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

Master of Science

(in Data Science and Engineering)

The Graduate School

The University of Maine

May, 2024

Advisory Committee:

Matthew P. Dube, Associate Professor of Computer Information Systems, Advisor

Kristen Gleason, Assistant Professor of Psychology

Sarah Walton, Assistant Professor of Sociology

RURALITY AND ROBUSTNESS: RURAL COMMUNITIES AND HOUSING INSECURITY RISK

By Steve Garcia

Thesis Advisor: Dr. Matthew P. Dube

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

This thesis explores rural housing insecurity through Swope and Hernandez's (2019) 4 C's of housing insecurity in rural areas. Little attention has been paid to rural areas in the conversation on housing insecurity and houselessness. To facilitate further discussion on this understudied issue, this exploratory study used unsupervised machine learning to group census tracts into risk levels across 7 axes of data from the American Community Survey. These were based on housing insecurity factors found in the literature. Multinomial logistic regression was used to determine variation between U.S. states based on how well state risk levels could be predicted with the national dataset. Furthermore, spatial autocorrelation analysis was employed to gauge the extent of spatial clustering within the identified risk levels and housing insecurity factors. 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 were not as highly spatially clustered as expected.

RURALITY AND ROBUSTNESS: RURAL COMMUNITIES AND HOUSING INSECURITY RISK

By Steve Garcia

Thesis Advisor: Dr. Matthew P. Dube

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

Keywords: housing insecurity, homelessness, data mining

This thesis delves into the challenges of insecure housing in rural areas, drawing from Swope and Hernandez's comprehensive 4 C's framework outlined in 2019. Remarkably, rural regions have been overlooked in conversations about housing insecurity and homelessness. To shed light on this neglected issue, this investigation utilized unsupervised machine learning techniques. It categorized census tracts into risk levels across seven key data axes sourced from the American Community Survey, focusing on factors commonly associated with housing insecurity as documented in existing literature.

By employing multinomial logistic regression, the study aimed to discern variations among U.S. states, determining how accurately state risk levels aligned with the national dataset. Furthermore, spatial autocorrelation analysis was employed to gauge the extent of spatial clustering within the identified risk levels and housing insecurity factors.

The findings uncover that numerous rural census tracts exhibit a moderate risk of housing insecurity, yet predicting these risk levels proves challenging. Intriguingly, the spatial autocorrelation analysis suggests that the housing insecurity variables didn't exhibit the anticipated high levels of spatial clustering.

DEDICATION

Dedicated to St. Thomas More and all who seek to build a better world.

ACKNOWLEDGEMENTS

I extend my heartfelt appreciation to the following individuals for their invaluable contributions to this research project:

My advisor, Dr. Matthew Dube, who provided unwavering guidance and feedback throughout this endeavor; my committee members, Drs. Kristen Gelason and Sarah Walton, for their insightful feedback and support; my beloved wife, Katie, whose encouragement has been a constant source of motivation; Drs. Megan Smith, Jeremy Mhire, and Dolliann Hurtig, for their mentorship and guidance during my academic journey; and my grandparents, Steve and Linda Mauldin, for their wisdom and support throughout life's challenges.

Your collective support has been instrumental in the success of this project, and I am sincerely grateful for your presence in my life."

Table of Contents

DEDICATION **iii**

ACKNOWLEDGEMENTS **iv**

1. INTRODUCTION **1**

1.1 *Rural Areas* 3
1.2 *Measures of Homelessness* 4
1.3 *Housing as Health* 6
1.4 *Theoretical Framework* 6
1.5 *Motivation* 7
1.6 *Approach* 8
1.7 *Major Results* 9
1.8 *Intended Audience* 10
1.9 *Structure of Thesis* 10

2. BACKGROUND **11**

2.1 *Housing Insecurity* 11
2.2 *The 4 Cs of Housing Insecurity* 12
2.2.1 *Cost* 12
2.2.2 *Conditions* 14
2.2.3 *Consistency* 15
2.2.4 *Context* 16
2.2.4.1 *Employment* 16
2.2.4.2 *Housing, race, education, and poverty* 18
2.2.4.3 *Housing Type* 19
2.2.4.4 *Household Factors* 20
2.3 *Challenges for Rural Areas* 21
2.4 *Data Mining Methods* 26
2.4.1 *Clustering* 27

2.4.2 *Association Rules* . . . 29
2.4.3 *Contiguity* . . . 31
2.4.4 *Spatial Autocorrelation* . . . 33
2.4.5 *Multinomial Logistic Regression* . . . 35
2.5 *Summary* . . . 37

3. ADDRESSING RURAL HOUSING INSECURITY **38**
3.1 *Defining Rurality* . . . 38
3.2 *Applying the 4 Cs* . . . 39
3.3 *Data* . . . 40
3.4 *Data processing* . . . 41
3.5 *Methods* . . . 42
3.5.1 *k-Medoid Clustering* . . . 42
3.5.2 *Identifying High Risk Census Tracts* . . . 43
3.5.3 *Association Rules Learning* . . . 43
3.5.4 *Moran's I* . . . 45
3.5.5 *Multinomial Logistic Regression* . . . 47

4. RESULTS **49**
4.1 *RUCA Distribution* . . . 50
4.2 *Cluster Analysis* . . . 50
4.2.1 *Employment* . . . 51
4.2.2 *Demographics* . . . 52
4.2.3 *Housing Cost* . . . 53
4.2.4 *Housing Quality* . . . 53
4.2.5 *Residential Mobility: Education* . . . 54
4.2.6 *Residential Mobility: Poverty* . . . 55
4.2.7 *Household Factors* . . . 55
4.2.8 *Housing Type* . . . 56
4.3 *Association Rules* . . . 57
4.4 *Moran's I* . . . 62
4.4.1 *Moran's I Outliers* . . . 64
4.5 *Multinomial Logistic Regression* . . . 65
4.5.1 *Probability* . . . 66
4.5.2 *Accuracy* . . . 69
4.6 *Rurality and Risk Levels* . . . 72

5. DISCUSSION AND FUTURE DIRECTIONS **76**

5.1 *Identifying and Analyzing Risk Levels* 76

5.2 *Patterns of Risk* 81

5.3 *Spatial Aspects of Housing Insecurity* 83

5.4 *The Predictability of Housing Insecurity Risks* 85

5.5 *Limitations* 86

6. CONCLUSION **87**

6.1 *Summary* 87

6.2 *Conclusions* 89

6.3 *Future Research* 90

REFERENCES **92**

Appendices

APPENDIX A – DESCRIPTIVE STATISTICS 117

APPENDIX B – HIGH-RISK RULES BY PERCENT OF CENSUS TRACTS PER STATE 120

APPENDIX C – CORRELATIONS 128

BIOGRAPHY OF THE AUTHOR 133

List of Tables

3.1 RUCA Codes and Descriptions 38
4.1 Median Values for Employment Diversity Clusters 52
4.2 Median Values for Demographic Diversity Clusters 53
4.3 Median Values for Housing Cost Clusters 53
4.4 Median Values for Housing Quality Clusters 54
4.5 Median Values for Residential Mobility: Education Clusters 54
4.6 Median Values for Residential Mobility: Poverty Clusters 55
4.7 Median Values for Household Factor Clusters 56
4.8 Median Values for Housing Type Clusters 56
4.9 High Risk Association Average Statistics 58
4.10 Low Risk Association Average Statistics 58
4.11 Low to High Risk Association Average Statistics 58
4.12 High to Low Risk Association Average Statistics 59
4.13 Moran's I Descriptive Statistics 63
4.14 Average Moran's I by Sector 63
4.15 Employment Confusion Matrix and Statistics 69
4.16 Demographics Confusion Matrix and Statistics 70
4.17 Housing Cost Confusion Matrix and Statistics 70
4.18 Housing Quality Confusion Matrix and Statistics 70
4.19 Residential Mobility: Education Confusion Matrix and Statistics 71
4.20 Residential Mobility: Poverty Confusion Matrix and Statistics 71
4.21 Household Factors Confusion Matrix and Statistics 71
4.22 Housing Type Confusion Matrix and Statistics 72
4.23 Local Morans I Risk-Level Results 72
4.24 High-Risk Census Tract RUCA Breakdown 73
4.25 High and Low-Risk Census Tract t-Test 75
A.1 Demographics Descriptive Statistics 117
A.2 RME Descriptive Statistics 117
A.3 RMP Descriptive Statistics 118

A.4 Housing Costs Descriptive Statistics 118
A.5 Housing Quality Descriptive Statistics 118
A.6 Housing Type Descriptive Statistics 118
A.7 Household Factors Descriptive Statistics 119
A.8 Employment Diversity Descriptive Statistics 119
B.1 Employment High-Risk Rules by Percent of Census Tracts Per State 120
B.2 Demographics High-Risk Rules by Percent of Census Tracts Per State 121
B.3 Housing Costs High-Risk Rules by Percent of Census Tracts Per State 122
B.4 Housing Quality High-Risk Rules by Percent of Census Tracts Per State 123
B.5 RME High-Risk Rules by Percent of Census Tracts Per State 124
B.6 RMP High-Risk Rules by Percent of Census Tracts Per State 125
B.7 Household Factors High-Risk Rules by Percent of Census Tracts Per State . . . 126
B.8 Housing Type High-Risk Rules by Percent of Census Tracts Per State 127

List of Figures

1.1	Stereotype of Urban Homelessness	2
1.2	Variation in the Rural Population	4
2.1	Myers-Briggs Type Indicator Breakdown	31
2.2	Types of Contiguity	32
2.3	Spatial Autocorrelation Example	34
4.1	State Census Tracts vs. State Neighbors Count	49
4.2	State Census Tracts vs. State Neighbors Count	50
4.3	Cluster Distribution by Sector	51
4.4	Scatter plot of Association Rules Statistics	59
4.5	Boxplot of Moran's *I* by Region	65
4.6	Boxplot of Moran's *I* by Sector	66
4.7	Boxplot of Probability by sector	67
4.8	Correlation Plot of Sector Probabilities	68
4.9	Bar Graph of Average Probabilities by State with Error Bars	68
4.10	Risk Level Across Sectors	74
4.11	High and Medium Risk Census Tracts	75
C.1	Demographics Correlation Plot	128
C.2	RME Correlation Plot	129
C.3	RME Correlation Plot	130
C.4	Housing Costs Correlation Plot	131
C.5	Housing Quality Correlation Plot	132

CHAPTER 1
INTRODUCTION

Homelessness research has undergone a significant transformation in recent years. Historically, the focus was on categorizing and describing different segments of the homeless population (Lee et al., 2021). Much of this research has focused on individuals rather than communities. A contemporary approach views homelessness as a spectrum rather than a binary condition (e.g., Desmond et al., 2015; Swope and Hernandez, 2019; Cox et al., 2019). This shift opens the door to a new approach in the study of housing insecurity. In data science terms, the housed/unhoused binary resembles logistic regression, where households are categorized as housed or unhoused. However, housing insecurity is far more complex, influenced by a multitude of interconnected factors. Data science offers promise in bridging gaps in our understanding of housing insecurity dynamics. By harnessing the power of data and analytics, we can develop more equitable, evidence-based solutions that address the root causes of housing insecurity and improve outcomes for vulnerable populations.

Four critical areas remain unaddressed in the literature that limit our understanding of housing insecurity. First, housing research has concentrated on urban settings resulting in an urban-centric view of social issues like poverty and homelessness. Figure 1.1 shows a stereotypical view of the urban unhoused population. Drawing comparisons to Hoovervilles of the great depression, the growing presence of tent cities in urban areas in the aftermath of the Great Recession became a notable social issue (Herring and Lutz, 2015). An often under-acknowledged problem is the idea of the hidden homeless. One study conducted in Los Angeles estimated that at least 20 percent of the homeless population was "hidden", meaning they were not counted based on the criteria set by the Greater Los Angeles Homeless Count (Agans et al., 2014).

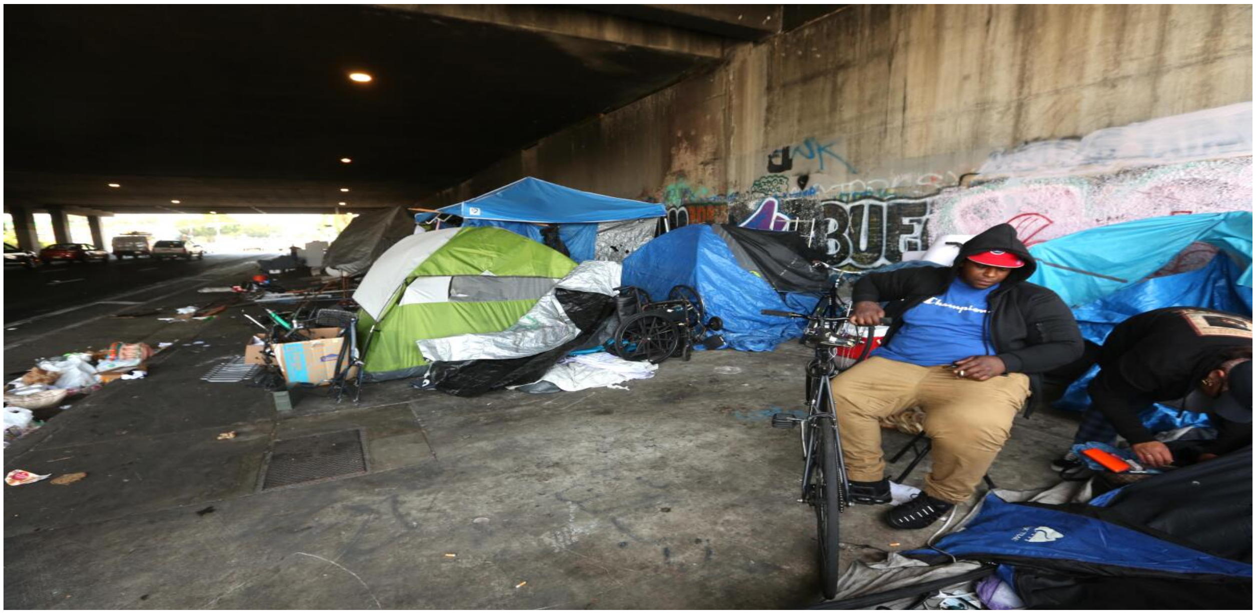

Figure 1.1: Stereotype of Urban Homelessness (Molina, 2023)

Those who do not meet our stereotypes of homeless (i.e. those sheltering in abandoned buildings, couch surfing, or living in a vehicle) are excluded from the conversation even in urban areas. The rural homeless population often remains unseen, not just because rural areas are sparsely populated, but also due to the absence of support services and amenities, like shelters, hostels, and drop-in centers, which are typically available in urban areas (Cloke et al., 2006). This is why the study of housing insecurity is a vital resource for rural areas. Without an appropriate infrastructure to deal with the unhoused, preventing literal homelessness should be a priority of researchers and policymakers. Second, measuring housing insecurity is challenging because of its dependence on circumstances and obstacles for both individuals and communities (Leifheit et al., 2022). Third, housing and homelessness in urban and rural areas necessitate a multi-disciplinary approach to properly capture the aspects that contribute to them, an approach rarely used in 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 on homelessness often focus on descriptive surveys of those accessing public services and providers of public services

(Robertson et al., 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*

Rural areas encompass a broad spectrum of places, including farms, ranches, villages, forested areas, 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 between rural communities as three general characteristics of rural places. 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 temporary 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). Despite this variation in rural areas, "rural" is often defined as "not urban" (National Coalition for the Homeless, 2009). In the study of housing, rural areas are often excluded from the conversation (Gkartzios and Ziebarth, 2017). Contributing to this problem is a wide variety of definitions of rurality used by governmental organizations, policymakers, and scholars (Yousey and Samudra, 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).

Rural people are distributed over a blend of the rural-urban continuum. They make up about 20 percent of the nation's population, 13 percent of the metropolitan population, 48 percent of the micropolitan population, and 75 percent of the noncore-base area population

(Isserman, 2005). As Figure 1.2 demonstrates, rural areas ranging from small towns to the most rural areas encompass a large mass of land with relatively few people. Deconstructing the urban-centric lens of housing research necessitates a novel approach that can accommodate the differences in rural areas. The size and variation of rural areas necessitate addressing rural issues differently because there can be no one-size-fits-all policy approach to improving conditions for rural people.

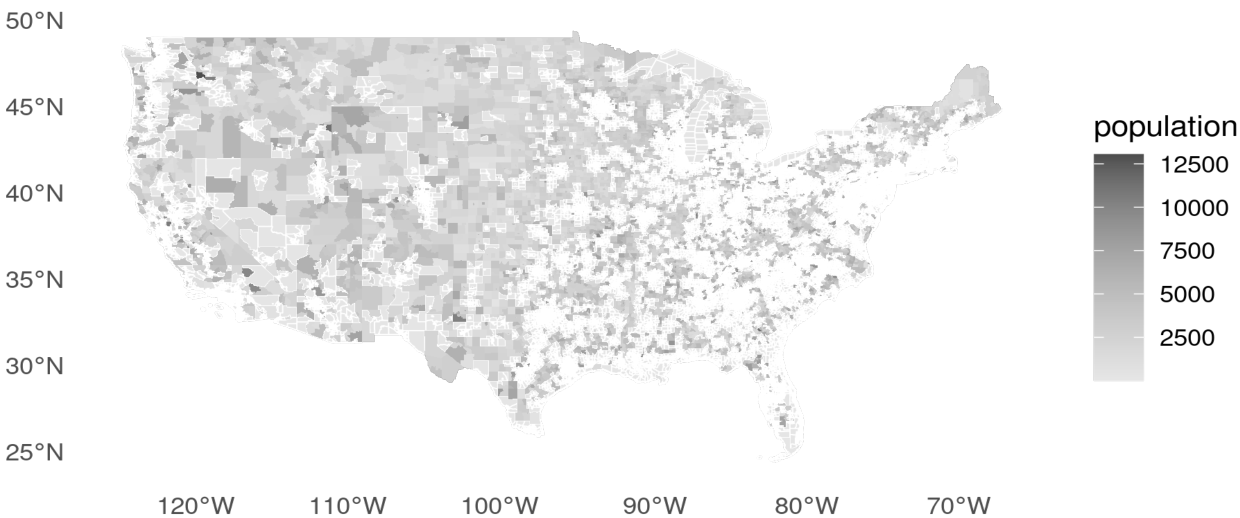

Figure 1.2: Variation in the Rural Population

1.2 *Measures of Homelessness*

For decades, scholars have debated if research should focus on the reasons why people become homeless or on the structural forces that create homelessness (Shlay and Rossi, 2003). Much research on homelessness focused on identifying and describing categories of homeless people (Lee et al., 2021). Researchers have given significant attention to 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 a household is housed and becomes unhoused. 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, 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. A common reactive program is a treatment program where an unhoused person is required to participate in short-term residential programs before being placed in more permanent housing (Evans et al., 2019). As homelessness is often seen as an urban problem, most intervention occurs in urban areas (Gleason et al., 2021). Significant federal action on homelessness began 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). The Homeless Emergency Assistance and Rapid Transition to Housing Act of 2009 expanded the definitions of homelessness for supported federal programs to expand those eligible beyond the literal homeless (Berg, 2013). These included those living in a place that is not meant for habitation, people who are expected to lose their residence within 14 days, families with children that are unstably housed, and people fleeing domestic violence (Evans et al., 2019). Federal action has significantly impacted the help available for households in need, the piece that is missing is a way to target rural areas in need of assistance.

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 have a significant impact on one's life. In the United States, housing is often a family's greatest expenditure, their 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 HUD. Housing is often 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, 17; Leifheit et al., 2022). A health disparity or health inequity is a difference in health or health outcomes relevant to social, political, and economic factors (Lutfiyya et al., 2012). One major factor that has been linked to health disparities is income, and this relationship exists across a wide range of socioeconomic factors (Canto et al., 2014). Part of acknowledging housing as health is moving beyond the housed and unhoused binary in order to better understand and intervene in households that are at risk of becoming 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 being housed or unhoused (DeLuca and Rosen, 2022).

1.4 *Theoretical Framework*

Housing insecurity has a variety of definitions across government organizations, but its domains can be characterized as encompassing housing stability, housing affordability, housing quality and safety, and neighborhood quality and safety (Cox et al., 2019). Housing insecurity arises when households face the extreme negatives of these concepts.To further refine these broad characteristics, this thesis follows the 4 Cs approach to housing insecurity. With little infrastructure for homelessness services in rural areas, the 4 Cs approach to housing insecurity proposed by Hernández and Swope (2019) 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 Cs include conditions: the quality of housing, cost: the affordability of housing, consistency: residential stability, and context: neighborhood opportunity. The 4 Cs of housing are an interconnected web of factors that impact health and encapsulate the "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. That multiple scholars have conceptualized a similar approach indicates that it may appropriately encapsulate housing insecurity within the limitations of our understanding.

1.5 *Motivation*

Three primary reasons motivate this thesis. The first motivation stems 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 by leveraging publicly available data sources intended to study social problems. The second motivation is to provide an algorithmic implementation of the 4 Cs model of housing insecurity in the rural United States. With the urban lens to housing insecurity, an adequate theoretical model must be capable of adapting to areas left out of the conversation. One study has applied the 4 Cs model to housing insecurity in the state of Maine (Gleason et al., 2021). Most applications of the 4 Cs have been to study the relationships between various health conditions and housing insecurity, but no studies have applied it broadly to rural housing insecurity. The final motivation is to provide policymakers and researchers with a 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 many people are involved in the policy-making process with no adequate mechanism for addressing housing insecurity in their constituency. The field of data science has a variety of techniques such as clustering and classification that can aid in targeting aid towards communities that have an elevated risk of housing insecurity.

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 under the 4 Cs model of housing insecurity from the data science perspective. Data science is a multidisciplinary field that employs scientific methods, algorithms, and systems to extract insights and knowledge from structured and unstructured data. It integrates expertise from statistics, mathematics, computer science, and domain-specific knowledge to analyze and interpret complex data sets. The goal is to enable informed decision-making and uncover valuable patterns within data sets. The field of data science is naturally suited to an exploratory analysis of housing insecurity. Risk factors across eight different axes are used to assign risk levels to rural census tracts. Each state is clustered with census tracts from other states within a 15-mile boundary to encapsulate how communities span across state lines. The clusters 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 represents a low, medium, or high risk of housing insecurity. These risk levels are used to highlight census tracts at a low, medium, and high risk of housing insecurity relative to other census tracts in their state. Association rules learning is used to identify common patterns between sector risk levels and note the considerable number of unexpected relationships. To better understand how factors of housing insecurity relate to space, local and global 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 spatial clustering of housing insecurity risk in rural areas. Finally, a multinomial logistic regression is used to determine how well each state's sector risk levels can be predicted, and a national model is generated for each sector's risk levels to analyze how well the risk levels created by this implementation of the 4 Cs model can be predicted nationally and state by state. Beginning to understand rural homelessness requires that several questions be answered:

- How can risk factors 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?
- When measuring housing insecurity across different dimensions, how often do the same features arise?
- What are the 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*

The work presented in this thesis presents a novel application of the 4 Cs of housing insecurity framework and applies it to rural areas. This framework allows for the identification of 280 census tracts with a high risk of housing insecurity and 1,692 census tracts with a medium risk of housing insecurity relative to their state and neighboring communities. The association rules learning results show that there are a notable number of unexpected relationships where a high risk in one sector is associated with a low risk in another sector and vice versa. It presents evidence that clustering of housing insecurity factors may not be as common in rural areas as they are in urban areas. It also shows that it is difficult to predict the risk levels of census tracts with state and national models. Due to its exploratory nature, the results are primarily intended to be used as a starting point for future research into rural housing insecurity.

1.8 *Intended Audience*

This thesis is intended for an audience with a significant interest in rural housing insecurity. Such an audience can include policymakers, economists, political scientists, community psychologists, rural sociologists, social data scientists, geographers, demographers, and many others concerned with housing insecurity and rural areas.

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 Cs of housing insecurity model and the study of data science as it relates to the methods used and previous applications of data science methods to housing insecurity-related factors. This chapter reviews pertinent literature on various facets of the model. Chapter 3 explains the methodology employed for data processing and analysis. It provides an in-depth explanation of the methodology for applying the theoretical framework and its execution. Chapter 4 presents the study's findings, offering a detailed analysis of the acquired results. Chapter 5 deliberates on 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 highlights potential avenues for future research and study.

CHAPTER 2

BACKGROUND

There are four primary areas where the extant literature must be analyzed. First, it is important to understand the distinction between homelessness and housing insecurity. Next, it is necessary to take a multi-disciplinary look at each pillar of the 4 Cs of housing insecurity. Third is the unique challenges that rural areas face in order to better inform the theoretical framework. Finally, the theoretical foundations of the methods and their previous applications in the study of housing insecurity tie together the literature with the methods used. These four aspects form the theoretical basis for exploring rural housing insecurity.

2.1 *Housing Insecurity*

Housing insecurity is a term that stems from the shift in homelessness research from focusing on only the housed and the unhoused. DeLuca and Rosen (2022) argue that the term housing insecurity is a more dynamic concept than the traditional housed and 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, 759). Homelessness is generally attributed to poverty and a lack of access to affordable housing (National Coalition for the Homeless, 2009). As researchers shifted away from the housed and unhoused binary, identifying characteristics that distinguish the housed from the unhoused led to identifying factors that occur on the path to homelessness (Phelan et al., 2010). The identification of these factors likely contributed to the rise of housing insecurity as a concept and as a field of study. DeLuca and Rosen (2022) argue that housing insecurity may be a more useful term than the housed and unhoused binary because it acknowledges housing hardship and housing risk as a continuum that affects a wider section of the population than literal homelessness. One issue with the study of housing insecurity is that, like rurality, a wide variety of definitions are used. Cox et al. (2019) analyzed 106 studies and found that current approaches to housing insecurity have three major issues: a lack of a uniform definition, it is often applied as an underdeveloped concept, and it is often measured inconsistently. An explanation for this variation is that housing insecurity operates through a variety of mechanisms, a relationship the 4 Cs model encapsulates (Leifheit et al., 2022). Rather than giving a strict definition, it may be more beneficial to look at the dimensions of housing insecurity. Cox et al. (2019) identifies seven dimensions of housing insecurity: housing stability, housing affordability, housing quality, housing safety, neighborhood safety, neighborhood quality, and literal homelessness. To adequately address housing insecurity, it is necessary to have a theoretical framework that can encompass these different dimensions to avoid these pitfalls.

2.2 *The 4 Cs of Housing Insecurity*

To understand housing insecurity in the context of the 4 Cs framework, the following subsections detail each pillar of housing insecurity under the model proposed by Hernández and Swope (2019). As each pillar forms a web rather than separate pieces, there is a significant amount of overlap between pillars. The pillars meet the dimensions identified by Cox et al. (2019): housing stability (consistency), housing affordability (cost), housing quality/ housing safety (conditions), and neighborhood safety/ neighborhood quality (context).

