The Food Desert Reduction Roadmap:

A Food-Energy-Water Nexus Analysis for the Appalachian Region

A Thesis Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Angela Michelle Price

May 2025

Copyright © 2025 by Angela M. Price

All rights reserved.

Dedication

I'd like to dedicate this thesis to my family and friends. To my friends, thank you for your love and listening ears as I pursued through this with you. To my family, who have given everything for me to have the opportunity to choose education, there are no words to describe my gratitude. A special dedication to my parents: My mom, who has shaped me into the person I am, and taught me to take up space in every room I find myself in. My dad, who supported every endeavor and whim with a smile on his face (and for driving me to school every day as a kid). Love you always.

Acknowledgements

I want to thank my advisor, Dr. Jie Zhuang, who had full confidence in my research ideas and allowed me to mold this project into work that I am proud to be associated with. I also would like to thank the National Science Foundation (NSF) for funding this research through the Food Energy Water Nexus (FEWSUS) program. Many thanks to my committee members: Dr. Emine Fidan, Dr. Charles Cao, and Dr. Qiusheng Wu, for their continued advice and help. I have enjoyed my time as your student.

Abstract

The United States Department of Agriculture has defined a food desert as a census tract that lacks access to fresh, nutritious, and affordable food. While food deserts define a geographic or financial barrier to a food source for a community, food insecurity indicates the uncertain availability of food within the home. These communities often see other socioeconomic factors influencing their status as a food desert and rates of food insecurity. The Food-Energy-Water nexus is used as it incorporates these factors to address the root causes of food deserts, allowing in-depth analysis of these communities and their needs.

As climate change worsens, significant shifts in temperature and precipitation are disrupting agriculture and, subsequently, the food supply chains. Appalachian communities often depend on local farmers to substitute for grocery stores or provide affordable food. Rural food desert communities in Appalachia are disproportionately impacted, with nearly 75% of total farmland in the region being small-scale and family-owned. Between 2017 and 2022, the region lost 2.45 million acres of family-owned farmland, which calculates to 46.6 acres lost per hour.

In 2020, the food insecurity rates in Appalachia ranged from 6.2% to 25.9% throughout the region, with higher rates seen in the Southern and South Central subregions. Between 2010 and 2022, food insecurity rates have remained stagnant throughout the region, with an average rate of 15.0%, with the greatest impact on food insecurity being the median household income. Through the decade, the number of census tracts identified as both geographic and financial food deserts have risen by 53, climbing from 886 to 939. The relationship analysis between food insecurity and food deserts yielded that if a household was food insecure, they were 1.50 times more likely to also be in a geographic and financial food desert. This study used the trends of food deserts, the loss of farmland, and socioeconomic data to predict the next 5 years of the

region. Based on the predictions and the relationship between deserts and insecurity, the number of food deserts is expected to continue to increase between 2023 and 2027 from 13.9% to 15.5%.

Table of Contents

Chapter 1 Introduction 1

1.1 Challenges in Appalachia 1

1.2 Housing Development and Regional Farmland Loss 2

1.3 Food Deserts and Food Insecurity in Appalachia 4

1.4 The Energy Industry and Utility Cost Burdens 6

1.5 Water Accessibility 8

1.6 Adopting the FEW Nexus in This Research 9

1.7 Review of Previous Projects and Aims of This Research 9

References 12

Chapter 2 Hotspot Analysis of Socioeconomic, Food Accessibility, and Farmland Trends in the Appalachian Region 15

Abstract 16

2.1 Introduction 17

2.2 Methods 20

2.3 Results 22

2.3.1 Poverty and Income 22

2.3.2 SNAP Recipients 23

2.3.3 Housing and Transportation 24

2.3.4 Disability and Insurance 26

2.3.5 Farmland and Food 26

2.4 Discussion 27

2.5 Conclusion 30

References 31
Appendix A 32
Chapter 3 Time-Series Trend Analysis of Appalachian Socioeconomic Factors 65
Abstract 66
3.2 Methods 68
3.3 Results 71
3.3.1 Yearly Linear Regressions and Principal Component Analysis 71
3.3.2 Random Forest Importance Ranking of Influential Food Insecurity Variables 76
3.3.3 Correlations between Food Insecurity and Food Deserts 77
3.4 Discussion 79
3.5 Conclusions 80
References 82
Appendix B 83
Chapter 4 Predictive Modeling of Food Insecurity Values 104
Abstract 105
4.2 Methods 107
4.3 Results 110
4.3.1 Single Timestep 110
4.3.2 Rolling Five Timesteps 111
4.4 Discussion 112
4.5 Conclusion 113
References 115
Appendix C 116

Chapter 5 Conclusions 122

5.1 Key Findings 122

5.2 Future Research Directions 123

Vita 125

List of Tables

Table 2.1. Socioeconomic variables chosen for this study. 20

Table A.1. Hotspot analysis summary statistics calculated by the Getis-Ord Gi* tool in ArcGIS Pro……..32

Table A.2. Appalachian Farmland Acreage from 2012-2022..... 33

Table A.3. Food Desert Trends in Appalachia from 2010-2019. 34

Table B.1. Contingency table calculations, food insecurity to food desert risk ratio……83

Table B.2. Yearly linear regression model summary statistics……..84

Table B.3. Variable Importance on Food Insecurity Ranked. 85

Table B.4. Regression statistics when variable is compared against food insecurity rates. 86

List of Figures

Figure A.1. Average overall poverty rate in Appalachia hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 35

Figure A.2. Average household poverty rate in Appalachia hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 36

Figure A.3. Median household income in Appalachia hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 37

Figure A.4. Average percentage of total Appalachian households receiving Supplemental Nutrition Assistance Program (SNAP) hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 38

Figure A.5. Average percentage of Appalachian households that receive Supplemental Nutrition Assistance Program (SNAP) that are below the poverty line hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 39

Figure A.6. Average percentage of total households that receive Supplemental Nutrition Assistance Program (SNAP) and are below the poverty line hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 40

Figure A.7. Average unemployment rate in Appalachia hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 41

Figure A.8. Average percentage of Appalachian households lacking complete plumbing hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 42

Figure A.9. Average percentage of households in Appalachia lacking transportation hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 43
Figure A.10. Average percentage of occupied rental units with excessive housing costs hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 44
Figure A.11. Average percentage of occupied mortgage units with excessive housing costs hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 45
Figure A.12. Average percentage of occupied housing units without a mortgage that have excessive housing costs hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 46
Figure A.13. Average disability rate in Appalachia hotspot analysis from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 47
Figure A.14. Average percentage of individuals in Appalachia lacking health insurance hotspot analysis from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022. 48
Figure A.15. Total farmland acreage in Appalachia in 2012. Data from the United States Department of Agriculture National Agricultural Statistics Service 2012. 49
Figure A.16. Total farmland acreage in Appalachia in 2017. Data from the United States Department of Agriculture National Agricultural Statistics Service 2017. 50
Figure A.17. Total farmland acreage in Appalachia in 2022. Data from the United States Department of Agriculture National Agricultural Statistics Service 2022. 51

Figure A.18. Family-owned farmland acreage in Appalachia in 2012. Data from the United States Department of Agriculture National Agricultural Statistics Service 2012..... 52
Figure A.19. Family-owned farmland acreage in Appalachia in 2017. Data from the United States Department of Agriculture National Agricultural Statistics Service 2017..... 53
Figure A.20. Family-owned farmland acreage in Appalachia in 2022. Data from the United States Department of Agriculture National Agricultural Statistics Service 2022..... 54
Figure A.21. Average food insecurity rates in Appalachia from 2010 to 2022. Data from the Feeding America organization's Map the Meal Gap Data Download for 2010 to 2022. 55
Figure A.22. Appalachian geographic food deserts in 2010. Data from the United States Department of Agriculture's Food Access Research Atlas in 2010. 56
Figure A.23. Appalachian geographic food deserts in 2015. Data from the United States Department of Agriculture's Food Access Research Atlas in 2015. 57
Figure A.24. Appalachian geographic food deserts in 2019. Data from the United States Department of Agriculture's Food Access Research Atlas in 2019. 58
Figure A.25. Appalachian financial food deserts in 2010. Data from the United States Department of Agriculture's Food Access Research Atlas in 2010. 59
Figure A.26. Appalachian financial food deserts in 2015. Data from the United States Department of Agriculture's Food Access Research Atlas in 2015. 60
Figure A.27. Appalachian financial food deserts in 2019. Data from the United States Department of Agriculture's Food Access Research Atlas in 2019. 61
Figure A.28. Appalachian geographic and financial food deserts in 2010. Data from the United States Department of Agriculture's Food Access Research Atlas in 2010..... 62

Figure A.29. Appalachian geographic and financial food deserts in 2015. Data from the United States Department of Agriculture's Food Access Research Atlas in 2015. 63
Figure A.30. Appalachian geographic and financial food deserts in 2019. Data from the United States Department of Agriculture's Food Access Research Atlas in 2019. 64
Figure B.1. Food insecurity rates in Appalachia from 2010 to 2022. Data from Feeding America's Mind the Meal Gap Database……...87
Figure B.2. Overall poverty rates in Appalachia from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022..... 88
Figure B.3. Household poverty rates in Appalachia from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. 89
Figure B.4. Median household incomes in Appalachia from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. 90
Figure B.5. Unemployment rate in Appalachia from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022..... 91
Figure B.6. Number of Appalachian households lacking transportation from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. 92
Figure B.7. Percentage of Appalachian households that receive SNAP from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. Supplemental Nutrition Assistance Program (SNAP). 93
Figure B.8. Percentage of Appalachian households that receive SNAP that are below the poverty line from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. Supplemental Nutrition Assistance Program (SNAP). 94

Figure B.9. Percentage of total Appalachian households that receive SNAP and are below the poverty line from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. Supplemental Nutrition Assistance Program (SNAP)..... 95
Figure B.10. Percentage of rental units with excessive housing costs from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. 96
Figure B.11. Percentage of housing units with a mortgage that have excessive housing costs from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022..... 97
Figure B.12. Percentage of housing units without a mortgage that have excessive housing costs from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. 98
Figure B.13. Socioeconomic variables ranked by influence on food insecurity. Calculated using a Random Forest model in R. 99
Figure B.14. Linear regression between median household income and food insecurity in Appalachia. 100
Figure B.15. Linear regression between rental units with excessive housing costs and food insecurity in Appalachia. 101
Figure B.16. Linear regression between households that lack transportation and food insecurity in Appalachia. 102
Figure B.17. Linear regression between household poverty rate and food insecurity in Appalachia. 103

Figure C.1. Loss and validation loss plots for the Long Short-Term Memory (LSTM) model built for food insecurity. Loss metrics used were mean absolute error (first graph) and mean standard error (second graph)……116

Figure C.2. Predicted versus actual food insecurity rates for single time step LSTM validation……..117

Figure C.3. Food insecurity rates in Appalachia from 2010 to 2027. Data from Feeding America and the U.S. Census. Years 2023 to 2027 were predicted using a single time step long short-term memory model. 118

Figure C.4. Loss and validation loss plots for the Long Short-Term Memory (LSTM) model built for food insecurity. Loss metrics used mean absolute error (first graph) and mean standard error (second graph)..... 119

Figure C.5. Predicted versus actual food insecurity rate for rolling five time step long short-term memory model validation. 120

Figure C.6. Food insecurity rates in Appalachia from 2010 to 2027. Data from Feeding America and the U.S. Census. Years 2023 to 2027 were predicted using a rolling five time step Long Short-Term Memory model. 121

Chapter 1

Introduction

1.1 Challenges in Appalachia

The Appalachian Region spans down the East Coast of the United States, including parts of twelve states: New York, Pennsylvania, Ohio, Maryland, Virginia, North Carolina, South Carolina, Tennessee, Kentucky, Georgia, Alabama, and Mississippi, as well as the entire state of West Virginia. The region is home to approximately 26.5 million people, with Southern Appalachia's population growth exceeding the national average (7.3%), with the rate climbing to 10.3% in 2021 and is expected to continue increasing (Appalachian Regional Commission, 2023). Despite the growing population, the infrastructure and industry in the region are falling behind, leading to complex and isolating economic struggles. The lack of long-term solutions in Appalachia stems from the decline in existing industries (mining, manufacturing, textiles, and lumber). As these issues are dissected, it is important to relate how socioeconomic struggles are intrinsically interconnected with the Food-Energy-Water nexus.

Through the rise and fall of industrialization throughout the United States, the struggle in the Appalachian region remains consistent and prominent. Nearly a century and a half ago, the region underwent massive industrialization that soon evolved into the U.S. relying on the Appalachian region to provide an unlimited supply of coal and lumber. When these resources began to dwindle, it was too late to turn back, and the lasting effects of the dying industries spread through the region.

Previous industries in the region ranged from mining to manufacturing, textiles, and lumber. Between 2010 and 2022, the number of people employed in the Appalachian coal mining industry fell from 56,000 to about 27,000 (Appalachian Regional Commission, 2023).

This level of unemployment is consistent throughout the region, with unemployment rates consistently above the national average. The unemployment rate throughout Appalachia ranges between 3.1%-15.7%, with the highest rates being in central Appalachia, Kentucky, and West Virginia, seeing a hit due to the loss of coal mining, compared to the U.S. average of 4.4% (Appalachian Regional Commission, 2023). The median household income in the region was $10,000 less than the United States average ($42,000) in other rural communities (Appalachian Regional Commission, 2023). This, coupled with the rampant poverty rates (ranging from 13%-20% throughout the region, the national average being 11.5%), there is no question that the region is experiencing a widespread hardship that isn't seen anywhere else in the country at this scale (Appalachian Regional Commission, 2023). Long-term income instability and a lack of governmental support lead to a rise in debt and homelessness.

1.2 Housing Development and Regional Farmland Loss

The United States Department of Housing and Urban Development (HUD) has its gold standard of monthly costs- the thirty percent rule. The HUD reports that for housing to be classified as affordable, the monthly housing costs must be no more than thirty percent of the household's monthly income (Office of Policy Development and Research, 2014). This standard came as a result of the Housing and Urban Development Act of 1965. This act and the Brooke Amendment capped the cost of public housing rent to twenty-five percent of the resident's income, which was then increased to thirty percent in 1981. The 1968 HUD Act is regarded as a major success that accomplished nearly everything it was designed to do in the face of a national urban crisis. Over the four years that it was fully enacted: around 400,000 new homes were built for low-income families, about 500,000 apartments were developed, discrimination in the sale or renting of housing was prohibited, and low-income families were able to save for a down

payment and apply for mortgages (United States Department of Housing and Urban Development, n.d.) This led to increasing homeownership and reducing long-term renting. In 1973, President Richard Nixon placed a moratorium on the program, which was not lifted until the passing of the Housing and Community Development Act of 1974. This act created "Section 8" housing, providing rental assistance to low-income households in the form of voucher payments to landlords (Morris, 1974). Section 8 was designed to be more productive than the 1968 Act, however, a widely agreed criticism of this act, especially in communities like Appalachia, is that landlord participation is voluntary, some states allow for income discrimination, and the national average waitlist for these vouchers is nearly three years (Hoffman, 2014).

Additionally, the region is home to almost a quarter of a million farms with more than 36 million acres dedicated to agriculture; however, between the years 2007 to 2017, the region lost farmland at a higher rate than the entire United States (Karp, et. al., 2022). Rural Central and Southern Appalachia lost the most farmland due to urbanization in comparison to the rest of the region and the country. Appalachian farms are characterized by their size, as the farms are typically utilized by small-scale family farmers only looking to provide for their families; the farms are much smaller than the average farm in the U.S. The main "for-profit" agriculture the region sees is through animal products (such as meat, milk, eggs, etc.), pulling in nearly 14.7 billion dollars in crop sales (Karp, et. al., 2022). This loss of agricultural land creates a shift in the region's food landscape and economic resilience. As local farmland continues to decrease, attributed to the rising urbanization, the number of food deserts and the likelihood of food insecurity increases. Losing farms that provide a close-to-home access point to fresh food

products, while also helping the local economy, worsens the overall problem throughout the region: limited access to food.

1.3 Food Deserts and Food Insecurity in Appalachia

Appalachia has historically high poverty rates, extremely rural communities, coupled with a rapid rate of urbanization, creating an uphill battle of socioeconomic issues. Specifically, Southern Appalachia. Specifically, has a population growth exceeding the national average (7.3%), with the rate climbing to 10.3% in 2021 and is expected to continue increasing (Appalachian Regional Commission, 2023). These population booms are due to the influx of people moving to the region for the lower cost of living, driving up the cost of living, and making it difficult for the people of the region to hold steadfast in the face of change. The growing population, poverty rates, and rural communities make the region a prime target for food insecurity. An average of 13% of the Appalachian population (around 4 million people) is experiencing food insecurity, with food access being severely limited, both geographically and financially (Appalachian Regional Commission, 2023). The majority of those who are living in food insecurity are also living in food deserts- meaning that it's unknown when they'll be able to afford food again within their household, *and* it is difficult to access affordable food sources within their communities.

The United States Department of Agriculture (USDA) defines a food desert as a census tract that lacks access to nutritious and affordable food. The United States has over 23 million people who live in a USDA-defined food desert, and nearly 50 million living in food insecurity due to financial hardship (Miller, et. at., 2016). Nearly 4 million of these people are located in the Appalachian mountains, with food insecurity rates being as high as 27% in some parts of the region in 2022 (Miller, et. at., 2016; Feeding America, 2022). Food insecurity is defined as a

household-level limitation that makes access to food unstable and unpredictable, and it often overlaps with food deserts to create both household and community-level challenges. These areas face compounded barriers, not only in terms of affordability and geographic accessibility to fresh food sources, but also in securing other essential resources. In many Appalachian communities, struggles extend beyond food, as access to clean water and reliable energy further threaten overall quality of life and public health (Miller, et. al., 2016).

The United States has tried several methods of addressing food deserts and food insecurity. The USDA has partnered with the Department of the Treasury and the Department of Health and Human Services to provide funding to projects working to increase the number of farmers' markets, food hubs, and universal school lunches. The Food Stamp Act of 1965 kicked off a long history of addressing food insecurity at the federal level. Food stamps were replaced by SNAP, Supplemental Nutrition Assistance Program, in 2008, which assists low-income households in buying food and food products through the use of electronic benefit transfer (EBT) as long as the head of the household is registered for work and meets current state requirements. The Emergency Food Assistance Program (TEFAP) provides emergency short-term food assistance through partnerships with local organizations, such as Feeding America. Federal food assistance comes in many different shapes and sizes, and while these programs help millions of people across the country, the reason for these programs' existence has struggled to be confronted. The main cause of this is the Pandora's box that lies beneath the surface of food assistance. As explored through the coming chapters, food inaccessibility is a multidimensional problem that requires attention to other socioeconomic issues such as poverty, housing costs, transportation, and so on. “Fixing” food accessibility goes much further than opening up food hubs or grocery stores; it is an unraveling of deeply intertwined hardships at the home and community level.

1.4 The Energy Industry and Utility Cost Burdens

The demand for coal is steadily decreasing as the cost of natural gas is declining. Between 2005 and 2015, overall coal production in the region fell by 45 percent. As this demand decreases and new industries are outsourced to other states, the industry supply chain and economy throughout the region are hit (Jackson, et. al., 2018). The appeal of Appalachian coal was how extensive the coal was throughout the mountains, making mining inexpensive and easy to access. Coal is becoming increasingly difficult to find throughout the Appalachian region after years of extensive mining. Most coal is mined in the western portion of the U.S., where mining is harder and more expensive (Hodge, 2016). Additionally, the manufacturers in Appalachia are small and lack the capabilities and access to get involved in the supply chains for renewable energy production. These manufacturers face difficulties with dated technology, old facilities, and a lack of employment and training (Hodge, 2016). This is coupled with the fact that the United States has seen an acceleration in energy efficiency, demand for renewable energy, and a surge in consumption. The country is seeing a change in how electricity is produced and consumed, and how the supply chains are being adapted to support these beneficial changes. Americans are seeing the creation of new jobs in renewable and clean energy while also advocating for new legislation protecting the quality of our air and water. The COVID-19 pandemic sparked an economic fallout not only in the United States but worldwide. In 2020, energy employment fell at a rapid rate, losing jobs faster than any other governmental field as a whole. Despite this, around 71% of the jobs lost have been recovered, and for the better (United States Department of Energy, 2023). The distribution of these jobs pushed employment toward newer, cleaner technologies such as solar and wind, accounting for nearly 87% of the new electric jobs (United States Department of Energy, 2023). The Department of Energy (DOE)

reported that the overall growth rate for clean energy jobs climbed almost 4% in 2022, outpacing every other U.S. workforce, creating over 8 million jobs in this sector (United States Department of Energy, 2023).

This work is pushing toward the DOE's goal to reach net-zero greenhouse gas emissions by 2050. Legislatively, the creation of new jobs can be attributed to the Partnerships for Opportunity and Workforce and Economic Revitalization (POWER) Initiative under the Obama Administration, and the Infrastructure Investment and Jobs Act signed into law in 2021 (United States Economic Development Administration, 2016). These are major advancements toward reviving economic growth and infrastructure change in the United States, but through this, the Appalachian region, once notorious for its coal industry ecosystem, has found itself slipping further and further away from the booming economy and revitalizing infrastructure that is being promised to the rest of the country.

The DOE has defined energy burdens as the high percentage of the gross household income that is spent on energy costs (United States Department of Energy, 2024). This primarily affects low-income households in rural America, making Appalachia a hotspot for energy burdens. Nationally, the energy burden in non-low-income urban households is around 3.1%; however, in low-income rural areas, the median energy burden is almost 9%. Of these low-income rural households, 41% had incomes below the federal poverty level (Drehobl, et. atl., 2018). This results in families spending large sums of their monthly income on necessities such as heating and cooling, especially as climate change causes harsher seasonal patterns (Drehobl, et. al., 2018). The DOE has put together resources for low-income households experiencing these high energy burdens. These resources, however, focus on planning for low-income households to meet the bare necessities of electricity. As an example, under the Responsible Utility Customer

Protection Act, the United States has laws stating that the electric company must allow residential sites to remain at or above 68°F and cannot legally shut off power when the temperature is below freezing (United States Department of Energy, 2024). While laws and regulations such as these are necessary and aid in protecting low-income residents, the bare minimum does not result in an adequate standard of living (United States Department of Energy, 2024).

1.5 Water Accessibility

Water security is the "ability to access and benefit from affordable, adequate, reliable, and safe water for well-being and a healthy life" (Wies, et. al., 2020). Water insecurity is when any of those factors are not met, and the well-being and health of a person are jeopardized due to their water quality. As expected, high poverty rates and water insecurity go hand-in-hand. Communities in Appalachia struggle to find clean water, and some don't have access to running water, toilets, or showers in their homes. Appalachia's water access and quality have been severely impacted over time due to industry (contamination from mines or factories) and the failing infrastructure. In areas that were used for coal mining, common contaminants found in the groundwater are heavy metals such as arsenic, copper, and lead (Wies, et. at., 2020). Long-term ingestion of these contaminants leads to cancer, heart damage, endocrine disruption, and serious neurological birth defects (Wies, et. at., 2020). Most infrastructure in Appalachia was built when the industry moved into the area. Many of the water systems are nearly a hundred years old and don't see regular maintenance (Wies, et. al., 2020). The rural communities being spaced out have made it difficult to ensure city water system access to those living outside of city limits. The American Society of Civil Engineers estimates that each Appalachian state will require billions for water infrastructure repair to ensure safe drinking water in each household (Wies, et. al.,

2020). Current projects that are aiming to update this infrastructure. For example, the recent Infrastructure Investment and Jobs Act is ensuring that West Virginia and Kentucky will receive over half a billion dollars each over a set 5-year plan. This money will go towards rebuilding the infrastructure and continuous testing of the water in the area. While more funding will be needed for additional states to complete the projects, this plan is the first step in rebuilding the crumbling infrastructure (Born, 2021).

1.6 Adopting the FEW Nexus in This Research

It is imperative to examine the issue of food insecurity and inaccessibility through the lens of the Food-Energy-Water (FEW) nexus. The Food-Energy-Water (FEW) nexus is a framework of how the three factors tie into each other, and decisions affecting one will have cascading effects on the other two. A unique piece of this framework is that it includes socioeconomic issues, whereas other frameworks focus solely on environmental issues. Socioeconomic issues impact the three food, energy, and water systems

It is imperative to examine the issue of food insecurity and inaccessibility through the lens of the FEW nexus. This framework allows the connection between food deserts, water insecurity, and energy and housing burdens to be made, as well as making the connection between these issues and the decline of farmland.

