of 6 percent African American population and 12 percent Hispanic and Latino population. Concentrations of African Americans are noteworthy because they are subjected to a variety of historical processes that put them at a higher risk of housing insecurity. Particularly concerning are concentrations of African Americans in the South, where the effects of segregation are still seen today. The presence of Hispanics and Latinos in the same cluster as the lowest white population and highest median black population is interesting because the migration of Hispanics and Latinos to rural areas has been indicated as a potential solution to the well-documented population issues facing rural areas (Source).

Next, the cluster medians indicate that while home ownership is widespread in rural areas, there are notable levels of renters that spend more than 30 percent of their income on housing. The overall average percent of renters with high housing costs is 15 percent compared to only 5 percent for homeowners with a mortgage that have high housing costs. This indicates that renters in rural areas may face similar issues as those in urban areas when it comes to the affordability of rental property. One issue noted not fully accounted for here is the presence of mobile home parks in rural areas. In areas categorized as high-risk for housing costs, an average of 15 percent of renters rent mobile homes while 60 percent rent single-unit homes. An average of 1 percent more renters rent mobile homes in low-risk housing cost census tracts. However, these results do not accommodate some of the extra expenses that come from mobile home renting or ownership such as lot rent and other fees that may come with living in a mobile home park.

The RME sector shows relatively low levels of transiency among those who have either a high school diploma or did not finish high school. Researchers should be concerned with the average of 8 percent of the population that are stable but do not have a high school education. The value of a high school diploma is well understood, so areas at high risk for RME instability require further attention. An area of concern is that households with low levels of education do not take advantage of public assistance, based on the average, only 1 percent of households with lower levels of education use public assistance compared to the rest of the population.

One notable observation from the RMP cluster analysis is that the number of households that live below the poverty level is higher than those who are slightly above the poverty level across all three risk levels. It is of great concern for researchers that poverty rates below the poverty level appear to be greater than the rate of households just above the poverty level. The rate of public assistance use is the same for this sector as it was for

RME. This indicates that in rural areas, the most vulnerable populations are not taking advantage of available public assistance that could improve their situations.

The cluster analysis shows 3 points of concern in rural areas for the household factors section. First, for households with no investment income and no other income, all cluster medians are above 30 percent. This indicates that the majority of households in rural areas are not building wealth through means outside of wages received from employment. importance of non-wage income. Second, a median of 5 percent of the population receives public assistance across all risk levels. This reflects that rural areas may not be fully taking advantage of assistance that could improve their living conditions. Previous research has documented the tendency for people to not use public assistance for various reasons (source). Third, the median Gini index for each risk level is between 0.42 and 0.44, indicating that income inequality may not be as prevalent of an issue in rural areas as it has been identified in rural areas.

The cluster analysis for housing type shows that single-unit renters and owners are the predominant means of housing in rural areas. One area of concern not accounted for here is the presence of mobile home parks. Research has shown that practices allowed in mobile home parks can put some at a higher risk of housing insecurity (source). For the high-risk housing type census tracts, an average of 13 and 15 percent of owners and renters live in mobile homes. These areas where there are higher rates of renters relative to owners than other sectors warrant further attention because there may be some unidentified factors in the community that contribute to the lower amounts of home ownership.

5.2 *Interpretation of Findings: RQ2*

When measuring housing insecurity across different dimensions, how often do the same features arise?

One important question when using the 4 C's model is how sector risk levels relate to each other. An area at an overall high risk of housing insecurity would at least have a high-risk level across more than one sector. Tables 4.10, 4.11, 4.12, 4.13 show the frequency of the most interesting relationships: high-risk-to-high-risk, low-risk-to-low-risk, low-risk-to-high-risk, and high-risk-to-high-risk. The most surprising result from this analysis is the similar levels of presence between all of these relationships. These indicate that areas with a high risk of housing insecurity in one sector may face a high risk of insecurity in another area, or it may be at a low risk of insecurity in another sector. The low support and fairly low confidence levels demonstrated across association rules indicate few areas that exhibit a risk of housing insecurity across multiple sectors. This is reflected by Figure 4.4 where the map of housing insecurity shows some pockets of red, but vast amounts of green and yellow indicated low and medium risk of housing insecurity across sectors for most rural areas.The lack of a significant number of rules with more than one element on the left-hand side provides further evidence for this hypothesis. This raises further questions related to the clustering of housing insecurity factors among rural areas as is often seen in urban areas.