2.2.1 *Cost*

Housing affordability is generally conceived as the amount of a household's budget that goes to housing, including rent or mortgage payments, utilities, and other expenses. A cost-to-income ratio is the most common way of measuring housing affordability. It is difficult to determine one number that determines when a household is spending too much on housing. The threshold for housing affordability has varied between 25 and 50 percent with the current standard set at 30 percent (Kropczynski and Dyk, 2012). Today, 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). Inherent to a cost-to-income ratio is the understanding that there are other expenses necessary for survival (Herbert et al., 2018).

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 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 communities and individuals. If a household cannot afford to live in their current place, they may be forced to relocate seeking more affordable housing voluntarily 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). 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). This is only one area where low-income households may have to compromise to maintain their fixed housing costs. Housing is often the biggest expense for low-income families, forcing them to make trade-offs between housing and other necessities (Desmond and Bell, 2015). The shortage of affordable housing drives lower-income families to substandard housing in worse neighborhoods (Braveman et al., 2011). This creates the potential for a spiral where housing instability cannot be escaped due to the added costs of moving. Kang (2021) characterizes housing instability as a by-product of the affordable housing shortage wherein households can be destabilized by minor financial shocks. 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 et al., 2015). Housing affordability is both influenced by and exerts influence on many other aspects of life, and its relationship to housing insecurity cannot be understated.

2.2.2 *Conditions*

Internal housing conditions have been identified 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 (Angel 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). Links to an increase in disease have been tied to poverty, poor housing, and degraded environments reflecting the interconnectedness of housing insecurity issues (Rauh et al., 2008). Angel and Bittschi (2014) 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 indoors (Palacios et al., 2021). The relationship between housing conditions, poverty, and the range of factors that contribute to housing conditions emphasize the importance of viewing housing as a matter of health. 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. The impacts of housing conditions on health means that adequate housing is a public health issue (Matte and Jacobs, 2000). 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). Without a sense of communal benefit towards housing, this leaves marginalized populations that are more likely to be exposed to harmful housing conditions without community support (Swope and Hernandez, 2020). Growing a sense of communal benefit towards housing would be beneficial for all aspects of housing.

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 two years (Bachmann and Cooper, 2014). Classic urban economic theories hold 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 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 et al., 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, millions of low-income households struggle to maintain housing stability (Phinney, 2013; Kang, 2019). Besides voluntary relocations forced relocation can be triggered by foreclosure, eviction, and condemnation (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 linked predictors (Kang, 2019). This limits our ability to analyze and reduce negative residential mobility. One group at a higher risk of housing instability are those who rent their housing. Renters are particularly vulnerable to relocating to worse neighborhoods than the ones they are exiting (Desmond et al., 2015). Residential instability is closely related to housing affordability, housing conditions, and context. These relationships reinforce the idea that housing insecurity is an interconnected web.

2.2.4 *Context*

Context revolves around neighborhood and community characteristics including demographics, green spaces, education, and healthcare among many other things. Context involves "the presence of positive or adverse health-relevant resources in the surrounding neighborhood" (Swope and Hernandez, 2020, 9). This makes it difficult to capture context in its entirety due to its wide scope. To best encapsulate context with the available data, this thesis focuses on demographics, employment, housing type, and household 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 selected 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, the productivity of the average worker has increased by 64.7 percent while hourly pay has only increased an average of 17.3 percent between 1979 and 2022 (Economic Policy Institute, 2022). Over this same period, HUD data show that the median price of a new single-family home increased from $60,600 ($232,091 adjusted for inflation) in 1979 to $369,800 in 2021 (U.S. Census Bureau and U.S. Department of Housing and Urban Development, 2023).

These shifts in the housing market are one of the underlying factors in the rise of the affordable housing shortage. As wages have failed to keep up with the price of housing, the current economic system under the risk regime places those with low incomes in a precarious situation for housing affordability and residential stability.

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). Economic conditions play a significant role in housing insecurity because adequate income is critical for all aspects of housing insecurity.

One significant cause of employment insecurity is a lack of economic diversity, generally caused by a lack of economic development. Sherrieb et al. (2010) identify three key elements connected to economic development: the level of economic resources, the level of equality in resource distribution, and the level of diversity in economic resources. Economic development alongside demographic change in rural areas have been linked to the quality and condition of local housing infrastructure (Barcus, 2011). How policies shape economic development has a direct effect on the overall housing insecurity risk of rural communities. Amid the recent major economic shifts, globalization and shifting employment sectors play a critical role in the development path of communities which has an inherent effect on the people who live there (Harrison et al., 2019). Demonstrating the interconnectedness of communities, regional economic development in one area can encourage economic stability of its neighboring regions as well so it is important to view communities as interrelated rather than separate entities (Chen, 2018). Deller and Watson (2016) highlights the importance of economic diversity, a vital aspect of economic development, finding that more diverse economies enhance economic stability. As an insulator against economic instability, employment diversity is a key factor that policymakers and scholars should consider as part of a holistic approach to housing insecurity.

2.2.4.2 *Housing, race, education, 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 a predictable fashion by those characteristics that define its members' position in the [social] stratification system" (Lee and Evans, 2020, 5). This vulnerability is driven by a combination of individual and socio-demographic factors. One major factor that has made minorities vulnerable to housing insecurity is discrimination in housing. 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 and enforcement of the law remained difficult until the Fair Housing Act of 1988 (Sharp and Hall, 2014). For example, the practice of redlining made it difficult for African Americans to receive mortgages under federal aid programs and created 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. A high concentration of poverty may exacerbate housing conditions 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). These sources of inequity are an important part of a holistic approach to housing insecurity.

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, the median rent in a poor neighborhood is higher compared to a middle-class or 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). Renters with high-cost housing are unable to increase household wealth through their means of housing. In addition to whether one rents or owns a home, the type of home can play a significant role in housing insecurity. Of particular concern 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, making an already vulnerable population more vulnerable (Wakin, 2005). In rural areas, mobile homes are often seen as an affordable option, but they come with certain risks not as common in traditional housing. Structural problems like poor construction and risks of air pollution and fire create a unique problem (MacTavish et al., 2006). 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 (MacTavish, 2007). Individuals who rent or own mobile homes, unconventional housing, face high rental costs or high mortgage costs are at a heightened risk of housing insecurity in rural areas and should be taken into consideration.

2.2.4.4 *Household Factors*

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" (The American Presidency Project). Since then, the patchwork of programs regulated at the federal, state, and local levels has expanded. A large part of the federal government's growth in the late 20th century is from the expansion of social welfare spending (Fishback, 2020). Today, 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 are closely related to one's poverty risk. To alleviate this poverty risk, social programs that 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 (U.S. Department of Agriculture, 2023). Transportation plays a large role in social and economic life. Access to everything from education to healthcare depends on the infrastructure and the ability to use available means of transportation. 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). This is another instance where opportunity is dependent on a household's income. Automobile ownership can be a crucial factor in avoiding residential instability (Kang, 2019). This concerns rural areas without public transportation where distances may be too far or too dangerous for alternate means of transportation due to a lack of proper road infrastructure, limiting access to adequate employment and opportunity. Areas at a high risk of housing insecurity are likely to exhibit many of these factors and they play a significant part in the context of housing.

2.3 *Challenges for 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. The Census Bureau defines metro areas as urban areas of 50,000 people or more, and urban clusters of 2,500 to 49,999 people with all other areas classified as rural; the Office of Management and Budget defines metro areas as urban cores with populations of 50,000 or more people, micro areas as urban cores of 10,000 to 49,999 people where micro areas and counties outside of metro and micro areas are considered rural (Health Resources & Services Administration, 2022). Several measures consider rurality a spectrum, such as the Rural-Urban Commuting Area Codes (RUCA) used in this research. RUCA codes form an urban-rural scale ranging from 1 to 10, accounting for population, density, urbanization, and commuting (U.S.D.A. Economic Research Service, 2023). Choosing an appropriate set of RUCA codes can more adequately capture the number of rural people that live in areas classified as urban identified by Isserman (2005). Improperly classifying rural areas as urban or wrapping all rural areas into a blanket group reduces ability to address housing insecurity appropriately (Yousey and Samudra, 2018).

Part of the blanket construct of rural areas is that they are cheaper to live in. However, Kurre (2003) notes that there is relatively little systematic data that supports this presumption. Zimmerman et al. (2008) found no consistent pattern of lower prices across rural counties in Pennsylvania. Rural areas face the same low per capita income and poverty problems faced by urban areas (Castle et al., 2011). 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). Although 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, scholars have identified both pockets of prosperity and pockets of deep poverty in rural areas. Concentrated poverty is "often the manifestation of an interactive and inter-generational 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" (Thiede et al., 2018, 7).

Poverty is acknowledged more in urban areas, but poverty rates are highest in both remote rural counties and in cities (Miller and Weber, 2003; Crandall and Weber, 2004). 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 (Burton et al., 2013). 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. A cluster analysis found 3,017 places with poverty rates around 20 percent above the national average that account for about 5 percent of the nation's population and almost 87 percent of this population lives in rural 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 non-metro 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 et al., 2007).

One potential explanation for the persistent effects of poverty in rural areas is the isolation from schools, services, social interactions, and labor market resources (Canto et al., 2014). 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).

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. 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 (Thiede et al., 2018). Spatial inequality expands concerns with stratification into the realm of geographic space (Lobao and Saenz, 2002). In rural areas where location determines many aspects of the community constructed on top of it, researchers cannot ignore the implications of spatial inequality. The presence of both highly prosperous and highly impoverished rural areas indicates a need for a better understanding of the role of spatial inequality in rural areas.

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, driven by the transition from a production to a consumption-based economy (Pendall et al., 2016). 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). 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. Preventing the amelioration of problems facing rural areas is the relatively uncoordinated approach to rural development that has occurred despite the active role the federal government has played in it (Wilson and Rahe, 2016). As a result, some rural regions have experienced periods of sustained growth while others have faced the previously mentioned issues (Johnson, 2012). One aspect of this is the friction that 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 many people move to urban areas for greater economic opportunities leaving rural towns 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 reduction and communities have little ability to control these processes that limit economic mobility and may perpetuate poverty (Zarecor et al., 2021).

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. One study found that of 746 counties facing population decline, 91 percent of them are considered rural although it should be noted that 31 percent of rural counties had population gains and this study followed the U.S. Office of Management and Budget definition of rural which is all non-metro counties (Johnson and Lichter, 2019).

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 (Thiede et al., 2018). Persistent problems faced by the rural poor include "physical isolation and poor public transportation, inadequate schools, and limited access to medical care and other basic public services while institutional support services are frequently limited or simply unavailable" (Lichter and Johnson, 2007, 333). Part of this is driven by the outflow from rural areas to urban areas. Rural areas have seen a population reduction, reducing the capabilities of public services to accommodate those in need (Bjerke and Mellander, 2019). Thiede et al. (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 African American population was by far the most disadvantaged. 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, 2012). Another demographic group that is significant to rural areas is Hispanics and Latinos, an increasing population despite the widespread population decline of rural areas (Lichter and Johnson, 2020). African Americans, Hispanics, and Latinos encounter comparable discrimination in the housing market, resulting in significantly diminished benefits from housing for these demographic groups (Krivo and Kaufman, 2004). 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.

Another problem rural areas face is gentrification. "Gentrification is the process by which higher-income households displace lower-income residents of a community, changing the essential character and flavor of that community" (Yagley et al., 2005, 1). Similar to housing insecurity, most gentrification research focuses on rural areas and leaves rural areas out of the conversation (Yagley et al., 2005). "Although there are many different types of gentrification, scholars generally agree that gentrification is fundamentally a process that involves the reinvestment of capital after a period of disinvestment, the production of an aestheticized landscape, and lower class displacement followed by middle class replacement" (Bryson, 2013, 578) The limited supply of housing gives rural areas a higher risk of severe outcomes from gentrification due to their limited housing supply, making housing prices increase faster and displacement move residents further than their urban counterparts (Golding, 2016). Rural gentrification often looks different in rural areas. In a case study, Yagley et al. (2005) found that rural gentrification in the counties studied led to an affordable housing shortage as low-income families were priced out of the housing market. Gentrification has a significant impact on rural areas and scholars should place greater emphasis on understanding where and why it happens in rural areas.

2.4 *Data Mining Methods*

This thesis uses the data mining framework to explore housing insecurity in rural areas. Data mining is the process of "extracting interesting information or patterns from large information repositories" (Zhao and Bhowmick, 2003, 1). Data mining has a long history with the use of the term tracing back to at least the 1980s (Coenen, 2011). Data mining encompasses a wide variety of methods and perspectives, but it can generally be thought of as a combination of traditional data analysis and statistical approaches drawn from a wide range of disciplines (Jackson, 2002). At a very general level data mining can be considered the search for hidden and interesting patterns in generic data (Chen et al., 2011). At the heart of data mining is machine learning which aims to increase the levels of automation in the knowledge discovery process (Jackson, 2002). This section analyzes how data mining methods have contributed to the study of housing insecurity and provides a brief overview of the history of the methods used.

2.4.1 *Clustering*

At its core, clustering is about grouping objects based on their similarities (Zhao and Bhowmick, 2003). Clustering has long played an important role in data analysis. It serves as a means of finding groups in a data set that are often characterized by similarity within groups and dissimilarity between groups (Sinaga and Yang, 2020). While not the first means of clustering data, the k-means clustering algorithm has had a significant amount of influence on the development of clustering as a methodology. The usage of the term traces back to Macqueen (1967, 283) where the procedure was informally described as "simply starting with K groups each of which consists of a single random point, and thereafter adding each new point to the group whose mean the new point is nearest. After a point is added to a group, the mean of that group is adjusted in order to take account of the new point". Steinhaus (1956) is attributed with an early version of k-means clustering in which the editors refer to the paper as the first formulation of the problem of partitioning by k-means, also referred to as nuées dynamiques or "dynamic clouds". Between early formulations of k-means clustering and the formalization of the standard algorithm of today, several other clustering algorithms were presented. One with similar underlying principles is the grouping for maximum homogeneity algorithm proposed by (Fisher, 1958). Ball and Hall (1965) proposed the Iterative Self-Organizing Data Analysis technique (ISODATA). The primary difference between ISODATA and k-means is that ISODATA is an unsupervised learning algorithm, meaning that K does not need to be specified as it does in k-means clustering. Both the Fisher and Ball and Hall algorithms follow similar principles of k-means clustering in that they aim to have similarity within clusters and dissimilarity between clusters. The standard k-means clustering formula was first published by Lloyd (1982) for utilization in pulse-code modulation. Since then, a plethora of clustering algorithms have been proposed for a variety of tasks. Jain (2010, 653) highlight three main purposes that clustering has been used for:

- Underlying structure: to gain insight into data, generate hypotheses, detect anomalies, and identify salient features
- Natural classification: to identify the degree of similarity among forms or organisms
- Compression: as a method for organizing the data and summarizing it through cluster prototypes

For a more detailed description of the various clustering algorithms see Xu and Tian (2015) or Ahmed et al. (2020).

k-medoid clustering was chosen to avoid the sensitivity to outliers that occurs when using k-means clustering (Kaur et al., 2014). k-medoid clustering avoids this problem by using the most centrally located data point in a cluster, known as the medoid. k-medoid clustering was first proposed by Kaufman and Rousseeuw (1987, 1) wherein "the representative object of a cluster is the object for which the average dissimilarity to all the objects of the cluster is minimal. This object is called the medoid of the cluster". The k-medoids algorithm aims to minimize the sum of dissimilarities between each object and its reference point (Kaur et al., 2014).

Madhulatha (2011, 67) outline the process of k-medoids clustering:

1. The algorithm begins with an arbitrary selection of the k objects as medoid points out of n data points
2. After the selection of the k medoid points, associate each data object in the given data set to the most similar medoid.

3. randomly select non-medoid object O'

4. compute total cost, S of swapping initial medoid object to O'

5. if S < O, then swap the initial medoid with the new one

6. Repeat steps 2 to 5 until there is no change in the medoid

It does not appear as though k-medoid clustering has been applied directly to housing insecurity, clustering has been used in applications related to housing. Peters (2009) use clustering to create a typology of American poverty at the minor civil division level. Yoder Clark et al. (2021) use k-means clustering to highlight significant factors leading to homelessness. Borders et al. (2018) use k-means clustering to analyze factors of food insecurity at the census tract level. Sarwosri et al. (2016) use k-means clustering to categorize levels of poverty as well. These previous uses lend credibility to clustering as a means of analyzing housing insecurity.

2.4.2 *Association Rules*

The next algorithm is the Apriori association rules algorithm. Rules-based methods aim to find regularities in the data that can be expressed in the form of an IF-THEN rule (Fürnkranz and Kliegr, 2015). The algorithm attempts to find the possibility of the simultaneous occurrence of items in the itemset and build relationships among them (Chen et al., 2011). The goal of association rules mining is to extract patterns or correlations occurring in the dataset (Zhao and Bhowmick, 2003). Association rules primarily developed out of the need to quickly analyze large amounts of data stored in databases. Agrawal et al. (1993) proposed an early version of association rules using R*-trees known as the AIS algorithm. The Apriori algorithm was first proposed by Agrawal and Srikant (1994). Apriori is considered the most well-known algorithm for association rules (Dunham et al., 2000). Since Agrawal and Srikant (1994) there have been a variety of association rules algorithms proposed including tree-based and fuzzy algorithms. It should be noted that association rules consider items to be dependent rather than independent. This means that the occurrence of one item influences the occurrence of another item. The concept of dependence is fundamental to association rule mining because it aims to identify relationships between items that occur together more frequently than would be expected if the items were independent of each other. See Solanki and Patel (2015) and Dunham et al. (2000) for surveys of association rules. The apriori algorithm works in two phases by first finding all frequent itemsets and then turning them into association rules based on confidence and support minimums (Fürnkranz and Kliegr, 2015). Association rules learning has untapped potential in the study of housing insecurity. Similar studies include one by Zhen and Mengxian (2020) who used association rules to determine the effectiveness of poverty alleviation programs in rural China and another by Tian et al. (2020) who used association rules and other machine learning algorithms to show that water reservoir density is related to poverty.

To better understand association rules learning, one can look at The Myers-Briggs Type Indicator (MBTI). It categorizes individuals into one of 16 personality types based on preferences in four dichotomous dimensions: Extraversion vs. Introversion, Sensing vs. Intuition, Thinking vs. Feeling, and Judging vs. Perceiving. Figure 2.1 shows the distribution of MBTI types in the United States. These personality types can be likened to items in association rules learning, where each type represents a specific combination of preferences. Association rules learning identifies patterns and relationships between variables in datasets. It generates rules that describe associations between items, akin to how MBTI types might be associated with certain behaviors or preferences. Metrics such as support and confidence quantify the strength of these associations, measuring the frequency of occurrence of specific itemsets (e.g., personality types) and the likelihood of one item occurring given the presence of another (e.g., behavior or preference). Again similar to association rules, MBTI types assume that the four dichotomies within each type are dependent on each other.

ISTJ 11.6%	ISFJ 13.8%	INFJ 1.5%	INTJ 2.1%
ISTP 5.4%	ISFP 8.8%	INFP 4.4%	INTP 3.3%
ESTP 4.3%	ESFP 8.5%	ENFP 8.1%	ENTP 3.2%
ESTJ 8.7%	ESFJ 12.3%	ENFJ 2.5%	ENTJ 1.8%

Figure 2.1: Myers-Briggs Type Indicator Breakdown (Varona and Capretz, 2011)

2.4.3 *Contiguity*

Contiguity is an important aspect of the present interaction with housing insecurity. First, one of the main methods used to better understand rural areas is spatial autocorrelation. To conduct spatial autocorrelation, you must create a weights matrix and the type of contiguity you choose can have a significant impact on results. Second, proximity-based contiguity is an important aspect of capturing the inter-state nature of rural communities. Queen's contiguity is used to build the spatial weights matrix for analyzing the spatial autocorrelation of housing insecurity factors in rural areas. Queen's Contiguity looks for polygons that share a border. There are three general types of contiguity: Rook, Bishop, and Queen. Figure 2.2 reflects the different boundaries considered by each type. As the figure shows, Queen contiguity allows for the highest number of shared borders. This application is well-suited for Queen's contiguity because it best encapsulates the interconnected nature of rural areas.

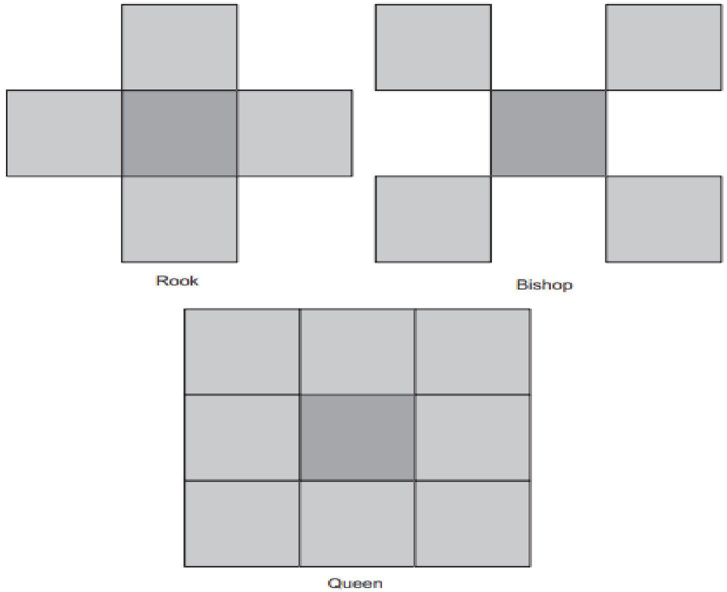

Figure 2.2: Types of Contiguity (Grubesic, 2008)

These contiguity measures are used to analyze the connectivity of polygonal features, such as census tracts, in a geographic area (Grubesic, 2008). Contiguity is an important part of creating the spatial weights matrix that underlies much of spatial analysis including Moran's I. The spatial weights matrix is one of the most used measures of the influence that places have on each other (Bavaud, 1998). Shortly after the proposal of Moran's I in 1950, the need to weight observations based on their influence was considered by the cartographer Arthur Robinson in 1956 (Getis, 2008). Since the 1960s, many researchers have attempted to properly represent spatial dependence with a spatial weights matrix, and the impact of one's definition of a neighbor on spatial results was recognized as well (Zhang, 2012; Getis and Aldstadt, 2004). Since then, a variety of ways to generate a spatial weights matrix have been created including spatially contiguous neighbors, the length of shared borders divided by the perimeter, ranked distances, and n nearest neighbors among others (Getis and Aldstadt, 2004). Spatial weights and the definition of neighbors are closely related to Tobler's First Law of Geography, which is detailed in the following subsection. It should be noted that the chessboard concept of contiguity in Figure 2.2 does not fully translate to the geographical space because of the inconsistencies in boundaries at different geographic levels.

In housing insecurity analyses that span state lines, overlooking neighboring communities can be unfair due to shared dependencies. To mitigate this, the analysis adopts proximity-based contiguity to encapsulate the interconnectedness of rural areas. Unlike the previously mentioned forms of contiguity, proximity-based contiguity defines adjacency based on proximity rather than a physical connection.

2.4.4 *Spatial Autocorrelation*

Spatial autocorrelation is a method that enables researchers to account for the spatial patterns inherent in spatial data. Spatial autocorrelation is similar to correlation but rather than showing relationships between variables, it shows the correlation across georeferenced space (Getis, 2008). Generally, the concept refers to similar or dissimilar values producing a detectable pattern when placed on a map, often referred to as spatial clustering (Griffith, 1992). In an analysis of crops, Fisher (1935, 74) noted that "patches in close proximity are commonly more alike... than those which are further apart". The idea known as spatial autocorrelation was not formalized until the proposal of Moran's I in 1950, when it was referred to as spatial correlation (Moran, 1950; Getis, 2008). Spatial autocorrelation is highly connected to Tobler's First Law of Geography. According to Tobler, "everything is related to everything else but near things are more related than distant things" (Tobler, 1970, 236). Shortly after the proposal of Tobler's First Law of Geography, Cliff and Ord published a work titled Spatial Autocorrelation in 1973 which explicated and generalized Moran's earlier work (Getis, 2008). Since then, a variety of global and local spatial autocorrelation techniques have been proposed. For a more detailed history of spatial autocorrelation see Getis (2010). Similar to other statistical tests, spatial autocorrelation is determined by rejecting the null hypothesis. De Jong et al. (1984) highlight two general null hypotheses underlying spatial autocorrelation:

- The observations are mutually independent with a known or unknown distribution
- Each permutation of the observations x_i is equally probable

When there is adequate information to reject the null hypothesis, an observation is said to be spatially autocorrelated. Figure 2.3 provides a visual representation of spatial autocorrelation.

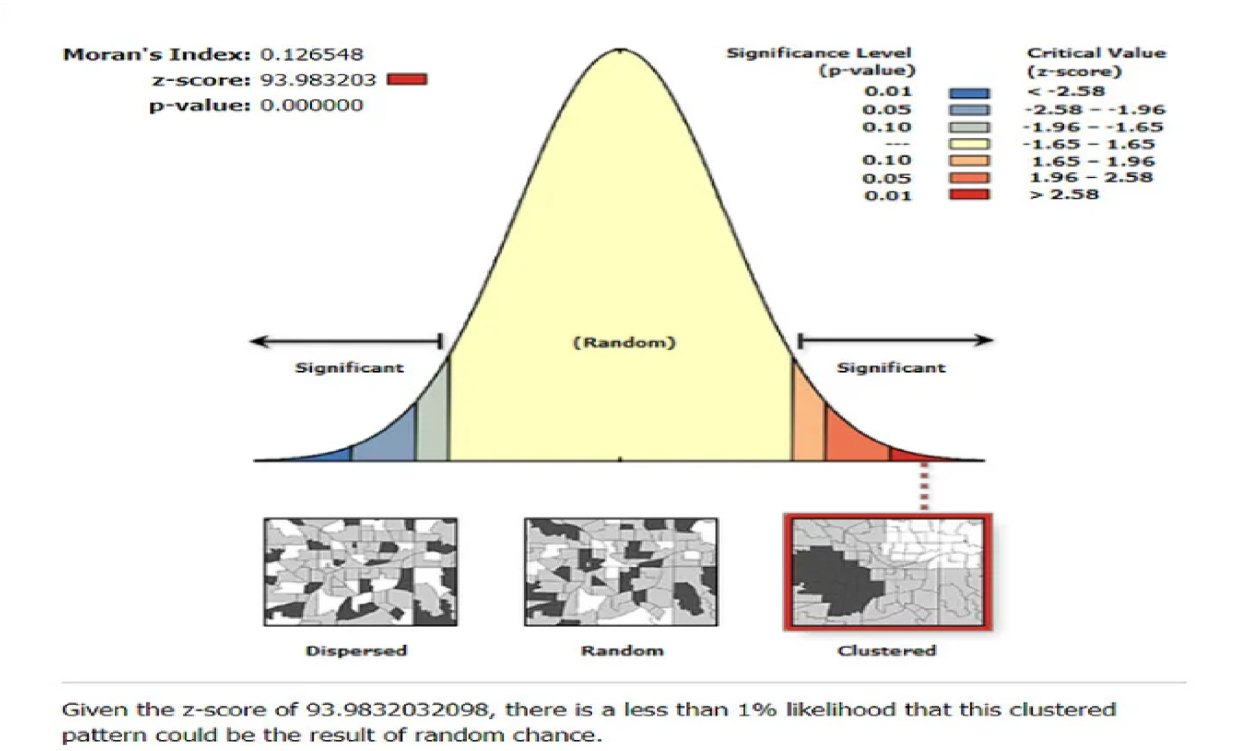

Figure 2.3: Spatial Autocorrelation Example (Asgary, 2020)

Both local and global Moran's I are used in this examination of housing insecurity. The primary difference between the two is that global Moran's I measures spatial autocorrelation across an entire dataset while local Moran's I assesses spatial autocorrelation at a local level. Three factors affect both global and local Moran's I statistics according to Boots and Tiefelsdorf (2000, 320) "the relative locations of i and j, the global average or local relative number of neighbours for i, and the observed values of x_i, $i \in \{1, \ldots, n\}$". Several studies have applied spatial autocorrelation to housing insecurity. Voss et al. (2006) used Moran's I to determine if there is spatial autocorrelation in county-level child poverty rates. Curtis et al. (2019) use spatial autocorrelation to analyze the clustering of industrial makeup (economic diversity) and poverty. Brooks (2019) used Moran's I and Local indicators of Spatial Association statistics to further spatial inequality research. Gleason et al. (2021) use Local Indicators of Spatial Autocorrelation to analyze community vulnerability in rural areas in the state of Maine. These examples demonstrate a precedent for the use of spatial autocorrelation in the study of housing insecurity.