1.7 Review of Previous Projects and Aims of This Research

These issues are widespread and intense, but not insurmountable. Through the Bipartisan Infrastructure Law (BIL), the ARC has launched its Appalachian Regional Initiative for Stronger Economies. This initiative has awarded around 50 million dollars to multi-state projects to invest in local infrastructure projects. Across the region, the ARISE project is focusing on moving forward into the green energy sector, getting Appalachia back to the powerhouse it once was.

Additionally, they have invested in projects funding green textile manufacturing, constructing and opening a textile hub in North Carolina that will later be used to establish a manufacturing supply chain that connects southern Appalachia. BIL has allowed for collaborative projects that are creating new jobs and supply chains that will connect Appalachian cities to other states, reconnecting the region to resources and opportunities (The United States White House, 2024; Appalachian Regional Commission, 2023).

Obtaining the funding to improve and reconnect infrastructure has been a monumental milestone for the ARC, however, the issues they are beginning to address are deeply intertwined. Many of these initiatives and projects fail or fall short of the ultimate goal. The primary issue seen within these projects is failing to understand the root of the problem, which ultimately leads to no progress. The population growth rate in Appalachia is higher than anywhere else in the country. This is due to the influx of people moving to the region for the low cost of living but ultimately leads to a widening gap between wealth classes and local people being driven out of their homes due to the increased costs. Without understanding how the cycle of poverty and the absence of industry work, these projects won't succeed for the Appalachian people. The Bipartisan Infrastructure Law is a great start to addressing and solving these issues, however, it's not enough. As an example, the BIL provided $200 million annually, it is estimated that fixing the water quality and infrastructure in West Virginia alone will cost close to $2 billion (Adams, 2023). These problems are caused by over a century of industrialization to the point of depletion and will take much longer than 5 years of funding to solve.

There is a lack of research and updated information on the issues in the Appalachian region. The majority of research is outdated or relies heavily on the Appalachian Regional Commission to publish its findings. Creating models of the socioeconomic issues in the Appalachian region

allows for an in-depth relationship analysis between the issues to propose more tailored solutions specific to the region and even local areas. Through historical FEW data and relationship analysis, a predictive neural network can be created to predict the coming years to see problem areas and get ahead of the curve. A model that can predict what will happen to food deserts, poverty rates, unemployment, etc., can be a tool as population growth continues and food insecurity rates rise.

References

Population and age in Appalachia. Appalachian Regional Commission. (2023, June 8). https://www.arc.gov/about-the-appalachian-region/the-chartbook/appalachias-population/#:~:text=Appalachia%20had%20a%20population%20of,the%20national%20average%20of%207.3%25.

United States Department of Energy. (2023). U.S. Energy & Employment Jobs Report (USEER). Energy.gov. https://www.energy.gov/policy/us-energy-employment-jobs-report-useer

Power initiative. POWER Initiative | U.S. Economic Development Administration. (2016). https://www.eda.gov/archives/2016/power/

Jackson, R., et.al. (2018). An Economic Analysis of the Appalachian Coal Industry Ecosystem: Summary Report. Appalachian Regional Commission. https://www.arc.gov/wp-content/uploads/2020/06/CoalIndustryPowerGenerationandSupplyChainReport.pdf

Hodge, D. (2016, March). Appalachian Coal Industry, Power Generation, and Supply Chain. Appalachian Regional Commission. https://www.arc.gov/wp-content/uploads/2020/06/CoalIndustryPowerGenerationandSupplyChainReport.pdf

Karp , K. et. al. (2022, April). *Agriculture and Local Food Economies in the Appalachian Region*. Appalachian Regional Commission. https://www.arc.gov/wp-content/uploads/2022/04/Agriculture-and-Local-Food-Economies-in-the-Appalachian-Region-April-2022.pdf

Population and age in Appalachia. Appalachian Regional Commission. (2023, June 8). https://www.arc.gov/about-the-appalachian-region/the-chartbook/appalachias-

population/#:~:text=Appalachia%20had%20a%20population%20of,the%20national%20average%20of%207.3%25.

Rental burdens: Rethinking affordability measures. Rental Burdens: Rethinking Affordability Measures | HUD USER. (2014, September 22). https://www.huduser.gov/portal/pdredge/pdr_edge_featd_article_092214.html

History of fair housing - HUD. HUD.gov / U.S. Department of Housing and Urban Development (HUD). (n.d.). https://www.hud.gov/program_offices/fair_housing_equal_opp/aboutfheo/history

Von Hoffman, A. (2014, October 8). *LBJ's biggest housing program that no one remembers*. LBJ's Biggest Housing Program that No One Remembers | Joint Center for Housing Studies. https://www.jchs.harvard.edu/blog/lbjs-biggest-housing-program-that-no-one-remembers

Miller, W. C., Rogalla, D., Spencer, D., Zia, N., Griffith, B. N., & Heinsberg, H. B. (2016). Community adaptations to an impending food desert in rural Appalachia, USA. Rural and Remote Health, 16(4), 1–12. https://search.informit.org/doi/10.3316/informit.206751467707439

Low-Income Community Energy Solutions. Office of State and Community Energy Programs. (2024). https://www.energy.gov/scep/slsc/low-income-community-energy-solutions

Ross, L., Drehobl, A., & Stickles, B. (2018, July). *The high cost of energy in rural America:* ACEEE Publications. https://www.aceee.org/sites/default/files/publications/researchreports/u1806.pdf

Wies, J.R., Mays, A., Collins, S.M. and Young, S.L. (2020), “As Long As We Have the Mine, We'll Have Water”: Exploring Water Insecurity in Appalachia. Annals of Anthropological Practice, 44: 65-76. https://doi.org/10.1111/napa.12134

Born, M. (2021, December 27). *How the infrastructure package could impact Appalachians living with failing water systems*. North Carolina Health News. https://www.northcarolinahealthnews.org/2021/12/28/how-the-infrastructure-package-could-impact-appalachians-living-with-failing-water-systems/

The United States Government. (2024, February 6). *Guidebook to the Bipartisan Infrastructure Law*. The White House. https://www.whitehouse.gov/build/guidebook/#:~:text=The%20Bipartisan%20Infrastructure%20Law%20makes,the%20nation’s%20lead%20service%20lines.

Bipartisan Infrastructure Law in Appalachia. Appalachian Regional Commission. (2023, November 15). https://www.arc.gov/BIL/#:~:text=On%20November%2015th%2C%202021%2C%20President,sustainable%2C%20resilient%2C%20and%20just.

Adams, M. (2023, November 17). *Virginia Tech study explores lack of water access in Appalachian communities*. West Virginia Public Broadcasting. https://wvpublic.org/virginia-tech-study-explores-lack-of-water-access-in-appalachian-communities/

Chapter 2

Hotspot Analysis of Socioeconomic, Food Accessibility, and Farmland Trends in the Appalachian Region

Abstract

The Appalachian Region spans over thirteen states stretching down the East Coast. Predominately rural, the region is experiencing elevated socioeconomic hardships, food insecurity rates, food desert expansion, and farmland loss. Using the U.S. Census Bureau's economic and demographic data, Feeding America's Mind the Meal Gap, and the U.S. Department of Agriculture's Food Research Atlas and Agriculture Survey, the factors of interest were mapped out in ArcGIS Pro from 2010 to 2022. These factors include: overall and household poverty rates where the household poverty rate is the distinction made for the entire household based on combined incomes and overall poverty rate being based on individual earners. Additional factors were, lacking a vehicle, disability rate, median household income, SNAP recipients, unemployment rates, lacking complete and reliable access to running water facilities, health insurance coverage, and the monthly costs of living exceeding the Department of Housing and Urban Development's thirty percent rule. After displaying the yearly data, the datasets were aggregated across the 2010 to 2022 period by averaging the values of each factor, which were then used to perform a hotspot analysis identifying areas with persistently high levels of vulnerability. The Hotspot Analysis (Getis-Ord Gi*) tool in ArcGIS Pro was used to evaluate the spatial dependence of various factors, helping to identify their regional impact and highlight areas of concern. It was found that the Central subregion had the most concentrated poor socioeconomic hotspots, and the Southern subregion had the highest acreage of farmland. Two of the socioeconomic factors, disability and lack of health insurance, were found to be spatially dependent. Food insecurity was determined to be spatially dependent, averaging 13.5%, with hotspots spanning across the entirety of the Southern and Central subregions. The number of combination geographic and financial food deserts has increased by 65 census tracts between

2010 and 2019, encompassing 15.8% of the total population, just over 4 million people. Concerningly, farmland has seen a sharp decline, total farmland increasing by 98,768 acres between 2012 and 2017, but then steeply declining by 2.42 million acres between 2017 and 2022. The connection between food deserts and food insecurity was confirmed through spatial dependency, with poverty holding notable influence on the two.

2.1 Introduction

The counties and inclusionary census tracts of the Appalachian Region are based on the Appalachian Regional Commission (ARC), a federal-state partnership that operates with the goal of socioeconomic progression in the region. The ARC definition of the Appalachian Region reflects an *economic* region, as opposed to *cultural* Appalachia. This broader economic region extends beyond the mountains to include nearby foothill areas that experience comparable socioeconomic hardships as the communities within the mountains. This study area includes 423 counties and, as of 2020, 7,241 census tracts across 13 states. Although predominantly rural, it is important to take note of the cities that are scattered throughout the region. These cities include Pittsburgh, PA; Asheville, NC; Knoxville, TN; Chattanooga, TN; Greenville, SC; Birmingham, AL; and Tuscaloosa, AL. Along with these cities, it's important to mention the effects of urban sprawl through proximity to other urban centers such as Albany, NY; Cincinnati, OH; Columbus, OH; Lexington, KY; and most notably Atlanta, GA. These cities and the subsequent urban sprawl drive the region's rapid urbanization.

These factors were chosen for this study through the lens of influence on food insecurity and food deserts. The socioeconomic factors selected were overall and household poverty rates where the household poverty rate is the distinction made for the entire household based on combined incomes and overall poverty rate being based on individual earners. Additional factors

were, lacking a vehicle, disability rate, median household income, SNAP recipients, unemployment rates, lacking complete and reliable access to running water facilities, health insurance coverage, and the monthly costs of living exceeding the Department of Housing and Urban Development's thirty percent rule. The hypothesis is that low-income households and communities would see a higher rate of food insecurity and food deserts. Economically disadvantaged communities are less likely to attract full-service grocery retailers, resulting in a reduced access to affordable and healthy food options. The prevalence in lack of transportation in low income communities further exacerbates this issue, making it more difficult to reach food sources. However, having the ability to travel to these food sources is not a fix, many grocery retailers also partake in dynamic pricing which includes customer and community based data. If the community that the grocery store is located in has a higher median income the store is likely to have higher prices. If someone from a low-income household or community travels to a food source, it does not guarantee that they will be able to afford the dynamic pricing. According to the United States Department of Agriculture (USDA), food deserts are more likely to occur in communities with higher poverty rates, and those living in poverty are more vulnerable to food insecurity due to financial and geographic limitations (USDA, 2009). Additionally, it is necessary to include factors that would likely cause living expenses to be higher than average, such as a lack of health insurance or reliable running water, as households have to dedicate more of their monthly income to living conditions instead of higher quality and healthier foods.

Total farmland was included to incorporate the impacts of the loss of farmland in these communities. It's assumed that the majority of the farms in the region are small-scale operations and predominantly family-owned. The disappearing farmland is mostly attributed to farmers needing to sell their land due to increasing unaffordability, as related to the previously mentioned

urbanization and rising cost of living. As farmland is lost, the surrounding area is forced to rely solely on grocery stores that have to source their food from across the country and import it internationally. The "food footprint" increases, which refers to the resources and impacts of food making its way through the supply chain from farm to table. As a result, food prices increase, contributing to inaccessibility in low-income communities.

By properly addressing the factors listed above and their role in food insecurity and present food deserts, it is reasonably hypothesized that the coexisting socioeconomic factors that are associated with food insecurity are spatially dependent. For instance, urban areas typically have lower poverty rates due to the higher median incomes found within the "white collar" city as opposed to the "blue collar" rural areas. Understanding the role that geography plays in this study is key to setting up the following research chapters, as it influences the decisions, predictions, and suggestions that will be made and given.

The research questions addressed in this chapter:

1. What are hotspots in Appalachia that consistently experience elevated rates of hardship, as identified by socioeconomic indicators?
2. What are the strongest positive and negative correlations between the individual socioeconomic factors analyzed and the prevalence of food insecurity within the same time frame?
3. How does the spatial distribution of food deserts across the Appalachian Region overlap with the identified areas of socioeconomic hardship and areas experiencing significant farmland loss between 2012 and 2022?

2.2 Methods

The data regarding the socioeconomic factors, seen in **Table 2.1**, were retrieved through the United States Census Bureau in their 5-year American Census Survey (ACS). The ACS-5 Year Survey was used as it utilizes a larger sample size over 5-year time spans and produces more accurate estimates with smaller margins of error in comparison to the one-year estimates. The TIGER/Line geodatabase was downloaded for each of the thirteen states on the census tract level. These geodatabases include specific socioeconomic and demographic data collections. As the geodatabase for 2022 had not been released by the time of this study, 2022 data were collected individually from the ACS 5-year survey detailed tables and manually included into a geodatabase. Each yearly factor was collected from 2010 to 2022, except for the disability and health insurance data. These two factors were not publicly available in the 2010 and 2011 datasets and were collected from 2012 to 2022. Once downloaded, the data for each year was filtered down to the regions' census tracts using the Appalachian Regional Commissions (ARC) county shapefile as a boundary on ArcGIS Pro. As step one of this chapter, all of the factors were mapped out using the normalized graduated color system to visualize the data.

Table 2.1. Socioeconomic variables chosen for this study.

Food Insecurity	Lacking a Vehicle
Overall Poverty Rate	Lack of Health Insurance
Household Poverty Rate	Disability Rate
Median Household Income	SNAP Recipients
Monthly Housing Costs	Unreliable Access to Water Facilities

The food insecurity data was retrieved from Feeding America, the largest nonprofit food organization in the United States. They collect and estimate food insecurity levels on a county-wide scale across the country each year. Their methodology is similar to that of the U.S. Census, sending out surveys to communities and supplementing their data with estimates based on poverty, unemployment, etc. Their data from 2010 to 2022 was provided upon request and then was disaggregated down from county to the census tract level for analysis. The disaggregation was done using the Tidyverse package in R, and merging the census tract identification codes with the Federal Information Processing Series (FIPS) geographic identifications given to each county.

Food desert data was obtained by the USDA's Food Access Research Atlas. The USDA identifies food deserts based on the distance to the nearest food retailer, using thresholds of half, one, ten, and twenty miles. These measurements consider both geographic accessibility and financial barriers to food access, as well as areas affected by a combination of the two. This data is only collected and calculated every 4-5 years (2010, 2015, 2019). Due to the limited amount of food desert data, it was not used in a hotspot analysis, but rather just mapped out for visual analysis and count calculations.

Farmland data was retrieved from the USDA National Agricultural Statistics Service (NASS) through their census survey. The data collected was the total acreage of farmland in the 13 states, as well as the total acreage of family-owned farmland. Similar to the food desert data, this information is only collected every 5 years (2012, 2017, 2022), thus it was mapped out for a yearly visual analysis and count calculations.

Following this, the factor tables were exported out of ArcGIS Pro as an Excel spreadsheet where they were polished, simplifying the spreadsheet to include only necessary information

such as the census tract FIPS codes and the selected socioeconomic factors, and averaged throughout the 13 years. Each of the factors were analyzed using the Hotspot Analysis (Getis-Ord Gi*) tool in ArcGIS Pro. Gi* was chosen over the Optimized Hotspot Analysis because it allows for greater control over parameters such as spatial relationships and distance thresholds, which is particularly important as Appalachia contains both densely populated urban centers and sparsely populated rural areas. The Getis-Ord Gi* tool calculates statistically significant hot and cold spots using Z-scores and P-values derived from the normal distribution. The number of neighbors (NNeighbors) was set to 75, ensuring that each feature was evaluated in the context of its 75 neighbors. Evaluating the features against its neighbors balances sensitivity seen in broad regional patterns. The Optimized Hotspot Analysis tool automates these settings, which obscures spatial distinctions in diverse geographic areas like Appalachia.

2.3 Results

2.3.1 Poverty and Income

All figures can be found in the Chapter Appendix below. The hotspot analysis, seen in **Figure A.1**, for the overall poverty rate showed that the regional average poverty rate over the thirteen years was 15.81% with a median of 13.98%. This result indicates that the majority of census tracts had a poverty rate closer to the median, with a few tracts pulling the data to a right-skewed distribution with higher rates of poverty over time. This is supported by the Z-score values. Both the mean and median Z-score were negative (-0.58 and -0.23, respectively), which confirms that overall there are more data points below the mean. In other words, many census tracts were located in areas where poverty is lower than expected, showing as “cold spots” in the hotspot analysis. The P-value, calculated by the Gi* statistic in ArcGIS Pro, for the dataset was 0.099, indicating that there was no statistical significance in support of the spatial dependence

hypothesis. Similarly, the household poverty rate overall had a mean of 7.91% and a median of 6.70%, which also indicated a skewed distribution with a few census tracts pulling the mean higher. This was also supported by the Z-score mean and median (-0.61 and -0.31, respectively) and the P value (0.17), which concludes there isn't enough evidence to confirm spatial dependency, albeit the results are geographically widespread shown in **Figure A.2**. The NNeighbors Fixed Value for the entire dataset was 75, ensuring that each census tract in the region had spatial consistency in comparison with the neighboring tracts.

The median household income between 2010 and 2022 was $48,510. The mean z score was 1.28, with the median being -0.47. These results suggest an asymmetric distribution amongst the hot and cold spots, where there are more cold spots than hot, but the hotspots that do exist are larger and more clustered. The P value was slightly above the statistical significance value (0.05) with a value of 0.07. Following the same trend was the average unemployment rate. With the average being 3.97% and the median of the dataset being 3.73%, unemployment followed the trend of skewness throughout the dataset. The Z-score mean and median were 0.14 and -0.039, respectively, indicating an asymmetrical distribution like income, which was also visually observed in the hotspot analysis, **Figure A.3** and **Figure A.7**. The p-value (0.17) confirmed that the unemployment rate is not spatially dependent.

2.3.2 SNAP Recipients

SNAP recipients were broken down into three groups: the percentage of households receiving SNAP (**Figure A.4**), the percentage of those households that were below the poverty line (**Figure A.5**), and the percentage of total households that received SNAP and were below the poverty line (**Figure A.6**). Of the percentage of total houses receiving SNAP, the mean was 14.57% with a median of 12.54%. The percentage of those households that were below the

poverty line had an average of 49.28% with a median of 51.16%. The percentage of total households that received SNAP and were below the poverty line had an average of 7.79% with a median of 6.1%. The percentage of total households receiving SNAP and the percent of total households receiving SNAP below the poverty line had medians that were lower than the averages, indicating the skewed distribution that is being seen throughout the dataset. The percentage of households that received SNAP that were also below the poverty line had a median that was higher than the average, which indicates that the data is negatively skewed, with more households on SNAP being below the poverty line than being above. All three of the datasets had negative z score means and medians, with the percentage of total households on SNAP being -0.60 and -0.23 respectively, the percent of households on SNAP that are below the poverty being -0.94 and -0.20 respectively, and the percent of total households on SNAP and below the poverty line being -0.64 and -0.33 respectively. These Z-scores confirm that there are more cold spots across the region overall, and while visual analysis of the hotspots confirms this, the cold spots are smaller and spread out as opposed to the larger and more clustered hotspots seen in the region. All of the P-values (0.12, 0.15, and 0.12, in the previous order) were not statistically significant, staying on trend with the rest of the factors in that there wasn't enough evidence over the averaged years to confirm spatial dependency.

2.3.3 Housing and Transportation

Monthly housing costs were separated by the type of housing. These groups were: renters whose monthly costs exceeded 30% of their monthly income (**Figure A.10**), those with mortgages over the 30% threshold (**Figure A.11**), and those without mortgages/a fully paid-off home (**Figure A.12**). The group that saw the most individuals paying more than 30% of their income towards living expenses was the renters. In approximately 2,909,323 rental units, the

average percent of renters paying over the threshold was 38.71% with a median of 35.58%. The Z-score mean and median were both positive (0.84 and 0.20), demonstrating a negative (or left-skewed) distribution in the data, showcasing more hotspots than cold spots throughout the region. Of housing units, around 4,039,986 had a mortgage, and individuals with a mortgage had an average of 26.40% paying over the threshold, with a median of 25.52. The Z-score showed a positively skewed data set with a mean of -0.0092 and a median of -0.30, following the trend of having more small cold spots with concentrated larger hotspots. Units without a mortgage are around 3,122,092, and individuals living in a fully paid-for unit saw the smallest percentage of threshold crossings, with a mean of 11.63% and a median of 10.90. This dataset had an asymmetrical distribution, with the Z-score mean being 0.36 and the median being -0.25. The P-values (in order: 0.13, 0.17, and 0.13) all followed the overall theme that excessive housing costs are not spatially dependent. Along with the housing costs, the average percent of households that did not have complete plumbing in their homes, creating a lack of reliable clean water, was 0.48%, with the median being 0.18%. The difference between the mean and the median is, again, the result of a few census tracts with elevated rates pulling the data to a right-skewed distribution. The negative Z-scores mean and median (-0.45 and -0.61), and the visual hotspot analysis had more cold spots with a few smaller and more concentrated hotspots, **Figure A.8**. The P-value (0.33) did not indicate a spatial dependency. Despite this small percentage, out of the total housing units, this average percentage still equates to nearly 48,342 homes that are currently occupied, primarily in the coal mining belt of the central subregion.

The percent of households that lacked access to transportation averaged to be 7.95% and had a median of 5.43%, which also indicated a asymmetrical distribution with a few census tracts pulling the mean higher. This was also supported by the Z-score mean and median (0.68 and -

0.23, respectively) and the P value (0.26), which concludes there isn't enough evidence to confirm spatial dependency. The hotspots for lacking transportation, shown in **Figure A.9**, were congregated in the urban areas of the region, most likely due to the lack of need for personal transportation with access to public transportation or walkability found in cities.

2.3.4 Disability and Insurance

Disability rates and lacking health insurance are similar in that they are spatially dependent (P-values being 0.029 and 0.0075). The mean and median disability rates were almost equal, being 16.43% and 16.10%, providing an almost normal distribution. The average lack of health insurance rate is 9.90%, and the median is 8.92%, demonstrating a slightly more skewed data compilation than disability. Both the Z-scores (-1.57 and -0.54; -0.33 and -0.24) followed par with the study; however, confirmed with both visual analysis and the P-values, the hotspots observed were large (spanning over the majority of the region) and more compacted than the cold spots, shown in **Figures A.13** and **Figure A.14**.

2.3.5 Farmland and Food

Farmland overall saw a decrease in the region. Family farmland makes up about 70% to 75% of farmland throughout the region. These farms are overwhelmingly small-scale, meaning that they make below $350,000 a year in revenue. Total farmland increased between 2012 and 2017 by 98,768 acres, and then rapidly decreased by 2.4 million acres between 2017 and 2022, a 6.43% loss in farmland overall (**Figures A.15**, **2.16**, and **2.17**; **Table A.2**). Family farmland lost acres between 2012 and 2017, losing 281,766 acres, and then saw the same rapid decline from 2017 to 2022, losing 2.7 million acres (**Figures A.18**, **2.19**, and **2.20**; **Table A.2**).

Lastly, food insecurity rates had an almost identical mean and median (13.51 and 13.2), the dataset being normally distributed, with a low P-value (0.00071), concluding that food

insecurity is spatially dependent. This was aided by the visual representation of rates, being drastically widespread, covering almost the entire southern and central subregions. There were more cold spots than hot (Z-scores being -2.04 and -0.86) simply because there was 1 hotspot, but it spanned over a significant portion of the region. The most widespread food deserts were seen as the financial food deserts, which mimic the distribution of food insecurity, spanning across the region entirely. Out of 6,189 census tracts, 2,931 of them were financial food deserts in 2010, 3,016 in 2015, and 2,966 in 2019 (**Figures A.25**, **2.26**, and **2.27**; **Table A.3**). Geographic food deserts were scattered throughout the region, with the number of geographic food deserts decreasing from 2,343 in 2010 to 2,324 in 2015 and then increasing to 2,347 in 2019 (**Figures A.22**, **2.23**, and **2.24**; **Table A.3**). Food deserts that were both geographic and financial were scattered throughout the region in smaller numbers but gradually increasing, with 952 tracts in 2010, 993 tracts in 2015, and 1,017 in 2019 (**Figures A.28**, **2.29**, **2.30**; **Table A.3**).