While the association rules did not find significant patterns of commonality between risk levels, there are a notable number of census tracts that warrant further investigation due to their overall level of risk. As Figure 4.5 demonstrated, there are 280 census tracts across 37 states and ? medium-risk census tracts across ? states. The reasoning behind the 12/24 and 15/24 thresholds for being high or medium-risk is mostly arbitrary, the primary goal being to high light areas that the cluster analysis show are very vulnerable to housing insecurity issues.

5.3 *Interpretation of Findings: RQ3*

Are there spatial relations between the different dimensions of housing insecurity?

Global Moran's I was calculated for each state and nationwide. Additionally, local Moran's I was used on the risk levels to determine their spatial autocorrelation at the census tract level. The first notable observation from the global Moran's I result is the strength of the key demographic variables. African Americans, Hispanics and Latinos, White, american indian and native Alaskan, and Other all have spatial autocorrelations greater than 0.5 nationwide. This offers strong support for previous research that has found pockets of minorities in rural areas (source). While the percentage of rural economies that manufacturing and agriculture, forestry, fishing, hunting, and mining make up has declined in recent years, there is a significant spatial autocorrelation to both of these variables. This is reflective of the amenities-based nature of these amenities and further enforces the role of single industry-based economies in rural areas. Economic diversity is generally seen as a good thing, so the high spatial autocorrelation of these two industries is concerning for the overall economic diversity of rural areas. Another concern is the relatively high spatial autocorrelation of households that did not move, but do not have a high school diploma. This could be reflective of areas where schools have suffered due to the declines facing rural areas (Source).

The most significant finding of the spatial analysis is the results of local Moran's I on the sector risk level variables. Table 4.16 shows that on average, there is no local spatial autocorrelation between medium-risk level census tracts across any sector. The strongest average spatial autocorrelation is for demographics, with a value of 0.14. Low-risk levels have some level of spatial autocorrelation, the highest being for economic diversity. This statistic is higher than the low-risk spatial autocorrelation for employment diversity indicating that while the industries that typically dominate an economy are spatially autocorrelated globally, there is almost the same amount of spatial autocorrelation for census tracts with both high and low economic diversity. Another concerning observation is the very high local spatial autocorrelation for demographics. This follows the trend of results indicating the existence of pockets of minority populations in rural areas.

5.4 *Interpretation of Findings: RQ4*

To what extent can this model of housing insecurity be used to predict risk levels across housing insecurity factors?

In total, ? multinomial logistic regression models were used to predict the risk levels of each sector for each state and nationally. The results are very low probabilities of the model predicting the correct cluster, and very low accuracy. Most notable for the national models shown in Chapter 4's confusion matrices is their tendency to over-classify census tracts as low-risk for every sector. One explanation for this is class imbalance. Figure 4.3 shows that there are class imbalances mostly due to high levels of lower risk levels. This could have led the models into overclassifying census tracts as low-risk when they are not. Most surprising is the lack of accuracy in the national models. Being tested using in-sample evaluation, the models should have performed notably better than the test-train split state models. The lack of accuracy of these models echoes what the literature has said: "rural" is not a blanket term but rather, it encompasses a wide-ranging and varying group of areas and people. The most common theme between rural areas may simply be that they are not urban.

5.5 *Limitations*

There are three significant limitations to this work. First, Due to the urban-centric lens toward housing insecurity, there is little previous research to compare to this study. Gleason et al. 2021 applied similar spatial techniques to census tracts in Maine and found that poverty, unemployment, and high housing costs are common in rural and urban areas of Maine but found these results to be inaccurate in a later study (Gleason et al 2022?). The second limitation is due the lower rate of ACS sampling in rural areas, the accuracy of the data is limited in how well it represents the real world. While the estimates are "likely reasonable approximations of the populations they represent", small area estimates like census tracts used here have issues with attribute uncertainty (Spielman 2014, ?). Despite this, it is currently the most detailed source of data available for rural areas. The final limitation is that as the risk-level assignment system is relative, it cannot be used to make definitive claims about the housing insecurity risk of an area.