2.4.5 *Multinomial Logistic Regression*

The final methodology used is multinomial logistic regression (MLR). The MLR is very similar to binomial logistic regression except the dependent variable contains more than two categories (Bayaga, 2010). There are multiple usages of MLR. First is to determine the amount of variance in the dependent variable explained by the independent variables, second is to rank the importance of independent variables, third is to assess the interaction effects between variables, and finally is to understand the impact of control variables (El-Habil, 2012). The use of regression dates back to the 19th century with the proposal of least-squared regression although the challenge of finding the parameters in the equation of a straight line that optimally matches three or more points in the (x, y) plane has been recognized since at least the time of Galileo (Harter, 1974). Three different scholars published the method between 1805 and 1809 with Adrien Marie Legendre publishing it first in 1805, followed by Robert Adrian in 1808 or 1809, and Carl Friedrich Gauss in 1809 although Gauss claims to have used the method since 1795 (Stigler, 1981). Legendre's application was to determine the orbit of a comet given three observations on its longitude and latitude (Legendre, 1805, 1). Gauss similarly used the method for astronomical purposes (Gauss, 2007). Adrian's version of regression is more probabilistic than that of Legendre or Gauss (Dutka, 1990). The logistic function underlying logistic regression was developed decades later by Pierre François Verhulst who developed it in a series of papers published between 1838 and 1847 (Cramer, 2003). Decades later, Pearl and Reed published the logistic curve, unaware of Verhulst's work (Kingsland, 1982). See Cramer (2003) for a more detailed history of logistic regression. The MLR model used in this thesis was proposed by Berkson in 1944 and further developed by Cox in 1958 (Msaouel et al., 2022). MLR is similar to linear regression except it uses a maximum likelihood method where the fitted regression coefficients are used to maximize the probability of the observed result (Jackson, 2002). It does so by first transforming the dependent variable into a logit variable which is the natural log odds of the expected outcome either occurring or not occurring (White, 2013). An example of the usage of multinomial logistic regression to housing insecurity could not be found, but regression is a popular technique among scholars. Peters (2009) used logistic regression in building their typology of American poverty. Crandall and Weber (2004) used an ordinary least squares regression to measure the spatial concentration of poverty and poverty dynamics. Wilson and Rahe (2016) use an ordinary least squares regression for modeling rural prosperity and federal expenditures. These similar applications demonstrate the usability of regression in the study of factors related to housing insecurity.

2.5 *Summary*

Throughout this chapter, the 4 Cs of housing insecurity have been covered. It is important to highlight the interconnected nature of the 4 Cs. There is a significant overlap between each pillar of housing insecurity. Housing costs, housing type, and housing conditions are directly linked to the economic conditions of a household. These economic conditions are linked to the household 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 live in. Rural areas face numerous issues, some that align with problems in urban areas and some that do not such as the presence of mobile homes, economies built around single amenities, and large pockets of persistent poverty and prosperity. Any discussion on housing insecurity must consider the historical forces affecting modern-day race and poverty dynamics, 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. This chapter has also detailed the history, development, and fundamentals of the methods used in this thesis.

CHAPTER 3

ADDRESSING RURAL HOUSING INSECURITY

This chapter develops a novel method for measuring housing insecurity in rural areas using the 4 Cs framework. To date, little research has been done into the presence or causes of housing insecurity in rural areas. The intent is to create a system that researchers and policymakers can use to better address housing insecurity in their constituencies. All analysis was conducted in the R statistical language[1].

3.1 *Defining Rurality*

The range of RUCA codes described in Table 3.1 was chosen to be inclusive rather than exclusive, including small towns with various levels of commuting to urban clusters and areas classified as rural. Small towns are included because they often serve as hubs for rural areas, playing an important role in rural areas and Isserman (2005) has identified a significant amount of rural people that live on the edge of urban places, like small towns. Rather than strictly defining rurality, this thesis uses the USDA Rural-Urban continuum. The following codes are used to encapsulate rurality:

Table 3.1: RUCA Codes and Descriptions (U.S.D.A. Economic Research Service, 2023)

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

[1]The code and data can be found at `https://github.com/steveg-ds/housing_insecurity`.

3.2 *Applying the 4 Cs*

Applying the four Cs 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. Priority is given to high-cost renters because studies have identified them as a high risk for housing instability. By targeting this demographic, interventions and policies can be tailored to alleviate the strain they experience, contributing significantly to the overall efforts to enhance housing stability within communities.

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. Housing conditions are measured by the lack of complete plumbing and kitchen facilities, with the assumption that if these are missing, there are likely other factors concerning housing quality factors. The risk of housing insecurity in an area therefore increases as the number of occupied and unoccupied housing without adequate basic needs increases. Priority is given to occupied housing as there will always be a certain amount of unoccupied housing not fit for habitation due to a lack of upkeep. Unoccupied housing is still important as the less housing that is available, the higher housing prices will be.

Consistency, or residential stability, 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 being housing insecure, the scope of residential mobility is limited to those who have moved in the past year with and without a high school degree and those who are either below or just above the poverty line. Emphasis is given to those who moved that are below the poverty line or do not have a high school degree as 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 Cs to capture because it encapsulates many individual, social, and political factors. As established in the literature review: demographics, employment, race, education, poverty, housing type, and household factors are all relevant to the study of housing insecurity. These sectors were chosen based on their relevance and the availability of data to measure them. Due to the influence of social, political, and historical processes, demographic diversity is used to account for 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 live in is a significant factor of context because mobile homes and unconventional 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*

A total of 85 2019 ACS 5-year variables at the census tract level are used to capture the 4 Cs of housing insecurity using indicators of housing insecurity identified in the literature to form eight different sectors. Each sector encapsulates a different aspect of housing insecurity. Appendix A shows the variables used for each sector and their descriptive statistics. These sectors are demographics, housing cost, housing quality, housing type, residential mobility: poverty, residential mobility: education, and household factors.

For demographic variables, seven variables including an "other" variable are used to account for race/ ethnicity and the number of people over or under 18 by sex. The economic diversity data is the number of people employed across 13 employment sectors. It was necessary to create three compound variables: high housing costs with a mortgage, high housing costs without a mortgage, and high renting costs to include all households spending more than 30 percent of their income on housing. There are four variables for housing conditions which include houses with an incomplete or insufficient kitchen or plumbing for occupied and unoccupied housing units. Due to the housing affordability and income inequality crises, those below the poverty level and those at 125 percent of the poverty level are included in residential mobility: poverty (RMP). For residential mobility: education (RME), 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. Household factors 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 three or more workers, and the household Gini index. The Gini index is a common measure of income inequality where zero represents perfect equality and one represents perfect inequality. For housing type, renters and owners of mobile homes, single-family residences, 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 data in one or more sectors were excluded from the analysis. These omitted tracts were assigned a risk level of zero to preserve the largest possible number of census tracts for analysis. To mitigate potential biases stemming from differences in population sizes and geographic areas, the data in each sector are scaled to a base unit. Demographic and economic diversity variables are proportional to the population size. Data on housing costs and housing type are proportional to the counts of homeowners and renters in a census tract. The household type data is proportional to the number of occupied houses, and housing condition indicators are proportional to the total count of occupied and unoccupied housing units. Household factors are proportional to the total number of households. All numerical values in the dataset are 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 global and local spatial autocorrelation and the Queen Contiguity spatial relationship algorithm to form and analyze the housing insecurity risk assignment system, and multinomial logistic regression is used to examine the predictive abilities of the risk assignment system.

3.5.1 *k-Medoid Clustering*

k-medoid 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-medoid uses data points within the dataset as cluster representatives. The key advantage of k-medoid lies in its robustness to outliers and noise due to the use of real data points. The objective of k-medoid clustering is to minimize the sum of dissimilarities within clusters. Each state, including neighboring census tracts, is clustered individually so that risk levels are relative to other communities in one state. This reduces the influence of outliers, allowing for a more equitable analysis of housing insecurity. The cluster medians are analyzed to determine which clusters have a high, medium, or low risk of housing insecurity based on the factors identified in the literature review. This approach highlights the areas that show the most vulnerability across sectors. The formula for k-medoids clustering is shown in Equation 3.1.

$$
\begin{aligned}
&\underset{S}{\text{minimize}} \quad \sum_{i=1}^{K} \sum_{x \in C_i} d(x, m_i) \\
&\quad \text{where:} \\
&\qquad S : \text{The set of clusters.} \\
&\qquad K : \text{The number of clusters.} \\
&\qquad i : \text{Index representing each cluster } (1 \leq i \leq K). \\
&\qquad C_i : \text{The } i\text{-th cluster containing data points.} \\
&\qquad x : \text{A data point within a specific cluster } (x \in C_i). \\
&\qquad m_i : \text{The medoid (representative) of the } i\text{-th cluster.} \\
&\qquad d(x, m_i) : \text{The dissimilarity (distance) between data point } x \text{ and medoid } m_i.
\end{aligned}
\tag{3.1}
$$

3.5.2 *Identifying High Risk Census Tracts*

Each sector now has a new categorical variable representing the risk level for that census tract with one being high-risk, two being medium-risk, and three being low-risk. This means that each census tract has a housing risk level on a scale of 8 to 24 with 8 being the highest level of risk and 24 being the lowest level of risk. To highlight areas of particular concern, a threshold of 12 out of 24 is used to identify high-risk census tracts. A threshold of 15 out of 24 is used to identify medium-risk census tracts. Census tracts with a total greater than 15 are considered to have a low risk of housing insecurity. This serves to highlight areas of concern for researchers and policymakers.

3.5.3 *Association Rules Learning*

Association Rules learning is a data mining technique used to uncover interesting relationships between itemsets 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. This methodology operates by analyzing transactions seeking statistically significant associations between different items. These associations are expressed as rules that outline the likelihood or dependency of one item's presence based on the occurrence of another. Association rules learning involves four 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.

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

Lift: Indicates how much more likely itemset A and B are to occur together compared to what would be expected if they were statistically independent.

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

Gain: Measures the improvement in predicting the presence of item B in a transaction when item A is also present.

$$\text{Gain}(A \to B) = \frac{\text{Confidence}(A \to B) - \text{Support}(B)}{1 - \text{Support}(B)}$$

Association rules are used to analyze the occurrences between sector risk levels. Of primary interest are high-risk-to-high-risk, low-risk-to-low-risk relationships, and unexpected relationships where a high-risk level is associated with a low-risk level and vice versa. This is used to examine if under the 4 Cs framework, census tracts commonly exhibit signs of risk or if there is variation in these relationships.

3.5.4 *Moran's I*

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 are 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 3.2.:

$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{3.2}$$

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 at the census tract level, 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 3.3. 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.

$$I_i = \frac{x_i - \bar{x}}{S^2} \sum_j w_{ij}(x_j - \bar{x}) \tag{3.3}$$

where:

- I_i is the local Moran's I statistic for location i.
- x_i and x_j are the values of the variable of interest at locations i and j.
- w_{ij} represents the spatial weight between locations i and j.
- $\bar{x}$ is the mean of the variable across all locations.
- S^2 is the variance of the variable.

Measuring spatial autocorrelation relies on the principle of contiguity discussed in Chapter 2. Contiguity defines which observations are considered neighbors or adjacent to each other, providing the basis for quantifying spatial relationships.

The borders considered in queen contiguity are shown in Equation 3.4.

$$\begin{aligned}
\text{Top-left:} \quad & (x-1, y-1) \\
\text{Top:} \quad & (x, y-1) \\
\text{Top-right:} \quad & (x+1, y-1) \\
\text{Left:} \quad & (x-1, y) \\
\text{Right:} \quad & (x+1, y) \\
\text{Bottom-left:} \quad & (x-1, y+1) \\
\text{Bottom:} \quad & (x, y+1) \\
\text{Bottom-right:} \quad & (x+1, y+1)
\end{aligned} \tag{3.4}$$

By identifying contiguous pairs of observations, contiguity guides the creation of the weights matrix, which represents the strength of connectivity between locations. In binary weights matrices, entries indicate whether pairs of observations are contiguous (with a value of 1) or not (with a value of 0), reflecting the presence or absence of spatial relationships.

3.5.5 *Multinomial Logistic Regression*

Cross-split validation is used where 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. These probabilities serve as indicators of the model's predictive accuracy, aiding in the assessment of classification performance for individual census tracts. Moreover, to gain deeper insights into the contributions of various factors to risk levels, separate national models are constructed for each sector using the entire dataset. If the models are fairly accurate this approach allows for an examination of model performance under optimal conditions and facilitates the measurement of prediction efficacy. If the models are not accurate, the confusion matrices can be analyzed to determine which risk level a model over-classifies as and it still presents valuable insights into the predictability of rural housing insecurity. The multinomial logistic regression model, employed for this analysis, is characterized by its ability to handle multiple classes and is governed by the equation depicted 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 6,362 rural census tracts with a RUCA code of seven or higher. Four states and Washington D.C. were 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 the 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. Washington D.C. was removed because it is completely urban. Figure 4.1 shows how the neighbor algorithm changed the state census tract counts. This allows the risk-level assignment to reflect the inter-state nature of communities.

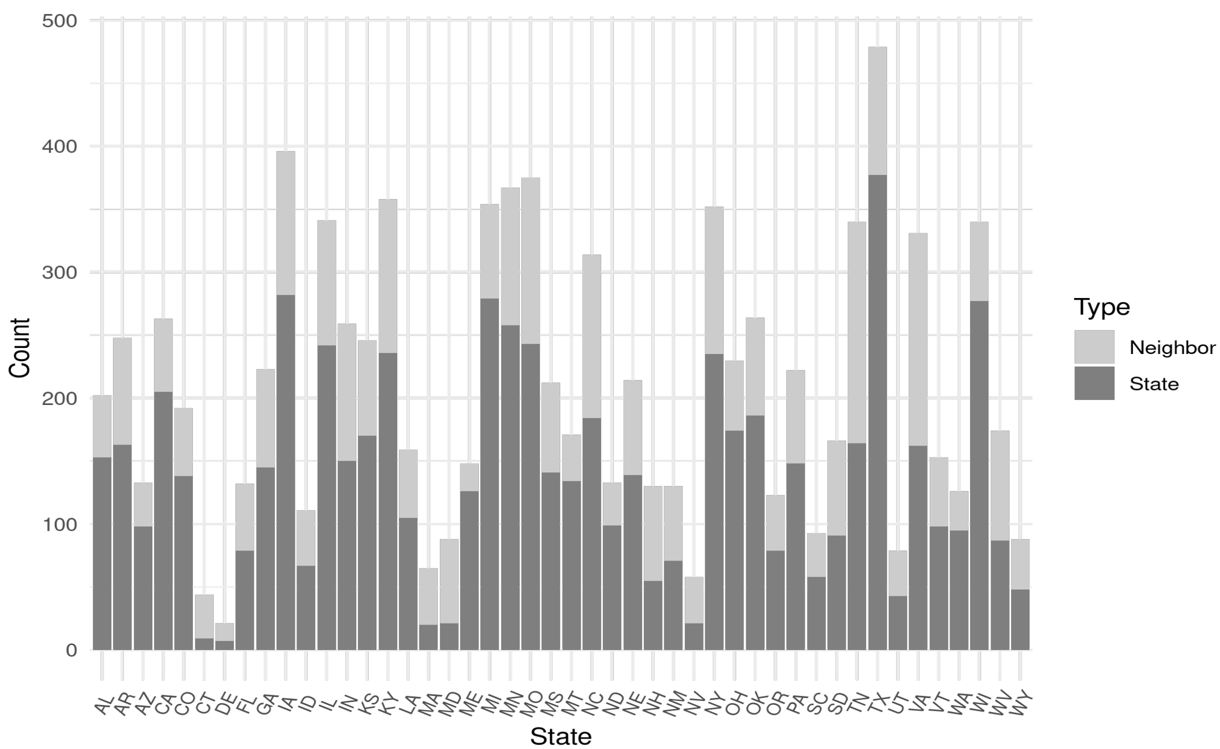

Figure 4.1: State Census Tracts vs. State Neighbors Count

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 (five percent). An area accounting for 22.4 million people was considered in the analysis.

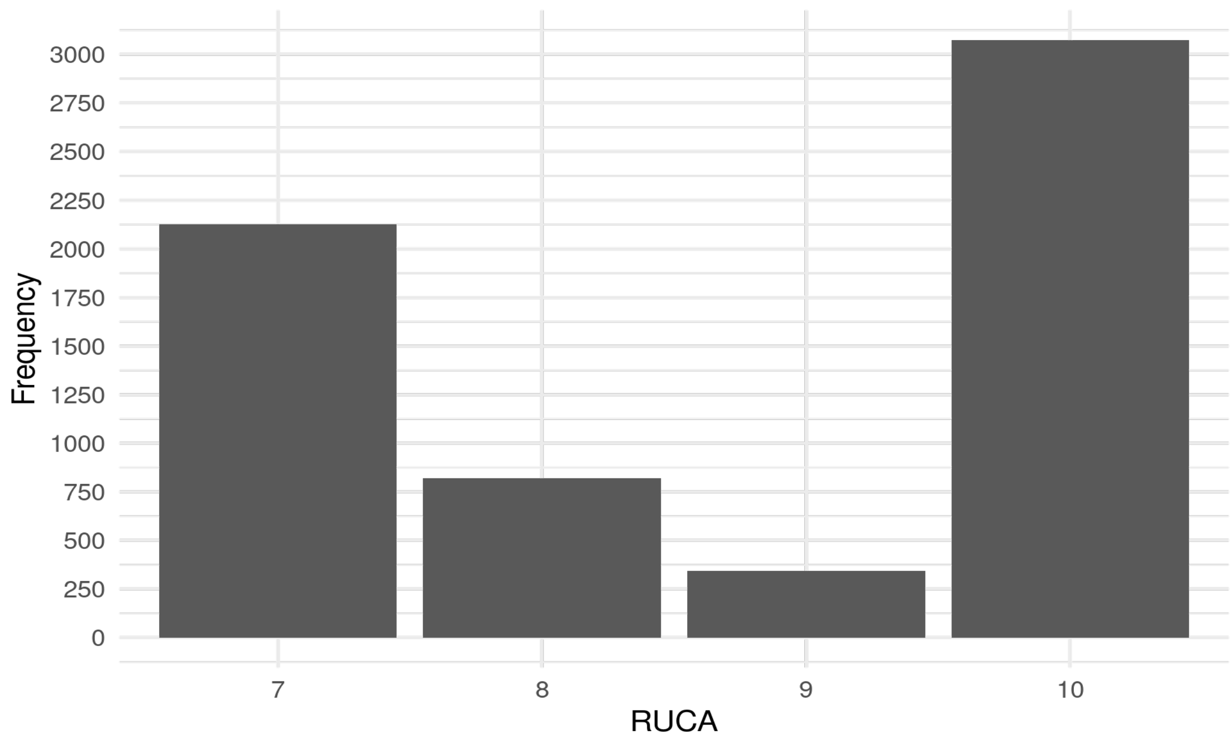

Figure 4.2: State Census Tracts vs. State Neighbors Count

4.2 *Cluster Analysis*

Here, the results of the cluster analysis are presented for each sector. All values are represented as a percentage corresponding to the base unit each sector is scaled to. Figure 4.3 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 medium-risk or high-risk level census tracts. Demographics is the only sector with notably more medium-risk than low-risk census tracts.

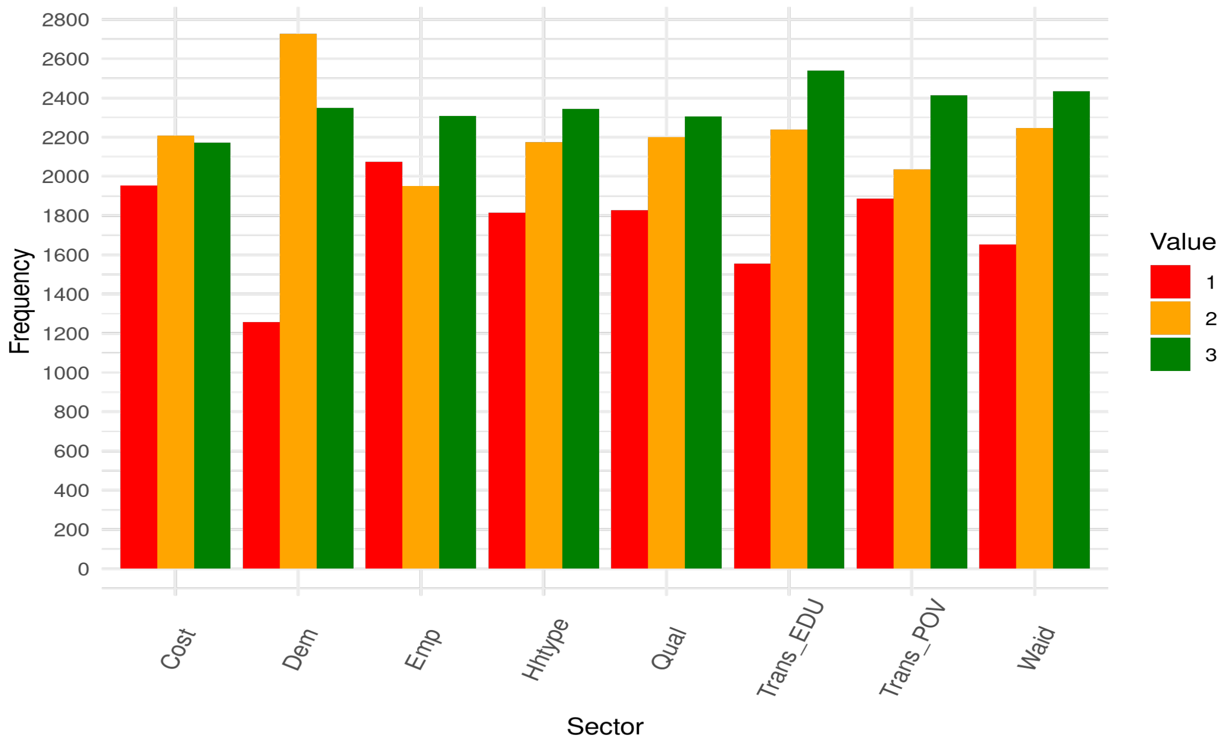

Figure 4.3: Cluster Distribution by Sector

4.2.1 *Employment*

Table 4.1 shows that cluster one had the lowest cluster medians in 61 percent of variables, cluster two had the highest cluster median for 53 percent of variables, and cluster three had the middle value for 69 percent of variables. Cluster one has the lowest level of economic diversity, cluster two has the highest 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. Cluster one becomes the high-risk level, cluster two becomes the low-risk level, and cluster three becomes the medium-risk level.

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

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. The risk levels of this sector are based on the median and average highest, lowest, and medium value counts as clusters two and three had almost the same cluster median counts. Table 4.2 shows that cluster one has the middle value for 90 percent of variables. Cluster two has the lowest values for 50 percent of variables. Cluster three has the highest number of highest values across means and medians with 55 percent of variables. Cluster three also has the largest African American and Hispanic/ Latino cluster medians. 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.

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***

Table 4.3 shows that cluster one has the highest value for high-cost mortgage. Cluster two has the lowest high-cost mortgage and high-cost rent cluster medians. Cluster three has the highest high-cost no-mortgage and high-cost rent cluster medians. 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 are of less concern than occupied housing. Table 4.4 shows that 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 low-risk level, cluster two becomes the medium-risk level, and cluster three becomes the high-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 RME, 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 three has the highest values for 71 percent of variables, including each of the less-than-high school education variables. Cluster one becomes the low-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 high-risk level.

Table 4.5: Median Values for Residential 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 RMP sector follows the criteria of residential RME closely with the variables for those who moved that are below the poverty level as the highest priority. Table 4.6 shows that cluster one has the lowest values for each variable. P1 represents below the poverty line variables and p2 represents the percentage of those at 125 percent of the poverty line. Cluster two has the medium value for 57 percent of variables. Cluster three has the highest values for 57 percent of variables including three 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.

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

4.2.7 *Household Factors*

For household factors the clusters with the highest number of maximum cluster medians determine the risk levels with particular attention given to households with no 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. Cluster one becomes the low-risk level, cluster two becomes the medium-risk level, and cluster three becomes the high-risk level.

Table 4.7: Median Values for Household Factor 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. Table 4.8 shows the values for this sector. Cluster one has the highest owner single and the lowest renter mobile home. Cluster two has the medium value for owner single and owner mobile home.Cluster three has the highest owner mobile and the medium value for renter mobile. 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

4.3 *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-riskassociations (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 4.9, 4.10, 4.11, and 4.12 show the average support, average confidence, coverage, and average lift for the different association rules. Figure 4.4 shows the overall trends in the association rules. Note that the observations with support of 20 percent or greater are empty on the left-hand side, meaning these points represent the presence of the risk levels in the dataset instead of associations between risk levels. Support is low with confidence below 0.2 for most of the rules. The figure shows a notable amount of clustering around the 0.35 confidence and ten percent 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 values. 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 across sectors.