2.4 Discussion

Among the factors analyzed, disability rates, lack of health insurance, and food insecurity were found to be spatially dependent, indicating statistically significant clustering in the Central subregion of Appalachia. These spatial patterns are critical to understanding regional disparities. The high rates of disability and the absence of health insurance were heavily concentrated in East Kentucky and Southern West Virginia, areas historically tied to the coal mining industry. Long-term exposure to the working conditions within the coal mines and the surrounding residential areas has been linked to chronic illnesses and disabilities such as respiratory problems like black lung, musculoskeletal disorders, and birth defects (Hendryx, 2008). The decline of the coal industry in Appalachia has led to unemployment and shifts to other industries, typically with less

protections as coal mining is heavily unionized, resulting in the loss of health insurance, job security, and worker rights (Hendryx, Ahern, 2009).

While food insecurity was spatially dependent, the socioeconomic factors that are associated with food insecurity, such as poverty and unemployment, were not. These factors were elevated across the region but their distribution lacked statistically significant clustering, meaning that the prevalence was widespread enough to resemble complete spatial randomness. In other words, the issues are not isolated to specific communities but are endemic across Appalachia. This distinction offers insight into how the geographic distribution of food insecurity compares to that of its contributing factors. These findings provide the foundation for following research into temporal and correlation analyses.

The central and southern subregions of Appalachia saw the highest concentration of troubling socioeconomic statistics. Across the board, the highest rates of poverty, SNAP recipients, disability, low income, and food insecurity were found in the lower subregions of Appalachia. Nearly all the factors (except for median household income and elevated housing costs for renters and non-mortgage units) were positively skewed, or right-leaning. This skewness aided in the visualization of the results, as it identified the presence of more statistically significant cold sports than hotspots across the region. The greater number of cold spots suggests that the majority of the region is experiencing lower-than-regional-average levels for these factors. The hotspots that appeared were highly concentrated and spanned across large portions of the Southern and South Central subregions, indicating localized clusters of intensified disparity. Recognizing the spatial clustering of these cold and hotspots ties into spatial dependency and reveals patterns in how socioeconomic factors interact with each other and how they compound onto regional hardship.

The factor that proved to be the common denominator was poverty rates. The average overall poverty rate over the thirteen years (15.81%) was higher than the national average of 14.3% (2010-2022, US Census Bureau). The other factors followed the same pattern as the poverty rates. Those who receive SNAP are more likely to be below the poverty line, with 49.28% of them falling below the line. Those who are experiencing food insecurity or food deserts are more likely to be low-income. As of 2019, 41.46% of the Appalachian region's population (roughly 25,535,826) was living in a geographic food desert, 43.69% were living in a financial food desert, and 15.80% were living in the combined deserts. The socioeconomic string between these statistics is low-income households, are tied to the answer for both the second and third research questions. Firstly, it's evident that poverty, food insecurity, and food deserts are linked. While poverty was not spatially dependent, high rates were seen throughout the entirety of the region; food insecurity and food deserts were linked to geography. Regarding how the factors tie into food insecurity and deserts, the areas that saw high rates of individual and household poverty also saw high rates of food insecurity and scattered food deserts. Not all food-insecure households or food desert communities are below the poverty line, but those below the poverty line are more likely to find themselves in food insecurity and a food desert.

Contributing to both food insecurity and poverty rates is the loss of farmland in the region. The area of concern mentioned above (Eastern Kentucky and Southern West Virginia) doesn't have significant portions of farmland; most of the region's farms are in Northern Mississippi and Alabama, East Tennessee, Western North Carolina, and Pennsylvania. These areas still saw poverty rate hotspots and widespread food insecurity. Although the area of concern doesn't have high acreages of farmland, the loss of farmland in the region alone is troublesome. As the region loses access to closer farms, the reliance on imported (domestic and

international) food increases. These foods tend to be more expensive as the cost has to include transportation expenses, contributing to the pre-existing cost barrier for low-income households and communities.

2.5 Conclusion

The central and southern subregions of the Appalachian Region are experiencing elevated socioeconomic hardships across the board. While it was found that food insecurity and food deserts were indeed dependent on geographic location, the contributing factors were not. This could point to the idea that while these factors are not *separately* spatially dependent, when observed together, they become spatially dependent, under the assumption that these factors are contributing to food insecurity. Poverty rates, which were consistently elevated across the subregions, emerged as the common denominator underlying the socioeconomic hardships and food inaccessibility in the subregions. Examined more thoroughly in the coming chapters, low-income households and communities are more likely to experience more food hardships that are linked to both financial struggle and geographic location. It was also found that the amount of farmland in the region is rapidly decreasing, with the region losing 2.4 million acres overall between 2017 and 2022. This loss of farmland, almost 75% of it being family-owned, is contributing to both food inaccessibility, as local communities are losing food sources, and to poverty rates, as farmers are losing income.

References

Hendryx, M. (2008). Mortality from heart, respiratory, and kidney disease in coal mining areas of Appalachia. *International Archives of Occupational and Environmental Health*, 82(2), 243–249.

Hendryx, M., & Ahern, M. M. (2009). Relations between health indicators and residential proximity to coal mining in West Virginia. *American Journal of Public Health*, 99(3), 550–552.

United States Department of Agriculture. (2009). Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences. Report to Congress. https://www.ers.usda.gov/webdocs/publications/42711/12716_ap036_1_.pdf

Appendix A

Table A.1. Hotspot analysis summary statistics calculated by the Getis-Ord Gi* tool in ArcGIS Pro.

Factor	Average	Median	Z-Score Mean	Z- Score Median	P-Value
Food Insecurity	13.51%	13.20%	-2.01	-0.86	0.00071
Poverty	15.81%	13.98%	-0.58	-0.23	0.099
Household Poverty	7.91%	6.70%	-0.61	-0.31	0.17
Income	X	$48,510.50	1.28	-0.47	0.07
Transportation	7.95%	5.43%	0.68	-0.23	0.26
Unemployment	3.97%	3.73%	0.14	-0.039	0.17
Lack of Complete Plumbing	0.48%	0.18%	-0.45	-0.61	0.33
Disability	16.43%	16.10	-1.57	-0.54	0.029
Lack of Health Insurance	9.90%	8.92%	-0.33	-0.24	0.0075
Receiving SNAP	14.27%	12.54%	-0.60	-0.23	0.12
Received SNAP under Poverty Line	49.28%	51.26%	-0.94	-0.20	0.15
Household receives SNAP and is in Poverty	7.79%	6.10%	-0.64	-0.33	0.12
Renters above 30%	38.71%	35.58%	0.84	0.20	0.13
Mortgage above 30%	26.40%	25.52%	-0.0092	-0.30	0.17
Non-Mortgage above 30%	11.63%	10.90%	0.36	-0.25	0.13

Table A.2. Appalachian Farmland Acreage from 2012-2022.

Year	Total Acres	Family Owned Acres
2012	37,676,568	27,952,814
2017	37,775,336	27,671,048
2022	35,253,726	25,218,727

Table A.3. Food Desert Trends in Appalachia from 2010-2019.

Year	Geographic Food Deserts	Financial Food Deserts	Combination Food Deserts
2010	2,343	2,931	952
2015	2,324	3,016	993
2019	2,347	2,966	1,017

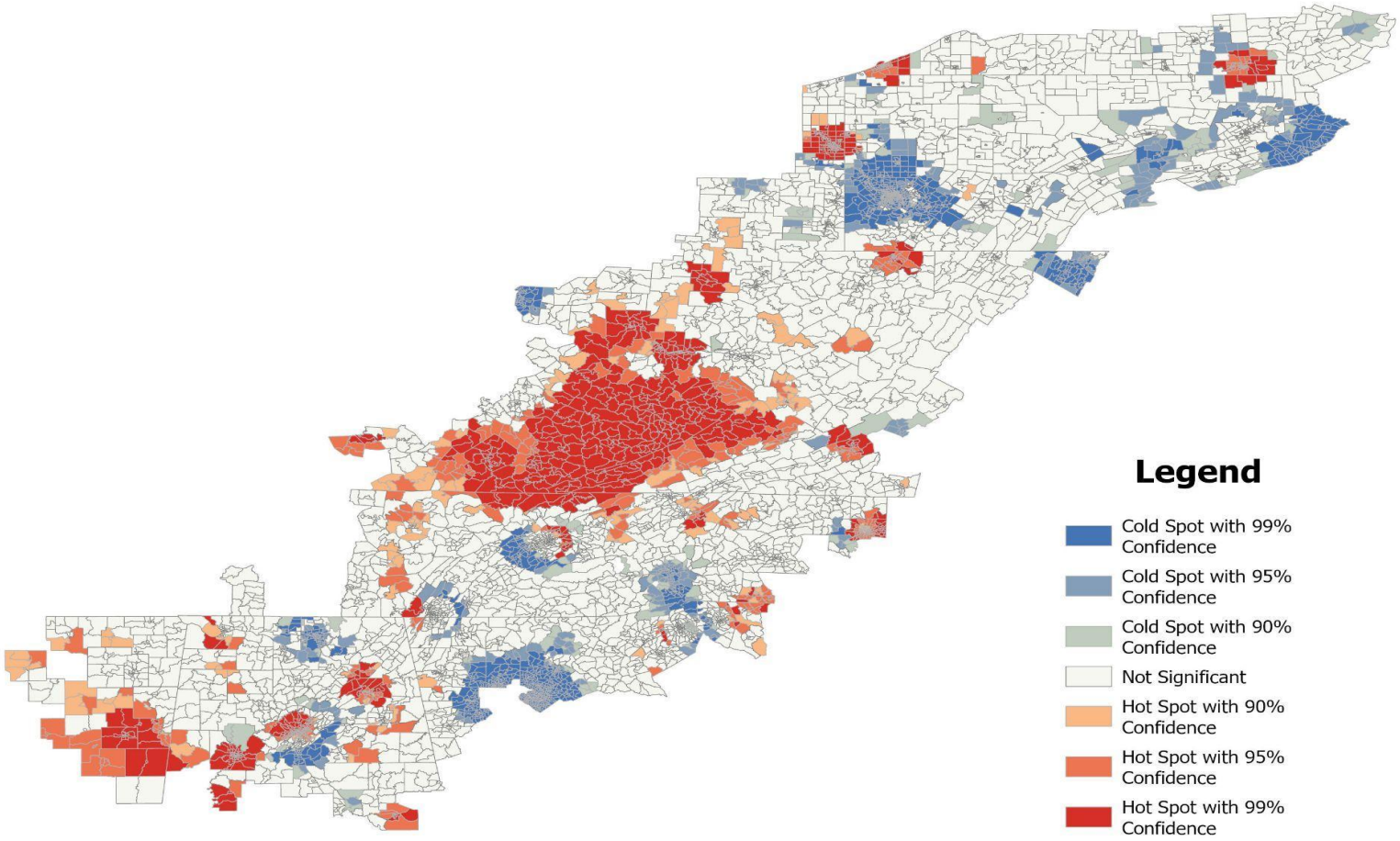


Figure A.1. Average overall poverty rate in Appalachia hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

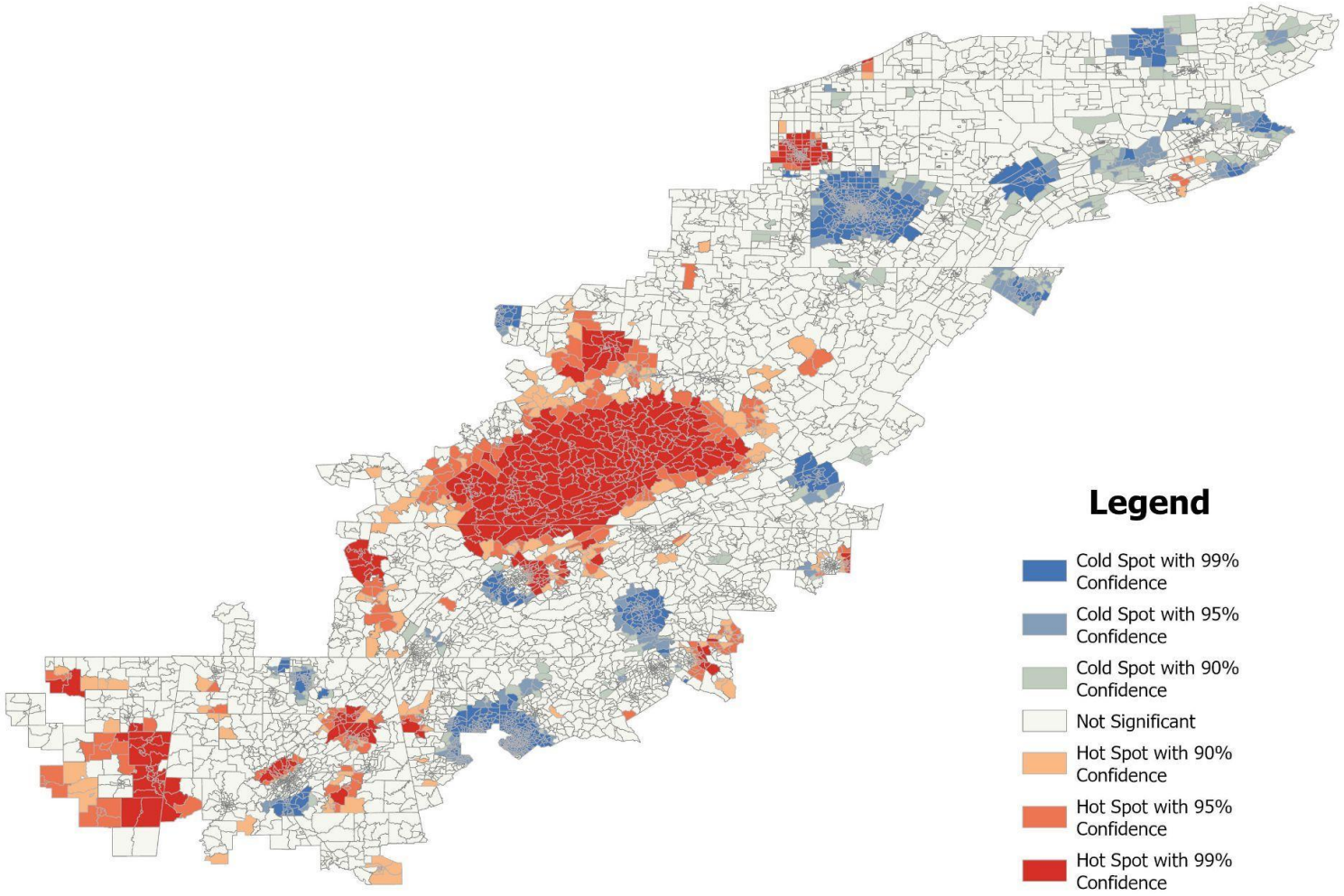


Figure A.2. Average household poverty rate in Appalachia hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

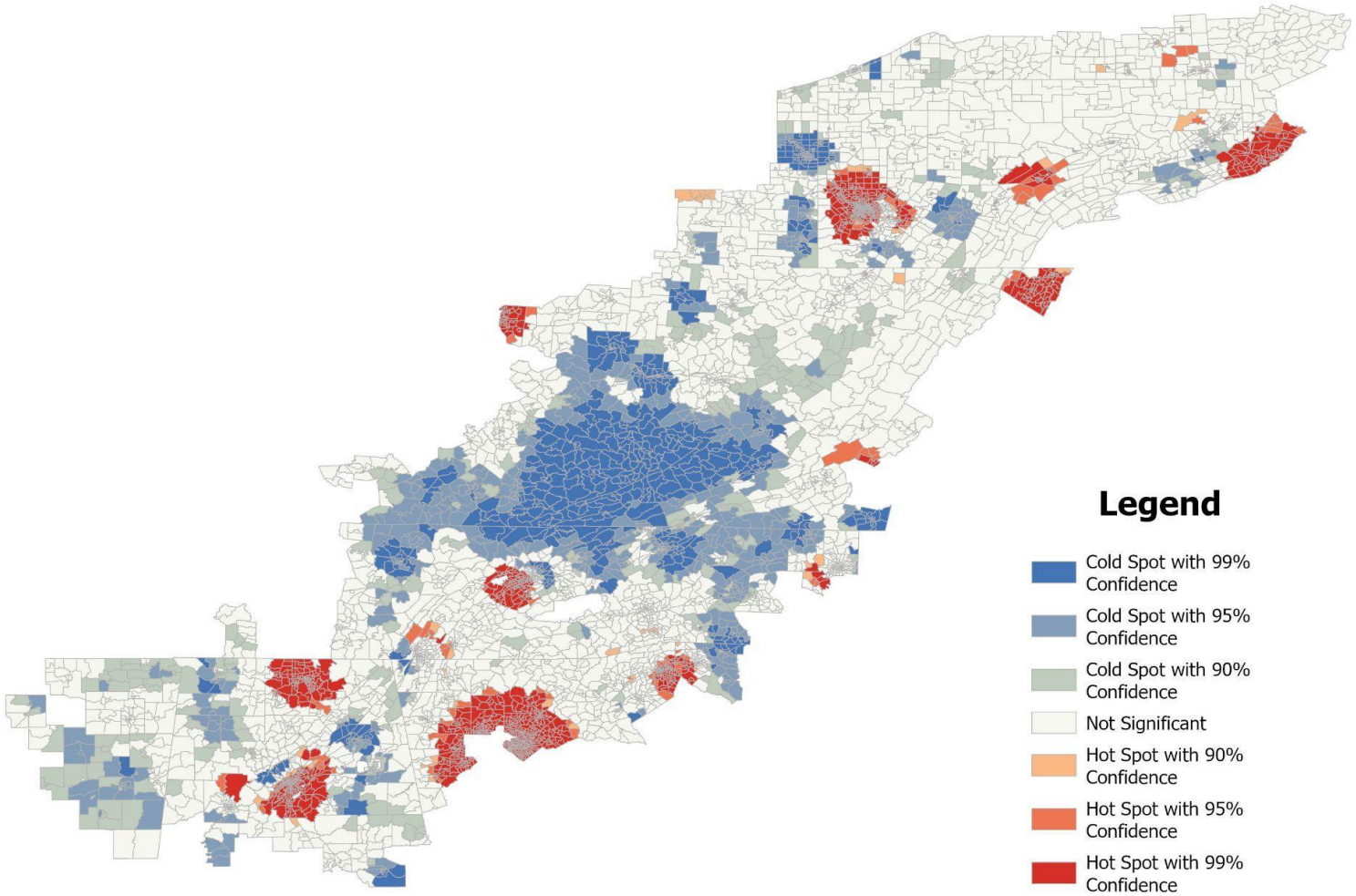


Figure A.3. Median household income in Appalachia hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

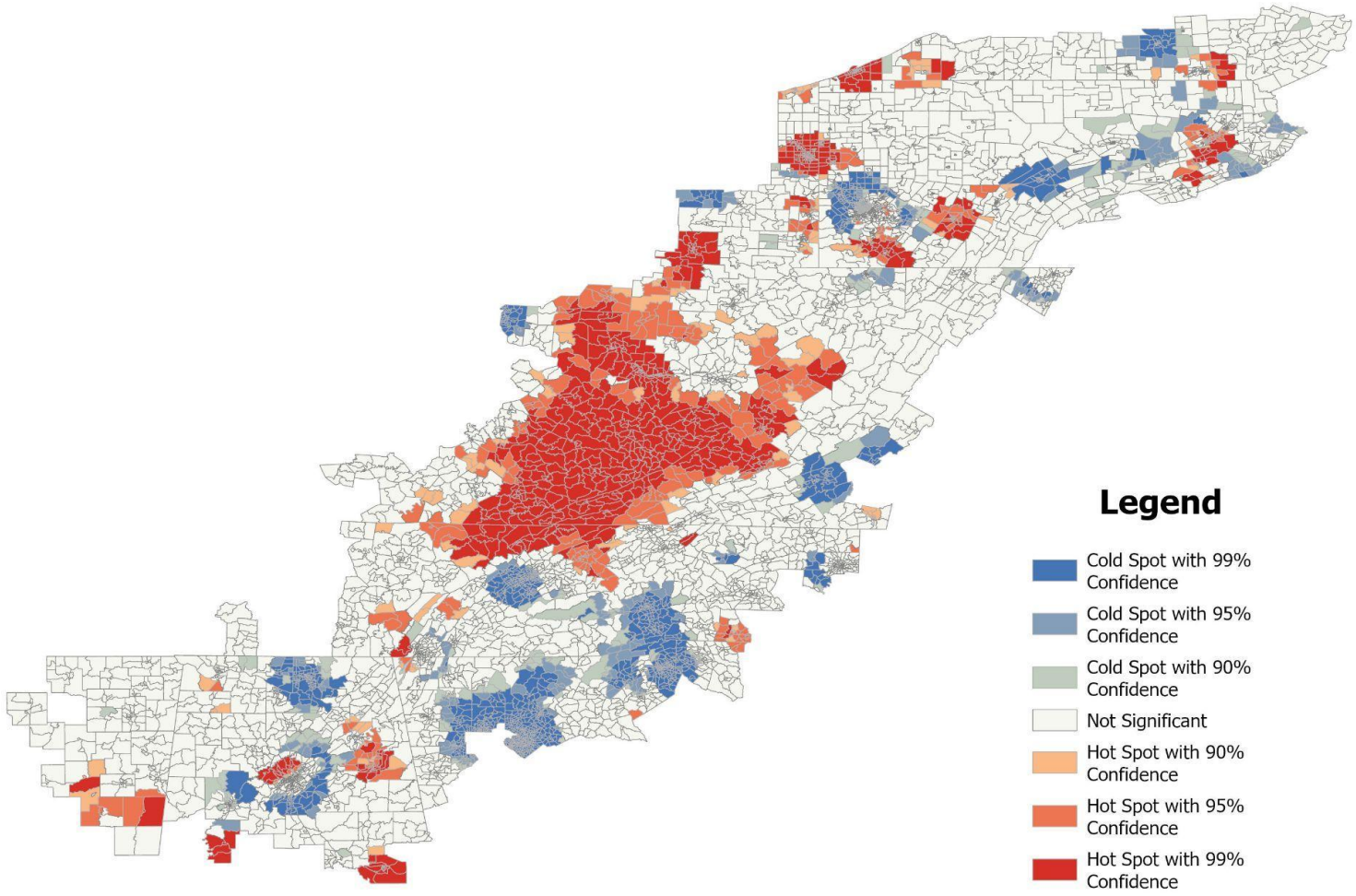


Figure A.4. Average percentage of total Appalachian households receiving Supplemental Nutrition Assistance Program (SNAP) hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

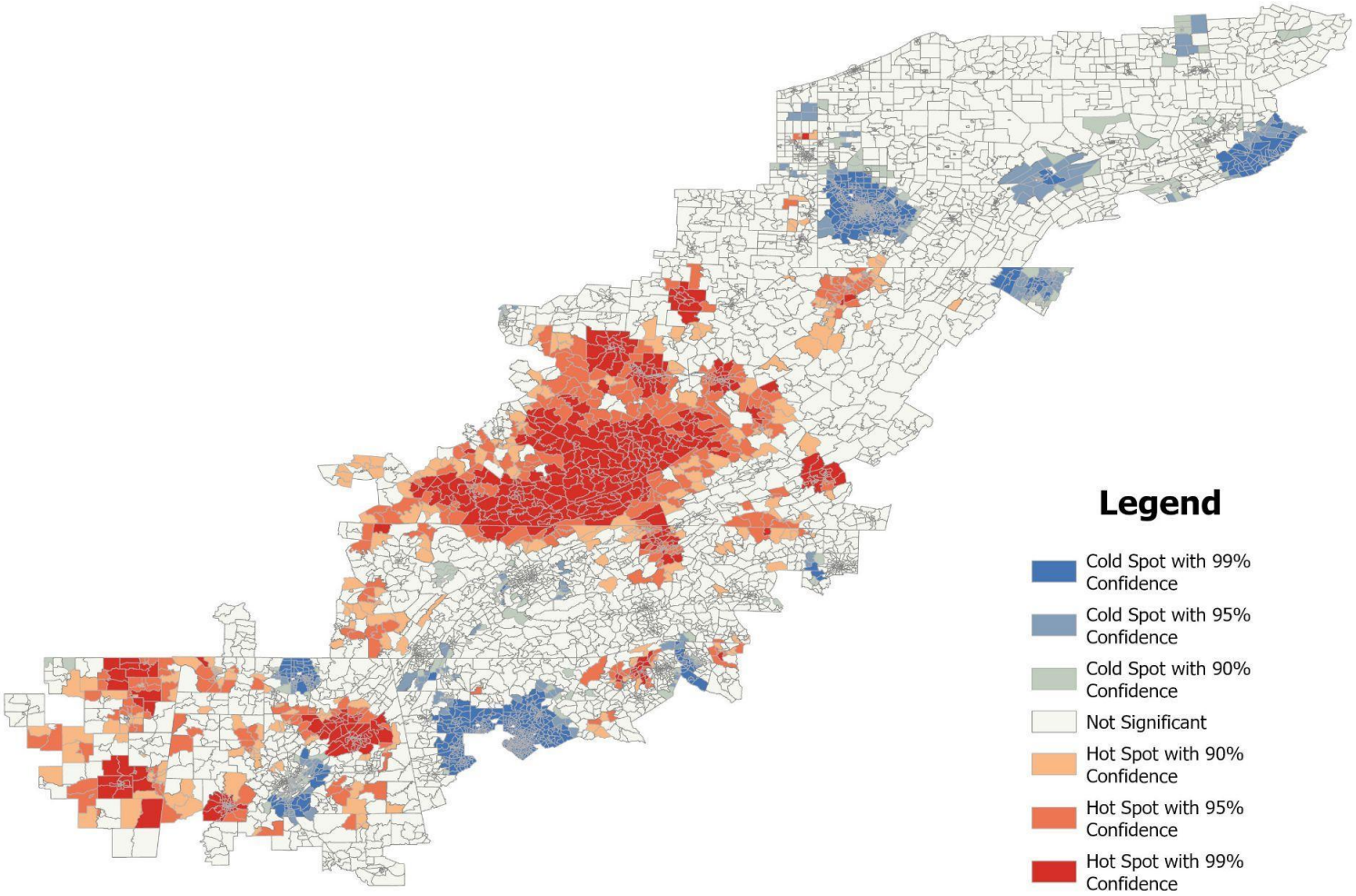


Figure A.5. Average percentage of Appalachian households that receive Supplemental Nutrition Assistance Program (SNAP) that are below the poverty line hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