5.6 *Future Research*

Future research should use this study as a starting point for giving housing insecurity and homelessness in rural areas adequate attention. The most important direction is to identify community-level risk factors unique to rural areas. Further studies should also use a wider range of data sources to capture sectors with few available variables such as housing conditions. Subsequent investigations should examine rural housing insecurity at a localized level. This will enable the refinement and enhancement of the existing model, providing more precise insights into the unique challenges faced by rural communities. Future endeavors should prioritize a closer examination of areas exhibiting unexpected high-risk-to-low-risk and low-risk-to-high-risk relationships, as identified through the association rules analysis. Understanding the underlying factors contributing to these unexpected relationships is essential for targeted interventions and policy recommendations. There is a need for in-depth research to discern how levels and trends in income inequality differ between urban and rural areas, shedding light on the specific socio-economic dynamics impacting housing insecurity in each setting. Future research should scrutinize the distinctions in poverty and housing cost dynamics between rural and urban areas, aiming to gain a deeper understanding of the factors at play in each context. The states highlighted in the Moran's outlier section warrant attention because they exhibit notably higher levels of spatial clustering of risk factors than other states. People in these states may be at a higher risk of housing insecurity relative to other states. By addressing these research gaps, researchers can better inform evidence-based policies and interventions that mitigate housing insecurity and advance the well-being of rural populations. By addressing these research gaps, researchers can better inform evidence-based policies and interventions that mitigate housing insecurity and advance the well-being of rural populations. As we strive to enhance housing security and social equity in both rural and urban landscapes, interdisciplinary collaboration and persistent research efforts will remain pivotal in driving meaningful societal change.

CHAPTER 6
CONCLUSION

Housing insecurity is difficult in several different ways. First, it is difficult to define. Until there is a full understanding of housing insecurity, which includes amending the gap between urban and rural housing insecurity research, we are limited in our ability to properly operationalize the meaning of the phrase. Second, it is difficult to study. As a concept that spans such a wide range of individual, social, and political factors is inherently difficult to study. Third, and most importantly, housing insecurity is difficult for those who experience it. In rural areas, these difficulties are compounded due to the urban-centric lens of housing insecurity that has developed over decades of primarily urban-oriented research. This initial exploration hopes to serve as a starting point for policy-makers and researchers to begin deconstructing the urban-centric lens and give those in rural populations the attention and resources they need and deserve.

REFERENCES

D. Augustus Anderson, Hye-Sung Han, and John Hisnanick. The Effect of Household Debt and Wealth on Subsequent Housing Tenure Choice. *City & Community*, 20(4):297–325, December 2021. ISSN 1535-6841, 1540-6040. doi: 10.1177/15356841211007757. URL `http://journals.sagepub.com/doi/10.1177/15356841211007757`.

Rüdiger Bachmann and Daniel Cooper. The ins and arounds in the U.S. housing market. Working Paper, Federal Reserve Bank of Boston, 2014.

Holly Barcus. Heterogeneity of rural housing markets. *Rural Housing, Exurbanization, and Amenity-Driven Development: Contrasting the 'Haves' and the 'Have Nots'*, pages 51–73, January 2011.

Seth A. Berkowitz and Deepak Palakshappa. Gaps in the welfare state: A role-based model of poverty risk in the U.S. *PLOS ONE*, 18(4):e0284251, April 2023. ISSN 1932-6203. doi: 10.1371/journal.pone.0284251. URL `https://dx.plos.org/10.1371/journal.pone.0284251`.

Evelyn Blumenberg and Gregory Pierce. Automobile Ownership and Travel by the Poor. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2320:28–36, 2012.