Table 4.9: 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
RME	0.07	0.3	0.25	1.1
RMP	0.09	0.3	0.3	1.1
Housing Costs	0.09	0.29	0.31	1.1
Housing Quality	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

Table 4.10: 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
RME	0.15	0.38	0.4	1
RMP	0.15	0.4	0.38	1.1
Housing Costs	0.13	0.37	0.34	1
Housing Quality	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 4.11: 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
RME	0.11	0.27	0.4	0.95
RMP	0.09	0.24	0.38	0.89
Housing Costs	0.09	0.27	0.34	0.99
Housing Quality	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

Table 4.12: 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
RME	0.09	0.35	0.25	0.95
RMP	0.11	0.35	0.3	0.96
Housing Costs	0.11	0.36	0.31	0.94
Housing Quality	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

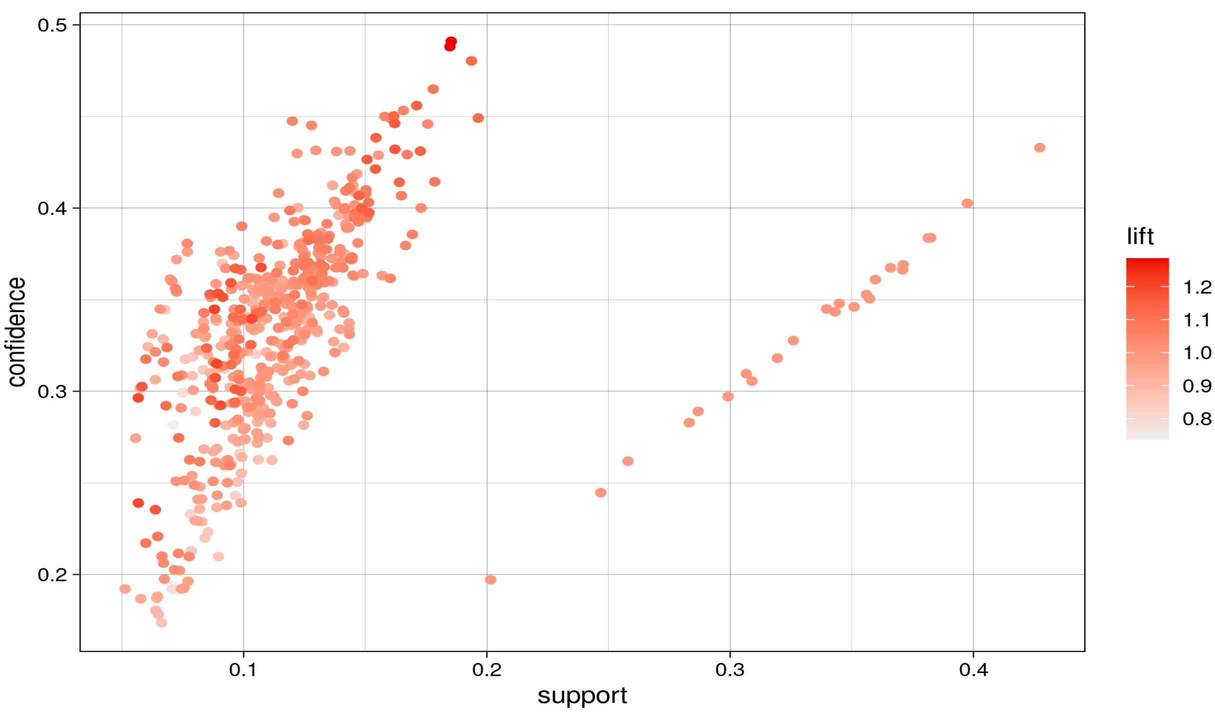

Figure 4.4: Scatter plot of Association Rules Statistics

Now that the trends of the association rules have been established it is time to analyze the rules themselves. Of the 528 association rules, there are 224 that are significant based on having a lift value greater than 1. This means that the occurrence of one item increases the likelihood of the occurrence of another item rather than the occurrence being attributed to chance. It should be noted that half are inverse relationships with only slightly different confidence values. This is because there is an imbalance in the number of each risk level in each sector, inversing the rules swaps the frequency of the precedent and antecedent, changing the confidence values.

There are 23 unique and significant high-risk-to-high-risk associations in the dataset with support ranging from 0.06 to 0.1 and confidence between 0.21 and 0.37. The employment sector has positive high-risk-to-high-risk associations with household factors (0.1, 0.29), RME (0.08, 0.25), and housing type (0.09, 0.29). The demographics sector has positive high-risk-to-high-risk associations with RME (0.06, 0.3), RMP (0.07, 0.33), household factors (0.06, 0.3), housing costs (0.06, 0.32), and housing quality (0.06, 0.31). There are only two census tracts with a high risk across each of these sectors, one in Utah and one in Wisconsin. The housing Cost sector has positive high-risk-to-high-risk associations with housing type (0.1, 0.34), housing quality (0.1, 0.32), RMP (0.1, 0.32), household factors (0.09, 0.3), RME (0.08, 0.25), and demographics (0.06, 0.21). The housing quality sector has positive high-risk-to-high-risk associations with housing costs (0.1, 0.35), RMP (0.09, 0.31), housing type (0.09, 0.3), household factors (0.08, 0.29), RME (0.08, 0.26), and demographics (0.06, 0.22). The RMP sector has positive high-risk-to-high-risk associations with housing costs (0.1, 0.33), household factors (0.09, 0.31), housing quality (0.09, 0.3), housing type (0.09, 0.3), RME (0.09, 0.29), and demographics (0.07. 0.22). The RME sector has positive high-risk-to-high-risk associations with RMP (0.09, 0.35), employment (0.08, 0.33), housing costs (0.08, 0.31), housing quality (0.08, 0.31), household factors (0.07, 0.29), household type (0.07, 0.29), and demographics (0.06, 0.24). Household factors have positive high-risk-to-high-risk associations with employment (0.1, 0.37), housing cost (0.09), RMP (0.09, 0.35), housing type (0.09, 0.35), housing quality (0.08, 0.32), RME (0.07, 0.27), and demographics (0.06, 0.23).

There are 32 unique and significant low-risk-to-low-risk associations in the dataset with support ranging from 0.13 to 0.19 and confidence between 0.39 and 0.49. The employment sector has positive high-risk-to-high-risk associations with housing quality (0.14, 0.4), household factors (0.15, 0.4), housing type (0.15, 0.41), RMP (0.15, 0.41), and RME (0.17, 0.45). The demographics sector has positive high-risk-to-high-risk associations with housing type (0.14, 0.38) and housing costs (0.13, 0.36). The housing costs sector has positive high-risk-to-high-risk associations with demographics (0.13, 0.39) and housing quality (0.13, 0.39). The housing quality sector has positive high-risk-to-high-risk associations with RME (0.15, 0.40), employment (0.14, 0.40), household factors (0.14, 0.39), RMP (0.14, 0.39), and housing costs (0.13, 0.36). The RME sector has positive high-risk-to-high-risk associations with RMP (0.17, 0.43), employment (0.17, 0.41), household factors (0.16, 0.41), and housing quality (0.15, 0.37). The RMP sector has positive high-risk-to-high-risk associations with household factors (0.19, 0.49), RMP (0.17, 0.46), housing type (0.15, 0.40), employment (0.15, 0.39), and housing quality (0.14, 0.37). The household factor sector has positive high-risk-to-high-risk associations with RMP (0.19, 0.49), housing type (0.16, 0.43), RME (0.16, 0.43), employment (0.15, 0.38), and housing quality (0.14, 0.37). The housing type sector has positive high-risk-to-high-risk associations with household factors (0.16, 0.45), RMP (0.15, 0.41), employment (0.15, 0.40), and demographics (0.14, 0.38).

there are 13 unique and significant low-risk-to-high-risk associations in the dataset with support ranging from 0.7 to 0.13 and confidence between 0.2 and 0.34. The employment sector has positive low-risk-to-high-risk associations with demographics (0.08, 0.21). The housing costs sector has positive low-risk-to-high-risk associations with employment (0.13, 0.34), RMP (0.12, 0.32), and RME (0.10, 0.26). The housing cost sector has positive low-risk-to-high-risk associations with employment (0.12, 0.35) and housing type (0.10, 0.29). The housing quality sector has positive low-risk-to-high-risk associations with housing type (0.11, 0.29) and household factors (0.10, 0.26). The RME sector has positive low-risk-to-high-risk associations with housing type (0.12, 0.29). RMP and household factors have no low-risk-to-high-risk associations. The housing quality sector has positive low-risk-to-high-risk associations with housing quality (0.11, 0.30), and RMP (0.09, 0.25).

there are 12 unique and significant high-risk-to-low-risk associations in the dataset with support ranging from 0.7 to 0.13 and confidence between 0.2 and 0.33. The employment sector has positive high-risk-to-low-risk associations with demographics (0.13, 0.38) and housing costs (0.12, 0.37). The demographics sector has positive high-risk-to-low-risk associations with employment (0.8, 0.38). The housing costs sector has no high-risk-to-low-risk associations. The housing quality sector has positive high-risk-to-low-risk associations with housing type (0.11, 0.38). The RMP sector has positive high-risk-to-low-risk associations with demographics (0.10, 0.39) and housing type (0.09, 0.37). The RMP sector has positive high-risk-to-low-risk associations with demographics (0.12, 0.4). The household factors sector has positive high-risk-to-low-risk associations with housing quality (0.1, 0.37). The housing type sector has positive high-risk-to-low-risk associations with RME (0.12, 0.41) and Housing Quality (0.11, 0.37).

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 the variables and risk levels. Moran's I is calculated for every state and the entire dataset. Table 4.13 shows the descriptive statistics for the significant Moran's I values. Table 4.14 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 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. Nationally, there are seven variables with notable statistically significant Moran's I values. These include the white population (0.66), American Indian and Native Alaskan (0.61), the catch-all ag_for_fish_hunt_mining variable (0.61), owners of mobile homes (0.56), individuals living in the same house with less than a high school education (0.57), owners of single-unit homes (0.55) and the "other" demographic variable (0.54). Two crucial variables, renters and owners of unconventional housing show almost no spatial autocorrelation at 0.04 and 0.08 respectively. Most of the variables with average Moran's I scores less than 0.1 are in the residential mobility sectors. The nationwide global spatial autocorrelation scores for the sector variables range from a low (0.17) to a medium strength spatial autocorrelation (0.35) with demographic risk levels being the most spatially clustered and housing costs being the least spatially clustered.

Table 4.13: Moran's I Descriptive Statistics

Statistic	N	Mean	SD	Min	Max
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

Table 4.14: Average Moran's I by Sector

sector	Morans I	SD	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
RME	0.23	4.08	0.00	-0.01	0.01
RMP	0.20	3.54	0.00	-0.01	0.01
RUCA	0.31	5.25	0.00	-0.01	0.00

4.4.1 *Moran's I Outliers*

Outliers based on the interquartile range (IRQ) method are calculated for the calculated Moran's I statistics to highlight areas that do not follow the overall trends in the data set. 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. Figure 4.6 shows the distribution of Moran's I for each sector. Demographics and household factors do not have any outliers based on the IRQ method. RMP has 13 outliers. The mean of all RMP 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 RMP risk levels variable. The average for these three states is 0.45 compared to 0.21 for the same variable overall. There are 4 outliers in the residential RME 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: 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 ag_for_fish_hunt_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 4.5 shows the distribution of Moran's I for each state by region.

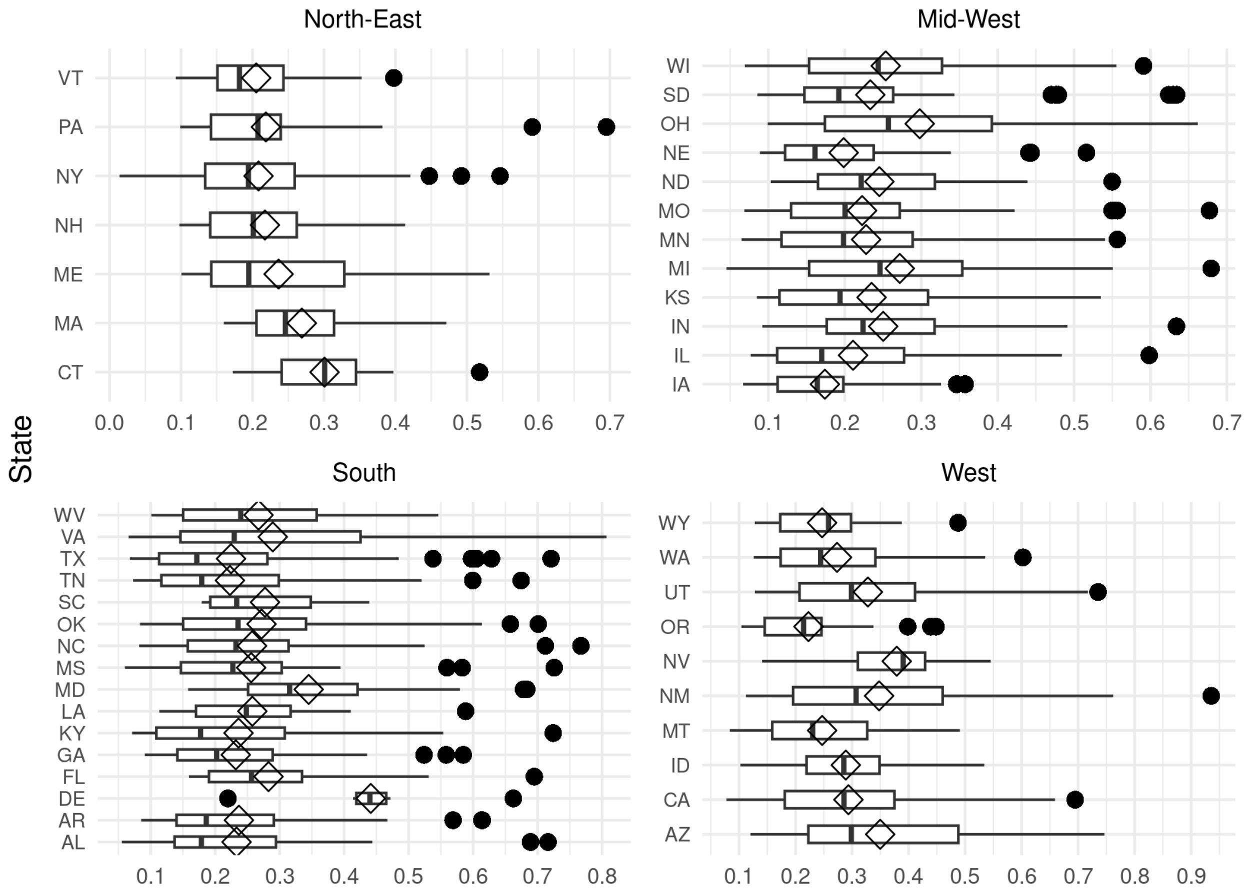

Figure 4.5: Boxplot of Moran's I by Region

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. National models using in-sample evaluation are used to measure how well a census tract's risk levels can be predicted.

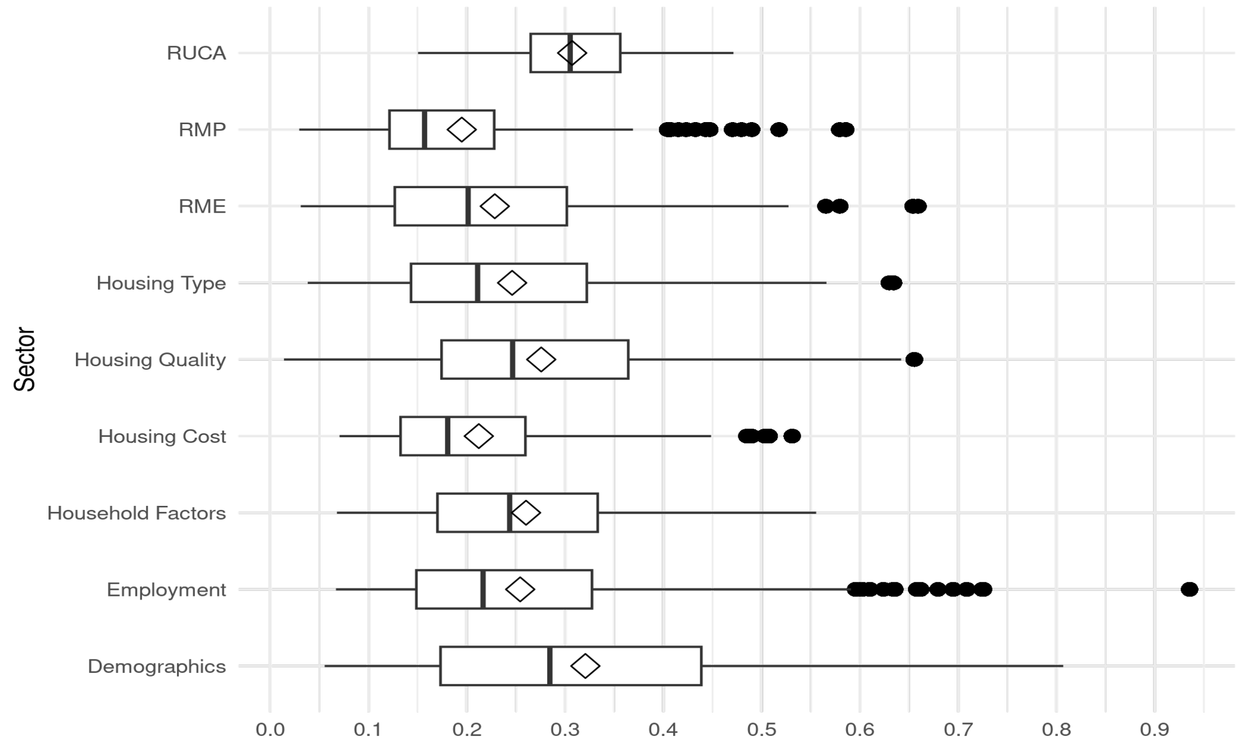

Figure 4.6: Boxplot of Moran's I by Sector

4.5.1 *Probability*

The average probability for all sectors was low as demonstrated by Figure 4.7. employment, housing quality, RMP and household factors had an average probability of 34 percent; housing type and housing cost had an average of 35 percent; RMP had an average of 36 percent; demographics had the highest average probability at 0.38. Demographics also had the highest standard deviation at 14 percent, indicating a high degree of variation in predictability. Utah had the best prediction results with an average of 41 percent, and Minnesota was the hardest to predict at 31 percent across 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 4.9 Shows the distribution of average probability for each state. With an average of 0.35, no states performed well across sectors. One last area of interest is any trends that may exist between the probabilities for each sector. Figure 4.8 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.

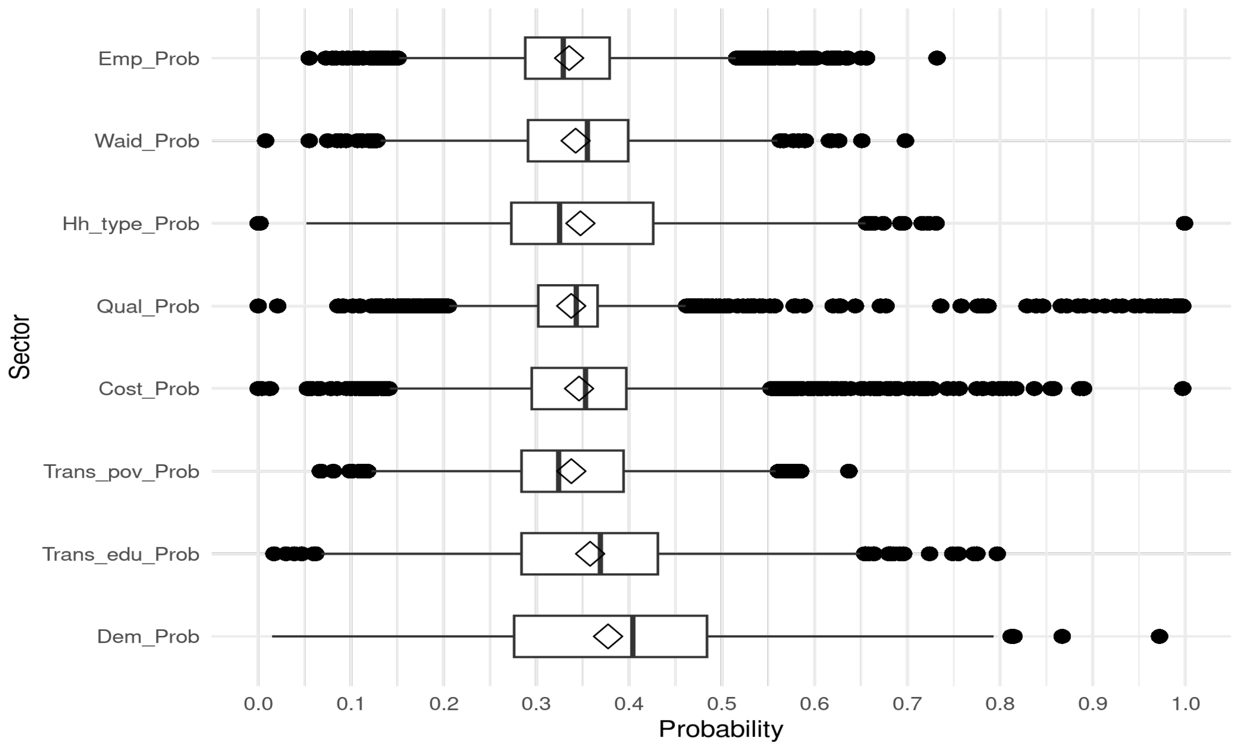

Figure 4.7: Boxplot of Probability by sector

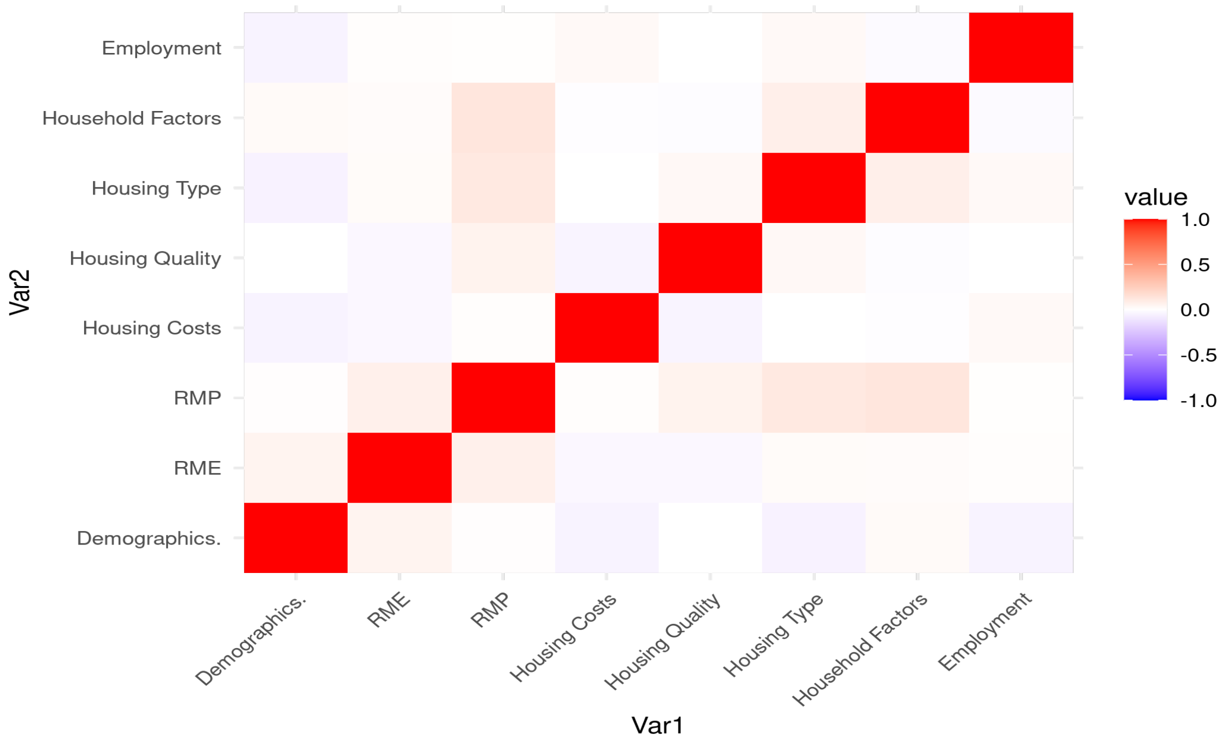

Figure 4.8: Correlation Plot of Sector Probabilities

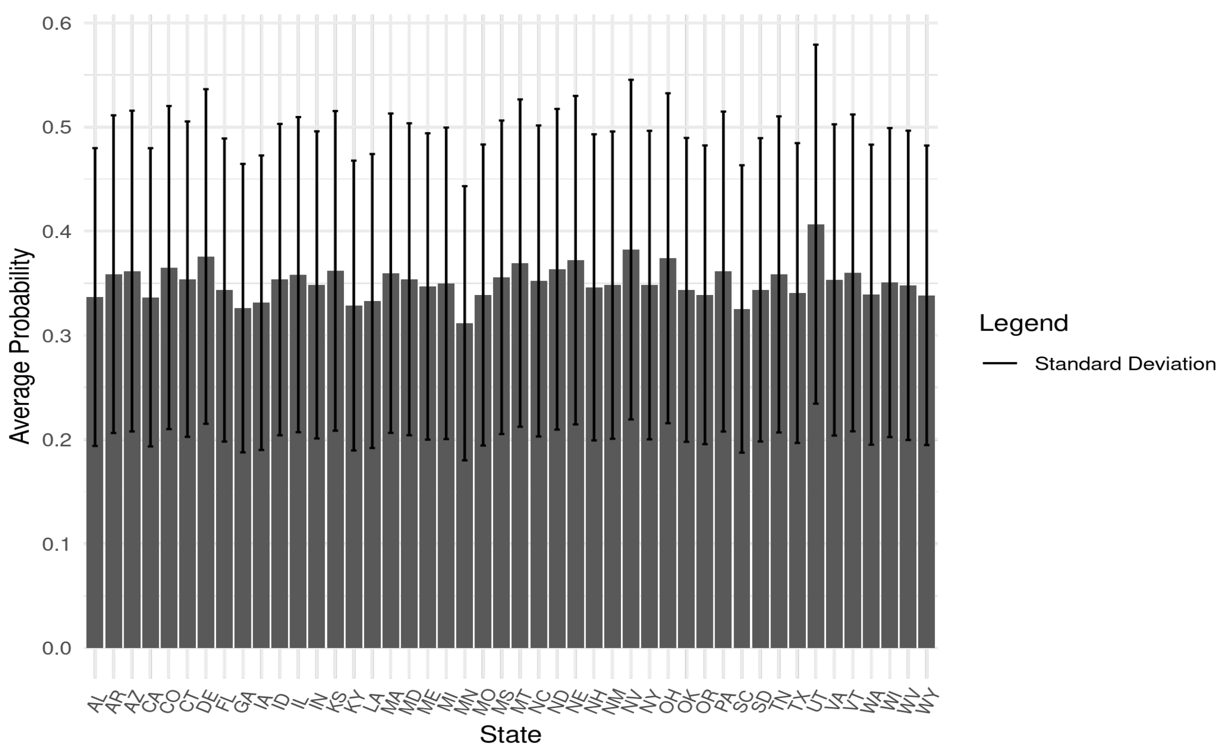

Figure 4.9: Bar Graph of Average Probabilities by State with Error Bars

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 harming 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 4.15 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. Overall, the state models were 34 percent accurate, and the national model was 41 percent accurate.

Table 4.15: 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 4.16 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 4.17 shows that the housing cost models struggled to classify all census tracts. They also struggled to differentiate between medium-risk levels and high-risk-level census tracts. The state models accurately predicted 34 percent of high-risk level census tracts, 44

Table 4.16: 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

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 4.17: 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 4.18 shows that the housing quality models significantly over-classified census tracts as low-risk levels. 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 4.18: 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

Table 4.19 shows that the RME models significantly over-classified census tracts as low-risk levels. They successfully predicted 14 percent of low-risk census tracts, 39 percent of medium-risk level census tracts, and 60 percent of low-risk census tracts. The state models had an accuracy of 46 percent, and the national model had an accuracy of 42 percent.