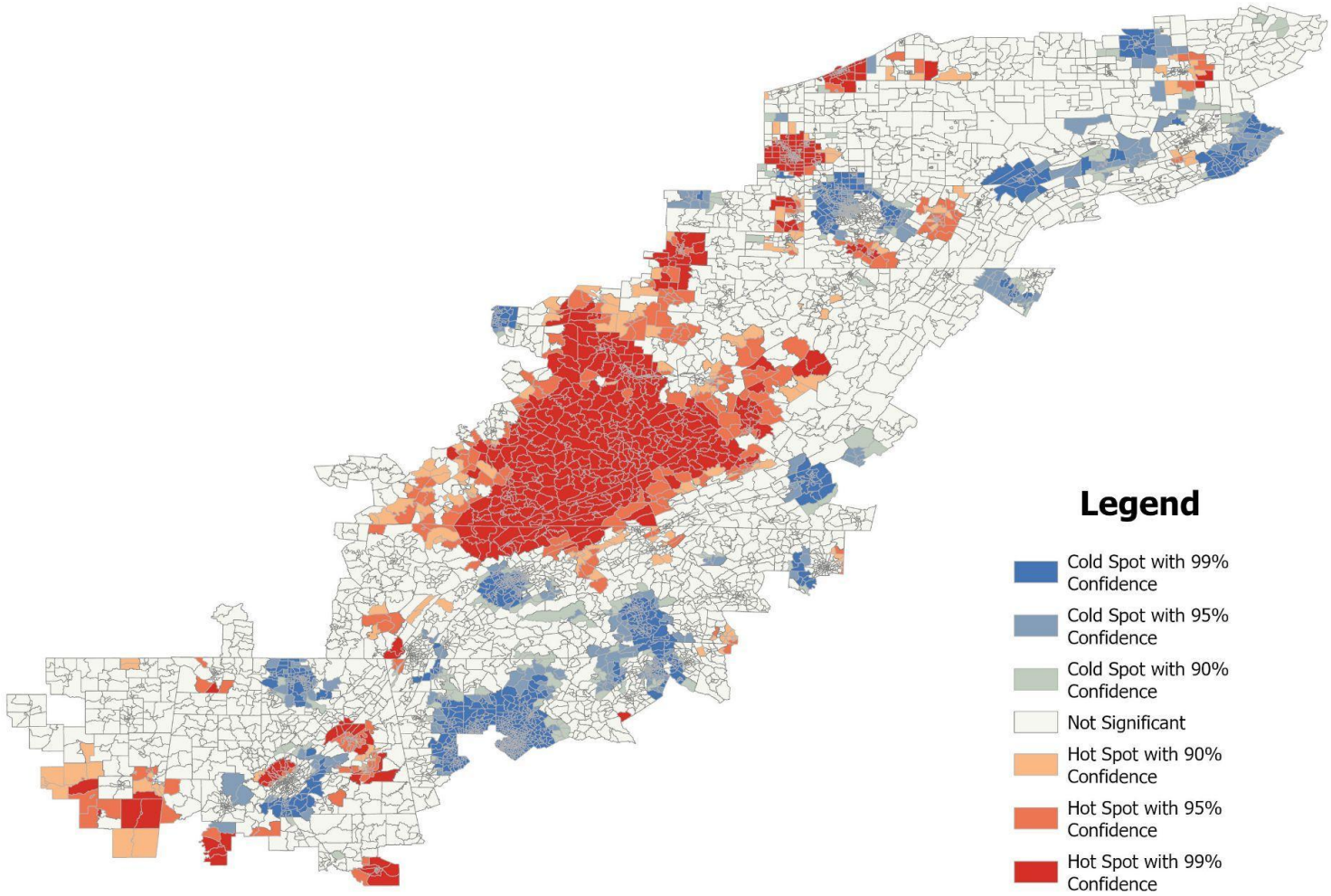


Figure A.6. Average percentage of total households that receive Supplemental Nutrition Assistance Program (SNAP) and are below the poverty line hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

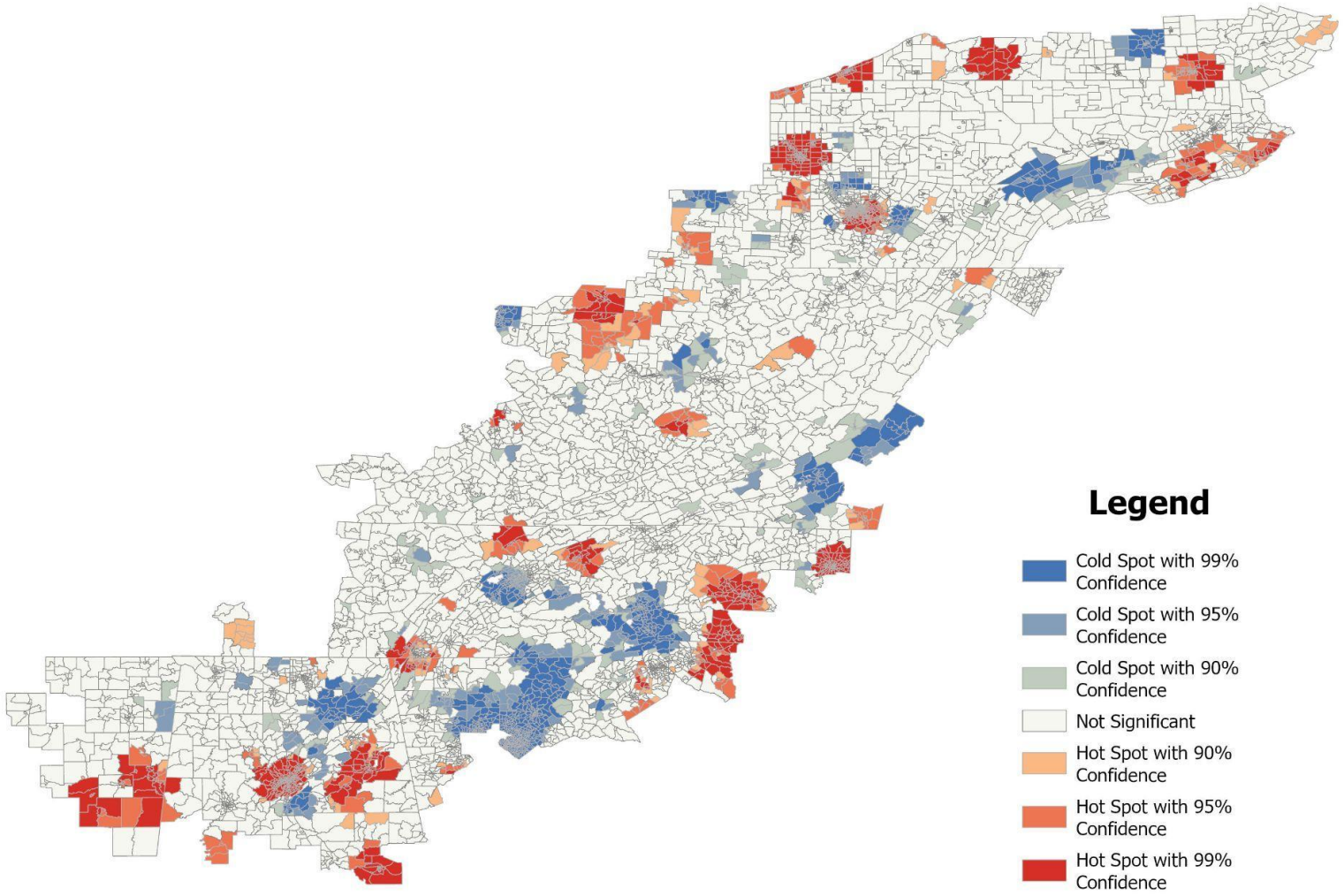


Figure A.7. Average unemployment rate in Appalachia hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

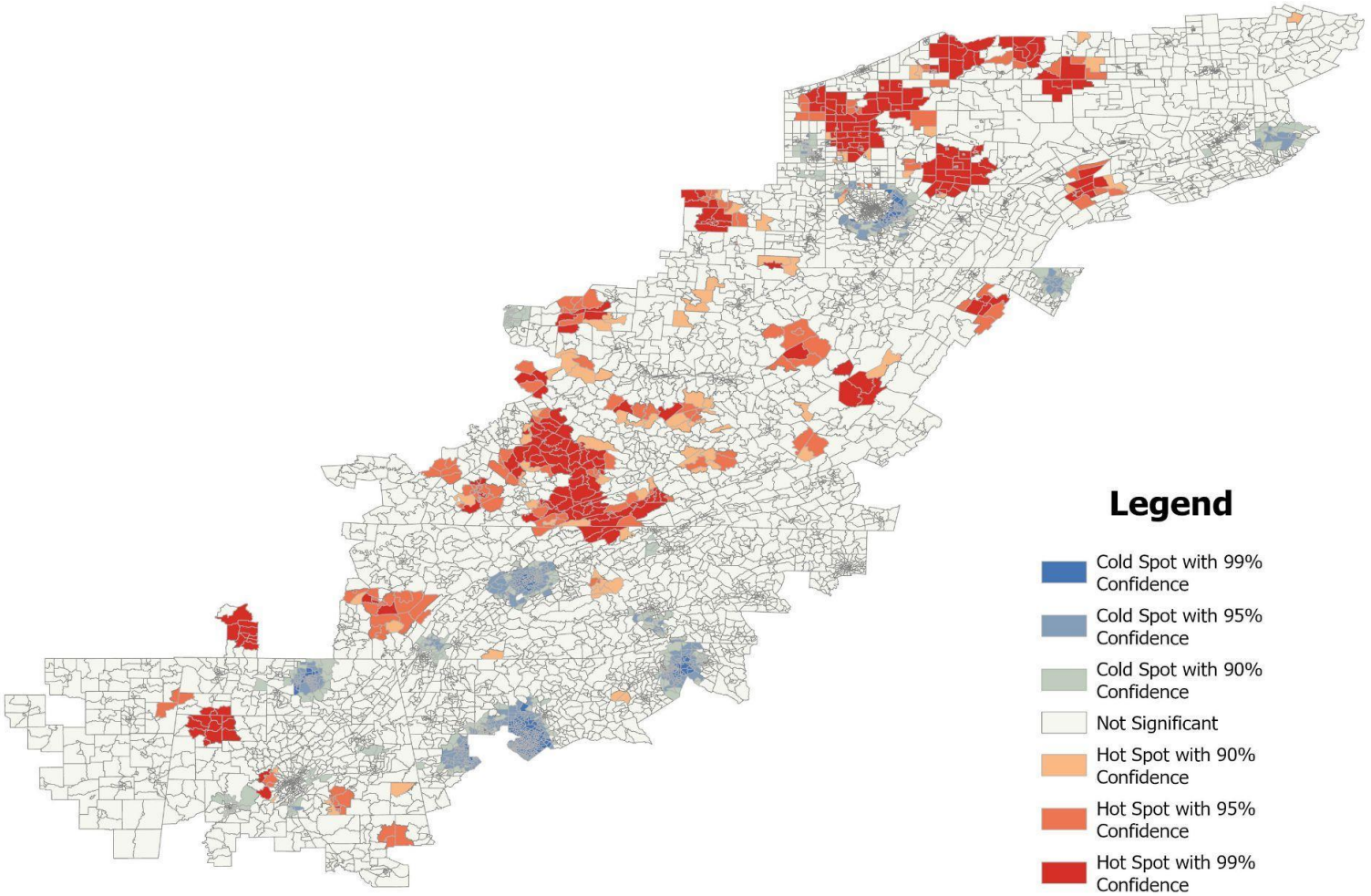


Figure A.8. Average percentage of Appalachian households lacking complete plumbing hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

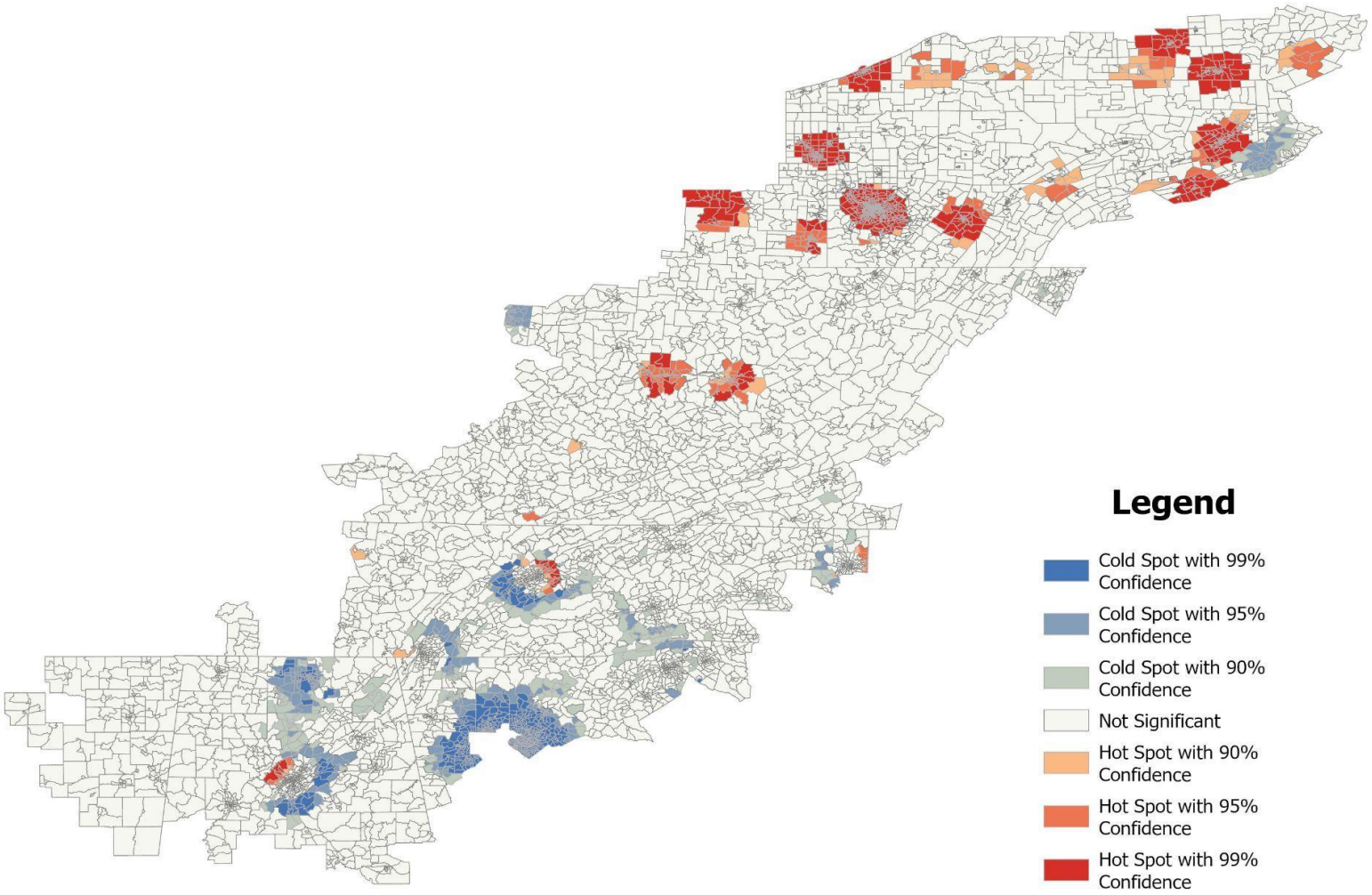


Figure A.9. Average percentage of households in Appalachia lacking transportation hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

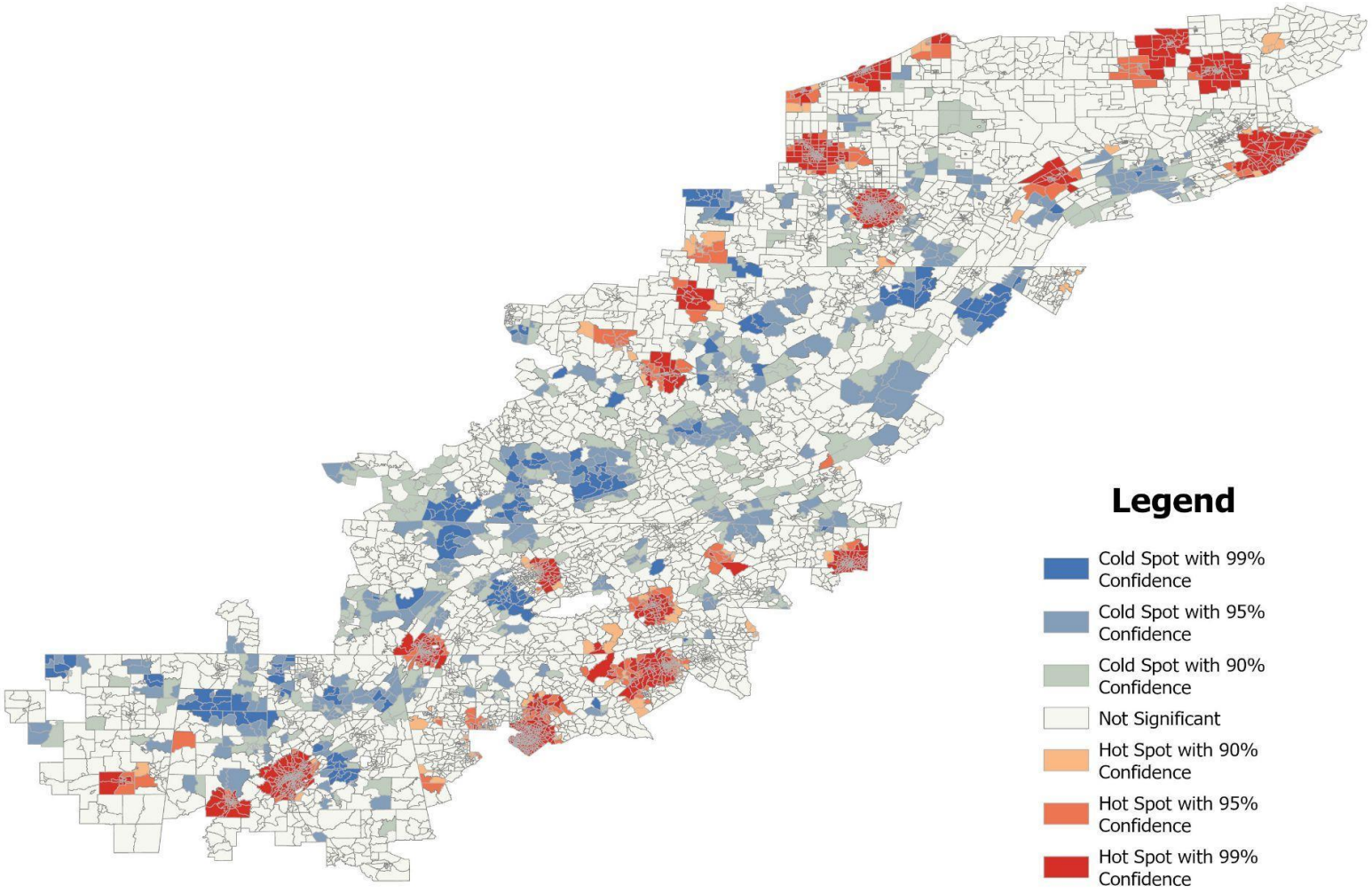


Figure A.10. Average percentage of occupied rental units with excessive housing costs hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

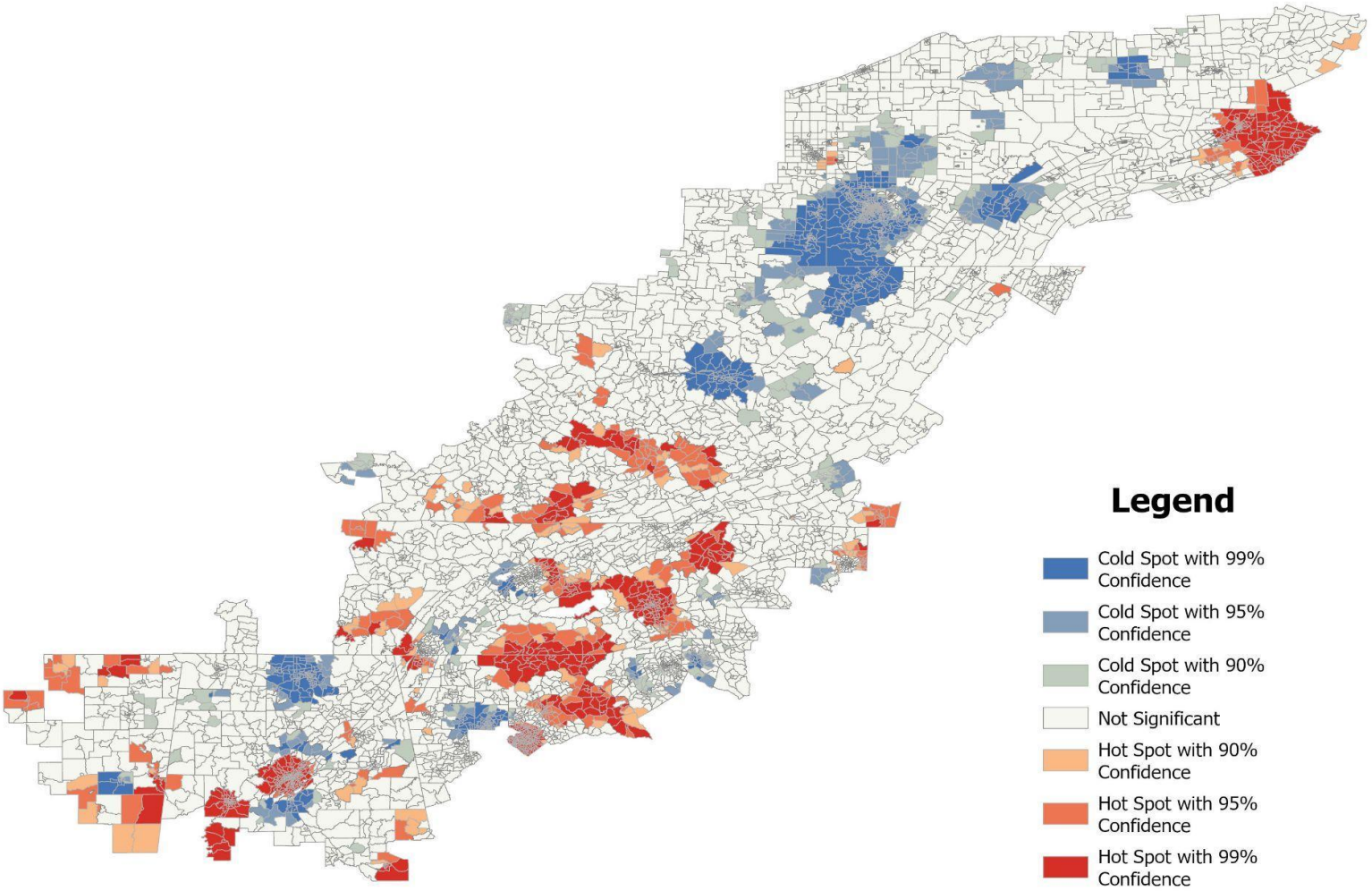


Figure A.11. Average percentage of occupied mortgage units with excessive housing costs hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