Evelyn Blumenberg, Anne Brown, and Andrew Schouten. Car-deficit households: determinants and implications for household travel in the U.S. *Transportation*, 47(3): 1103–1125, June 2020. ISSN 0049-4488, 1572-9435. doi: 10.1007/s11116-018-9956-6. URL `http://link.springer.com/10.1007/s11116-018-9956-6`.

Paula Braveman, Mercedes Dekker, Susan Egerter, Tabashir Sadegh-Nobari, and Craig Pollack. Housing and Health, 2011.

Sarah K. Bruch, Joseph Van Der Naald, and Janet C. Gornick. Poverty Reduction through Federal and State Policy Mechanisms: Variation over Time and across the United States. *Social Service Review*, 97(2):270–319, June 2023. ISSN 0037-7961, 1537-5404. doi: 10.1086/724556. URL `https://www.journals.uchicago.edu/doi/10.1086/724556`.

Emery N. Castle, JunJie Wu, and Bruce A. Weber. Place Orientation and Rural–Urban Interdependence. *Applied Economic Perspectives and Policy*, 33(2):179–204, June 2011. ISSN 2040-5790, 2040-5804. doi: 10.1093/aepp/ppr009. URL `https://onlinelibrary.wiley.com/doi/abs/10.1093/aepp/ppr009`.

Jing Chen. Economic Diversity and Regional Economic Performance: A Methodological Concern from Model Uncertainty. Working Paper, West Virginia University, 2018.

John Cromartie and Shawn Bucholtz. Defining the "Rural" in Rural America. *Amber Waves*, 6(3), 2008.

Steven Deller and Philip Watson. Spatial variations in the relationship between economic diversity and stability. *Applied Economics Letters*, 23(7):520–525, May 2016. ISSN 1350-4851, 1466-4291. doi: 10.1080/13504851.2015.1085630. URL `http://www.tandfonline.com/doi/full/10.1080/13504851.2015.1085630`.

Stefanie DeLuca and Eva Rosen. Housing Insecurity Among the Poor Today. *Annual Review of Sociology*, 48(1):343–371, July 2022. ISSN 0360-0572, 1545-2115. doi: 10.1146/annurev-soc-090921-040646. URL `https://www.annualreviews.org/doi/10.1146/annurev-soc-090921-040646`.

Matthew Desmond and Monica Bell. Housing, Poverty, and the Law. *Annual Review of Law and Social Science*, 11(1):15–35, November 2015. ISSN 1550-3585, 1550-3631. doi: 10.1146/annurev-lawsocsci-120814-121623. URL `https://www.annualreviews.org/doi/10.1146/annurev-lawsocsci-120814-121623`.

Matthew Desmond and Carl Gershenson. Housing and Employment Insecurity among the Working Poor. *Social Problems*, 63(1):46–67, February 2016. ISSN 0037-7791, 1533-8533. doi: 10.1093/socpro/spv025. URL `https://academic.oup.com/socpro/article-lookup/doi/10.1093/socpro/spv025`.

Matthew Desmond and Kristin L. Perkins. Housing and Household Instability. *Urban Affairs Review*, 52(3):421–436, May 2016. ISSN 1078-0874, 1552-8332. doi: 10.1177/1078087415589192. URL `http://journals.sagepub.com/doi/10.1177/1078087415589192`.

Matthew Desmond and Tracey Shollenberger. Forced Displacement From Rental Housing: Prevalence and Neighborhood Consequences. *Demography*, 52(5):1751–1772, October 2015. ISSN 0070-3370, 1533-7790. doi: 10.1007/s13524-015-0419-9. URL `https://read.dukeupress.edu/demography/article/52/5/1751/169475/Forced-Displacement-From-Rental-Housing-Prevalence`.

Matthew Desmond and Nathan Wilmers. Do the Poor Pay More for Housing? Exploitation, Profit, and Risk in Rental Markets. *American Journal of Sociology*, 124(4): 1090–1124, January 2019. ISSN 0002-9602, 1537-5390. doi: 10.1086/701697. URL `https://www.journals.uchicago.edu/doi/10.1086/701697`.

Rachel Bogardus Drew. Believing in Homeownership: Behavioral Drivers of Housing Tenure Decisions. Technical report, Joint Center for Housing Studies, Harvard Univerity, 2014.