Table 4.19: 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 4.20 shows that the RMP models significantly over-classified census tracts as low-risk levels. 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 4.20: 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 4.21 shows that the household factor models significantly over-classified census tracts as low-risk levels. 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 4.21: 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 4.22 shows that the housing type models significantly over-classified census tracts as low-risk levels. They successfully predicted 54 percent of low-risk level census tracts, 33 percent of medium-risk level census tracts, and 11 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 4.22: 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

4.6 *Rurality and Risk Levels*

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

Table 4.23: Local Morans I Risk-Level Results

sector	Cluster 1	Cluster 2	Cluster 3
Employment	0.83	0.00	0.89
Demographics	2.09	0.14	0.58
RME	1.28	0.10	0.25
RMP	1.16	0.05	0.48
Housing costs	0.79	-0.00	0.57
Housing Quality	1.06	0.00	0.74
Housing Type	0.91	-0.00	0.68
Household Factors	1.52	0.09	0.22

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. Table 4.24 shows the percentage of each RUCA code that has a high-risk level for each sector. Figure 4.10 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 4.24: High-Risk Census Tract RUCA Breakdown

sector	RUCA	Pct
Housing Quality	10	0.5
Employment	10	0.49
Demographics	10	0.49
Housing Costs	10	0.49
Household Factors	10	0.49
RMP	10	0.48
RME	10	0.46
Household Type	10	0.43
RME	7	0.39
Household Type	7	0.37
Household Factors	7	0.35
Employment	7	0.34
RMP	7	0.34
Housing Costs	7	0.33
Housing Quality	7	0.33
Demographics	7	0.32
Household Type	8	0.15
Demographics	8	0.14
RMP	8	0.13
Housing Costs	8	0.13
Housing Quality	8	0.13
Employment	8	0.12
Household Factors	8	0.11
RME	8	0.1
Employment	9	0.05
Demographics	9	0.05
RMP	9	0.05
Housing Costs	9	0.05
Housing Quality	9	0.05
Household Type	9	0.05
RME	9	0.04
Household Factors	9	0.04

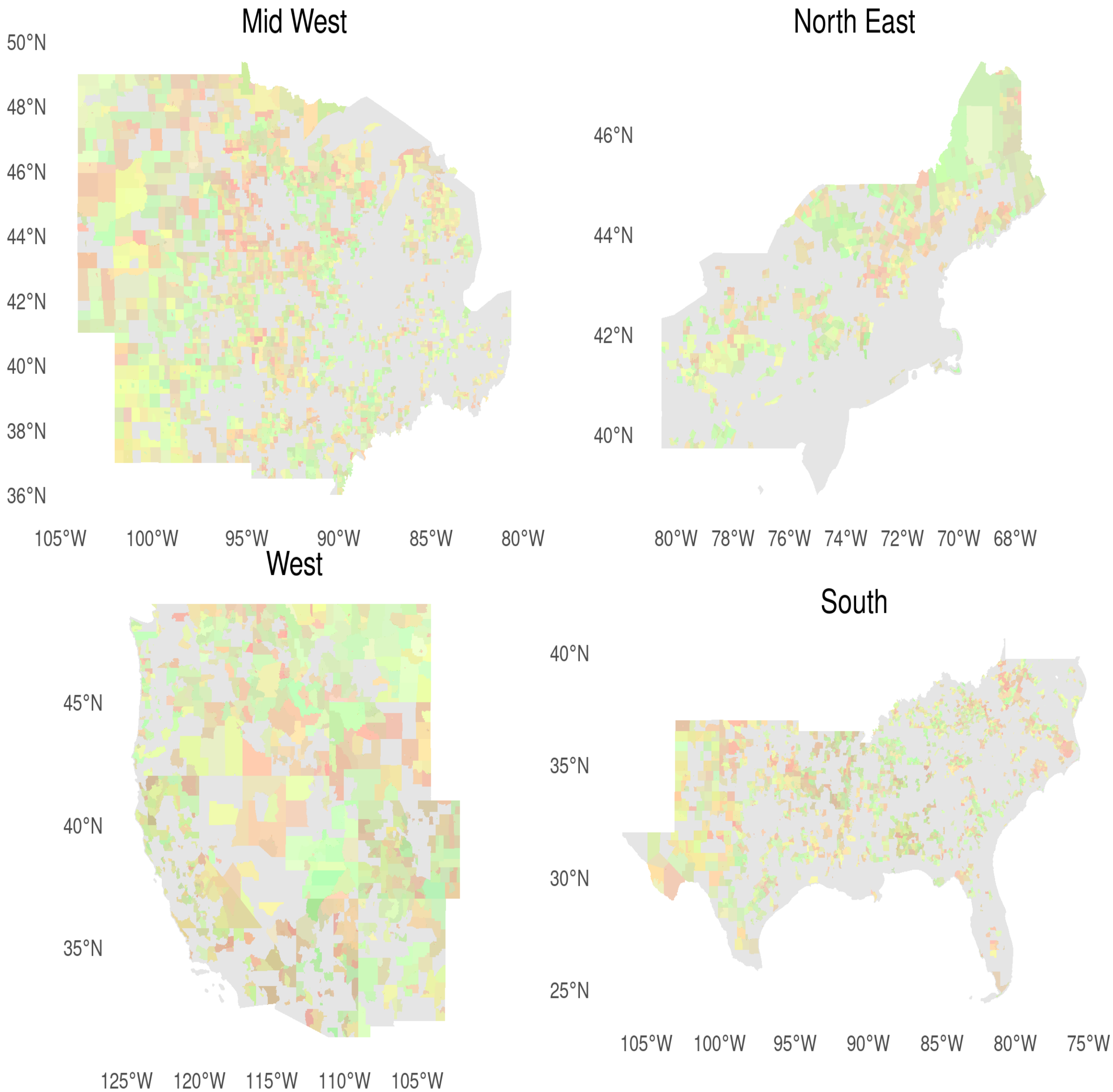

Figure 4.10: Risk Level Across Sectors

The census tract risk threshold results in 280 census tracts (four percent) labeled as high-risk, 1,692 (27 percent) labeled as medium-risk, and 4,361 (69 percent) labeled as low-risk based on the sum of their risk level variables. Figure 4.11 highlights the high-risk areas in red, and the medium-risk levels in yellow. The 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. There is a significant clustering of high-risk areas in Arizona. Outside of Arizona, there is little clustering of high-risk level census tracts with some clustering of medium-risk level census tracts. It should be noted that a standard t-test found no statistically significant differences in variable averages between high and low-risk census tracts 4.25.

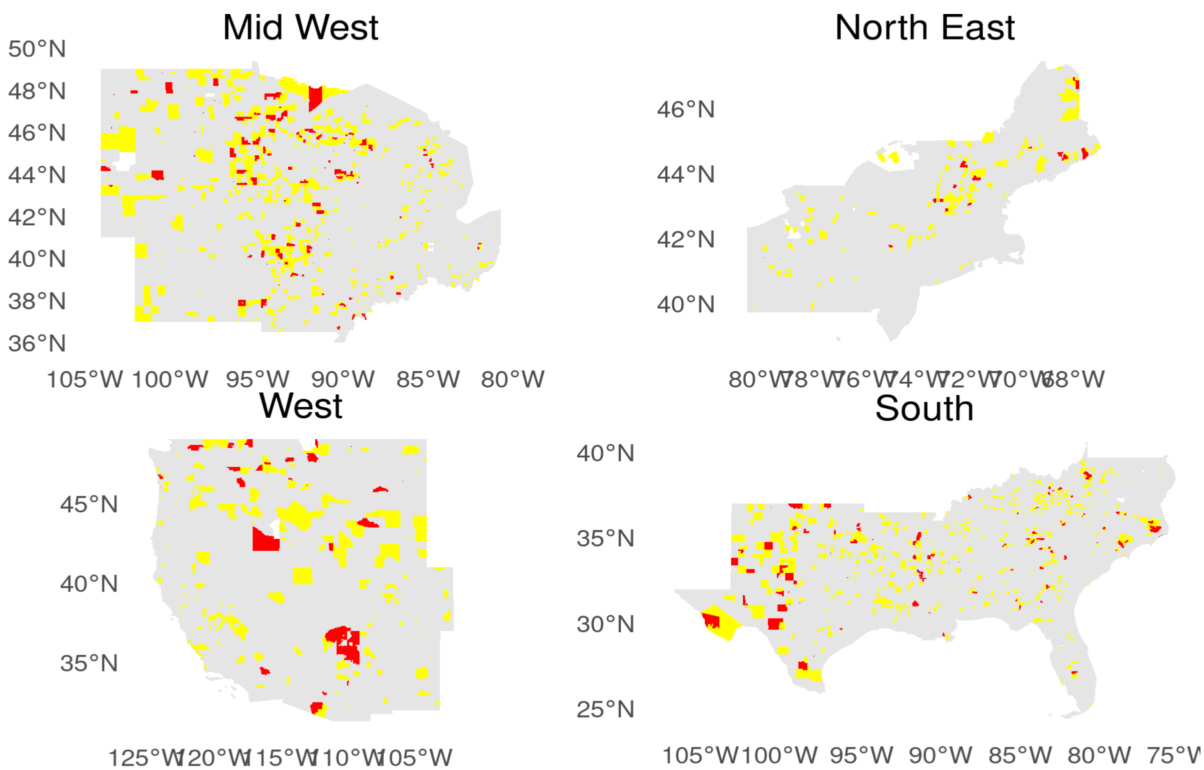

Figure 4.11: High and Medium Risk Census Tracts

Table 4.25: High and Low-Risk Census Tract t-Test

	t
Test statistic	0.01159873
DF	185.95
p value	0.9907582
Alternative hypothesis	two sided

Welch Two Sample t-test: High Risk Census Tracts and Low Risk Census Tracts

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 a lack of uniform definitions, persistent poverty, and hardships from economic changes (Yousey and Samudra, 2018; Crandall and Weber, 2004; Pendall et al., 2016; Kropczynski and Dyk, 2012). 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 policymakers. 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. The results show a notable amount of census tracts at a high or medium risk of housing insecurity and provide evidence that there is great variation in the housing insecurity risk in rural areas at the census tract level.

5.1 *Identifying and Analyzing Risk Levels*

Risk factors of housing insecurity were used to identify levels of housing insecurity risk while accounting for the variation in rural areas with a combination of unsupervised machine learning and a spatial neighbors algorithm. k-medoid clustering was used to cluster census tracts with similar values across eight 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 policymakers and researchers can use in their states. 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.

Previous research has identified both pockets of "prosperous" rural areas and rural areas that are considered pockets of poverty (Isserman et al., 2009; Miller and Weber, 2003). A similar process is used whereby a threshold was established to differentiate between high, medium, and low-risk census tracts in terms of housing insecurity across sectors using supervised machine learning. The cluster analysis highlighted 280 census tracts identified as high-risk and 1,692 census tracts identified as medium-risk (see Figure 4.11). These are areas of concern as they all have an increased amount of high and medium risk levels across sectors relative to other census tracts in their state. These census tracts show consistent signs of housing insecurity relative to other census tracts in their state.

The cluster analysis shows several important observations about rural areas. First is the importance of education, health, and social work employment in rural areas (See Table 4.1). The cluster medians for each risk level is nine 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 affected by the decreases in funding caused by several processes affecting rural areas (Blank, 2005). 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 (Robertson et al., 2007). Public administration is the most stable sector across clusters with values remaining almost identical across risk levels. One unexpected result is the significant presence of retail trade, which has higher cluster medians than most of the job sectors considered "rural". The clustering results provide further evidence of the shift away from traditional rural economies that scholars have identified (Pendall et al., 2016; Blank, 2005).

The demographic variable cluster results align with 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. Concentrations of minorities are espectially 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 the 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 (Lichter and Johnson, 2020). The clustering results also provide further evidence of the aging population of rural areas, with the number of males and females over 18 being almost three times higher than the number of males and females under 18 across all three risk levels. This reflects the trend of rural areas aging faster than urban areas (Cohen and Greaney, 2022).

For housing costs, the cluster medians indicate that while home ownership is widespread in rural areas, there are generally more high-cost renters than owners. The cluster medians for high-cost renters range from 14 to 17 percent while homeowners with a mortgage and high housing costs range from 4 to 5 percent. 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. High housing costs can create a vicious cycle where low-income households must move frequently, often to worse neighborhoods with decreasing housing conditions (Desmond et al., 2015). Segregation in rural areas is a concern for homeownership as well, Krivo and Kaufman (2004) found that Hispanics and African Americans face discrimination in the housing market that suppresses the accumulation of home equity. It must be remembered that the ability to access affordable housing is largely determined by demographic characteristics which in turn are influenced by historical forces such as discrimination (Yadavalli et al., 2020, Hernández and Swope, 2019).

The housing quality cluster analysis does not show any concerning results. Between 17 and 25 percent of unoccupied housing have incomplete kitchens or incomplete plumbing based on the cluster medians. While it is expected that some housing will not be habitable due to a lack of maintenance, it is unclear if these numbers are high or to be expected. Occupied houses with these problems are much less common with all cluster medians less than one percent.

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 cluster medians between 7 to 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 with low levels of high school graduation require further attention. Rural areas are particularly vulnerable to these problems in education in areas facing population loss and economic problems because they are left with a lower tax base and less funding for schools, potentially increasing the likelihood of higher dropout rates (Blank, 2005).

The most notable observation from the RMP cluster analysis is that the number of households that live below the poverty level is higher than those slightly above it across all three risk levels. This is concerning due to the well-documented impacts of poverty. At the extreme end, severe poverty can lead to literal homelessness and households experiencing poverty often face housing insecurity (Evans et al., 2019; Cox et al., 2019).

The cluster analysis shows three 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 a significant number of households in rural areas are not building wealth through means outside of wages received from employment. This is concerning given the rise of economic insecurity (Desmond and Gershenson, 2016). Second, a median of five 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 (Lichter and Johnson, 2007). Third, the median Gini index for each risk level is between 0.42 and 0.44, whereas the national Gini index is 0.488 in 2022 (US Census Bureau, 2023). Although income inequality is often considered an urban problem, there is only slightly less income inequality in rural areas compared to the nation.

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 (MacTavish et al., 2006). For the high-risk housing type census tracts, 10 percent of owners and renters each live in mobile homes based on the cluster medians. These areas where there are higher rates of renters relative to owners than other areas warrant further attention because there may be some unidentified factors in the community that contribute to the lower amounts of home ownership.

The cluster analysis of rural areas reveals several significant findings. Education, health, and social work employment emerge as vital sectors, contrasting with declining traditional rural industries like agriculture and manufacturing. Demographically, clusters with high minority populations highlight historical processes impacting housing insecurity, with concentrations of African Americans and Hispanics in particular areas. Housing affordability challenges persist, with high-cost renters more prevalent than owners, reflecting urban-like housing dynamics. Maintenance issues in unoccupied housing suggest ongoing challenges in housing quality. Additionally, low high school graduation rates and high poverty levels underscore economic insecurity in rural communities. Limited wealth-building opportunities and relatively low public assistance usage further contribute to rural economic vulnerability. The presence of mobile homes presents unique challenges, especially in high-risk areas. Overall, these findings underscore the complexity of rural housing dynamics, requiring multifaceted solutions to address economic, demographic, and housing quality challenges.

5.2 *Patterns of Risk*

Under the 4 Cs model, there is an implicit assumption that areas at a high level of risk in one sector may have a higher level of risk in another sector because the pillars are an interconnected web. The association rules analysis reveals several interesting findings about the co-occurrence of housing insecurity risk. First, An area at an overall high risk of housing insecurity would at least have a high-risk level across more than one sector. Tables 4.9, 4.10, 4.11, 4.12 show the frequency of the most general trends of the association rules: 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 similar levels of presence between these relationships. The trends in association rules indicate that few areas exhibit a risk of housing insecurity across multiple sectors. This is reflected by Figure 4.10 where the map of housing insecurity shows some pockets of red, but vast amounts of green and yellow indicating a 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. The most interesting finding from the risk analysis is the high presence of high-risk census tracts in the most rural areas. 40 to 50 percent of high-risk census tracts are classified as the most rural across all sectors. Second to the most rural areas are small towns, with 30 to 40 percent of high-risk census tracts across all sectors. Only four to 15 percent of high-risk census tracts are classified as an eight or a nine on the Rural-Urban Commuting Area Codes scale. These results show that the most rural areas are at the highest risk of housing insecurity, next to the most urban areas considered in this dataset.

As the goal of this research is not just to provide an exploratory analysis of rural housing insecurity but also to serve as a tool for further study, it's important to highlight areas that have a high risk of housing insecurity. The first way this was accomplished was through the risk threshold of 12 and 15 out of 24. As these are ultimately arbitrary metrics, this system is not meant to make any definitive claims but rather to highlight areas of interest. The 208 high-risk census tracts and 1,692 census tracts provide a starting point for further analyzing housing insecurity. There are several important observations to be made from this thresholding method. First, high and medium-risk census tracts have higher African American and Hispanic/Latino populations as well as smaller white populations. There are more people with a high school diploma and fewer people below the poverty line in low-risk census tracts. High and medium-risk census tracts have similar levels of high-cost mortgage and high-cost renter households but different levels of high-cost households without a mortgage. There is slightly greater usage of public assistance and supplemental security index usage in high and medium-risk census tracts.

A second way that this research serves as a tool for researchers is the breakdown of high-risk-to-high-risk association rules by state found in Appendix B. This provides a succinct overview of the presence of census tracts found to be at a high risk of housing insecurity and their relationships to different risk levels.

5.3 ***Spatial Aspects 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 identified pockets of minorities in rural areas (Lichter and Johnson, 2020). 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 some rural economies and further enforces the role of single industry-based economies in rural areas. Economic diversity is 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 (Blank, 2005). The low average Moran's I for residential mobility at the national level indicates that there is not significant spatial clustering of residential mobility for high-risk households. One last observation is the fairly high spatial autocorrelation for individuals in the same house with less than a high school education (0.56). This further reinforces the potential issues facing rural areas when it comes to education.

The most significant finding of the spatial analysis is the results of local Moran's I on the sector risk level variables. Table 4.23 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 relatively high local spatial autocorrelation for demographics. This follows the trend of results indicating the existence of pockets of minority populations in rural areas.

The spatial analysis conducted in this study yielded several significant findings. Global Moran's I calculations revealed strong spatial autocorrelations for key demographic variables nationwide, highlighting the presence of minority populations in rural areas. Despite declines in manufacturing and agriculture's share of rural economies, these industries still exhibit significant spatial autocorrelation, underscoring the persistence of single-industry-based economies in rural regions. Concerns arise from the high autocorrelation of households lacking high school diplomas, possibly indicating educational challenges in areas facing rural decline. Additionally, low residential mobility autocorrelation suggests limited spatial clustering of high-risk households' movements nationally. Local Moran's I analysis examined the relationship between risk levels, showing negligible spatial autocorrelation among medium-risk level census tracts across sectors but significant spatial autocorrelation between high and low-risk census tracts. These findings collectively underscore the multifaceted nature of spatial patterns in rural areas, highlighting the need for targeted interventions to address economic, demographic, and educational disparities.

The spatial analysis conducted in this study has shed light on significant patterns of spatial autocorrelation at both national and local levels. The presence of strong spatial autocorrelations for key demographic variables underscores the prevalence of minority populations in rural areas, while the persistence of spatial autocorrelation in industries such as manufacturing and agriculture highlights the enduring influence of single-industry-based economies in rural regions. Concerns about educational challenges in areas facing rural decline were also raised, particularly evident in the high autocorrelation of households lacking high school diplomas. Furthermore, the limited spatial clustering of high-risk households' movements nationally suggests a nuanced understanding of rural migration patterns. Particularly noteworthy are the findings from local Moran's I analysis, which revealed significant spatial autocorrelation between high and low-risk census tracts across sectors, underscoring the complex spatial dynamics at play in rural areas. These findings emphasize the importance of targeted interventions to address economic, demographic, and educational disparities in rural communities.

5.4 *The Predictability of Housing Insecurity Risks*

An important question in the study of housing insecurity is the extent to which the risk level of a census tract can be predicted. This provides insights into the variation of rural areas identified in the literature (Cromartie and Bucholtz, 2008). MLR regression models were used to predict the risk levels of each sector for each state and nationally. The results show 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. Class imbalance in multinomial logistic regression occurs when there are unequal proportions of observations across the different outcome categories. 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 (Cromartie and Bucholtz, 2008).

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. The second limitation is due to 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 et al., 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.

CHAPTER 6
CONCLUSION

This chapter presents a synthesis of the work discussed in this thesis on "Rurality and Robustness: Rural Communities and Housing Insecurity Risk". A summary of the preceding chapters highlights how the analysis of rural housing insecurity was systematically developed. Section 6.2 presents the three major conclusions related to the housing insecurity risk assignment system developed, the connectivity of housing insecurity risk, and future research possibilities.

6.1 *Summary*

The goal of this thesis is to establish a baseline for further study into rural housing insecurity. Scholars must validate this exploration's findings before we can begin to understand rural housing insecurity and homelessness. The exploratory analysis was performed using supervised and unsupervised machine learning, and spatial analysis techniques. This is an expansion of previous research into rural areas that has used similar threshold measurements to divide places into prosperous or high-poverty places. Rural areas vary greatly between and within themselves, so a methodology that can address housing insecurity in a relative rather than absolute way is necessary and important to give rural areas the attention they deserve.

Chapter 2 builds up the 4 Cs of housing insecurity framework, describes the nuance of housing insecurity as an alternative to the housed and unhoused binary, presents some of the challenges faced by rural areas, and details the development and fundamentals of the applied methods. It demonstrates the significant overlap between each pillar of housing insecurity. Each pillar of housing insecurity reflects a vital and interrelated part of housing insecurity. Housing costs need to be affordable to have consistency. Housing conditions play a vital role in health and households may be driven to housing with worse conditions by residential instability. The context in which one lives defines much of their opportunity in life and has a major impact on their housing insecurity risk. As the rural challenges section demonstrates, rural areas are far more than simply "not-urban" as they are often defined. These include problems with poverty, geographic isolation, a growing economic divide between urban and rural places, and racial segregation. Rural areas encompass a wide range of people, social structures, and communities. Necessary to any study of housing insecurity or homelessness is the need to acknowledge the historical factors influencing the struggles of people, especially marginalized populations, to understand the underlying processes.

Chapter 3 presents a method to apply the 4 Cs of housing insecurity to rural areas using machine learning and spatial analysis techniques. First is generating risk levels relative to other census tracts on a state-by-state basis and then analyzing those clusters to identify the clusters with a high, medium, and low risk of housing insecurity relative to each other. Association rule learning is then used to analyze common relationships between risk levels and pillars of the 4 Cs. Then, local and global Moran's I are used to improve our understanding of the spatial nature of housing insecurity in rural areas. Finally, multinomial logistic regression is used to discover the predictability of risk levels.

Chapter 4 presents and examines the results of the risk level assignment system. 12 percent of census tracts are identified as having a medium or high risk of housing insecurity based on a threshold of risk levels. Each census tract has a maximum risk of 8 and a minimum risk of 24. They are considered a high risk if their total is less than or equal to 12 and a medium risk if it is between 12 and 15. All other census tracts are considered at a low risk of housing insecurity. These thresholds were adopted to highlight the areas with the most urgent housing insecurity levels. The most important insights come from the cluster analysis and spatial autocorrelation results. The cluster analysis highlighted segregation in rural areas, the importance of employment in education, health, and social work in rural areas, a notable presence of high-cost renters, and that there is a concerning number of individuals who did not move but either did not have a high school diploma or are below the poverty level. The association rules show that there is an average probability of about 30 percent that a census tract with a high risk in one sector has a high risk in another sector. The global Moran's I analysis shows that spatial autocorrelations across variables are generally low at the state and national level with outliers that future work should investigate.

Chapter 5 synthesizes the results of the analysis with the literature, highlights areas where the results align with the literature, and points out interesting observations that future research should consider. Rural areas may share some of the same issues as urban areas when it comes to rural housing insecurity. One example of this is a significant number of rent-burdened renters, another is the high levels of racial segregation especially between Whites and African Americans. The association rules show that while there are about 10 percent of census tracts with a high-risk level in at least 2 sectors, there are a similar number of inverse high-risk-to-low-risk and low-risk-to-high-risk relations. The spatial analysis reveals that unlike what is generally expected in rural areas, the variables considered here had generally low spatial autocorrelations except for the 7 variables: the white population, American Indian and Native Alaskan, the catch-all agriculture, forestry, fishing, hunting, and mining variable, owners of mobile homes, individuals living in the same house with less than a high school education, owners of single-unit homes and the "other" demographic variable. While the variables themselves were not highly spatially autocorrelated, local Moran's I reveals that there is notable spatial autocorrelation of high and medium-risk census tracts. The multinomial logistic regression shows that rural areas are highly unpredictable, with both national and state-by-state models performing poorly.

6.2 *Conclusions*

Housing insecurity is difficult in several different ways. First, it is difficult to define. Until there is a greater 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 thesis helps to bridge this gap with its three major conclusions. First, it identifies 280 census tracts that show significant signs of housing insecurity and 1,692 census tracts that show a medium risk of housing insecurity for researchers and policymakers to address. Second, it serves as a starting point for researchers to investigate housing insecurity at a smaller level to refine and improve this methodology. Third and most importantly, it provides significant evidence that housing insecurity may not experience the same clustering in rural areas as it does in urban areas. This initial exploration hopes to serve as a starting point for policymakers and researchers to begin deconstructing the urban-centric lens and give those in rural populations the attention and resources they need and deserve.

6.3 *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 this 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 I 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. Researchers should also experiment with applying different clustering algorithms such as hierarchical clustering which would allow for the number of risk levels derived from the data rather than explicitly chosen. 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.

REFERENCES

Robert P. Agans, Malcolm T. Jefferson, James M. Bowling, Donglin Zeng, Jenny Yang, and Mark Silverbush. Enumerating the Hidden Homeless: Strategies to Estimate the Homeless Gone Missing From a Point-in-Time Count. Journal of Official Statistics, 30 (2):215–229, June 2014. ISSN 2001-7367. doi: 10.2478/jos-2014-0014. URL https://www.sciendo.com/article/10.2478/jos-2014-0014.

Rakesh Agrawal and Ramakrishnan Srikant. Fast Algorithms for Mining Association Rules. In Proceedings of the 20th VLDB Conferance, Santiago, Chile, 1994.

Rakesh Agrawal, Christos Faloutsos, and Arun Swami. Efficient similarity search in sequence databases. In Gerhard Goos, Juris Hartmanis, and David B. Lomet, editors, Foundations of Data Organization and Algorithms, volume 730, pages 69–84. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993. ISBN 978-3-540-57301-2 978-3-540-48047-1. doi: 10.1007/3-540-57301-1_5. URL http://link.springer.com/10.1007/3-540-57301-1_5. Series Title: Lecture Notes in Computer Science.