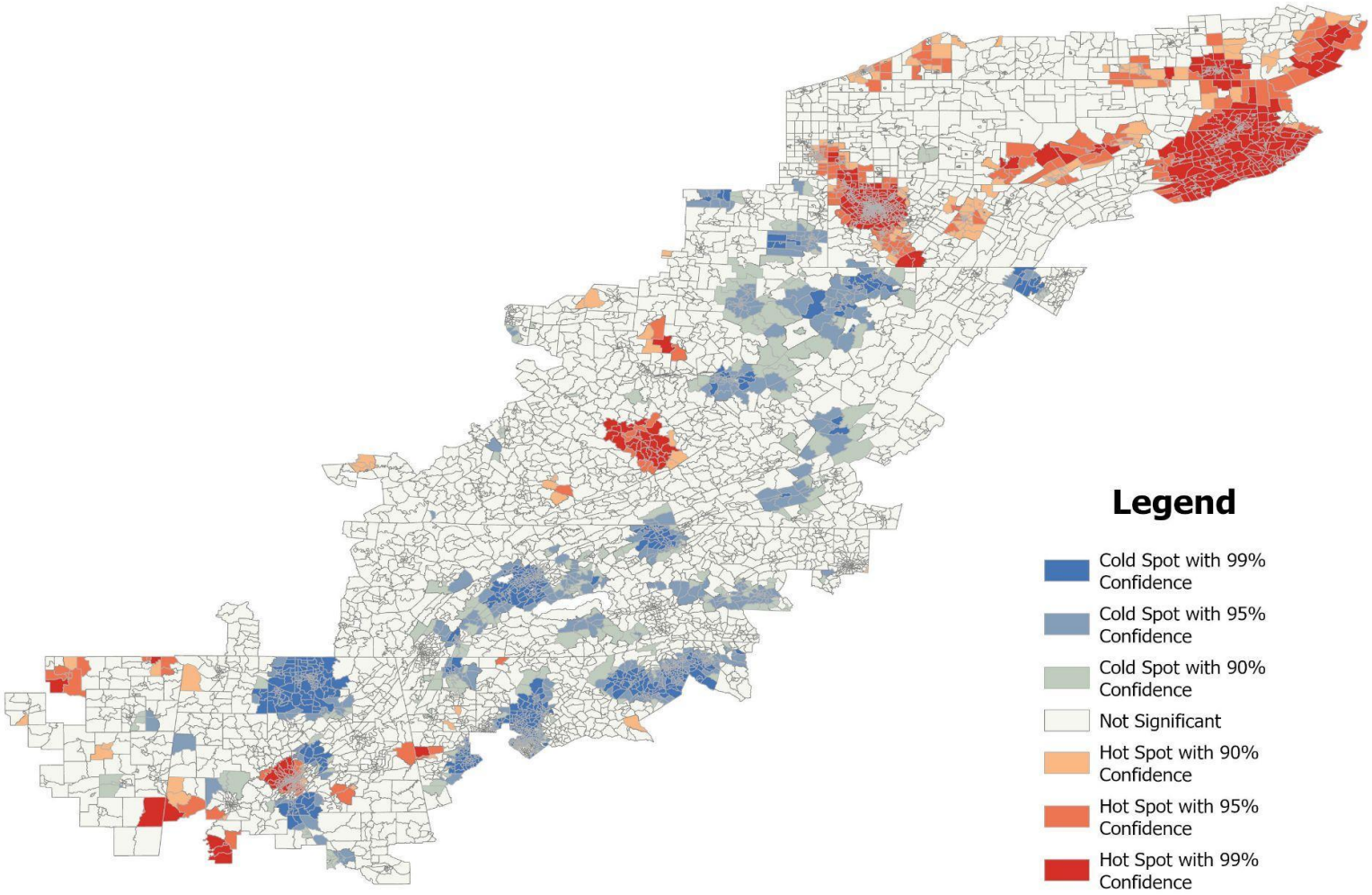


Figure A.12. Average percentage of occupied housing units without a mortgage that have excessive housing costs hotspot analysis 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

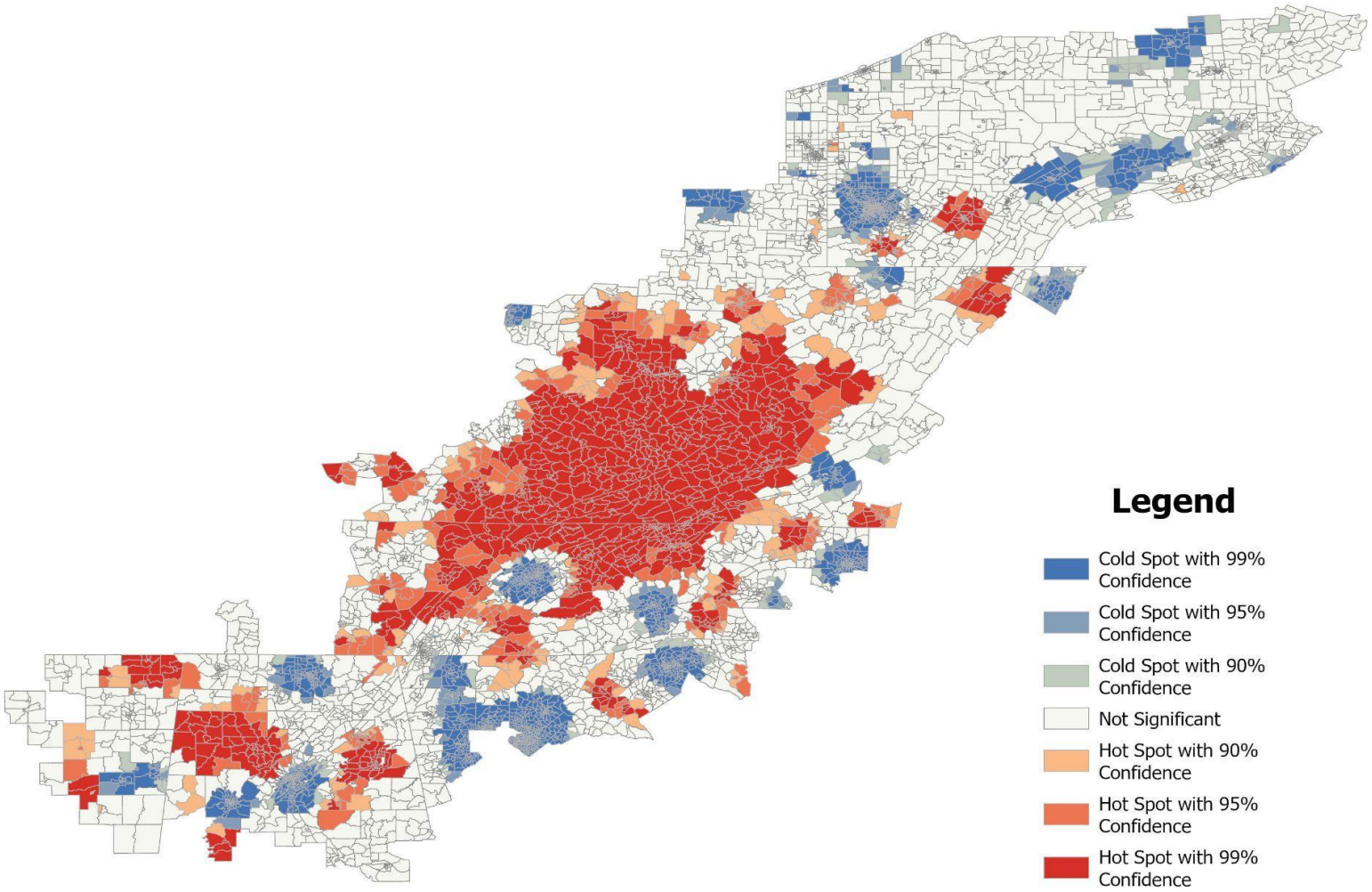


Figure A.13. Average disability rate in Appalachia hotspot analysis from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

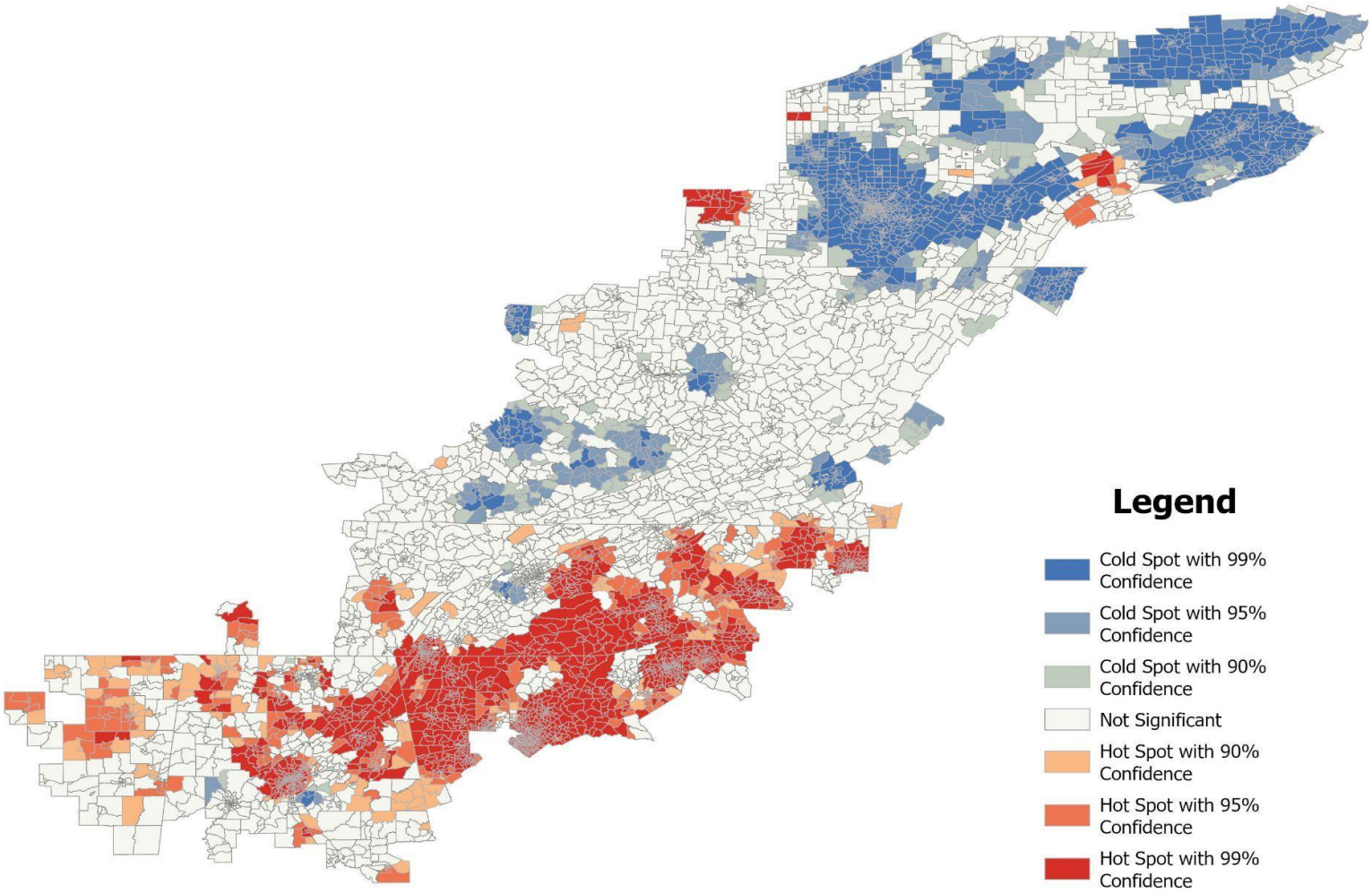


Figure A.14. Average percentage of individuals in Appalachia lacking health insurance hotspot analysis from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey from 2010 to 2022.

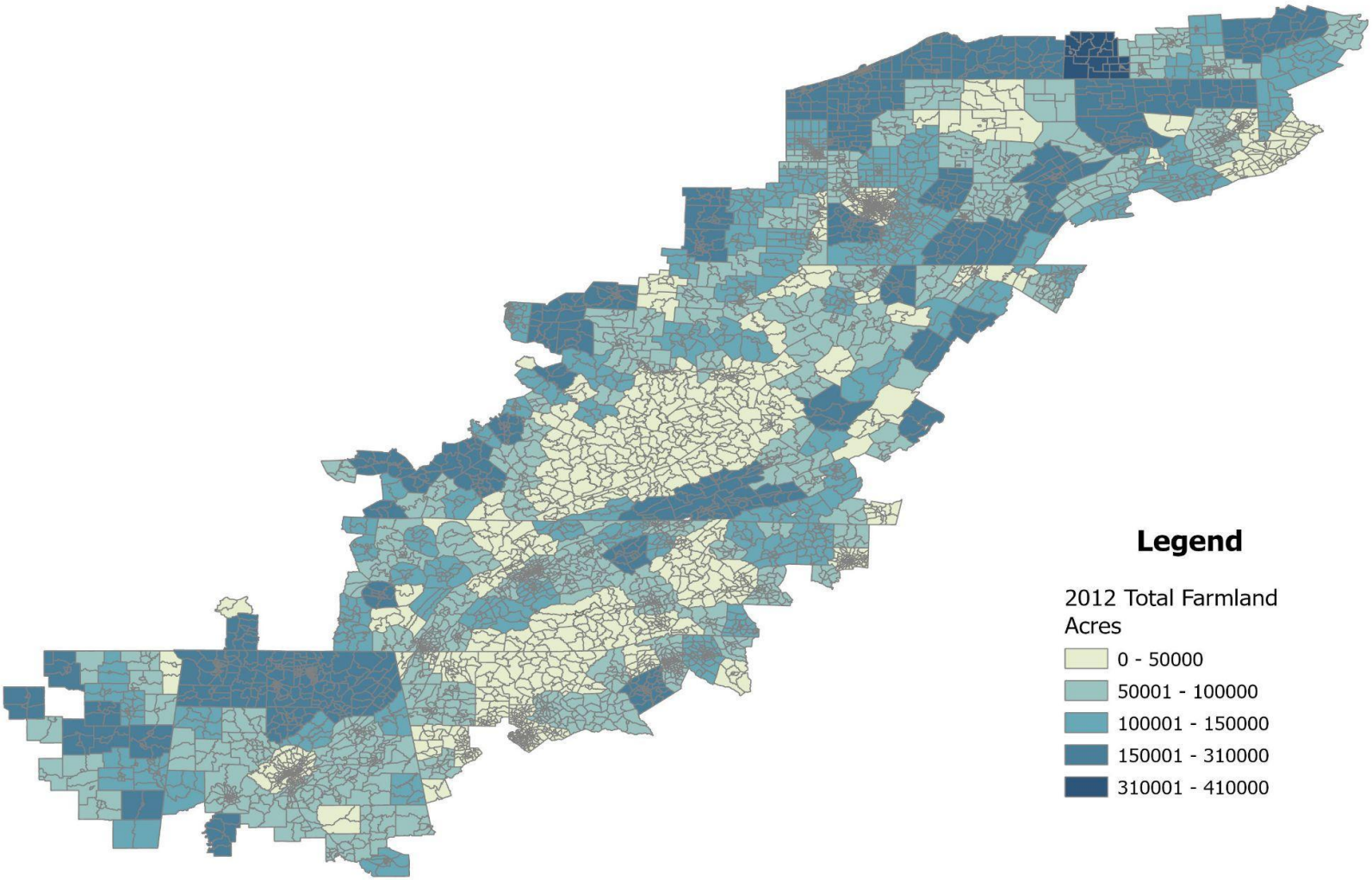


Figure A.15. Total farmland acreage in Appalachia in 2012. Data from the United States Department of Agriculture National Agricultural Statistics Service 2012.

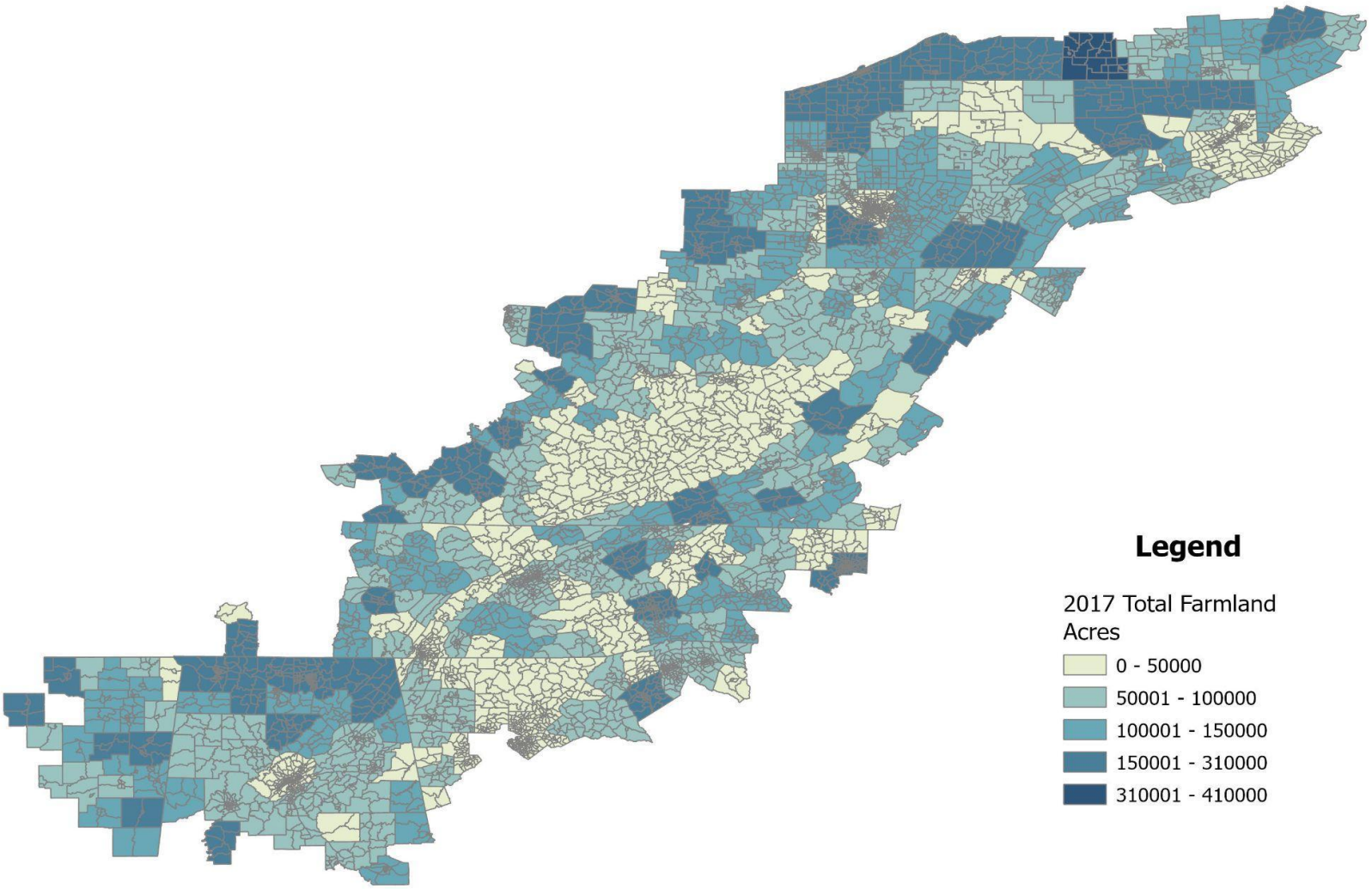


Figure A.16. Total farmland acreage in Appalachia in 2017. Data from the United States Department of Agriculture National Agricultural Statistics Service 2017.

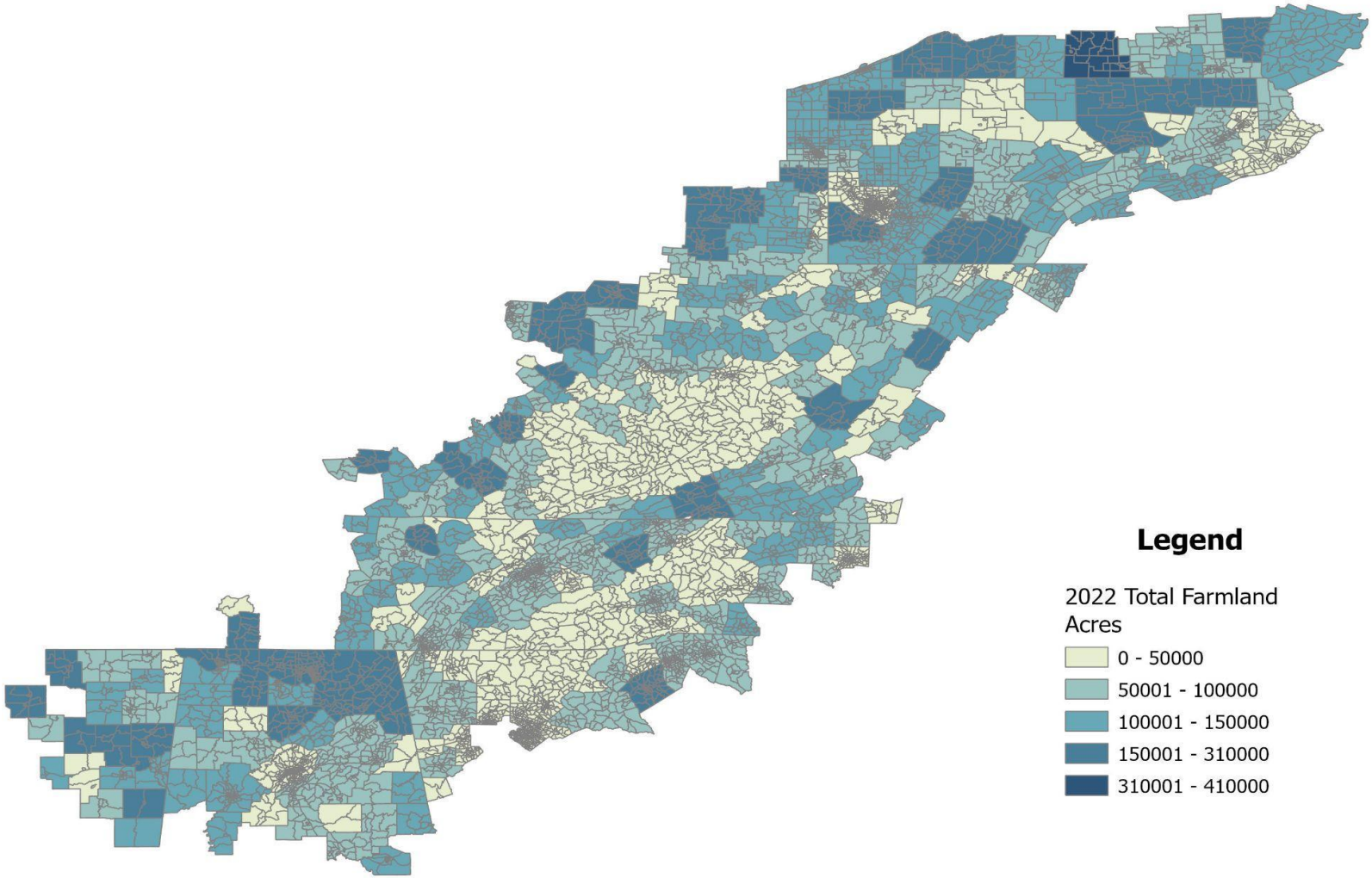


Figure A.17. Total farmland acreage in Appalachia in 2022. Data from the United States Department of Agriculture National Agricultural Statistics Service 2022.