D D'Alessandro and L Appolloni. Housing and health: an overview. *Ann Ig*, 32(5):17–26, 2020.

Price V. Fishback. Social Insurance and Public Assistance in the Twentieth-Century United States. *The Journal of Economic History*, 80(2):311–350, June 2020. ISSN 0022-0507, 1471-6372. doi: 10.1017/S0022050720000200. URL `https://www.cambridge.org/core/product/identifier/S0022050720000200/type/journal_article`.

Jason M. Fletcher, Tatiana Andreyeva, and Susan H. Busch. Assessing the effect of changes in housing costs on food insecurity. *Journal of Children and Poverty*, 15(2): 79–93, September 2009. ISSN 1079-6126, 1469-9389. doi: 10.1080/10796120903310541. URL `http://www.tandfonline.com/doi/abs/10.1080/10796120903310541`.

MEnelaos Gkartzios and Ann Ziebarth. Housing: A Lens to rural Inequalities. *International Handbook of Rural Studies*, November 2017.

Kristin Gleason, Matthew Dube, and Jennifer Martin. Using Geographic Information Systems to Assess Community-Level Vulnerability to Housing Insecurity in Rural Areas. *Journal of Community Psychology*, 50(4):1993–2012, 2021.

Christopher Herbert, Alexander Hermann, and Daniel McCue. Measuring Housing Affordability: Assessing the 30-Percent of Income Standard, 2018.

Diana Hernández and Carolyn B. Swope. Housing as a Platform for Health and Equity: Evidence and Future Directions. *American Journal of Public Health*, 109(10):1363–1366, October 2019. ISSN 0090-0036, 1541-0048. doi: 10.2105/AJPH.2019.305210. URL `https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2019.305210`.

Lingqian Hu and Liming Wang. Housing location choices of the poor: does access to jobs matter? *Housing Studies*, 34(10):1721–1745, November 2019. ISSN 0267-3037, 1466-1810. doi: 10.1080/02673037.2017.1364354. URL `https://www.tandfonline.com/doi/full/10.1080/02673037.2017.1364354`.

David E. Jacobs. Environmental Health Disparities in Housing. *American Journal of Public Health*, 101(S1):S115–S122, December 2011. ISSN 0090-0036, 1541-0048. doi: 10.2105/AJPH.2010.300058. URL `https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2010.300058`.

Seungbeom Kang. Why Low-Income Households Become Unstably Housed: Evidence From the Panel Study of Income Dynamics. *Housing Policy Debate*, 29(4):559–587, July 2019.

ISSN 1051-1482, 2152-050X. doi: 10.1080/10511482.2018.1544161. URL https://www.tandfonline.com/doi/full/10.1080/10511482.2018.1544161.

Seungbeom Kang. Severe and persistent housing instability: examining low-income households' residential mobility trajectories in the United States. *Housing Studies*, pages 1–26, October 2021. ISSN 0267-3037, 1466-1810. doi: 10.1080/02673037.2021.1982871. URL https://www.tandfonline.com/doi/full/10.1080/02673037.2021.1982871.

Huiyun Kim, Sarah A. Burgard, and Kristin S. Seefeldt. Housing Assistance and Housing Insecurity: A Study of Renters in Southeastern Michigan in the Wake of the Great Recession. *Social Service Review*, 91(1):41–70, March 2017. ISSN 0037-7961, 1537-5404. doi: 10.1086/690681. URL https://www.journals.uchicago.edu/doi/10.1086/690681.

Lauren J. Krivo and Robert L. Kaufman. Housing and wealth inequality: Racial-ethnic differences in home equity in the United States. *Demography*, 41(3):585–605, August 2004. ISSN 0070-3370, 1533-7790. doi: 10.1353/dem.2004.0023. URL https://read.dukeupress.edu/demography/article/41/3/585/170331/Housing-and-wealth-inequality-Racial-ethnic.