Mohiuddin Ahmed, Raihan Seraj, and Syed Mohammed Shamsul Islam. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. Electronics, 9(8): 1295, August 2020. ISSN 2079-9292. doi: 10.3390/electronics9081295. URL https://www.mdpi.com/2079-9292/9/8/1295.

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.

Stefan Angel and Benjamin Bittschi. Housing and Health. SSRN Electronic Journal, 2014. ISSN 1556-5068. doi: 10.2139/ssrn.2518364. URL http://www.ssrn.com/abstract=2518364.

Ali Asgary. What Does Spatial Autocorrelation Tell Us about the COVID-19 Spread in the USA?, March 2020. URL https://medium.com/@ali.asgary/what-does-spatial-autocorrelation-tell-us-about-the-covid-19-spread-in-the-usa-233b5d

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

Geoffrey H Ball and David J Hall. ISODATA, A NOVEL METHOD OF DATA ANALYSIS AND PATTERN CLASSIFICATION. Technical report, Stanford Research Institute, Menlo Park, California, 1965.

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.

François Bavaud. Models for Spatial Weights: A Systematic Look. Geographical Analysis, 30(2):153–171, April 1998. ISSN 0016-7363, 1538-4632. doi: 10.1111/j.1538-4632.1998.tb00394.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.1998.tb00394.x.

Anass Bayaga. MULTINOMIAL LOGISTIC REGRESSION: USAGE AND APPLICATION IN RISK ANALYSIS. Journal of Applied Quantitative Methods, 5(2): 288–297, 2010.

Steve Berg. The HEARTH Act. Cityscape, 15(1):317–323, 2013. ISSN 1936-007X. URL https://www.jstor.org/stable/41958974. Publisher: US Department of Housing and Urban Development.

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.

Lina Bjerke and Charlotta Mellander. Mover stayer winner loser. Cities, 130:103850, 2019. ISSN 02642751. doi: 10.1016/j.cities.2022.103850. URL https://linkinghub.elsevier.com/retrieve/pii/S026427512200289X.

Rebecca M. Blank. Poverty, Policy, and Place: How Poverty and Policies to Alleviate Poverty Are Shaped by Local Characteristics. International Regional Science Review, 28 (4):441–464, October 2005. ISSN 0160-0176, 1552-6925. doi: 10.1177/0160017605278999. URL http://journals.sagepub.com/doi/10.1177/0160017605278999.

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.

Barry Boots and Michael Tiefelsdorf. Global and local spatial autocorrelation in bounded regular tessellations. Journal of Geographical Systems, 2(4):319–348, December 2000. ISSN 1435-5930. doi: 10.1007/PL00011461. URL http://link.springer.com/10.1007/PL00011461.

Stephen Borders, Blake Ferris, Mary Jo Beeby, and Pattijean McCahill. West Michigan Food Insecurity: Using Cluster Analysis and Geographic Information Systems to Identify Needs. Progress in Community Health Partnerships: Research, Education, and Action, 12(2):187–197, 2018. ISSN 1557-055X. doi: 10.1353/cpr.2018.0039. URL https://muse.jhu.edu/article/704596.

Paula Braveman, Mercedes Dekker, Susan Egerter, Tabashir Sadegh-Nobari, and Craig Pollack. Housing and Health. Brief 7, Robert Wood Johnson Foundation, 2011.

Matthew M. Brooks. The Advantages of Comparative LISA Techniques in Spatial Inequality Research: Evidence from Poverty Change in the United States. Spatial Demography, 7(2-3):167–193, October 2019. ISSN 2364-2289, 2164-7070. doi: 10.1007/s40980-019-00052-4. URL `http://link.springer.com/10.1007/s40980-019-00052-4`.

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`.

Jeremy Bryson. The Nature of Gentrification. Geography Compass, 7(8):578–587, August 2013. ISSN 1749-8198, 1749-8198. doi: 10.1111/gec3.12056. URL `https://compass.onlinelibrary.wiley.com/doi/10.1111/gec3.12056`.

Linda M. Burton, Daniel T. Lichter, Regina S. Baker, and John M. Eason. Inequality, Family Processes, and Health in the "New" Rural America. American Behavioral Scientist, 57(8):1128–1151, August 2013. ISSN 0002-7642, 1552-3381. doi: 10.1177/0002764213487348. URL `http://journals.sagepub.com/doi/10.1177/0002764213487348`.

Amber Canto, Laura E. Brown, and Steven C. Deller. Rural Poverty, Food Access, and Public Health Outcomes. Choices, 29(2):1–5, 2014. ISSN 0886-5558. URL `https://www.jstor.org/stable/choices.29.2.04`. Publisher: Agricultural & Applied Economics Association.

Emery N. Castle. A Conceptual Framework for the Study of Rural Places. American Journal of Agricultural Economics, 80(3):621–631, 1998. ISSN 0002-9092. doi: 10.2307/1244569. URL `https://www.jstor.org/stable/1244569`. Publisher: [Agricultural & Applied Economics Association, Oxford University Press].

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.

Jiangping Chen, Yanan Chen, Jie Yu, and Zhaohui Yang. Comparisons with spatial autocorrelation and spatial association rule mining. In Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, pages 32–37, Fuzhou, China, June 2011. IEEE. ISBN 978-1-4244-8352-5. doi: 10.1109/ICSDM.2011.5969000. URL http://ieeexplore.ieee.org/document/5969000/.

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

Paul J. Cloke, Terry Marsden, and Patrick H. Mooney, editors. Handbook of rural studies. SAGE, London ; Thousand Oaks, Calif, 2006. ISBN 978-0-7619-7332-4. OCLC: ocm57639108.

Frans Coenen. Data mining: past, present and future. The Knowledge Engineering Review, 26(1):25–29, February 2011. ISSN 0269-8889, 1469-8005. doi: 10.1017/S0269888910000378. URL https://www.cambridge.org/core/product/identifier/S0269888910000378/type/journal_article.

Steven A. Cohen and Mary L. Greaney. Aging in Rural Communities. Current Epidemiology Reports, 10(1):1–16, November 2022. ISSN 2196-2995. doi: 10.1007/s40471-022-00313-9. URL https://link.springer.com/10.1007/s40471-022-00313-9.

Robynn Cox, Benjamin Henwood, Seva Rodnyansky, Eric Rice, and Suzanne Wenzel. Road Map to a Unified Measure of Housing Insecurity. Cityscape, 21(2):93–127, 2019. URL https://www.huduser.gov/portal/periodicals/cityscpe/vol21num2/ch5.pdf.

J.S. Cramer. The Origins of Logistic Regression. SSRN Electronic Journal, 2003. ISSN 1556-5068. doi: 10.2139/ssrn.360300. URL http://www.ssrn.com/abstract=360300.

Mindy S. Crandall and Bruce A. Weber. Local Social and Economic Conditions, Spatial Concentrations of Poverty, and Poverty Dynamics. American Journal of Agricultural Economics, 86(5):1276–1281, 2004. ISSN 0002-9092. URL https://www.jstor.org/stable/3697941. Publisher: [Agricultural & Applied Economics Association, Oxford University Press].

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

Katherine J. Curtis, Junho Lee, Heather A. O'Connell, and Jun Zhu. The Spatial Distribution of Poverty and the Long Reach of the Industrial Makeup of Places: New Evidence on Spatial and Temporal Regimes. Rural Sociology, 84(1):28–65, March 2019. ISSN 0036-0112, 1549-0831. doi: 10.1111/ruso.12216. URL https://onlinelibrary.wiley.com/doi/10.1111/ruso.12216.

P. De Jong, C. Sprenger, and F. Van Veen. *On Extreme Values of Moran's* I *and Geary's* c. Geographical Analysis, 16(1):17–24, January 1984. ISSN 0016-7363, 1538-4632. doi: 10.1111/j.1538-4632.1984.tb00797.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.1984.tb00797.x.

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.

Matthew Desmond, Carl Gershenson, and Barbara Kiviat. Forced Relocation and Residential Instability among Urban Renters. Social Service Review, 89(2):227–262, June 2015. ISSN 0037-7961, 1537-5404. doi: 10.1086/681091. URL https://www.journals.uchicago.edu/doi/10.1086/681091.

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

Margaret H. Dunham, Yongqiao Xiao, Le Gruenwald, and Zahid Hossain. Mining Association Rule. ACM Survey Journal, 2000.

Jacques Dutka. Robert Adrain and the method of least squares. Archive for History of Exact Sciences, 41(2):171–184, 1990. ISSN 0003-9519, 1432-0657. doi: 10.1007/BF00411864. URL http://link.springer.com/10.1007/BF00411864.

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

Economic Policy Institute. The Productivity-Pay Gap. Technical report, Economic Policy Institute, 2022. URL https://www.epi.org/productivity-pay-gap/.

Abdalla M El-Habil. An Application on Multinomial Logistic Regression Model. Pakistan Journal of Statistics and Operation Research, 8(2):271, March 2012. ISSN 2220-5810, 1816-2711. doi: 10.18187/pjsor.v8i2.234. URL http://www.pjsor.com/index.php/pjsor/article/view/234.

William Evans, David Philips, and Krista Ruffini. Reducing and Preventing Homelessness: A Review of the Evidence and Charting a Research Agenda. Technical Report w26232, National Bureau of Economic Research, Cambridge, MA, September 2019. URL http://www.nber.org/papers/w26232.pdf.

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.

R.A. Fisher. The design of experiments. Oliver & Boyd, 1935. URL http://tankona.free.fr/fisher1935.pdf.

Walter D. Fisher. On Grouping for Maximum Homogeneity. Journal of the American Statistical Association, 53(284):789–798, December 1958. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.1958.10501479. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1958.10501479.

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.

Johannes Fürnkranz and Tomáš Kliegr. A Brief Overview of Rule Learning. In Nick Bassiliades, Georg Gottlob, Fariba Sadri, Adrian Paschke, and Dumitru Roman, editors, Rule Technologies: Foundations, Tools, and Applications, volume 9202, pages 54–69. Springer International Publishing, Cham, 2015. ISBN 978-3-319-21541-9 978-3-319-21542-6. doi: 10.1007/978-3-319-21542-6_4. URL http://link.springer.com/10.1007/978-3-319-21542-6_4. Series Title: Lecture Notes in Computer Science.

Carl Friedrich Gauss. Theory Of The Motion Of The Heavenly Bodies Moving About The Sun In Conic Sections: A Translation Of Gauss's Theoria Motus. Kessinger Publishing, 2007.

Arthur Getis. A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective. Geographical Analysis, 40(3):297–309, July 2008. ISSN 0016-7363, 1538-4632. doi: 10.1111/j.1538-4632.2008.00727.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.2008.00727.x.

Arthur Getis. Spatial Autocorrelation. In Manfred M. Fischer and Arthur Getis, editors, Handbook of Applied Spatial Analysis, pages 255–278. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-03646-0 978-3-642-03647-7. doi: 10.1007/978-3-642-03647-7_14. URL https://link.springer.com/10.1007/978-3-642-03647-7_14.

Arthur Getis and Jared Aldstadt. Constructing the Spatial Weights Matrix Using a Local Statistic. Geographical Analysis, 36(2):90–104, July 2004. ISSN 0016-7363, 1538-4632. doi: 10.1111/j.1538-4632.2004.tb01127.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.2004.tb01127.x.

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.

Shaun A. Golding. Gentrification and Segregated Wealth in Rural America: Home Value Sorting in Destination Counties. Population Research and Policy Review, 35(1):127–146, February 2016. ISSN 0167-5923, 1573-7829. doi: 10.1007/s11113-015-9374-9. URL http://link.springer.com/10.1007/s11113-015-9374-9.

Daniel A. Griffith. What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. L'Espace géographique, 21(3):265–280, 1992. ISSN 0046-2497. doi:

10.3406/spgeo.1992.3091. URL
https://www.persee.fr/doc/spgeo_0046-2497_1992_num_21_3_3091.

Tony H. Grubesic. Zip codes and spatial analysis: Problems and prospects. Socio-Economic Planning Sciences, 42(2):129–149, June 2008. ISSN 00380121. doi: 10.1016/j.seps.2006.09.001. URL
https://linkinghub.elsevier.com/retrieve/pii/S0038012106000516.

Jane L. Harrison, Claire A. Montgomery, and P. Wilner Jeanty. A spatial, simultaneous model of social capital and poverty. Journal of Behavioral and Experimental Economics, 78:183–192, February 2019. ISSN 22148043. doi: 10.1016/j.socec.2018.09.001. URL
https://linkinghub.elsevier.com/retrieve/pii/S2214804318304166.

W. Leon Harter. The Method of Least Squares and Some Alternatives: Part I. International Statistical Review / Revue Internationale de Statistique, 42(2):147, August 1974. ISSN 03067734. doi: 10.2307/1403077. URL
https://www.jstor.org/stable/1403077?origin=crossref.

Health Resources & Services Administration. Defining Rural Population, 2022. URL
https://www.hrsa.gov/rural-health/about-us/what-is-rural.

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.

Chris Herring and Manuel Lutz. The roots and implications of the USA's homeless tent cities. City, 19(5):689–701, September 2015. ISSN 1360-4813, 1470-3629. doi:

10.1080/13604813.2015.1071114. URL
http://www.tandfonline.com/doi/full/10.1080/13604813.2015.1071114.

Housing Assistance Council. Race & Ethnicity in Rural America. Research Brief, Housing Assistance Council, Washington D.C., 2012. URL https://ruralhome.org/wp-content/uploads/storage/research_notes/rrn-race-and-ethnicity-web.pdf.

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.

Andrew M. Isserman. In the National Interest: Defining Rural and Urban Correctly in Research and Public Policy. International Regional Science Review, 28(4):465–499, October 2005. ISSN 0160-0176, 1552-6925. doi: 10.1177/0160017605279000. URL
http://journals.sagepub.com/doi/10.1177/0160017605279000.

Andrew M. Isserman, Edward Feser, and Drake E. Warren. Why Some Rural Places Prosper and Others Do Not. International Regional Science Review, 32(3):300–342, July 2009. ISSN 0160-0176, 1552-6925. doi: 10.1177/0160017609336090. URL
http://journals.sagepub.com/doi/10.1177/0160017609336090.

Joyce Jackson. Data Mining; A Conceptual Overview. Communications of the Association for Information Systems, 8, 2002. ISSN 15293181. doi: 10.17705/1CAIS.00819. URL
https://aisel.aisnet.org/cais/vol8/iss1/19.

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.

Anil K. Jain. Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31 (8):651–666, June 2010. ISSN 01678655. doi: 10.1016/j.patrec.2009.09.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0167865509002323.

Kenneth Johnson. Rural demographic change in the new century: slower growth, increased diversity. Technical report, 2012. URL https://scholars.unh.edu/carsey/159.

Kenneth Johnson and Daniel Lichter. Rural Depopulation in a Rapidly Urbanizing America. Brief 139, Durham, 2019. URL https://scholars.unh.edu/carsey/358.

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.

Leonard Kaufman and Peter Rousseeuw. Clustering by means of medoids. In Proceedings of the statistical data analysis based on the L1 norm conference, volume 31, Neuchatel, Switzerland, 1987. URL https://wis.kuleuven.be/stat/robust/papers/publications-1987/kaufmanrousseeuw-clusteringbymedoids-l1norm-1987.pdf.

Noor Kamal Kaur, Usvir Kaur, and Dr Dheerendra Singh. K-Medoid Clustering Algorithm- A Review. International Journal of Computer Application and Technology, 1 (1):42–45, 2014.

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.

Sharon Kingsland. The Refractory Model: The Logistic Curve and the History of Population Ecology. The Quarterly Review of Biology, 57(1):29–52, 1982. ISSN 0033-5770. URL https://www.jstor.org/stable/2825134. Publisher: University of Chicago Press.

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.

James A. Kurre. Is The Cost Of Living Less In Rural Areas? International Regional Science Review, 26(1):86–116, January 2003. ISSN 0160-0176. doi: 10.1177/0160017602238987. URL https://doi.org/10.1177/0160017602238987. Publisher: SAGE Publications Inc.

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.

A.M. Legendre. Nouvelles méthodes pour la détermination des orbites des comètes. F. Didot, 1805. URL https://books.google.com/books?id=FRcOAAAAQAAJ.

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.

S. Lloyd. Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2):129–137, March 1982. ISSN 0018-9448. doi: 10.1109/TIT.1982.1056489. URL http://ieeexplore.ieee.org/document/1056489/.

Linda Lobao and Rogelio Saenz. Spatial Inequality and Diversity as an Emerging Research Area*. Rural Sociology, 67(4):497–511, December 2002. ISSN 0036-0112, 1549-0831. doi: 10.1111/j.1549-0831.2002.tb00116.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1549-0831.2002.tb00116.x.

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

Travis Scott Lowe. Perceived Job and Labor Market Insecurity in the United States: An Assessment of Workers' Attitudes From 2002 to 2014. Work and Occupations, 45(3): 313–345, August 2018. ISSN 0730-8884, 1552-8464. doi: 10.1177/0730888418758381. URL http://journals.sagepub.com/doi/10.1177/0730888418758381.

May Nawal Lutfiyya, Joel Emery McCullough, Irina V. Haller, Stephen C. Waring, Joseph A. Bianco, and Martin Stephen Lipsky. Rurality as a Root or Fundamental Social Determinant of Health. Disease-a-Month, 58(11):620–628, November 2012. ISSN 00115029. doi: 10.1016/j.disamonth.2012.08.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0011502912001563.

J Macqueen. SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS. volume 1 of Berkeley Symposium on Mathematical Statistics and Probability, pages 281–297, Berkley, CA, 1967. University of California Press.

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.

Tagaram Soni Madhulatha. Comparison between K-Means and K-Medoids Clustering Algorithms. In David C. Wyld, Michal Wozniak, Nabendu Chaki, Natarajan Meghanathan, and Dhinaharan Nagamalai, editors, Advances in Computing and Information Technology, volume 198, pages 472–481. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-22554-3 978-3-642-22555-0. doi: 10.1007/978-3-642-22555-0_48. URL http://link.springer.com/10.1007/978-3-642-22555-0_48. Series Title: Communications in Computer and Information Science.

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.

Kathleen K. Miller and Bruce A. Weber. PERSISTENT POVERTY ACROSS THE RURAL-URBAN CONTINUUM. 2003. doi: 10.22004/AG.ECON.18910. URL https://ageconsearch.umn.edu/record/18910. Publisher: Unknown.

Genaro Molina, 2023. URL https://ca-times.brightspotcdn.com/dims4/default/4f39bdd/2147483647/strip/true/crop/5760x3840+0+0/resize/1200x800!/format/

webp/quality/75/?url=https%3A%2F%2Fcalifornia-times-brightspot.s3.amazonaws.com%2F71%2F0e%2Fad1bfbd3481198adb245a359ec50%2F1319023-me-0706-conservatives-say-homelessness-paralyzing-west-gem-121.jpg. Source: Los Angeles Times.

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.

P. A. P. Moran. Notes on Continuous Stochastic Phenomena. Biometrika, 37(1/2):17, June 1950. ISSN 00063444. doi: 10.2307/2332142. URL https://www.jstor.org/stable/2332142?origin=crossref.

Pavlos Msaouel, Paula Jimenez-Fonseca, Bora Lim, Alberto Carmona-Bayonas, and Giancarlo Agnelli. Medicine before and after David Cox. European Journal of Internal Medicine, 98:1–3, April 2022. ISSN 09536205. doi: 10.1016/j.ejim.2022.02.022. URL https://linkinghub.elsevier.com/retrieve/pii/S0953620522000875.

National Coalition for the Homeless. Rural Homelessness. Technical report, National Coalition for the Homeless, 2009. URL https://homelesshub.ca/sites/default/files/xzzomc0w.pdf.

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.

Juan Palacios, Piet Eichholtz, Nils Kok, and Erdal Aydin. The impact of housing conditions on health outcomes. Real Estate Economics, 49(4):1172–1200, December

2021. ISSN 1080-8620, 1540-6229. doi: 10.1111/1540-6229.12317. URL https://onlinelibrary.wiley.com/doi/10.1111/1540-6229.12317.

Mark D. Partridge and Dan S. Rickman. Persistent Pockets of Extreme American Poverty and Job Growth: Is There a Place-Based Policy Role? Journal of Agricultural and Resource Economics‘, 32(1), 2007. doi: 10.22004/AG.ECON.8599. URL https://ageconsearch.umn.edu/record/8599. Publisher: Unknown.

Rolf Pendall, Laurie Goodman, Jun Zhu, and Amanda Gold. The Future of Rural Housing. Research Report, The Urban Institute, Washington D.C., 2016.

David J. Peters. Typology of American Poverty. International Regional Science Review, 32 (1):19–39, January 2009. ISSN 0160-0176, 1552-6925. doi: 10.1177/0160017608325795. URL http://journals.sagepub.com/doi/10.1177/0160017608325795.

Jo C. Phelan, Bruce G. Link, and Parisa Tehranifar. Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications. Journal of Health and Social Behavior, 51(1_suppl):S28–S40, March 2010. ISSN 0022-1465, 2150-6000. doi: 10.1177/0022146510383498. URL http://journals.sagepub.com/doi/10.1177/0022146510383498.

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.

Marjorie Robertson, Natalie Harris, Nancy Fritz, Rebecca Noftsinger, and Pamela Fischer. Rural Homelessness. 2007. doi: 10.13140/RG.2.1.3433.7683. URL `http://rgdoi.net/10.13140/RG.2.1.3433.7683`. Publisher: Unpublished.

Sarwosri, Dwi Sunaryono, Rizky Januar Akbar, and Risky Dwi Setiyawan. Poverty classification using Analytic Hierarchy Process and k-means clustering. In 2016 International Conference on Information & Communication Technology and Systems (ICTS), pages 266–269, Surabaya, Indonesia, 2016. IEEE. ISBN 978-1-5090-1381-4. doi: 10.1109/ICTS.2016.7910310. URL `http://ieeexplore.ieee.org/document/7910310/`.

Gregory Sharp and Matthew Hall. Emerging Forms of Racial Inequality in Homeownership Exit, 1968–2009. Social Problems, 61(3):427–447, August 2014. ISSN 00377791, 15338533. doi: 10.1525/sp.2014.12161. URL `https://academic.oup.com/socpro/article-lookup/doi/10.1525/sp.2014.12161`.

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.

Kristina P. Sinaga and Miin-Shen Yang. Unsupervised K-Means Clustering Algorithm. IEEE Access, 8:80716–80727, 2020. ISSN 2169-3536. doi: 10.1109/ACCESS.2020.2988796. URL `https://ieeexplore.ieee.org/document/9072123/`.

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

Surbhi K. Solanki and Jalpa T. Patel. A Survey on Association Rule Mining. In 2015 Fifth International Conference on Advanced Computing & Communication Technologies, pages 212–216, Haryana, India, February 2015. IEEE. ISBN 978-1-4799-8488-6. doi: 10.1109/ACCT.2015.69. URL http://ieeexplore.ieee.org/document/7079081/.

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.

Seth E. Spielman, David Folch, and Nicholas Nagle. Patterns and causes of uncertainty in the American Community Survey. Applied Geography, 46:147–157, January 2014. ISSN 01436228. doi: 10.1016/j.apgeog.2013.11.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0143622813002518.

H Steinhaus. Sur la division des corps materiels en parties. Bulletin de L'Academie Polonaise des Sciences, 4(12):801–804, 1956. URL http://www.laurent-duval.eu/Documents/Steinhaus_H_1956_j-bull-acad-polon-sci_division_cmp-k-means.pdf.

Stephen M. Stigler. Gauss and the Invention of Least Squares. The Annals of Statistics, 9 (3):465–474, 1981. ISSN 0090-5364. URL https://www.jstor.org/stable/2240811. Publisher: Institute of Mathematical Statistics.

Engelbert Stockhammer. Some Stylized Facts on the Finance-Dominated Accumulation Regime. Competition & Change, 12(2):184–202, June 2008. ISSN 1024-5294. doi: 10.1179/102452908X289820. URL https://journals.sagepub.com/doi/abs/10.1179/102452908X289820. Publisher: SAGE Publications Ltd.

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=0BF7CE998C4949B5AD4DB9DB6CE6CA02DF1448E46028A9D722AB36EBD1473C1B404CCF479C8632CC8BEF36 originRegion=us-east-1&originCreation=20230118070028.

The American Presidency Project. Annual Message to the Congress on the State of the Union. | The American Presidency Project. URL https://www.presidency.ucsb.edu/documents/annual-message-the-congress-the-state-the-union-25.

Brian Thiede, Hyojung Kim, and Matthew Valasik. The Spatial Concentration of America's Rural Poor Population: A Postrecession Update: Concentrated Poverty in Rural America. Rural Sociology, 83(1):109–144, March 2018. ISSN 00360112. doi: 10.1111/ruso.12166. URL https://onlinelibrary.wiley.com/doi/10.1111/ruso.12166.

Fuyou Tian, Bingfang Wu, Hongwei Zeng, Shukri Ahmed, Nana Yan, Ian White, Miao Zhang, and Alfred Stein. Identifying the Links Among Poverty, Hydroenergy and Water Use Using Data Mining Methods. Water Resources Management, 34(5):1725–1741, March 2020. ISSN 0920-4741, 1573-1650. doi: 10.1007/s11269-020-02524-5. URL http://link.springer.com/10.1007/s11269-020-02524-5.

W. R. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography, 46:234, June 1970. ISSN 00130095. doi: 10.2307/143141. URL https://www.jstor.org/stable/143141?origin=crossref.

US Census Bureau. Income Inequality Down Due to Drops in Real Incomes at the Middle and Top, But Post-Tax Income Estimates Tell a Different Story, 2023. URL https://www.census.gov/library/stories/2023/09/income-inequality.html. Section: Government.

U.S. Census Bureau and U.S. Department of Housing and Urban Development. Median Sales Price of Houses Sold for the United States, 2023. URL https://fred.stlouisfed.org/series/MSPUS. Publisher: FRED, Federal Reserve Bank of St. Louis.

U.S. Department of Agriculture. USDA Rural Development Summary of Programs, 2023.

U.S.D.A. Economic Research Service. Rural-Urban Commuting Area Codes, 2023. URL https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/.

Daniel Varona and Luiz Capretz. Comparing Cuban and Brazilian software engineers. World Transactions on Engineering and Technology Education, 9:109–113, January 2011.

Paul R. Voss, David D. Long, Roger B. Hammer, and Samantha Friedman. County child poverty rates in the US: a spatial regression approach. Population Research and Policy Review, 25(4):369–391, November 2006. ISSN 0167-5923, 1573-7829. doi: 10.1007/s11113-006-9007-4. URL http://link.springer.com/10.1007/s11113-006-9007-4.

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.

Jeffry L White. Logistic regression Model Effectiveness: Proportional Chance Criteria and Proportional Reduction in Error. Journal of Contemporary Research in Education, 2(1), 2013.

Noé M. Wiener. Labor Market Segmentation and Immigrant Competition: A Quantal Response Statistical Equilibrium Analysis. Entropy, 22(7):742, July 2020. ISSN

1099-4300. doi: 10.3390/e22070742. URL
https://www.mdpi.com/1099-4300/22/7/742.