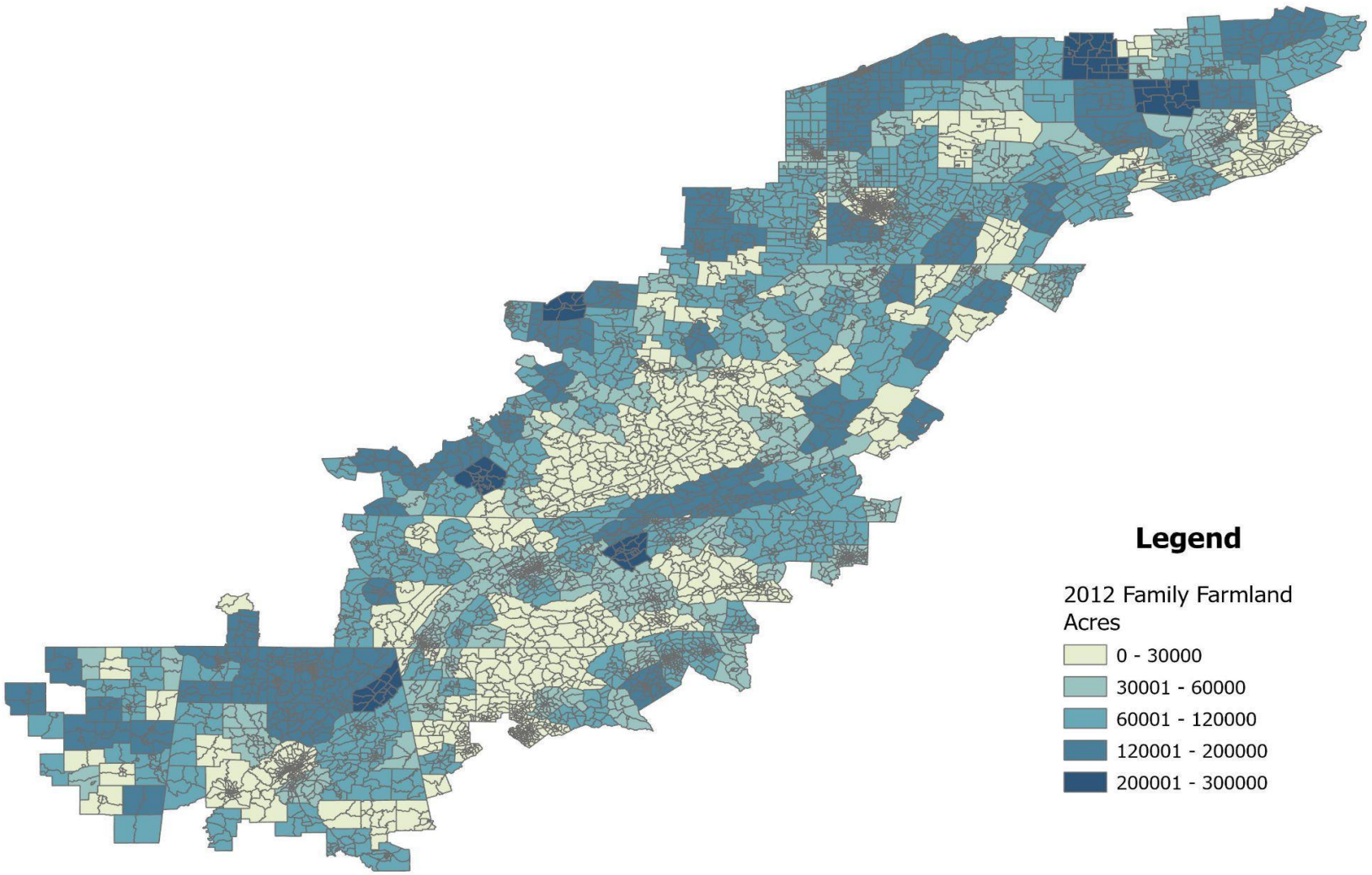


Figure A.18. Family-owned farmland acreage in Appalachia in 2012. Data from the United States Department of Agriculture National Agricultural Statistics Service 2012.

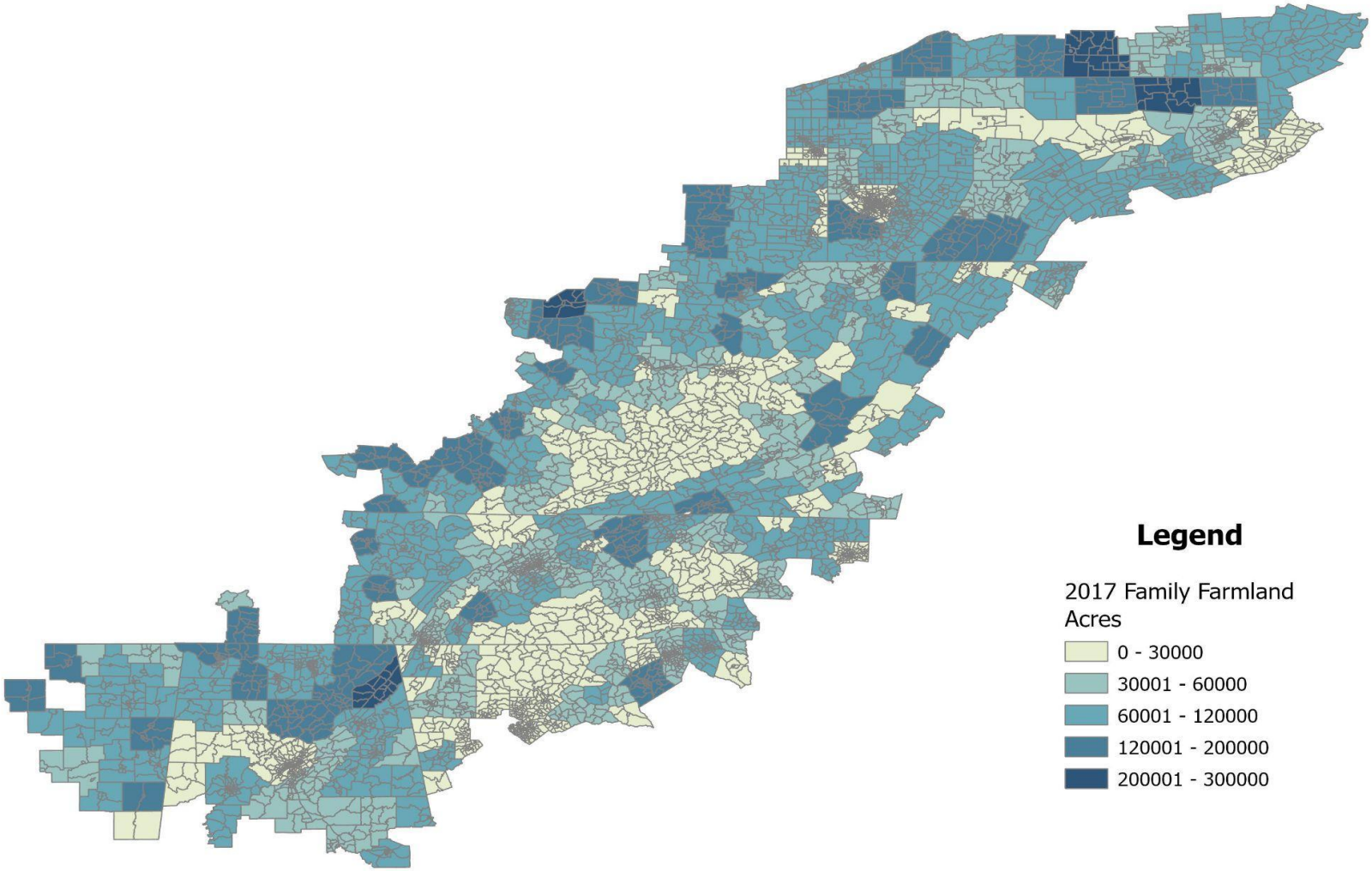


Figure A.19. Family-owned farmland acreage in Appalachia in 2017. Data from the United States Department of Agriculture National Agricultural Statistics Service 2017.

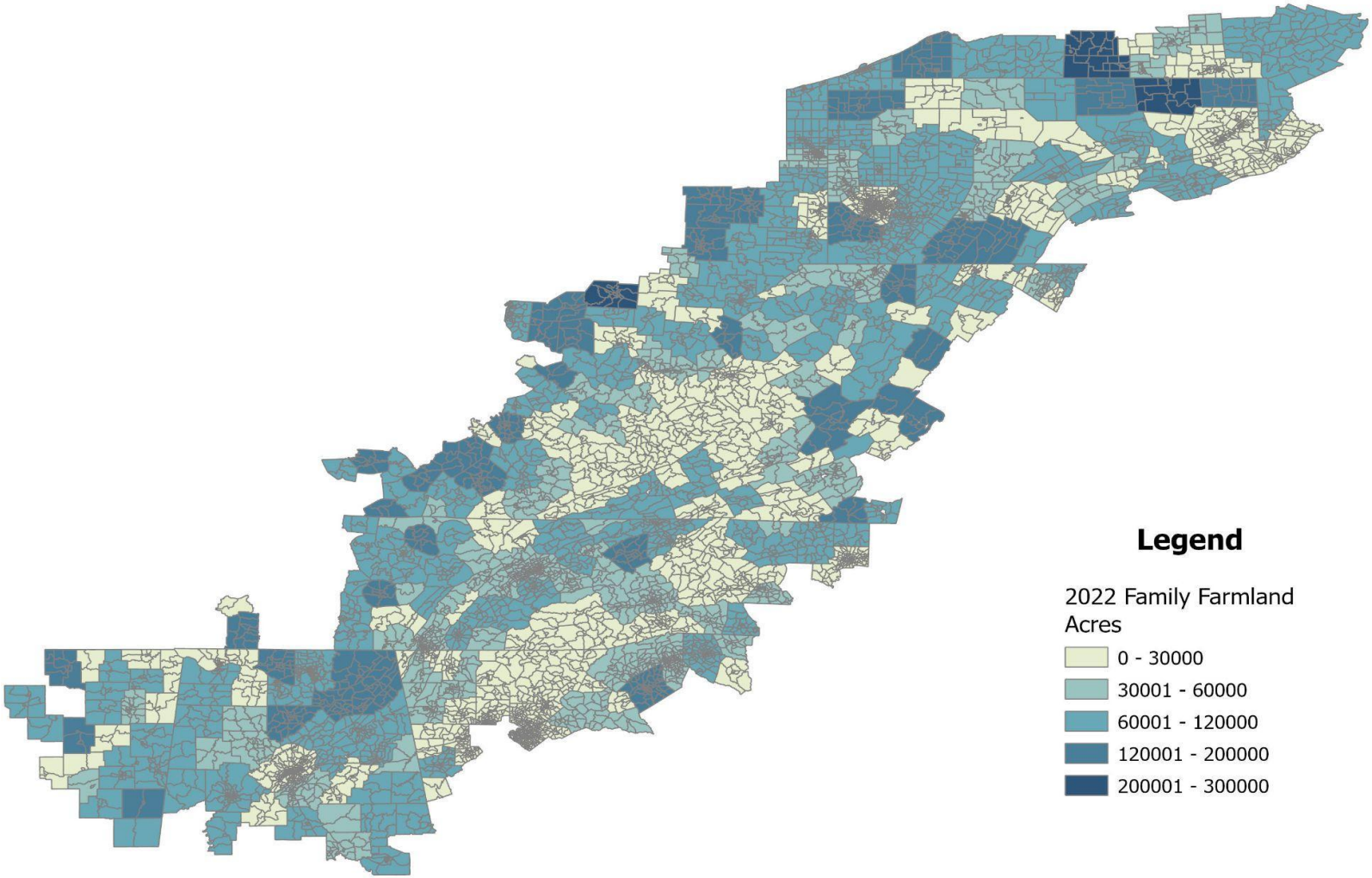


Figure A.20. Family-owned farmland acreage in Appalachia in 2022. Data from the United States Department of Agriculture National Agricultural Statistics Service 2022.

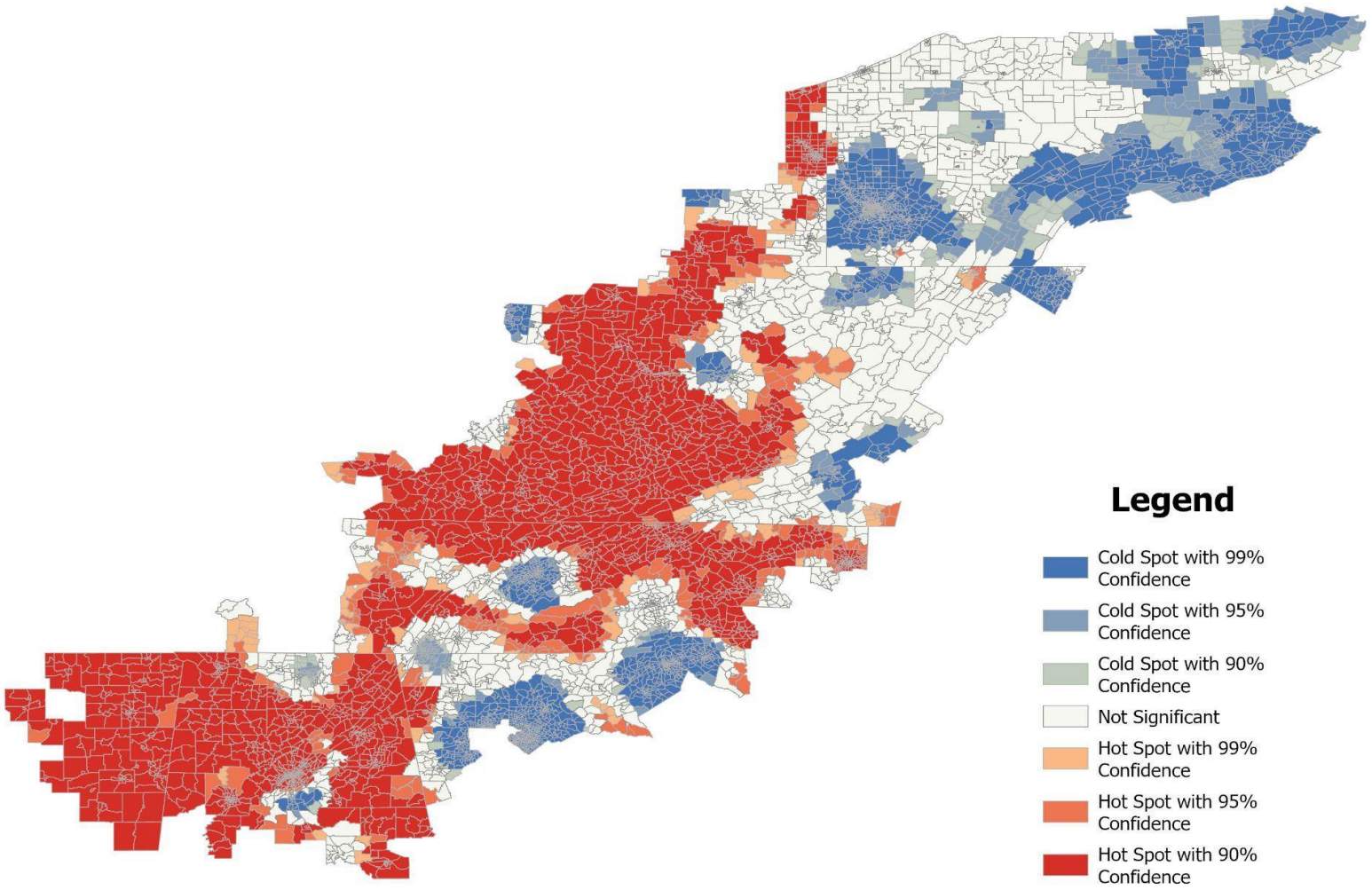


Figure A.21. Average food insecurity rates in Appalachia from 2010 to 2022. Data from the Feeding America organization's Map the Meal Gap Data Download for 2010 to 2022.

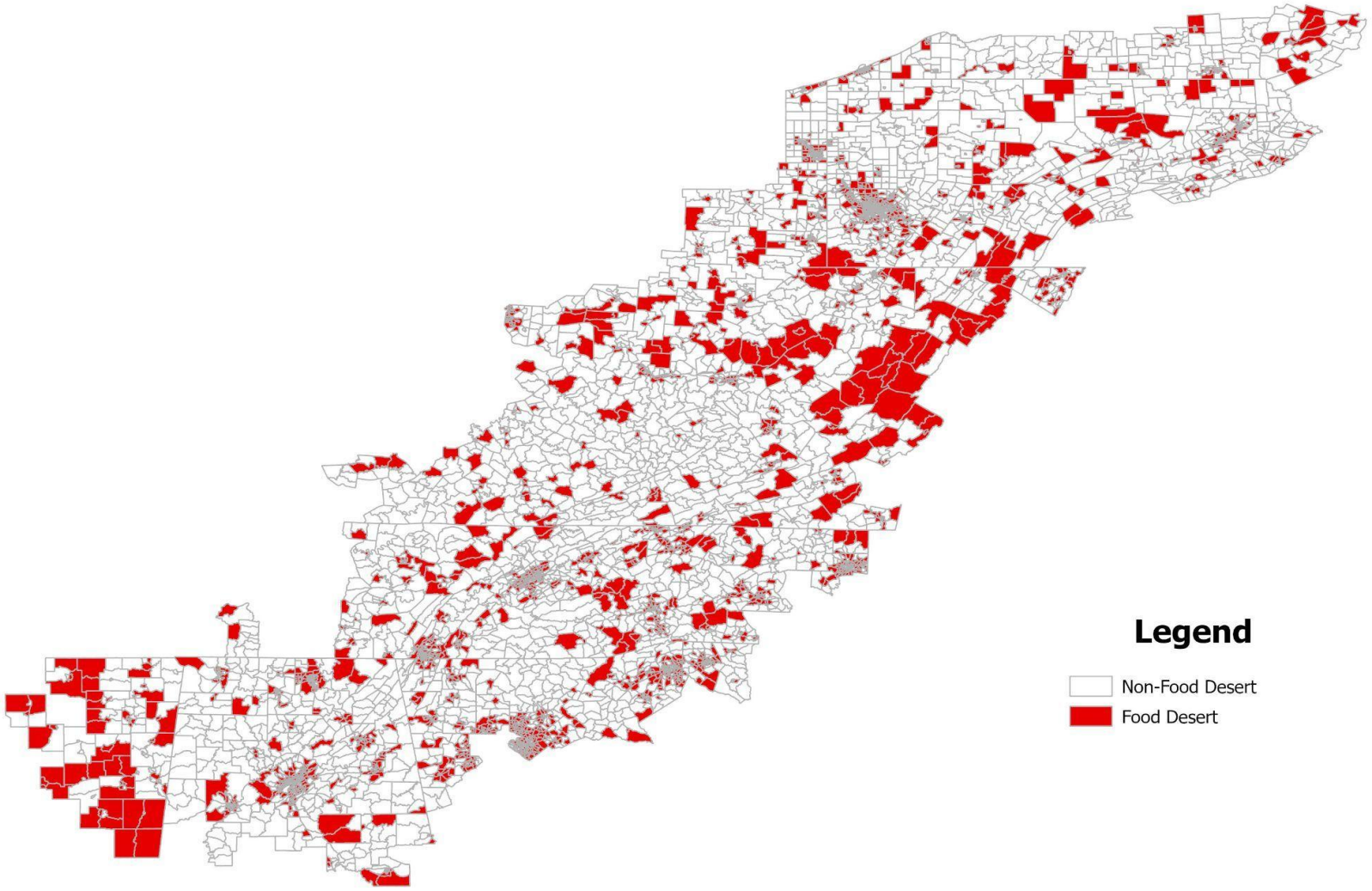


Figure A.22. Appalachian geographic food deserts in 2010. Data from the United States Department of Agriculture's Food Access Research Atlas in 2010.

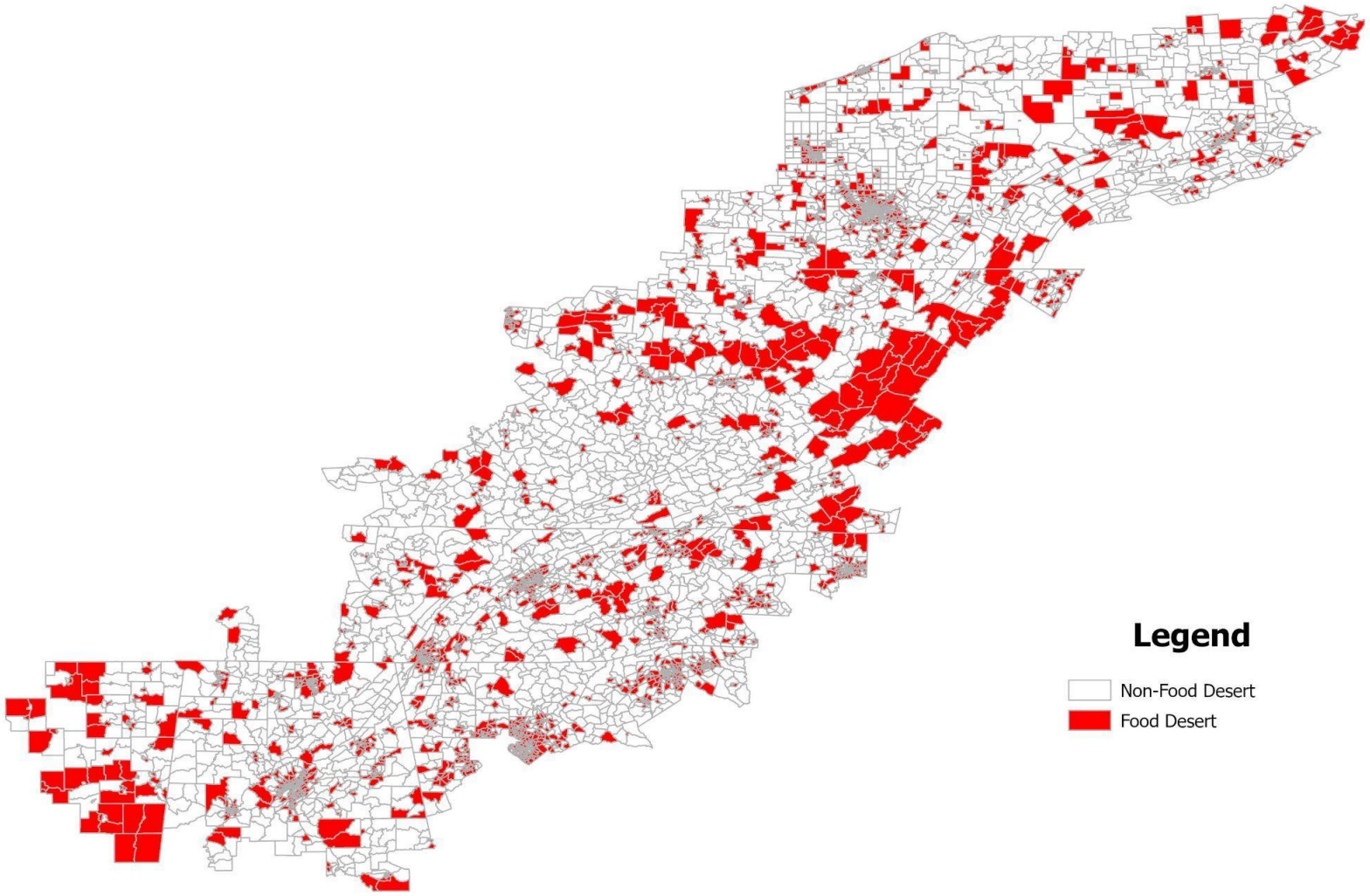


Figure A.23. Appalachian geographic food deserts in 2015. Data from the United States Department of Agriculture's Food Access Research Atlas in 2015.