Jessica N. Kropczynski and Patricia H. Dyk. Insights into Housing Affordability for Rural Low-Income Families. *Housing and Society*, 39(2):125–148, January 2012. ISSN 0888-2746, 2376-0923. doi: 10.1080/08882746.2012.11430603. URL http://www.tandfonline.com/doi/full/10.1080/08882746.2012.11430603.

Barrett A. Lee and Megan Evans. Forced to move: Patterns and predictors of residential displacement during an era of housing insecurity. *Social Science Research*, 87:102415, March 2020. ISSN 0049089X. doi: 10.1016/j.ssresearch.2020.102415. URL https://linkinghub.elsevier.com/retrieve/pii/S0049089X20300132.

Barrett A. Lee, Marybeth Shinn, and Dennis P. Culhane. Homelessness as a Moving Target. *The ANNALS of the American Academy of Political and Social Science*, 693(1): 8–26, January 2021. ISSN 0002-7162, 1552-3349. doi: 10.1177/0002716221997038. URL `http://journals.sagepub.com/doi/10.1177/0002716221997038`.

Kathryn M Leifheit, Gabriel L Schwartz, Craig Evan Pollack, and Sabriya L Linton. Building health equity through housing policies: critical reflections and future directions for research. *Journal of Epidemiology and Community Health*, 76(8):759–763, August 2022. ISSN 0143-005X, 1470-2738. doi: 10.1136/jech-2021-216439. URL `https://jech.bmj.com/lookup/doi/10.1136/jech-2021-216439`.

Daniel T. Lichter and David L. Brown. Rural America in an Urban Society: Changing Spatial and Social Boundaries. *Annual Review of Sociology*, 37(1):565–592, August 2011. ISSN 0360-0572, 1545-2115. doi: 10.1146/annurev-soc-081309-150208. URL `https://www.annualreviews.org/doi/10.1146/annurev-soc-081309-150208`.

Daniel T. Lichter and Kenneth M. Johnson. The Changing Spatial Concentration of America's Rural Poor Population*. *Rural Sociology*, 72(3):331–358, September 2007. ISSN 00360112. doi: 10.1526/003601107781799290. URL `http://doi.wiley.com/10.1526/003601107781799290`.

Daniel T. Lichter and Kenneth M. Johnson. A Demographic Lifeline? Immigration and Hispanic Population Growth in Rural America. *Population Research and Policy Review*, 39(5):785–803, October 2020. ISSN 0167-5923, 1573-7829. doi: 10.1007/s11113-020-09605-8. URL `https://link.springer.com/10.1007/s11113-020-09605-8`.

Daniel T. Lichter, David L. Brown, and Domenico Parisi. The rural–urban interface: Rural and small town growth at the metropolitan fringe. *Population, Space and Place*, 27(3), April 2021. ISSN 1544-8444, 1544-8452. doi: 10.1002/psp.2415. URL https://onlinelibrary.wiley.com/doi/10.1002/psp.2415.

Sarah A Low. Rural Manufacturing at a Glance, 2017 Edition. *Plastics and rubber*, 2017.

Katherine MacTavish, Michelle Eley, and Sonya Salamon'. POLICY AND PRACTITIONER PERSPECTIVES. *Georgetown Journal on Poverty Law & Policy*, 13 (1):95–117, 2006.

Katherine A. MacTavish. The Wrong Side of the Tracks: Social Inequality and Mobile Home Park Residence. *Community Development*, 38(1):74–91, March 2007. ISSN 1557-5330, 1944-7485. doi: 10.1080/15575330709490186. URL http://www.tandfonline.com/doi/abs/10.1080/15575330709490186.

Thomas D. Matte and David E. Jacobs. Housing and Health- Current Issues and Implications for Research and Programs. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 7(1):7–25, 2000.

Molly W Metzger and Amy T Khare. Fair Housing and Inclusive Communities. *American Academy of Social Work and Social Welfare*, 2017.

Raven Molloy, Christopher L Smith, and Abigail Wozniak. Internal Migration in the United States. *Journal of Economic Perspectives*, 25(3):173–196, August 2011. ISSN 0895-3309. doi: 10.1257/jep.25.3.173. URL https://pubs.aeaweb.org/doi/10.1257/jep.25.3.173.