Bev Wilson and Mallory L. Rahe. Rural prosperity and federal expenditures, 2000–2010. Regional Science Policy & Practice, 8(1-2):3–26, March 2016. ISSN 1757-7802, 1757-7802. doi: 10.1111/rsp3.12070. URL
https://rsaiconnect.onlinelibrary.wiley.com/doi/10.1111/rsp3.12070.

Dongkuan Xu and Yingjie Tian. A Comprehensive Survey of Clustering Algorithms. Annals of Data Science, 2(2):165–193, June 2015. ISSN 2198-5804, 2198-5812. doi: 10.1007/s40745-015-0040-1. URL
http://link.springer.com/10.1007/s40745-015-0040-1.

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.

James Yagley, Lance George, Cequyna Moore, and Jennifer Pinder. They Paved Paradise...Gentrification in Rural Communities. Technical report, Housing Assistance Council, 2005. URL https://ruralhome.org/wp-content/uploads/storage/documents/gentrification.pdf.

Andrea Yoder Clark, Nicole Blumenfeld, Eric Lal, Shikar Darbari, Shiyang Northwood, and Ashkan Wadpey. Using K-Means Cluster Analysis and Decision Trees to Highlight Significant Factors Leading to Homelessness. Mathematics, 9(17):2045, August 2021. ISSN 2227-7390. doi: 10.3390/math9172045. URL
https://www.mdpi.com/2227-7390/9/17/2045.

Amelia Yousey and Rhucha Samudra. Defining Homelessness in the Rural United States. Online Journal of Rural Research & Policy, 13(4), December 2018. ISSN 1936-0487. doi: 10.4148/1936-0487.1094. URL https://newprairiepress.org/ojrrp/vol13/iss4/1.

Kimberly E. Zarecor, David J. Peters, and Sara Hamideh. Rural Smart Shrinkage and Perceptions of Quality of Life in the American Midwest. In Javier Martinez, Claudia Andrea Mikkelsen, and Rhonda Phillips, editors, Handbook of Quality of Life and Sustainability, pages 395–415. Springer International Publishing, Cham, 2021. ISBN 978-3-030-50539-4 978-3-030-50540-0. doi: 10.1007/978-3-030-50540-0_20. URL `http://link.springer.com/10.1007/978-3-030-50540-0_20`. Series Title: International Handbooks of Quality-of-Life.

Changping Zhang. Spatial Weights Matrix and its Application. Journal of Regional Development Studies, 2012.

Qiankun Zhao and Zhao Bhowmick. Association Rule Mining: A Survey. Technical Report 2003116, Nanyang Technological University, Singapore, 2003.

Zeng Zhen and Zhu Mengxian. Poverty Groups Identification and Assessment of Poverty Alleviation Programs in Rural China. In Proceedings of the 6th International Conference on Humanities and Social Science Research (ICHSSR2020), Hangzhou, China, 2020. Atlantis Press. ISBN 978-94-6252-960-1. doi: 10.2991/assehr.k.200428.040. URL `https://www.atlantis-press.com/article/125939300`.

Julie N. Zimmerman, Sunny (Seonok) Ham, and Sarah Michelle Frank. Does it or Doesn't it? Geographic Differences and the Costs of Living. Rural Sociology, 73(3):463–486, September 2008. ISSN 0036-0112, 1549-0831. doi: 10.1526/003601108785766561. URL `https://onlinelibrary.wiley.com/doi/10.1526/003601108785766561`.

APPENDICES

APPENDIX A

DESCRIPTIVE STATISTICS

Table A.1: Demographics Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
white	6,362	0.859	0.189	0.000	1.000
black	6,362	0.067	0.150	0.000	1.000
am_in_ala_nat	6,362	0.028	0.115	0.000	0.992
asian	6,362	0.006	0.012	0.000	0.283
haw_pac	6,362	0.001	0.003	0.000	0.133
other	6,362	0.017	0.043	0.000	0.575
hisp_lat	6,362	0.076	0.133	0.000	0.995
male_u18	6,362	0.110	0.032	0.000	0.276
female_u18	6,362	0.105	0.031	0.000	0.396
male_o18	6,362	0.391	0.055	0.000	1.000
female_o18	6,362	0.394	0.045	0.000	1.000

Table A.2: RME Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
same_house_less_than_hs	6,362	0.084	0.048	0.000	0.495
same_house_hs	6,362	0.231	0.060	0.000	0.667
moved_in_county_less_than_hs	6,362	0.005	0.007	0.000	0.079
moved_in_county_hs	6,362	0.012	0.010	0.000	0.083
moved_diff_county_less_than_hs	6,362	0.004	0.008	0.000	0.134
moved_diff_county_hs	6,362	0.008	0.011	0.000	0.169
moved_diff_state_less_than_hs	6,362	0.001	0.003	0.000	0.090
moved_diff_state_hs	6,362	0.003	0.005	0.000	0.080

Table A.3: RMP Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
same_house_p1	6,362	0.124	0.073	0.000	0.758
same_house_p2	6,362	0.090	0.043	0.000	0.374
moved_in_county_p1	6,362	0.015	0.018	0.000	0.190
moved_in_county_p2	6,362	0.008	0.012	0.000	0.154
moved_diff_county_p1	6,362	0.007	0.011	0.000	0.115
moved_diff_county_p2	6,362	0.004	0.007	0.000	0.073
moved_diff_state_p1	6,362	0.004	0.008	0.000	0.128
moved_diff_state_p2	6,362	0.002	0.006	0.000	0.169

Table A.4: Housing Costs Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
mortgage_high_cost	6,342	0.052	0.029	0.000	1.000
no_mortgage_high_cost	6,342	0.026	0.017	0.000	0.611
rent_high_cost	6,342	0.162	0.081	0.000	1.000

Table A.5: Housing Quality Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
all_incomplete_plumb	6,347	0.263	0.317	0.000	16.500
all_incomplete_kitchen	6,347	0.292	0.304	0.000	9.000
occ_incomplete_plumb	6,347	0.008	0.019	0.000	0.329
occ_incomplete_kitchen	6,347	0.010	0.018	0.000	0.300

Table A.6: Housing Type Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
owner_single	6,343	0.860	0.124	0.000	1.000
owner_2to4	6,343	0.005	0.018	0.000	1.000
owner_5plus	6,343	0.003	0.018	0.000	1.000
owner_mobile	6,343	0.130	0.122	0.000	1.000
owner_unconvent	6,343	0.002	0.015	0.000	1.000
renter_single	6,343	0.604	0.191	0.000	1.000
renter_2to4	6,343	0.127	0.122	0.000	1.000
renter_5plus	6,343	0.111	0.123	0.000	1.000
renter_mobile	6,343	0.157	0.163	0.000	1.000
renter_unconvent	6,343	0.002	0.011	0.000	0.348

Table A.7: Household Factors Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
hh_no_wage	6,362	0.143	0.048	0.000	0.667
hh_no_other_income	6,362	0.366	0.050	0.121	1.000
hh_no_investment_income	6,362	0.328	0.050	0.058	1.000
hh_public_assistance	6,362	0.058	0.035	0.000	0.333
hh_ssi	6,362	0.027	0.017	0.000	0.162
hh_3plus_worker	6,362	0.019	0.011	0.000	0.080
hh_worker_no_vehicle	6,362	0.018	0.016	0.000	0.165
hh_no_vehicle	6,362	0.026	0.021	0.000	0.257
gini_index	6,362	0.437	0.054	0.004	0.786

Table A.8: Employment Diversity Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
ag_for_fish_hunt_mining	6,362	0.034	0.037	0.000	0.385
construction	6,362	0.033	0.017	0.000	0.160
manufacturing	6,362	0.056	0.041	0.000	0.280
wholesale_trade	6,362	0.010	0.009	0.000	0.108
retail_trade	6,362	0.048	0.019	0.000	0.283
trans_warehouse_util	6,362	0.023	0.013	0.000	0.139
information	6,362	0.005	0.005	0.000	0.058
fin_re_insur	6,362	0.017	0.012	0.000	0.126
prof_sci_mgmt_waste	6,362	0.025	0.017	0.000	0.316
edu_health_social	6,362	0.097	0.034	0.000	0.712
arts_rec_food	6,362	0.036	0.029	0.000	0.601
othersvcs	6,362	0.020	0.011	0.000	0.278
public_admin	6,362	0.022	0.016	0.000	0.579

APPENDIX B

HIGH-RISK RULES BY PERCENT OF CENSUS TRACTS PER STATE

Table B.1: Employment High-Risk Rules by Percent of Census Tracts Per State

State	CTs	Employment	Household Factors	Housing Type	RMP
AL	153	30.72	6.54	13.07	11.11
AR	163	43.56	3.68	15.95	14.72
AZ	98	37.76	10.20	10.20	3.06
CA	205	12.68	0.49	3.90	4.39
CO	138	37.68	8.70	0	6.52
CT	9	33.33	0	0	0
DE	7	42.86	0	14.29	14.29
FL	79	36.71	2.53	10.13	7.59
GA	145	31.03	11.03	4.83	8.97
IA	282	26.60	10.64	3.55	4.26
ID	67	50.75	7.46	11.94	16.42
IL	242	33.88	8.68	5.37	2.89
IN	150	35.33	14.67	5.33	7.33
KS	170	37.65	11.76	5.88	14.71
KY	236	26.27	7.63	2.54	5.08
LA	105	32.38	5.71	1.90	12.38
MA	20	35	0	15	5
MD	21	19.05	4.76	0	4.76
ME	126	36.51	9.52	3.17	11.11
MI	279	35.48	6.45	12.90	2.87
MN	258	44.96	20.54	20.93	21.32
MO	243	30.86	10.29	6.58	13.58
MS	141	36.17	9.22	16.31	0
MT	134	40.30	13.43	20.15	16.42
NC	184	46.74	21.74	3.80	14.67
ND	99	50.51	18.18	15.15	16.16
NE	139	30.22	7.91	2.88	10.79
NH	55	50.91	21.82	7.27	9.09
NM	71	28.17	11.27	9.86	4.23
NV	21	66.67	23.81	23.81	4.76
NY	235	32.34	7.66	8.51	0.85
OH	174	26.44	3.45	4.02	2.87
OK	186	26.88	9.68	11.83	9.14
OR	79	27.85	3.80	10.13	10.13
PA	148	22.30	6.76	11.49	5.41
SC	58	36.21	0	17.24	0
SD	91	23.08	9.89	4.40	3.30
TN	164	7.93	0.61	1.22	4.27
TX	377	29.71	10.88	10.61	8.22
UT	43	27.91	0	4.65	2.33
VA	162	35.80	0	12.35	10.49
VT	98	44.90	22.45	20.41	14.29
WA	95	15.79	5.26	6.32	3.16
WI	277	33.21	15.52	16.61	9.03
WV	87	40.23	10.34	27.59	4.60
WY	48	45.83	18.75	18.75	6.25

Table B.2: Demographics High-Risk Rules by Percent of Census Tracts Per State

State	CTs	Demographics	RMP	Housing Costs	Housing Quality	HH Factors	RME
AL	153	5.88	1.31	0.65	1.96	1.31	1.96
AR	163	14.11	8.59	5.52	1.23	6.13	3.07
AZ	98	30.61	28.57	28.57	19.39	26.53	0
CA	205	7.80	3.41	0.49	2.44	0.49	2.44
CO	138	0.72	0.72	0	0	0	0
CT	9	11.11	0	0	0	0	0
DE	7	14.29	14.29	14.29	0	0	0
FL	79	27.85	12.66	13.92	20.25	1.27	3.80
GA	145	37.24	8.97	9.66	22.76	15.86	16.55
IA	282	5.32	2.48	2.13	1.77	1.77	3.19
ID	67	41.79	19.40	10.45	10.45	8.96	5.97
IL	242	5.79	0.41	1.24	1.65	2.48	3.72
IN	150	42	10.67	10	1.33	17.33	12.67
KS	170	7.06	1.76	1.76	0.59	0	1.18
KY	236	63.98	21.19	13.56	14.83	17.80	20.34
LA	105	44.76	34.29	5.71	13.33	1.90	11.43
MA	20	0	0	0	0	0	0
MD	21	19.05	0	0	0	4.76	9.52
ME	126	39.68	2.38	15.08	15.87	21.43	17.46
MI	279	5.02	4.30	3.58	3.94	1.79	1.79
MN	258	34.11	14.34	16.28	17.05	3.88	18.60
MO	243	2.88	0	0	0.41	0	0.41
MS	141	2.84	2.13	0.71	0.71	2.13	2.84
MT	134	8.96	5.22	0	2.24	5.97	1.49
NC	184	29.89	16.85	8.70	5.98	14.67	7.61
ND	99	14.14	5.05	5.05	10.10	5.05	5.05
NE	139	0.72	0.72	0	0	0	0
NH	55	41.82	1.82	3.64	10.91	30.91	9.09
NM	71	35.21	4.23	9.86	0	15.49	11.27
NV	21	4.76	0	0	0	0	0
NY	235	16.17	4.68	6.38	0	3.40	1.70
OH	174	10.92	4.02	0	0.57	5.17	4.02
OK	186	10.22	4.30	4.84	3.23	2.15	5.38
OR	79	10.13	1.27	6.33	1.27	0	1.27
PA	148	4.05	0.68	1.35	0.68	0.68	2.03
SC	58	15.52	1.72	5.17	5.17	0	0
SD	91	48.35	4.40	10.99	4.40	19.78	9.89
TN	164	15.24	4.88	4.27	4.27	4.88	3.66
TX	377	39.26	7.69	12.73	10.34	7.69	12.47
UT	43	2.33	2.33	2.33	2.33	2.33	2.33
VA	162	20.37	2.47	11.73	5.56	4.32	6.79
VT	98	21.43	4.08	5.10	8.16	10.20	5.10
WA	95	40	8.42	1.05	13.68	10.53	7.37
WI	277	23.83	7.58	14.44	17.69	5.42	1.44
WV	87	9.20	2.30	2.30	3.45	2.30	1.15
WY	48	2.08	0	0	2.08	2.08	0

Table B.3: Housing Costs High-Risk Rules by Percent of Census Tracts Per State

State	CTs	Cost	Housing Type	Housing Quality	RMP	HH Factors	RME	Demographics
AL	153	18.95	4.58	5.23	7.19	3.27	3.27	0.65
AR	163	30.06	18.40	4.91	15.34	7.98	6.75	5.52
AZ	98	42.86	5.10	19.39	28.57	28.57	2.04	28.57
CA	205	33.17	5.37	7.80	16.10	8.78	9.27	0.49
CO	138	26.09	4.35	18.84	5.80	2.17	7.25	0
CT	9	11.11	11.11	0	0	0	0	0
DE	7	71.43	42.86	0	14.29	0	14.29	14.29
FL	79	27.85	10.13	16.46	6.33	0	5.06	13.92
GA	145	33.10	5.52	16.55	5.52	10.34	9.66	9.66
IA	282	36.17	8.16	13.12	10.28	12.06	8.87	2.13
ID	67	29.85	10.45	5.97	17.91	11.94	5.97	10.45
IL	242	36.78	9.50	11.16	4.55	13.64	6.20	1.24
IN	150	19.33	6.67	0	6.67	12	4.67	10
KS	170	25.88	4.71	2.94	2.35	7.06	7.65	1.76
KY	236	19.49	4.66	5.93	8.47	8.90	3.39	13.56
LA	105	30.48	16.19	12.38	7.62	15.24	14.29	5.71
MA	20	25	5	0	10	10	5	0
MD	21	9.52	4.76	0	0	4.76	0	0
ME	126	38.89	7.14	11.90	6.35	20.63	18.25	15.08
MI	279	38.71	19.35	19	16.49	7.17	6.09	3.58
MN	258	32.56	5.04	19.38	13.18	4.26	14.34	16.28
MO	243	38.27	20.16	9.05	22.63	16.05	10.29	0
MS	141	30.50	17.02	11.35	11.35	18.44	1.42	0.71
MT	134	34.33	13.43	18.66	7.46	16.42	11.94	0
NC	184	32.07	5.43	7.61	18.48	15.76	9.78	8.70
ND	99	26.26	9.09	18.18	8.08	8.08	10.10	5.05
NE	139	25.90	7.91	10.07	3.60	4.32	10.07	0
NH	55	12.73	5.45	3.64	1.82	1.82	5.45	3.64
NM	71	26.76	7.04	0	1.41	5.63	5.63	9.86
NV	21	14.29	4.76	4.76	4.76	9.52	0	0
NY	235	20.85	5.96	0	8.09	5.11	3.83	6.38
OH	174	33.33	10.34	0	10.92	0	0	0
OK	186	46.24	24.19	16.67	8.06	15.05	16.13	4.84
OR	79	25.32	5.06	1.27	7.59	3.80	6.33	6.33
PA	148	28.38	12.16	0	7.43	6.08	8.11	1.35
SC	58	31.03	12.07	12.07	10.34	0	0	5.17
SD	91	23.08	17.58	4.40	6.59	13.19	8.79	10.99
TN	164	25	8.54	11.59	15.24	1.22	7.32	4.27
TX	377	32.89	15.92	7.69	5.84	13.26	8.75	12.73
UT	43	9.30	4.65	2.33	2.33	2.33	4.65	2.33
VA	162	43.83	6.79	7.41	6.17	2.47	9.26	11.73
VT	98	29.59	13.27	9.18	2.04	13.27	11.22	5.10
WA	95	28.42	22.11	9.47	7.37	14.74	7.37	1.05
WI	277	29.24	4.33	18.05	14.44	8.30	2.89	14.44
WV	87	33.33	22.99	12.64	4.60	2.30	5.75	2.30
WY	48	25	12.50	14.58	6.25	0	10.42	0

Table B.4: Housing Quality High-Risk Rules by Percent of Census Tracts Per State

State	CTs	Housing Quality	Housing Costs	RMP	Housing Type	HH Factors	RME	Demographics
AL	153	35.29	14.38	18.95	14.38	7.84	9.80	1.96
AR	163	15.95	3.68	7.98	3.68	1.84	4.29	1.23
AZ	98	19.39	0	19.39	0	19.39	0	19.39
CA	205	21.95	4.88	4.88	4.88	9.27	6.34	2.44
CO	138	57.97	7.25	10.14	7.25	21.74	18.12	0
CT	9	11.11	0	0	0	0	0	0
DE	7	0	0	0	0	0	0	0
FL	79	41.77	16.46	13.92	16.46	1.27	8.86	20.25
GA	145	48.28	10.34	11.72	10.34	17.93	10.34	22.76
IA	282	40.78	5.32	10.99	5.32	12.41	9.57	1.77
ID	67	20.90	7.46	10.45	7.46	5.97	8.96	10.45
IL	242	20.25	8.68	2.89	8.68	4.96	4.55	1.65
IN	150	1.33	0	1.33	0	0	1.33	1.33
KS	170	11.76	2.94	0.59	2.94	5.29	5.29	0.59
KY	236	22.88	4.24	11.44	4.24	8.90	4.66	14.83
LA	105	36.19	14.29	11.43	14.29	14.29	16.19	13.33
MA	20	5	0	0	0	5	0	0
MD	21	4.76	0	0	0	0	0	0
ME	126	35.71	6.35	3.97	6.35	17.46	14.29	15.87
MI	279	46.24	12.90	21.51	12.90	7.89	9.32	3.94
MN	258	51.16	15.50	19.38	15.50	13.57	20.93	17.05
MO	243	20.16	8.23	7.41	8.23	5.35	4.12	0.41
MS	141	39.72	15.60	15.60	15.60	9.22	1.42	0.71
MT	134	46.27	17.91	6.72	17.91	8.96	17.16	2.24
NC	184	23.37	4.35	10.33	4.35	10.33	5.98	5.98
ND	99	52.53	14.14	17.17	14.14	16.16	18.18	10.10
NE	139	33.09	4.32	5.04	4.32	9.35	11.51	0
NH	55	16.36	1.82	0	1.82	9.09	5.45	10.91
NM	71	9.86	5.63	9.86	5.63	0	9.86	0
NV	21	9.52	0	0	0	4.76	0	0
NY	235	2.13	1.28	0	1.28	0	0	0
OH	174	0.57	0	0.57	0	0.57	0.57	0.57
OK	186	33.33	17.20	7.53	17.20	11.83	12.37	3.23
OR	79	13.92	1.27	3.80	1.27	2.53	0	1.27
PA	148	15.54	2.70	2.70	2.70	2.70	8.11	0.68
SC	58	44.83	22.41	13.79	22.41	0	1.72	5.17
SD	91	16.48	8.79	2.20	8.79	6.59	4.40	4.40
TN	164	39.02	11.59	20.12	11.59	3.66	10.37	4.27
TX	377	22.81	9.55	5.31	9.55	7.16	6.90	10.34
UT	43	2.33	2.33	2.33	2.33	2.33	2.33	2.33
VA	162	14.81	3.09	0.62	3.09	1.85	4.94	5.56
VT	98	36.73	16.33	3.06	16.33	18.37	9.18	8.16
WA	95	30.53	13.68	8.42	13.68	12.63	8.42	13.68
WI	277	49.82	16.25	18.05	16.25	13.36	4.69	17.69
WV	87	39.08	20.69	10.34	20.69	14.94	3.45	3.45
WY	48	45.83	18.75	8.33	18.75	16.67	10.42	2.08

Table B.5: RME High-Risk Rules by Percent of Census Tracts Per State

State	CTs	RME	RMP	Employment	cost	waid	hhtype	Demographics
AL	153	25.49	11.76	11.11	3.27	4.58	7.84	1.96
AR	163	25.77	15.95	14.72	6.75	2.45	3.68	3.07
AZ	98	6.12	0	3.06	2.04	2.04	1.02	0
CA	205	26.34	6.83	4.39	9.27	2.93	7.32	2.44
CO	138	22.46	3.62	6.52	7.25	12.32	7.25	0
CT	9	11.11	0	0	0	11.11	0	0
DE	7	14.29	0	14.29	14.29	0	14.29	0
FL	79	25.32	1.27	7.59	5.06	0	6.33	3.80
GA	145	30.34	11.72	8.97	9.66	10.34	4.14	16.55
IA	282	25.18	12.06	4.26	8.87	4.96	6.03	3.19
ID	67	26.87	14.93	16.42	5.97	10.45	8.96	5.97
IL	242	14.46	2.07	2.89	6.20	6.20	4.55	3.72
IN	150	30.67	11.33	7.33	4.67	16	4.67	12.67
KS	170	29.41	1.76	14.71	7.65	13.53	2.35	1.18
KY	236	32.20	5.93	5.08	3.39	0.42	11.86	20.34
LA	105	39.05	16.19	12.38	14.29	10.48	9.52	11.43
MA	20	10	0	5	5	0	5	0
MD	21	14.29	0	4.76	0	4.76	0	9.52
ME	126	40.48	7.14	11.11	18.25	19.84	8.73	17.46
MI	279	15.41	8.96	2.87	6.09	1.79	3.94	1.79
MN	258	46.12	21.32	21.32	14.34	12.40	18.99	18.60
MO	243	29.22	13.17	13.58	10.29	10.29	7.41	0.41
MS	141	3.55	2.13	0	1.42	2.84	1.42	2.84
MT	134	35.82	8.21	16.42	11.94	13.43	14.18	1.49
NC	184	27.72	16.85	14.67	9.78	10.33	4.35	7.61
ND	99	31.31	13.13	16.16	10.10	12.12	8.08	5.05
NE	139	33.09	7.19	10.79	10.07	9.35	4.32	0
NH	55	23.64	7.27	9.09	5.45	7.27	5.45	9.09
NM	71	35.21	23.94	4.23	5.63	8.45	11.27	11.27
NV	21	9.52	0	4.76	0	0	4.76	0
NY	235	14.47	7.23	0.85	3.83	5.96	5.96	1.70
OH	174	4.02	1.15	2.87	0	4.02	0	4.02
OK	186	33.87	12.37	9.14	16.13	10.75	10.22	5.38
OR	79	21.52	7.59	10.13	6.33	2.53	3.80	1.27
PA	148	33.11	6.76	5.41	8.11	4.73	6.08	2.03
SC	58	5.17	5.17	0	0	3.45	0	0
SD	91	19.78	5.49	3.30	8.79	10.99	12.09	9.89
TN	164	27.44	14.63	4.27	7.32	4.88	10.37	3.66
TX	377	26.26	7.69	8.22	8.75	7.16	9.28	12.47
UT	43	4.65	2.33	2.33	4.65	2.33	4.65	2.33
VA	162	24.69	1.23	10.49	9.26	1.85	1.23	6.79
VT	98	30.61	2.04	14.29	11.22	7.14	17.35	5.10
WA	95	17.89	7.37	3.16	7.37	10.53	9.47	7.37
WI	277	12.64	9.03	9.03	2.89	7.58	7.22	1.44
WV	87	11.49	3.45	4.60	5.75	3.45	8.05	1.15
WY	48	18.75	8.33	6.25	10.42	0	8.33	0

Table B.6: RMP High-Risk Rules by Percent of Census Tracts Per State

	CTs	RMP	Housing Costs	HH Factors	Housing Quality	Housing Type	RME	Demographics
AL	153	52.29	7.19	15.69	18.95	20.92	11.76	1.31
AR	163	50.92	15.34	5.52	7.98	19.63	15.95	8.59
AZ	98	31.63	28.57	26.53	19.39	0	0	28.57
CA	205	40.98	16.10	11.71	4.88	6.83	6.83	3.41
CO	138	28.26	5.80	3.62	10.14	0	3.62	0.72
CT	9	0	0	0	0	0	0	0
DE	7	14.29	14.29	0	0	0	0	14.29
FL	79	32.91	6.33	3.80	13.92	7.59	1.27	12.66
GA	145	28.97	5.52	9.66	11.72	8.28	11.72	8.97
IA	282	32.27	10.28	4.96	10.99	6.38	12.06	2.48
ID	67	52.24	17.91	16.42	10.45	16.42	14.93	19.40
IL	242	10.74	4.55	2.89	2.89	3.31	2.07	0.41
IN	150	26	6.67	11.33	1.33	5.33	11.33	10.67
KS	170	7.65	2.35	2.35	0.59	1.76	1.76	1.76
KY	236	31.78	8.47	14.41	11.44	5.51	5.93	21.19
LA	105	48.57	7.62	5.71	11.43	6.67	16.19	34.29
MA	20	15	10	5	0	10	0	0
MD	21	0	0	0	0	0	0	0
ME	126	15.08	6.35	8.73	3.97	3.17	7.14	2.38
MI	279	42.65	16.49	4.66	21.51	15.05	8.96	4.30
MN	258	38.76	13.18	15.89	19.38	14.34	21.32	14.34
MO	243	41.98	22.63	17.28	7.41	18.52	13.17	0
MS	141	43.97	11.35	10.64	15.60	17.02	2.13	2.13
MT	134	23.13	7.46	10.45	6.72	2.24	8.21	5.22
NC	184	50	18.48	18.48	10.33	7.07	16.85	16.85
ND	99	26.26	8.08	8.08	17.17	7.07	13.13	5.05
NE	139	20.86	3.60	7.91	5.04	2.88	7.19	0.72
NH	55	7.27	1.82	1.82	0	0	7.27	1.82
NM	71	36.62	1.41	8.45	9.86	15.49	23.94	4.23
NV	21	9.52	4.76	4.76	0	0	0	0
NY	235	34.04	8.09	15.74	0	14.89	7.23	4.68
OH	174	23.56	10.92	1.72	0.57	8.05	1.15	4.02
OK	186	18.82	8.06	4.84	7.53	4.84	12.37	4.30
OR	79	27.85	7.59	5.06	3.80	8.86	7.59	1.27
PA	148	18.24	7.43	8.11	2.70	10.14	6.76	0.68
SC	58	27.59	10.34	3.45	13.79	6.90	5.17	1.72
SD	91	13.19	6.59	6.59	2.20	8.79	5.49	4.40
TN	164	45.73	15.24	5.49	20.12	17.68	14.63	4.88
TX	377	16.98	5.84	6.37	5.31	6.63	7.69	7.69
UT	43	2.33	2.33	2.33	2.33	2.33	2.33	2.33
VA	162	10.49	6.17	2.47	0.62	3.09	1.23	2.47
VT	98	8.16	2.04	4.08	3.06	3.06	2.04	4.08
WA	95	21.05	7.37	9.47	8.42	13.68	7.37	8.42
WI	277	42.24	14.44	17.69	18.05	13.72	9.03	7.58
WV	87	25.29	4.60	14.94	10.34	9.20	3.45	2.30
WY	48	25	6.25	4.17	8.33	4.17	8.33	0