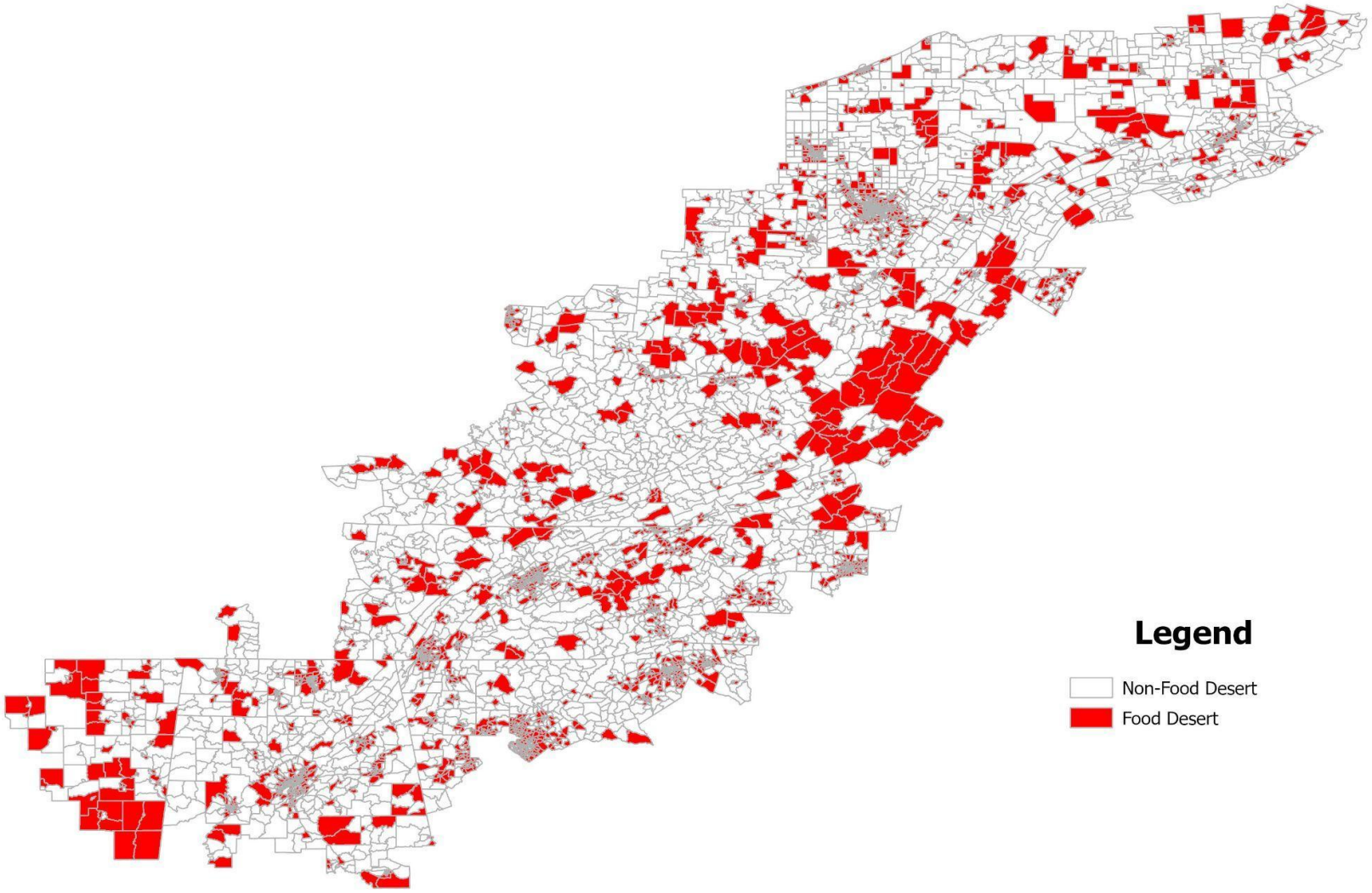


Figure A.24. Appalachian geographic food deserts in 2019. Data from the United States Department of Agriculture's Food Access Research Atlas in 2019.

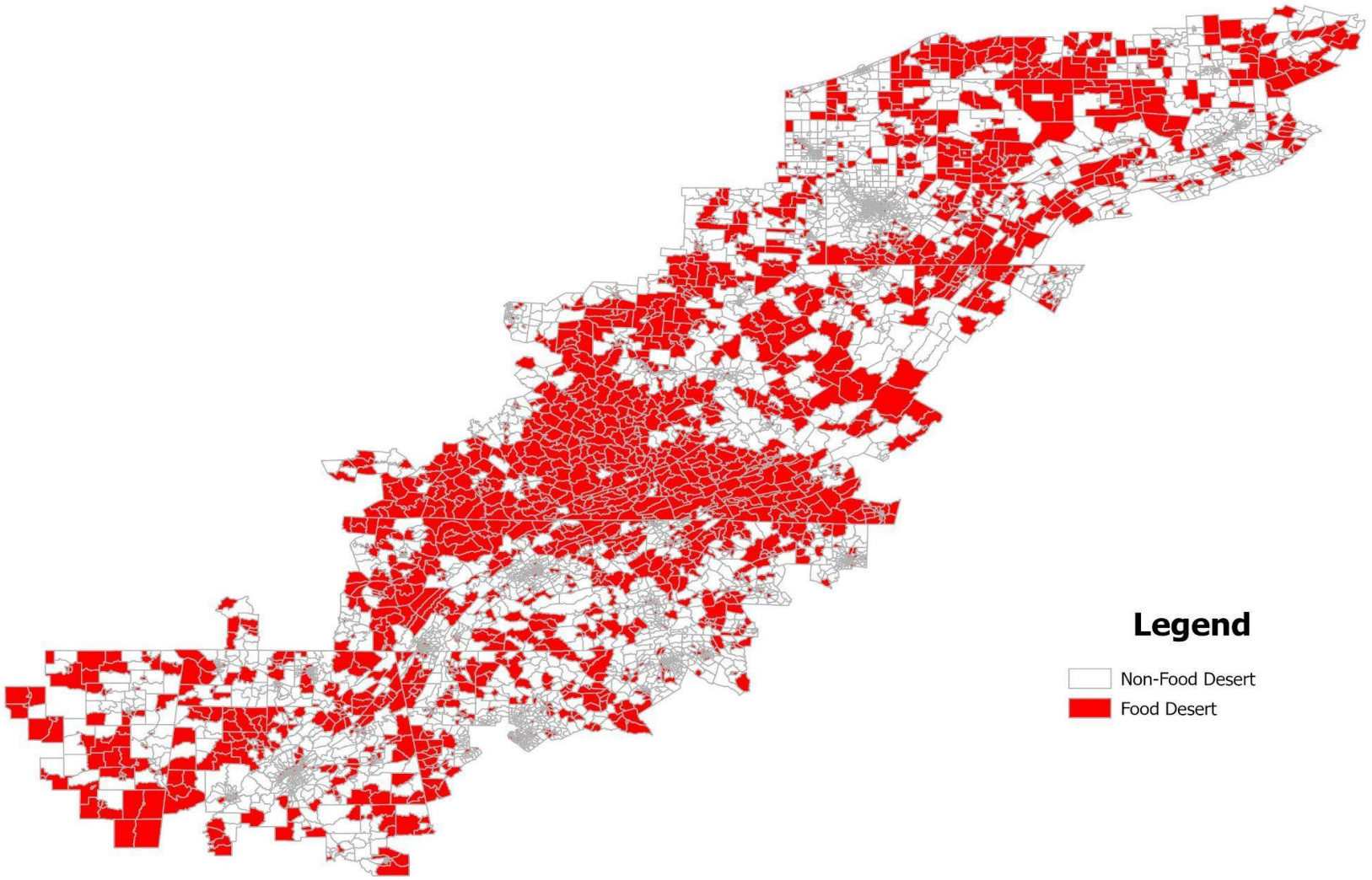


Figure A.25. Appalachian financial food deserts in 2010. Data from the United States Department of Agriculture's Food Access Research Atlas in 2010.

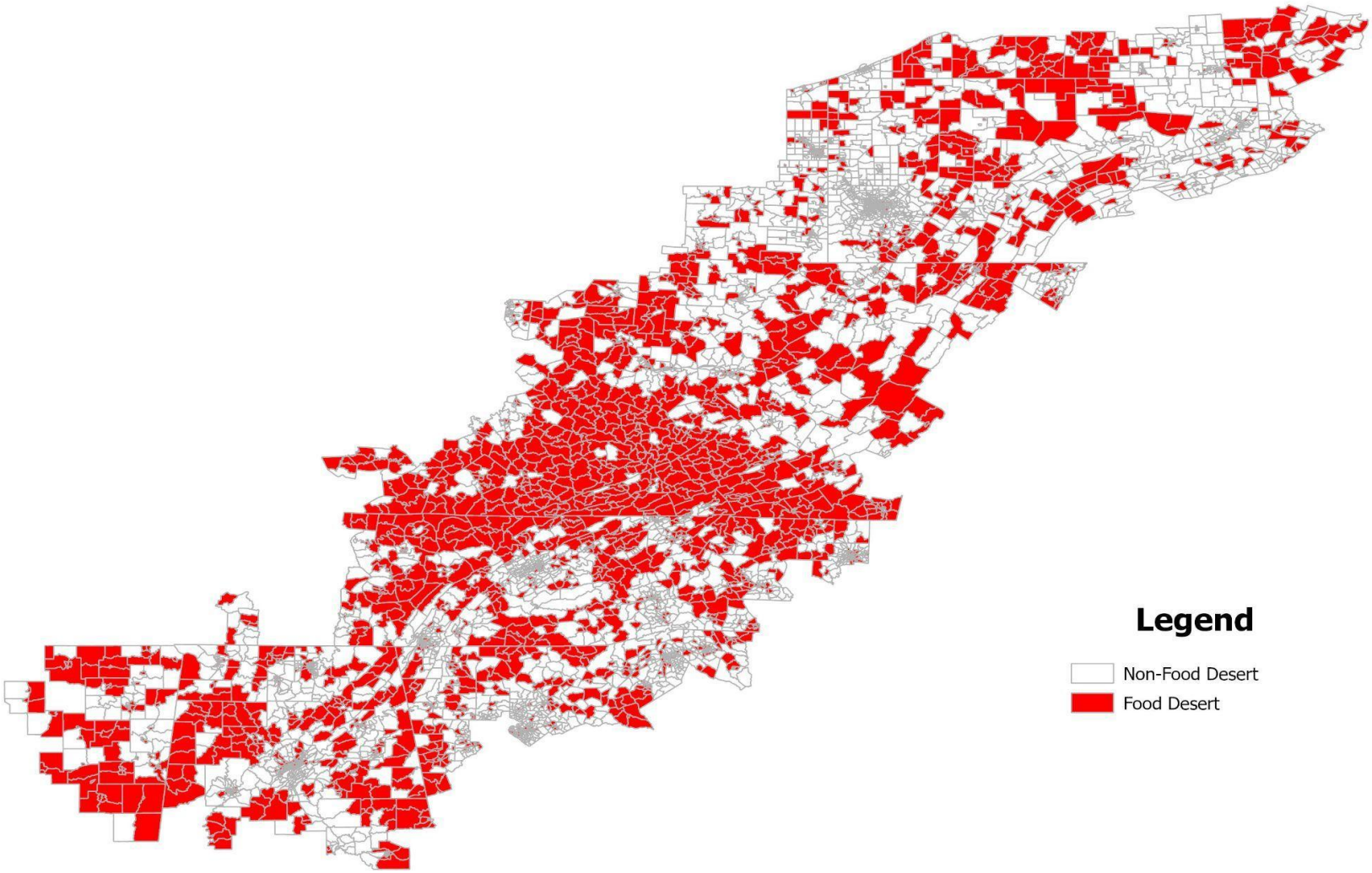


Figure A.26. Appalachian financial food deserts in 2015. Data from the United States Department of Agriculture's Food Access Research Atlas in 2015.

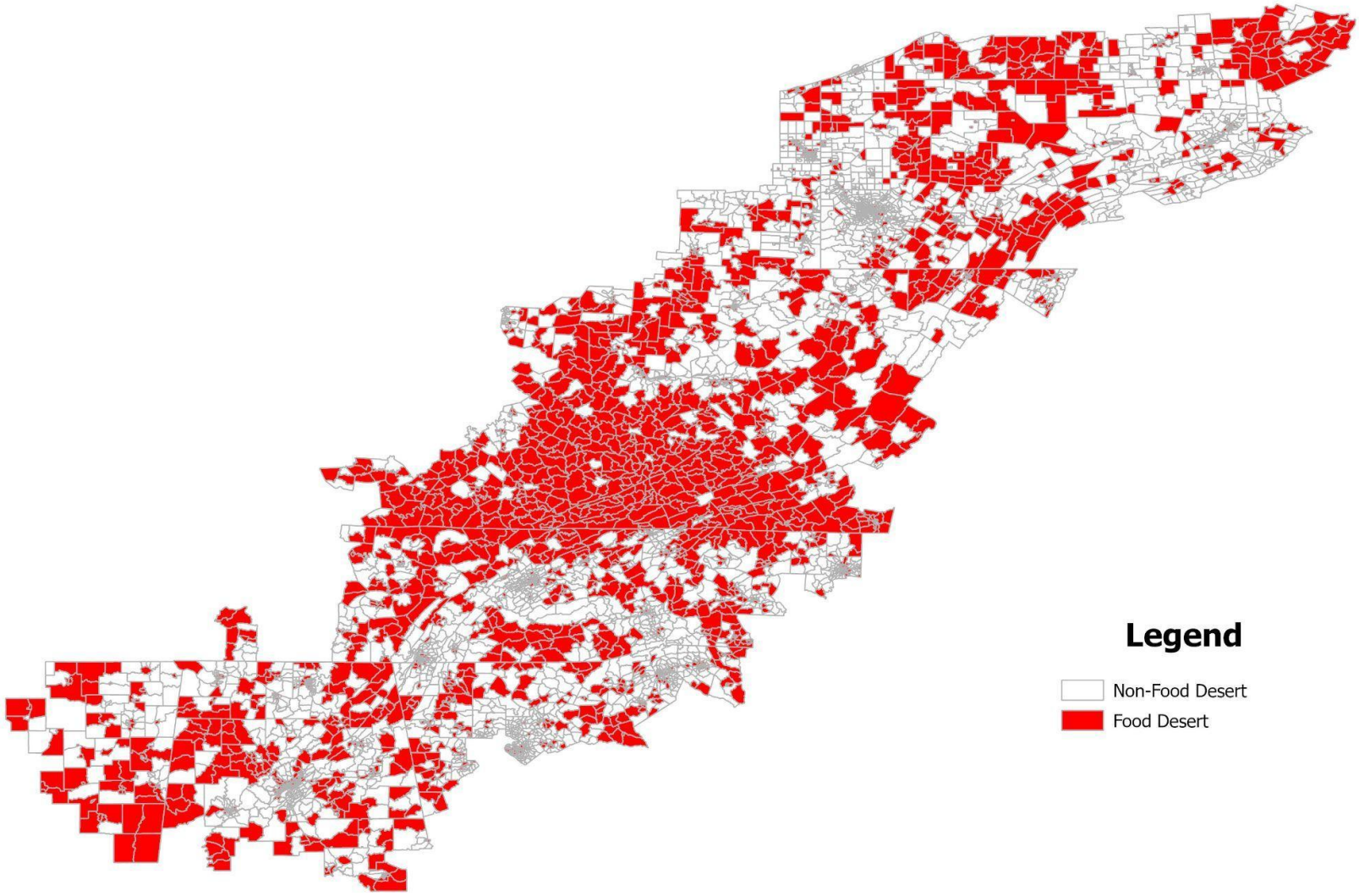


Figure A.27. Appalachian financial food deserts in 2019. Data from the United States Department of Agriculture's Food Access Research Atlas in 2019.

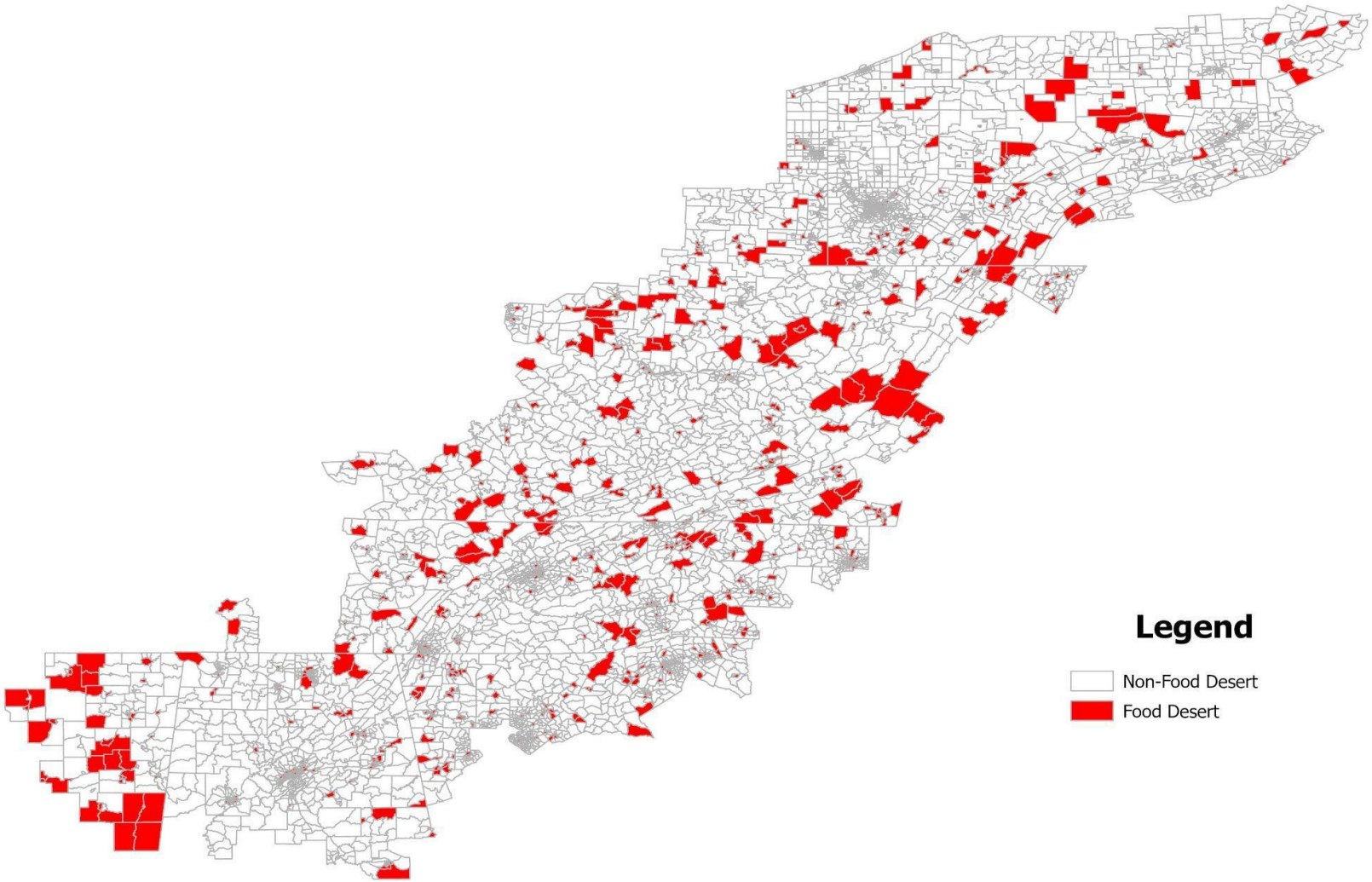


Figure A.28. Appalachian geographic and financial food deserts in 2010. Data from the United States Department of Agriculture's Food Access Research Atlas in 2010.

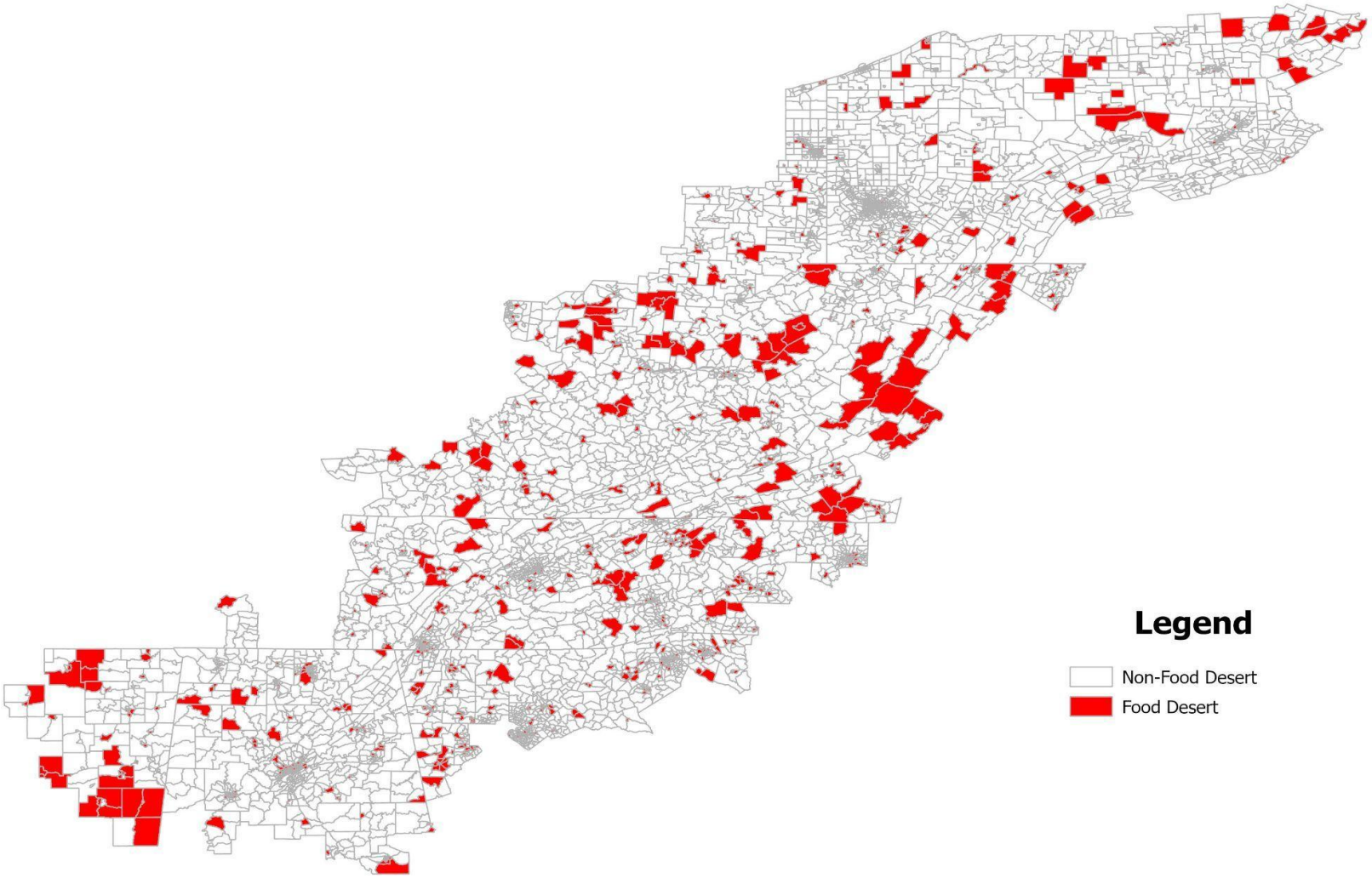


Figure A.29. Appalachian geographic and financial food deserts in 2015. Data from the United States Department of Agriculture's Food Access Research Atlas in 2015.

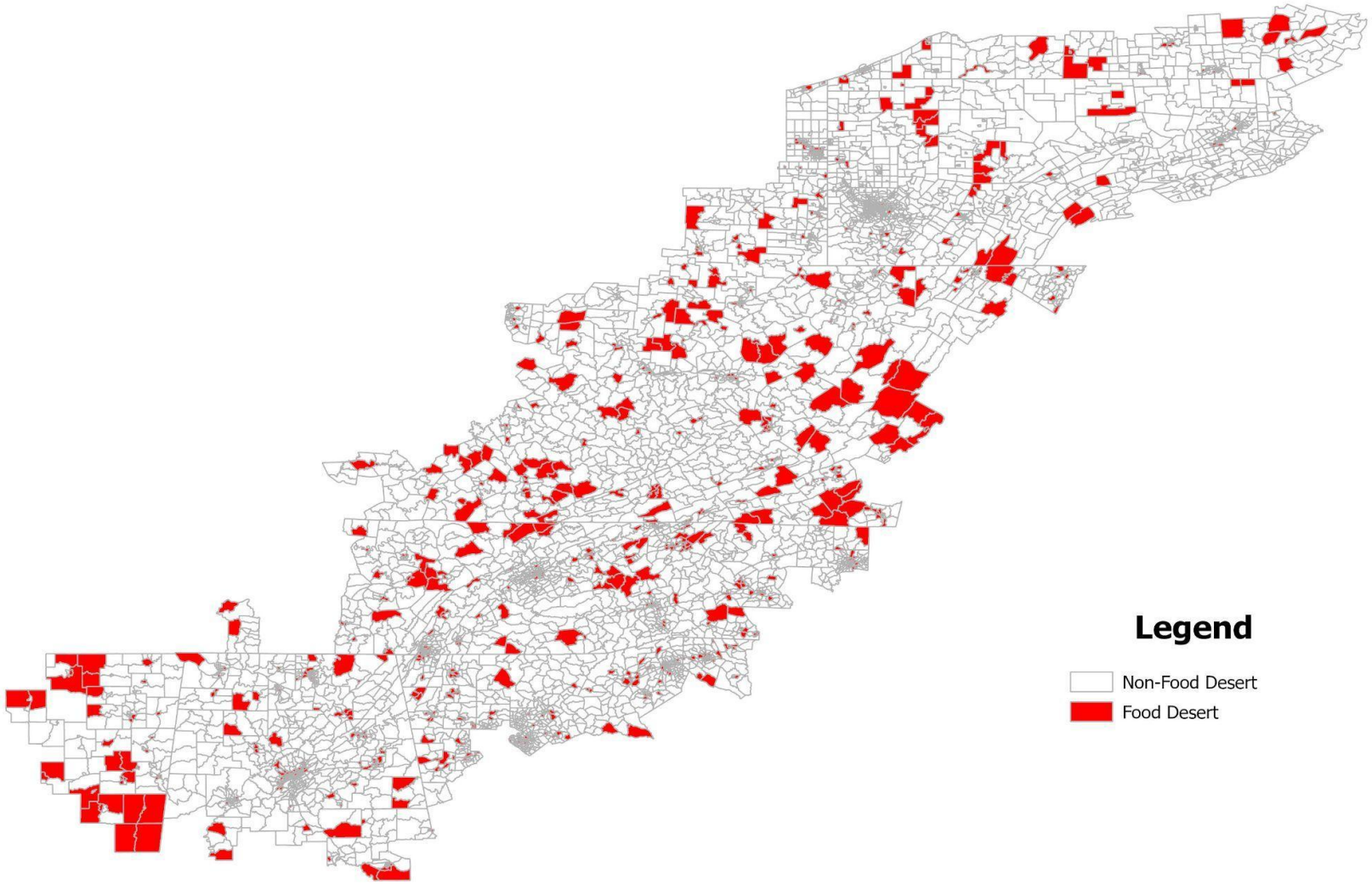


Figure A.30. Appalachian geographic and financial food deserts in 2019. Data from the United States Department of Agriculture's Food Access Research Atlas in 2019.