Katherine M. O'Regan, Ingrid Gould Ellen, and Sophie House. How to Address Homelessness: Reflections from Research. *The ANNALS of the American Academy of Political and Social Science*, 693(1):322–332, January 2021. ISSN 0002-7162, 1552-3349. doi: 10.1177/0002716221995158. URL http://journals.sagepub.com/doi/10.1177/0002716221995158.

Robin Phinney. Exploring Residential Mobility among Low-Income Families. *Social Service Review*, 87(4):780–815, December 2013. ISSN 0037-7961, 1537-5404. doi: 10.1086/673963. URL https://www.journals.uchicago.edu/doi/10.1086/673963.

Mark R. Rank and Thomas A. Hirshl. Welfare Use as a Life Course Event: Toward a New Understanding of the U.S. Safety Net. *Social Work*, 47(3):237–248, 2002.

Virginia A. Rauh, Philip J. Landrigan, and Luz Claudio. Housing and Health: Intersection of Poverty and Environmental Exposures. *Annals of the New York Academy of Sciences*, 1136(1):276–288, July 2008. ISSN 00778923. doi: 10.1196/annals.1425.032. URL http://doi.wiley.com/10.1196/annals.1425.032.

Kathleen Sherrieb, Fran H. Norris, and Sandro Galea. Measuring Capacities for Community Resilience. *Social Indicators Research*, 99(2):227–247, November 2010. ISSN 0303-8300, 1573-0921. doi: 10.1007/s11205-010-9576-9. URL http://link.springer.com/10.1007/s11205-010-9576-9.

Anne Shlay and Peter Rossi. Social Science Research and Contemporary Studies of Homelessness. *Annual Review of Sociology*, 18:129–160, 2003. doi: 10.1146/annurev.so.18.080192.001021.

Lilly Shoup and Becca Houma. Principles for Improving Transportation Options in Rural and Small Town Communities. White Paper, Transportation for America, 2010.

Thomas C Siskar. *Who Is Forced to Move?* Master of Arts, Pennsylvannia State University, 2019.

P. Johnelle Sparks, Corey S. Sparks, and Joseph J. A. Campbell. Poverty Segregation in Nonmetro Counties: A Spatial Exploration of Segregation Patterns in the US. *Spatial Demography*, 1(2):162–177, June 2013. ISSN 2364-2289, 2164-7070. doi: 10.1007/BF03354896. URL https://link.springer.com/10.1007/BF03354896.

Carolyn B. Swope and Diana Hernandez. Housing as a determinant of health equity_ A conceptual model. *Soc Sci Med*, 243, 2020. doi: 10.1016/j.socscimed.2019.112571. URL `https://reader.elsevier.com/reader/sd/pii/S0277953619305659?token=0BF7CE998C4949B5AD4DB9DB6CE6CA02DF1448E46028A9D722AB36EBD1473C1B404CCF479C8632CC8BEF36originRegion=us-east-1&originCreation=20230118070028`.

U.S. Census Bureau and U.S. Department of Housing and Urban Development. Median Sales Price of Houses Sold for the United States, January 1963. URL `https://fred.stlouisfed.org/series/MSPUS`. Publisher: FRED, Federal Reserve Bank of St. Louis.

Michele Wakin. Not Sheltered, Not Homeless: RVs as Makeshifts. *American Behavioral Scientist*, 48(8):1013–1032, April 2005. ISSN 0002-7642, 1552-3381. doi: 10.1177/0002764204274197. URL `http://journals.sagepub.com/doi/10.1177/0002764204274197`.

John C. Weicher. Housing conditions and homelessness. *Gender Issues*, pages 35–53, 2006.

Anita Yadavalli, Brenna Rivett, James Brooks, Christiana K McFarland, and David Hardiman. A Comprehensive Look at Housing Market Conditions Across America's Cities. *Cityscape*, 22(2):111–132, 2020.