Table B.7: Household Factors High-Risk Rules by Percent of Census Tracts Per State

State	CTs	HH Factors	Employment	Housing Costs	RMP	Housing Type	Housing Quality	RME	Demographics
AL	153	20.26	6.54	3.27	15.69	7.19	7.84	4.58	1.31
AR	163	11.66	3.68	7.98	5.52	9.20	1.84	2.45	6.13
AZ	98	32.65	10.20	28.57	26.53	1.02	19.39	2.04	26.53
CA	205	25.85	0.49	8.78	11.71	3.90	9.27	2.93	0.49
CO	138	31.88	8.70	2.17	3.62	4.35	21.74	12.32	0
CT	9	11.11	0	0	0	0	0	11.11	0
DE	7	0	0	0	0	0	0	0	0
FL	79	3.80	2.53	0	3.80	0	1.27	0	1.27
GA	145	27.59	11.03	10.34	9.66	4.14	17.93	10.34	15.86
IA	282	29.79	10.64	12.06	4.96	5.32	12.41	4.96	1.77
ID	67	23.88	7.46	11.94	16.42	14.93	5.97	10.45	8.96
IL	242	20.66	8.68	13.64	2.89	6.20	4.96	6.20	2.48
IN	150	53.33	14.67	12	11.33	11.33	0	16	17.33
KS	170	21.76	11.76	7.06	2.35	2.35	5.29	13.53	0
KY	236	26.27	7.63	8.90	14.41	2.12	8.90	0.42	17.80
LA	105	26.67	5.71	15.24	5.71	17.14	14.29	10.48	1.90
MA	20	20	0	10	5	5	5	0	0
MD	21	28.57	4.76	4.76	0	9.52	0	4.76	4.76
ME	126	42.06	9.52	20.63	8.73	7.94	17.46	19.84	21.43
MI	279	17.56	6.45	7.17	4.66	3.94	7.89	1.79	1.79
MN	258	28.68	20.54	4.26	15.89	15.50	13.57	12.40	3.88
MO	243	28.81	10.29	16.05	17.28	11.93	5.35	10.29	0
MS	141	31.91	9.22	18.44	10.64	23.40	9.22	2.84	2.13
MT	134	32.09	13.43	16.42	10.45	12.69	8.96	13.43	5.97
NC	184	40.22	21.74	15.76	18.48	3.80	10.33	10.33	14.67
ND	99	29.29	18.18	8.08	8.08	8.08	16.16	12.12	5.05
NE	139	23.74	7.91	4.32	7.91	2.16	9.35	9.35	0
NH	55	50.91	21.82	1.82	1.82	5.45	9.09	7.27	30.91
NM	71	26.76	11.27	5.63	8.45	11.27	0	8.45	15.49
NV	21	28.57	23.81	9.52	4.76	4.76	4.76	0	0
NY	235	25.53	7.66	5.11	15.74	11.91	0	5.96	3.40
OH	174	5.75	3.45	0	1.72	0	0.57	4.02	5.17
OK	186	38.17	9.68	15.05	4.84	24.73	11.83	10.75	2.15
OR	79	13.92	3.80	3.80	5.06	1.27	2.53	2.53	0
PA	148	14.86	6.76	6.08	8.11	12.84	2.70	4.73	0.68
SC	58	3.45	0	0	3.45	0	0	3.45	0
SD	91	40.66	9.89	13.19	6.59	20.88	6.59	10.99	19.78
TN	164	11.59	0.61	1.22	5.49	3.05	3.66	4.88	4.88
TX	377	33.16	10.88	13.26	6.37	12.73	7.16	7.16	7.69
UT	43	2.33	0	2.33	2.33	2.33	2.33	2.33	2.33
VA	162	7.41	0	2.47	2.47	4.32	1.85	1.85	4.32
VT	98	38.78	22.45	13.27	4.08	19.39	18.37	7.14	10.20
WA	95	28.42	5.26	14.74	9.47	16.84	12.63	10.53	10.53
WI	277	26.35	15.52	8.30	17.69	14.80	13.36	7.58	5.42
WV	87	22.99	10.34	2.30	14.94	12.64	14.94	3.45	2.30
WY	48	37.50	18.75	0	4.17	16.67	16.67	0	2.08

Table B.8: Housing Type High-Risk Rules by Percent of Census Tracts Per State

State	CTs	Housing Type	Housing Costs	Employment	Household Factors	RMP	Housing Quality	RME
AL	153	33.33	4.58	13.07	3.92	20.92	14.38	7.84
AR	163	44.17	18.40	15.95	12.27	19.63	3.68	3.68
AZ	98	24.49	5.10	10.20	2.04	0	0	1.02
CA	205	20	5.37	3.90	1.46	6.83	4.88	7.32
CO	138	7.97	4.35	0	0	0	7.25	7.25
CT	9	22.22	11.11	0	0	0	0	0
DE	7	42.86	42.86	14.29	0	0	0	14.29
FL	79	29.11	10.13	10.13	10.13	7.59	16.46	6.33
GA	145	18.62	5.52	4.83	7.59	8.28	10.34	4.14
IA	282	15.96	8.16	3.55	1.77	6.38	5.32	6.03
ID	67	29.85	10.45	11.94	11.94	16.42	7.46	8.96
IL	242	30.17	9.50	5.37	1.24	3.31	8.68	4.55
IN	150	18.67	6.67	5.33	7.33	5.33	0	4.67
KS	170	14.12	4.71	5.88	6.47	1.76	2.94	2.35
KY	236	19.92	4.66	2.54	13.14	5.51	4.24	11.86
LA	105	27.62	16.19	1.90	4.76	6.67	14.29	9.52
MA	20	30	5	15	0	10	0	5
MD	21	14.29	4.76	0	4.76	0	0	0
ME	126	12.70	7.14	3.17	3.97	3.17	6.35	8.73
MI	279	33.33	19.35	12.90	1.79	15.05	12.90	3.94
MN	258	31.78	5.04	20.93	10.08	14.34	15.50	18.99
MO	243	32.51	20.16	6.58	0.82	18.52	8.23	7.41
MS	141	43.97	17.02	16.31	1.42	17.02	15.60	1.42
MT	134	26.87	13.43	20.15	0.75	2.24	17.91	14.18
NC	184	18.48	5.43	3.80	3.80	7.07	4.35	4.35
ND	99	28.28	9.09	15.15	6.06	7.07	14.14	8.08
NE	139	16.55	7.91	2.88	0	2.88	4.32	4.32
NH	55	14.55	5.45	7.27	7.27	0	1.82	5.45
NM	71	35.21	7.04	9.86	9.86	15.49	5.63	11.27
NV	21	33.33	4.76	23.81	0	0	0	4.76
NY	235	24.68	5.96	8.51	2.98	14.89	1.28	5.96
OH	174	16.67	10.34	4.02	0	8.05	0	0
OK	186	48.39	24.19	11.83	2.15	4.84	17.20	10.22
OR	79	26.58	5.06	10.13	2.53	8.86	1.27	3.80
PA	148	27.03	12.16	11.49	1.35	10.14	2.70	6.08
SC	58	34.48	12.07	17.24	8.62	6.90	22.41	0
SD	91	37.36	17.58	4.40	15.38	8.79	8.79	12.09
TN	164	34.15	8.54	1.22	3.05	17.68	11.59	10.37
TX	377	40.85	15.92	10.61	17.77	6.63	9.55	9.28
UT	43	6.98	4.65	4.65	2.33	2.33	2.33	4.65
VA	162	25.31	6.79	12.35	6.79	3.09	3.09	1.23
VT	98	48.98	13.27	20.41	10.20	3.06	16.33	17.35
WA	95	52.63	22.11	6.32	18.95	13.68	13.68	9.47
WI	277	34.66	4.33	16.61	3.25	13.72	16.25	7.22
WV	87	44.83	22.99	27.59	3.45	9.20	20.69	8.05
WY	48	33.33	12.50	18.75	2.08	4.17	18.75	8.33

APPENDIX C

CORRELATIONS

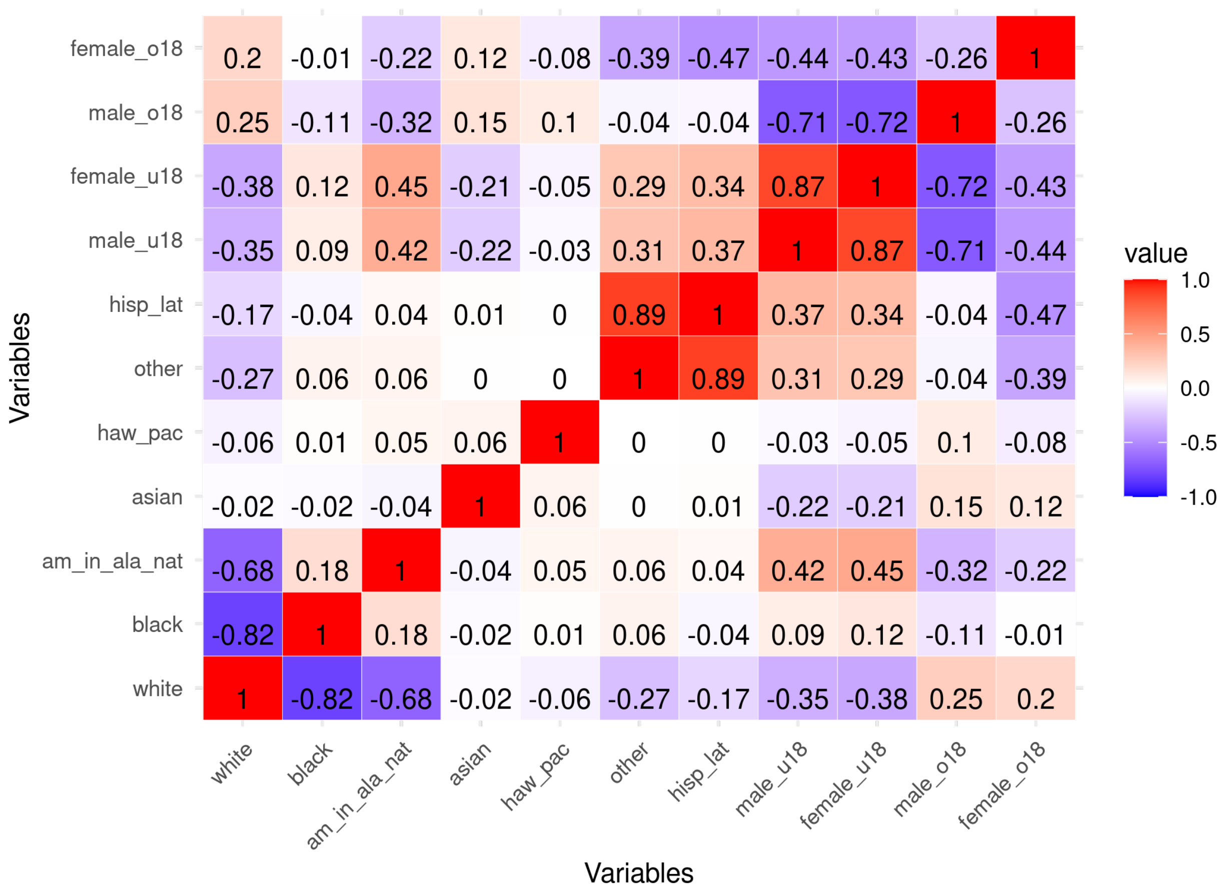

Figure C.1: Demographics Correlation Plot

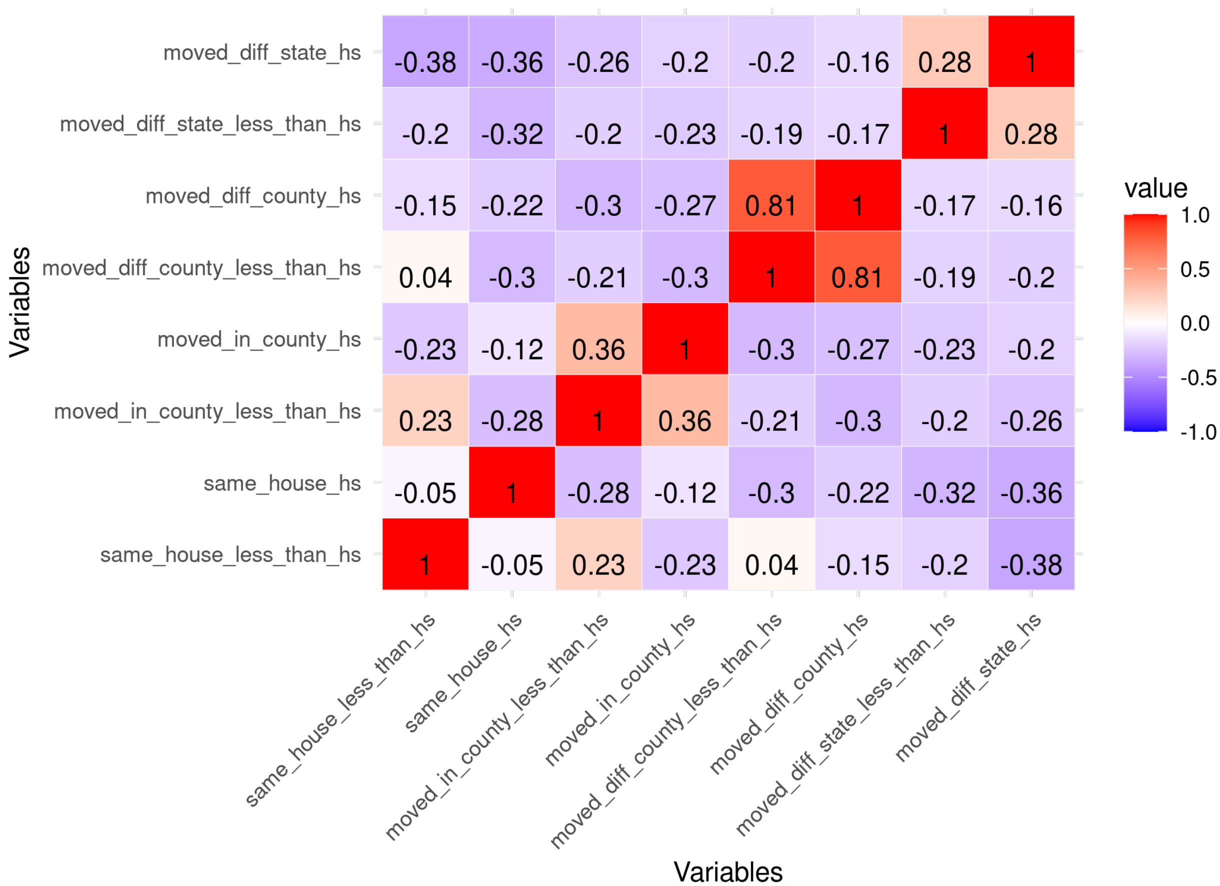

Figure C.2: RME Correlation Plot

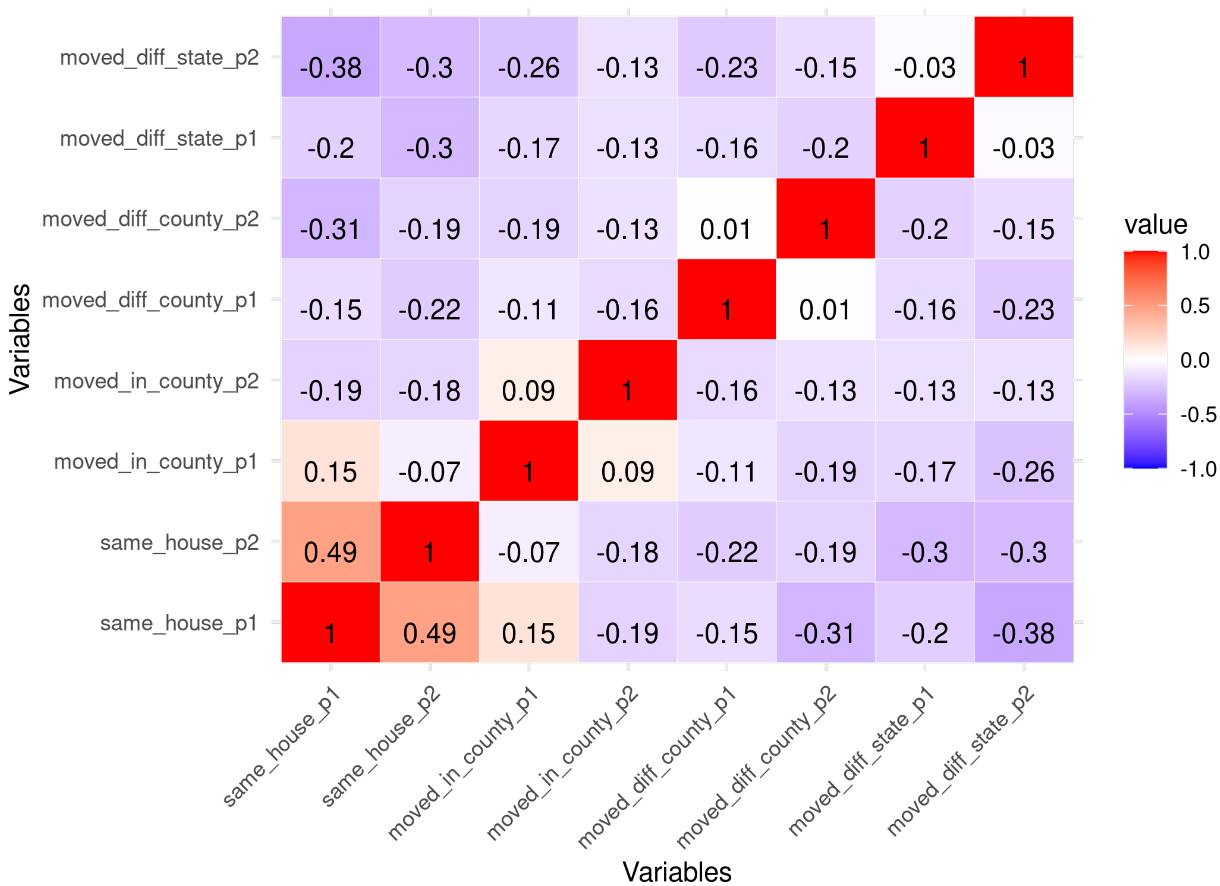

Figure C.3: RME Correlation Plot

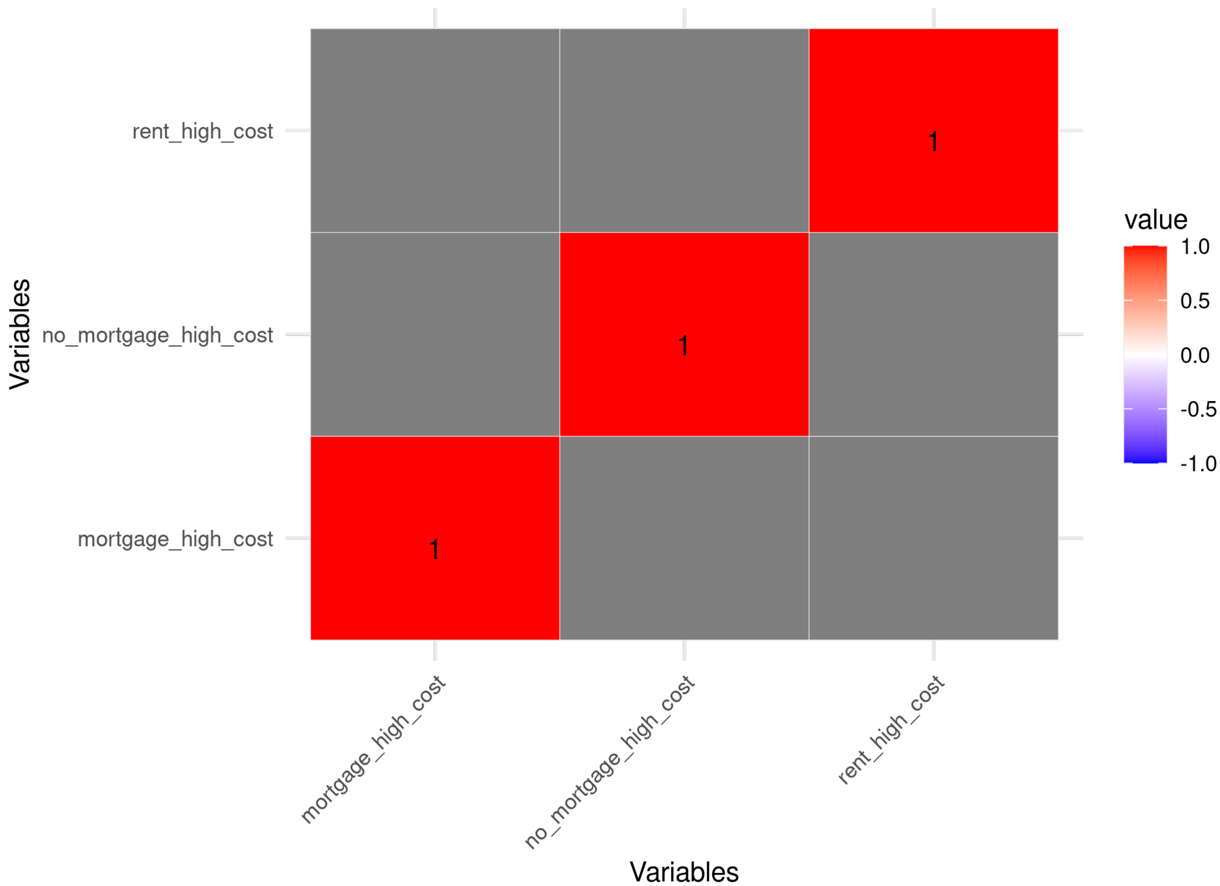

Figure C.4: Housing Costs Correlation Plot

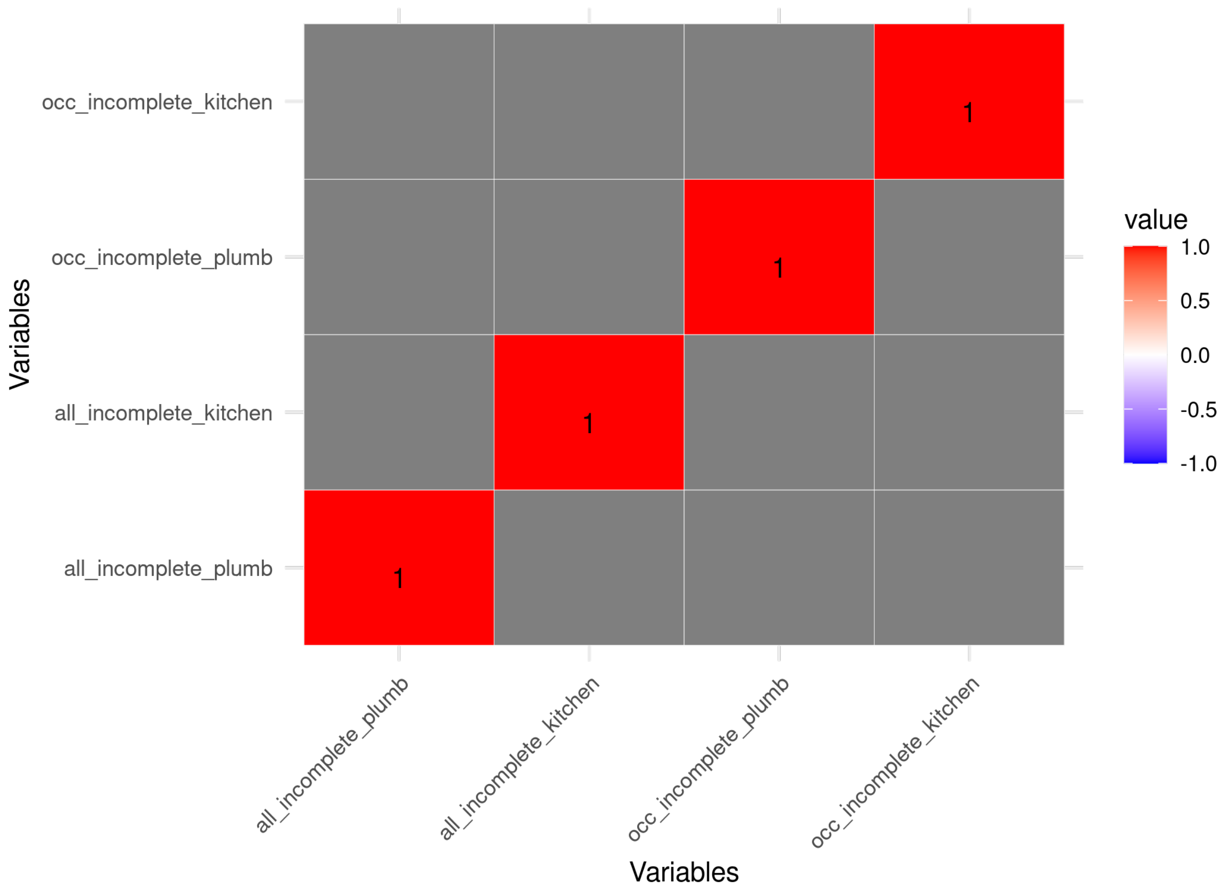

Figure C.5: Housing Quality Correlation Plot

BIOGRAPHY OF THE AUTHOR

Steve Garcia was born on April 15, 1997, in Dallas Texas. He obtained his GED in 2015. He attended Bossier Parish Community College from 2016 to 2019 where he obtained an Associate of Applied Science in Criminal Justice. He attended Lousiana Tech University and received a Bachelor of Arts in Political Science in 2021. He then enrolled in the Master of Science in Data Science and Engineering program in the Department of Spatial Information Science and Engineering at the University of Maine in September 2021. He competed in the International Public Debate Association for 5 years.

Steve Garcia is a candidate for the Master of Science degree in Data Science and Engineering from the University of Maine in May, 2024.