APPENDICES

APPENDIX A

VARIABLE AVERAGES BY RUCA CODE

Table A.1: Variable Average by RUCA Code

	R7	R8	R9	R10
white	0.82	0.87	0.87	0.88
black	0.10	0.07	0.07	0.04
am_in_ala_nat	0.02	0.03	0.02	0.04
asian	0.01	0.00	0.00	0.01
haw_pac	0.00	0.00	0.00	0.00
other	0.02	0.01	0.01	0.01
hisp_lat	0.10	0.05	0.04	0.07
male_u18	0.11	0.11	0.11	0.11
female_u18	0.11	0.10	0.10	0.10
male_o18	0.38	0.40	0.39	0.40
female_o18	0.40	0.39	0.39	0.39
same_house_less_than_hs	0.08	0.09	0.09	0.08
same_house_hs	0.21	0.25	0.27	0.23
moved_in_county_less_than_hs	0.01	0.00	0.00	0.00
moved_in_county_hs	0.02	0.01	0.01	0.01
moved_diff_county_less_than_hs	0.00	0.00	0.00	0.00
moved_diff_county_hs	0.01	0.01	0.01	0.01
moved_diff_state_less_than_hs	0.00	0.00	0.00	0.00
moved_diff_state_hs	0.00	0.00	0.00	0.00
same_house_p1	0.13	0.13	0.12	0.12
same_house_p2	0.09	0.09	0.09	0.09
moved_in_county_p1	0.02	0.01	0.01	0.01
moved_in_county_p2	0.01	0.01	0.01	0.01
moved_diff_county_p1	0.01	0.01	0.01	0.01
moved_diff_county_p2	0.00	0.00	0.00	0.00
moved_diff_state_p1	0.00	0.00	0.00	0.00
moved_diff_state_p2	0.00	0.00	0.00	0.00
mortgage_high_cost	0.05	0.05	0.05	0.05
no_mortgage_high_cost	0.02	0.03	0.03	0.03
rent_high_cost	0.19	0.14	0.14	0.15
all_incomplete_plumb	0.24	0.30	0.30	0.27
all_incomplete_kitchen	0.31	0.31	0.31	0.27
occ_incomplete_plumb	0.00	0.01	0.01	0.01
occ_incomplete_kitchen	0.01	0.01	0.01	0.01
owner_single	0.89	0.81	0.82	0.86
owner_2to4	0.01	0.00	0.00	0.00
owner_5plus	0.00	0.00	0.00	0.00
owner_mobile	0.10	0.19	0.18	0.13
owner_unconvent	0.00	0.00	0.00	0.00
renter_single	0.54	0.62	0.61	0.64
renter_2to4	0.18	0.07	0.07	0.11
renter_5plus	0.17	0.05	0.05	0.09
renter_mobile	0.10	0.25	0.26	0.16
renter_unconvent	0.00	0.00	0.00	0.00
hh_no_wage	0.14	0.14	0.14	0.15
hh_no_other_income	0.37	0.36	0.35	0.37
hh_no_investment_income	0.34	0.32	0.32	0.32
hh_public_assistance	0.07	0.05	0.05	0.05
hh_ssi	0.03	0.03	0.03	0.02
hh_3plus_worker	0.02	0.02	0.02	0.02
hh_worker_no_vehicle	0.03	0.01	0.01	0.02
hh_no_vehicle	0.04	0.02	0.02	0.02
gini_index	0.44	0.43	0.43	0.44
ag_for_fish_hunt_mining	0.02	0.03	0.03	0.04
construction	0.03	0.03	0.04	0.04
manufacturing	0.06	0.06	0.07	0.05
wholesale_trade	0.01	0.01	0.01	0.01
retail_trade	0.05	0.05	0.05	0.05
trans_warehouse_util	0.02	0.02	0.03	0.02
information	0.01	0.00	0.00	0.01
fin_re_insur	0.02	0.02	0.02	0.02
prof_sci_mgmt_waste	0.03	0.02	0.02	0.03
edu_health_social	0.10	0.10	0.09	0.09
arts_rec_food	0.04	0.03	0.03	0.04
othersvcs	0.02	0.02	0.02	0.02
public_admin	0.02	0.02	0.02	0.02