Chapter 3

Time-Series Trend Analysis of Appalachian Socioeconomic Factors

Abstract

The yearly trends for the socioeconomic variables show an overwhelmingly steady trend with very little variation throughout the years. Using a Random Forest model in R, the variables were ranked based on their influence on food insecurity. The variables that had the greatest impact were median household income, renters with excessive housing costs, lack of transportation, and the household poverty rate. The notable feature about these variables is the unifying theme of low-income households and the role that income status has on both financial but also geographic accessibility to food. As of 2019, those living in food insecurity rates higher than 10.5% (the national food insecurity rate in 2019) are 1.5 times more likely to also be living in a combination geographic and financial food desert. Food insecurity rates saw the most variation out of the variables, bouncing between 15.2% in 2010 and 11.9% in 2021 and then back up to 14.7% in 2022, compared to the 2022 national food insecurity rate of 12.8% (Feeding America, 2022). The median household income remained stagnant, leading up to 2017, when a gradual increase began. The median household income in the region in 2010 was $43,295, and $63,116 at the end of 2022. Comparatively, the median household income in the United States in 2010 was around $50,000 and $74,500 in 2022 (U.S. Census Reports, 2010 and 2022). The median household poverty rate was 8.1% in 2010 and saw slight fluctuations throughout the decade, ending at 7.0% in 2022. Comparatively, the national household poverty rate was 10.1% in 2010 and 8.9% in 2022. The common theme seen in Appalachia is that the variables of interest remain elevated and stagnant over time, suggesting that very little has been done, or has been unsuccessful, to address these issues.

3.1 Introduction

While this long-term aggregation offers a comprehensive visualization of the different factors and their spatial variability (see Chapter 2), it is also essential to perform annual analyses to evaluate how these factors influence food insecurity over time and to capture temporal variability. Such longitudinal assessments are critical for understanding how factors change over time, identifying consistent trends, and detecting potential disruptions or anomalies that may signal shifts in underlying conditions.

The goal of the yearly linear regressions was to understand the temporal trends of each of the variables within the Appalachian Region. This analysis identifies what socioeconomic challenges are worsening, improving, or unchanging over the past thirteen years. Using the change in time as a function of change in socioeconomic factors. The slope of the regression line indicates the rate of change, and the statistical significance dictates whether the change is meaningful or random variation. Following the temporal regressions, a Random Forest model was completed with the goal of ranking each variable and it's influence on food insecurity. Random Forest models are used in research such as this because of their built-in importance metrics, such as the mean decrease in accuracy (permutation) and the mean decrease in Gini impurity. The Random Forest's ability to handle large and noisy datasets makes it the ideal choice for studies that are examining parts of human behavior, such as food insecurity (Qi, 2012). After analyzing the temporal trends of the socioeconomic factors and ranking them by their influence, the second linear regression was done to understand the potential drivers of food insecurity. These analyses determine which of the socioeconomic factors have the strongest statistical association with the changes in food insecurity across the region. Similar studies have been conducted, primarily in other countries, but this has not yet been done in Appalachia. For

instance, a study was done to examine the socioeconomic impacts of COVID-19 on food security in Arab countries. Their results were found to be similar to the results of this study, primarily the influence of income on food stability and the use of a Random Forest importance feature mode (Qasrawi, et al., 2023).

The research questions addressed in this chapter include:

1. How have socioeconomic conditions and food insecurity in Appalachia changed over the past 13 years?
2. What factors have driven that change the most and what is the relationship between those factors and food insecurity?
3. What are the relationships between food insecurity and the differing food deserts in Appalachia?

3.2 Methods

The first portion of this chapter is focused on plotting the yearly trends amongst each of the variables, except for farmland and food deserts, as they have a select number of data years, as mentioned previously. The first was done to observe how each variable changed over time and how much of the trend variation is due to time passing. The yearly trends allow a further examination of how these factors tie into the target variable: food insecurity. Food insecurity was also examined to see its trends over time and how the region reacted to the potential rise and fall of insecurity over the study time.

The second portion of this chapter ties both the last chapter and the first portion of this chapter together, as it is a Random Forest study model to rank the importance of each factor. This will identify the most influential inputs for further consideration and notability in the next chapter. Although the Random Forest model can be used for predictions, it was not used for

prediction as it doesn't take into account the spatial relationships that were determined. The Random Forest model was used as it can average the data across the decision trees and give an estimation of the influence of the factors. As opposed to the Long Short-Term Memory Model used in the next chapter, the Random Forest didn't take into account any assumptions or training about the relationships between the individual factors and food insecurity, allowing for a flexible and impartial ranking.

The third portion of this chapter focused on the correlation between the higher-rated factors of influence and food insecurity using the ranking that was determined using the Random Forest. Dissecting the relationship, if any, between food insecurity allows for further insight into what is driving food insecurity in the region. However, it is also important to note that poverty and food insecurity are intertwined. While individual factors may not impact food insecurity, they may impact poverty, which indirectly impacts food insecurity. Therefore, a linear regression was calculated for each factor against food insecurity. Additionally, calculations were made for each factor and poverty to look for any indirect causes that may have been overlooked by previous research or indirect effects.

The data used in this chapter is the same as the previous chapter. The socioeconomic datasets were collected from the U.S. Census Bureau's 5-Year American Community Survey. The food insecurity data was resourced from the national organization, Feeding America, through their Mind the Meal Gap database. As the variables were shown to have relationships with one another, the null values in the dataset were dropped from modeling to avoid the possibility of modifying or influencing the existing relationships.

Displaying and analyzing the trends of each of the variables and their statistics was done through calculating a linear regression between each socioeconomic variable and using the year

as a factor. This was done in R using the lm() function, before plotting the regression results using the ggplot package for boxplots. The statistics for these analyses were gathered through the summary of the predictive yearly models. These statistics included the residuals to determine both the functionality of the model to capture and predict the socioeconomic variables accurately. The R^2 coefficient was used to determine the percentage of variance in the model that was explained by the yearly changes.

The goal of the yearly linear regressions was to understand the temporal trends of each of the variables within the Appalachian Region. This analysis identifies what socioeconomic challenges are worsening, improving, or unchanging over the past thirteen years. Using the change in time as a function of change in socioeconomic factors. The slope of the regression line indicates the rate of change, and the statistical significance dictates whether the change is meaningful or random variation. After analyzing the temporal trends of the socioeconomic factors and ranking them by their influence, the second linear regression was done to understand the potential drivers of food insecurity. These analyses determine which of the socioeconomic factors have the strongest statistical association with the changes in food insecurity across the region.

The statistical significance was classified using both the P-values and the F-statistics, as R classifies significance into 4 categories (highly, moderately, marginally significant, and not significant) while using the benchmark of 0.05.

This same process was done for the second round of linear regressions, using the variables that were found to be the most influential. Instead of using the year as a secondary factor, food insecurity was used to determine the amount of influence the factors had on food insecurity changes.

Determining the importance of each variable and how those variables play into food insecurity and poverty rates was done through a Random Forest model. Using the Random Forest package on R, the dataset was cleared of all null values. The cleaned dataset was then grouped based on the GEOID of the census tracts and summarized based on the numeric values of the variables. The model was trained using all available data, and the importance was determined using 500 trees and an mtry hyperparameter of 3. The data was not split for training or testing, as the main objective was to rank the factors' importance. This was decided because the model was not being used to predict future trends, as Random Forest models aren't equipped to predict time or spatial-based data, however, Random Forest was decided on because of its ability to handle nonlinear relationships between factors while calculating importance (Random Forests with Feature Importance, 2024). The importance of the factors was based on the permutation importance, and the mean squared error was used to observe which values had the greatest impact on food insecurity when removed from the dataset. After the model was completed, results were plotted using the ggplot2 package in R for a bar graph, and statistics were analyzed using the summary of the model. The relationship between food insecurity and food deserts was measured through contingency tables calculated for available food desert years (2010, 2015, and 2019).

3.3 Results

3.3.1 Yearly Linear Regressions and Principal Component Analysis

The yearly linear regression analysis for food insecurity revealed both declining and increasing trends over time. From 2010 to 2014, the median food insecurity rate remained relatively steady, followed by a notable decrease through 2018. The rate stabilized until a slight increase in 2022, ultimately resulting in an overall increase from the baseline. The residuals

ranged from a minimum of -8.40 to a maximum of 18.15, with a median residual of -0.35. The near-zero median residual suggests that the model has low overall bias, although the high maximum residual indicates the presence of several significant outliers.

The R^2 value for the model was 0.1045, meaning that approximately 10.45% of the variation in food insecurity can be explained by the progression of years alone. This relatively low value suggests that additional factors beyond time are influencing food insecurity trends. The results support the alternative hypothesis, that significant differences exist across the years, indicated by the consistently statistically significant p-values (all < 0.05) and a high F-statistic value of 812.2. As illustrated in **Figure B.1**, the trend is non-linear, with periods of both decline and increase. These fluctuations reinforce the idea that multiple interacting factors contribute to changes in food insecurity across the region. All figures for this chapter can be found in the chapter appendix below.

Neither individual nor household poverty rates were widely changing. The overall poverty rate experienced very little variation over the thirteen years. The range in residuals was -17.68 to 82.54, a larger range, which could lead to the model not capturing the full picture. This was aided by the median residual of -2.05, which points to the model slightly overpredicting poverty rates each year. The R^2 value was very small (0.008364), meaning that the model was hardly capturing the variation in the dataset year-to-year. The P-values were mostly highly statistically significant, except 2011 and 2018 being marginally significant, with both years having the value of 0.0129. The year 2019 was not significant and had a P-value of 0.993. As shown in **Figure B.2**, the poverty rates had very little to almost no variation from 2018. The F-statistics were 59.51, supplementing the slow-moving fluctuations in the trends. Comparatively, the household poverty rate had a residual range of -9.43 to 91.20, with a median of -1.33. The range was large

enough to consider that the model was not able to accommodate the full prediction based on year alone, with the median pointing to a slight amount of over-prediction. The R^2 value was 0.01404, which was larger than the overall poverty rate, approximating that 1.404% of the variation is attributable to the yearly change. All the years were considered to have statistically significant variation, with the majority of years having a highly significant variation, except 2011 and 2017 being marginally significant (P-values of 0.04091 and 0.04600). The variation and the trend line were paired with an F-statistic of 98.56, which does verify the variation in years but also agrees with the visual plot that the variation between years is quite small and slow-moving (**Figure B.3**).

As shown in **Figure B.4**, median household incomes did not change until the year 2017, when incomes saw a gradual increase going into 2022. The residual range was -$55,616 to $189,815, suggesting that there is a large variation in the accuracy of predictions on a year-to-year basis, as this is just an income-to-year relationship. The median household income overall was $43,295.90, while the median residual was -2,961. A negative median suggests that, on average, the model is overpredicting incomes. For instance, within the range, the largest overprediction was a household income of $55,616, and the largest underprediction was $189,815. The R^2 value was 0.08721, approximately 8.721% of the variation in income is explained by changes over the years, which is in line with the adjustment of wages to fit inflation and the link between income and other socioeconomic variables. The P-values were largely highly statistically significant, confirming that there is a change on a year-to-year basis, with the exceptions of 2011 being marginally significant (0.015) and 2012 being statistically significant (0.0016). The F-statistic (662.7) was high enough to confirm the fluctuations in the data and its statistical significance.

Unemployment rates across the thirteen years remained fairly constant (**Figure B.5**). The residual range was -5.86 to 65.00, with the median of -0.46. A median that is nearly zero suggests that the unemployment model is relatively unbiased and isn't substantially overpredicting the unemployment rate. The R^2 value was 0.1329, approximately 13.29% of the variation in unemployment is related to yearly changes. The P-values were all highly statistically significant, with the one exception of 2015, which had a P-value of 0.555. The F-statistic confirmed the statistical significance of the yearly fluctuations with a value of 1063.

The percentage of households that lack access to transportation had almost no variation (**Figure B.6**). However, the extremely low P-value and F-statistics suggest that the variation that did occur was meaningful. The P-values had a variety of significance, with 2010, 2020, and 2021 being highly significant and 2013 and 2022 being moderately significant. The rest of the years had larger P-values and were not significant, concluding that the variation between years, while there was some scattered throughout, was not enough to suggest that the yearly changes had any effect on accessibility to transportation. This suggests that while changes occurred, they were likely influenced by factors other than time alone or were only prominent in specific years, rather than showing a consistent upward or downward trend. The F-statistic was 8.19, and the R^2 was 0.001182, both due to the lack of variation. The residual range was -8.66 to 92.34, with a median of -2.53. The larger range and larger median indicate that the model cannot capture the true trend in the data based on the relationship with the year, while also overpredicting each year.

The percentage of households receiving SNAP benefits saw a small and brief increase between 2013-2017 before declining back down (**Figure B.7**). The residual range was -16.68 to 87.76, with the median (-2.18) slightly overpredicting the percent of households. The R^2 value (0.014) indicates that 1.4% of the variation was due to the yearly changes. The P-values were all

highly statistically significant, and the F-statistic (98.52) was high enough to agree with the statistical significance, while also being low enough to couple with the very gradual changes seen in the trends. The percentage of households that received SNAP benefits that were under the poverty line saw a steady and slow decline overall (**Figure B.8**). The residual range was a tad larger, -56.27 to 54.20, which could be the result of the number of outliers or the absence of other factors in the calculation. The small positive median (1.73) indicates that the model slightly underpredicted the number of impoverished households that had received SNAP. The low value of R^2 (0.02285) suggests that a small amount of variation was due to the yearly change. The P-values were all statistically significant (P-values < 0.05). The statistical significance of the yearly variations was aided by the larger F-statistic of 162.2. Lastly, the percentage of total households that were both receiving SNAP and under the poverty line followed the same pattern as the total number of households receiving SNAP, a slight increase between 2013 and 2017 before declining back down (**Figure B.9**). The residual range was smaller, -9.48 to 65.60, pointing to a smaller number of outliers or an increase in predictability with the model. The median was -2.00, suggesting that the model was slightly overpredicting the percentage of households. Similarly to the percentage of households receiving SNAP, the R^2 value (0.01127) indicated that only 1.1% of the variation was caused by the yearly changes. The P-values were mostly highly statistically significant; however, 2020-2021 were both statistically insignificant, and 2022 was marginally significant (0.4416, 0.1003, and 0.0181, respectively).

The percentage of rental units that had housing costs exceeding 30% of their monthly incomes saw slight variability over time, gradually increasing and decreasing (**Figure B.10**). The residual range was large, -41.01 to 62.73, and the median was -0.048, which suggests that this model is at risk of outliers within the dataset or the model not being able to capture all the trend

just based on yearly trends. The R^2 value (0.007261) was low, pairing with the mentioned inability to capture the yearly trend. Most P-values were highly statistically significant; however, 2016-2017 were insignificant (p = 0.3586 and 0.2346, respectively). This pattern was seen with the F-statistic as well, 50.73.

The percentage of housing units with a mortgage saw a slight decline in the percentage of units with excessive housing costs beginning in 2015 (**Figure B.11**). The residual range was again large, -31.28 to 93.98, with the same reasoning. The median was -1.44, which suggests the model slightly overpredicted the percentage of units over time. The R^2 value (0.06099) was higher than the rental units; approximately 6.099% of the variation was due to the yearly shifts. Most P-values were statistically significant, except 2011-2012 being insignificant (0.4289 and 0.5217). The F-statistic (450.5) was much larger, confirming the variation between the years and the decrease overall. Lastly, the percentage of housing units without a mortgage that had excessive housing costs was fairly unchanging (**Figure B.12**). The residual range was similar to the rental and mortgage units, -13.37 to 89.27, following the same pattern of reasoning. The median (-1.13) suggests that the model was also slightly overpredicting the data. Approximately only 1.125% of the variance was related to the yearly trend. The P-values were highly significant, except for 2011 (0.0919), which was insignificant, and 2012 (0.0038), which was moderately significant.

3.3.2 Random Forest Importance Ranking of Influential Food Insecurity Variables

The Random Forest model's accuracy was determined by both the R^2 and the Root Mean Squared Error (RSME) used in the training set. The R^2 value was 0.8948, indicating a high level of accuracy. The RSME was 0.8759, indicating a low prediction error in the training set. These two values give confidence to the results of the model. The results of analysis using the Random

Forest importance ranking (**Table B.2** and **Figure B.13**) show that the variable that had the most influence on food insecurity was median income, with its increase of mean squared error (MSE) being 58.24%. This large increase indicates that income has a significant influence and relationship with food insecurity. The percentage of increase in MSE is the percentage of increase in model error when the data points for food insecurity are predicted; the higher the percentage increase, the greater the importance. The model calculated that renters with excessive housing costs (57.41%) and lacking transportation (57.05%) had a significant influence on food insecurity. After these three variables, there was a drop in influence, with the next variable, household poverty, having an MSE of 42.15%. The household poverty rate was deemed more important than the overall poverty rate (42.15% versus 34.34%, 4th versus 7th ranking). Whether or not a household received SNAP plays a large role in food insecurity, with the percentage of households that received SNAP that were also below the poverty line having an increase of MSE of 36.43%. Those receiving SNAP, and the percentage of total households that received SNAP and were below the poverty line, ranked 6th and 8th (35.50% and 33.363). Compared to the percentage of renters, units with and without a mortgage ranked considerably lower (33.35% and 25.21%). The unemployment rate had a surprisingly less influence on food insecurity than was originally thought, ranking at 10th in importance with an increase in MSE of 29.06%. Lastly, population changes and lack of completed plumbing had the least amount of influence on food insecurity (18.44% and 14.43%, respectively).

3.3.3 Correlations between Food Insecurity and Food Deserts

Analyses of the top influential factors and their roles in food insecurity were calculated using linear regression. Median household income and food insecurity have a negative correlation; as household income decreases, the likelihood of food insecurity increases. The

correlation model between food insecurity and median household income yielded a small residual range and median (-9.15 to 18.52 and -0.26), indicating that the predictive model for the two factors was concise. The R^2 (0.1428) indicates that the household income accounts for 14.28% of the variance seen in the model (**Figure B.14**). The relationship between income and food insecurity was statistically significant, with the F-statistic being 1.37×10^4.

Renters with excessive housing costs are more likely to be food insecure. The residual range and median were -8.70 to 19.42 and -0.19, therefore, the model is accurate and can see the extent of the relationship. However, the R^2 and the F-statistic (3.919×10^{-6} and 0.3228, respectively) were quite low. This, as well as the visual relationship shown in **Figure B.15**, suggests that there was very little variation in the data and that the variation in the data is only partially due to the relationship between the two factors. The percentage of households that lack access to transportation (**Figure B.16**) saw a similar result to the relationship with rental units, with a closer residual range (-8.77 to 19.37) and a low median (-0.20). The relationship garnered a smaller R^2 (0.001803) and F-statistic (148.8); there was some variation between the two factors, but little of this variation was explained by the relationship.

The relationship between the household poverty rate and food insecurity was more akin to the relationship with income. The model was predicting fairly concisely and was accurate with its prediction (residual range of -17.35 to 18.63, and median of -0.090. Approximately 10.50% of the variation in the model (**Figure B.17**) was explained by the relationship between the two variables, and there was a significant change in food insecurity based on the household poverty rate observed.

The relationship between food insecurity and food deserts was calculated by using a contingency table, **Table 3.1**. The table was based on national food insecurity values for the

years available for food desert data (2010, 2015, and 2019). This was done to determine the risk ratio of census tracts with food insecurity rates over the national rate and how likely they were to also be in one of the types of food deserts. The risk of being above the national food insecurity rate and being in a geographic food desert has steadily decreased over time, but the risk of being in a financial food desert has increased, and the risk of being in both has increased between 2015 and 2019.

3.4 Discussion

The yearly trends for each of the factors were found to be significant overall. The determined over- and under-predictions of the variables were slim and expected to have no large impact on the impartialness of the model itself. The significance, based on the P-value and F-statistic, across the board for each factor was majority ruled by highly statistically significant, suggesting the idea that the yearly changes are responsible for some of the variance in the dataset overall. The amount of variance due to the year was slight, indicated by the value of R^2, with the highest value being only 13.29% (unemployment rate), but with the majority of the coefficients being below 10%. The changes were related to the year and significant enough to be notable; the passing of time year-to-year was the minority reason for change. The yearly changes yielding incredibly small R^2 values for many of the factors, however the relationship analyses between the importance factors and food insecurity showed that while individually and temporally the factors do not have much influence, but when paired together or paired against food insecurity they were major drivers of change. The majority of reasons for change were the relationship that existed between these factors. For example, while poverty rates do fluctuate slightly over time, the main reason these rates change is due to factors such as unemployment or household income changes.

Median household income has the biggest influence on food insecurity and plays a crucial role in household poverty as well. The regression plot in **Figure B.14** showed a clear relationship between household income and food insecurity, with the rates of insecurity rapidly dropping as income increased. This ties into the household poverty regression analysis, which shows that the food insecurity rate dropped as incomes crossed the poverty threshold and that once incomes exceeded the "at-risk" homes (homes bordering on the poverty line), the food insecurity rate was almost nonexistent.

The relationship between rental units with excessive housing costs and food insecurity rates was proven to be substantial in the importance ranking, but there was very low variation in the relationship regression model. This is explained through observing the food insecurity for the rental units across the board. Any rental unit with housing costs exceeding 30% of their monthly income is more likely to see a steady trend in food insecurity rate. This finding suggests that while the percentage of rental units in one tract increases, the food insecurity rate in that rental unit is nearly the same as in other units with excessive costs. This reasoning could be applied to the lack of transportation. Households lacking transportation while also being in food insecurity are not going to see a significant difference in their access to transportation on a year-to-year basis, resulting in very little variation but a strong relationship, as those lacking access to transportation were disproportionately more likely to be experiencing food insecurity.

3.5 Conclusions

The year-to-year data shows a variance that cannot be captured on a still map. These linear regressions, combined with the Random Forest model to determine variable importance, showed that the factors most impacting food insecurity, and in turn food deserts, were household income, renters with excessive housing costs, lack of transportation, and household poverty rates.

While it may seem obvious that these factors would have a strong relationship with food insecurity, determining the strength and the reliance of the relationship is vital when it comes to predicting the future of the region and providing insight into what socioeconomic link needs to be addressed to start reducing the number of food deserts and the rate of food insecurity. For example, household income has the greatest impact on food insecurity rates, and remained unchanged until 2017 when it began gradually increasing. The unemployment rate in the region has remained elevated and constant, indicating that the employment opportunities have stayed the same. These two observations, paired together, conclude that an issue in the region is the lack of opportunities and wage stagnation, which influences the rates of food insecurity (Qasrawi, et al., 2023; Deléglise, 2022). The influence that variables have on food insecurity also influences the presence of food deserts, as concluded by the contingency tables. This research found that in 2019, census tracts with food insecurity over the national average, 10.5%, were 1.67 times more likely to be a financial food desert, 0.77 times less likely to be in a geographic food desert, and 1.50 times more likely to be in both geographic and financial deserts.

References

(2024, April 5). *Feature Importance with Random Forests*. GeeksforGeeks. https://www.geeksforgeeks.org/feature-importance-with-random-forests/

Qi, Y., 2012. Random forest for bioinformatics, in: Ensemble machine learning. Springer, pp. 307–323.

Qasrawi, R., Hoteit, M., Tayyem, R., Bookari, K., Sabbah, H. A., Kamel, I., Dashti, S., Allehdan, S., Bawadi, H., Waly, M., Ibrahim, M. O., CORONA COOKING Survey Group, T. R., Polo, S. V., & Al-Halawa, D. A. (2023). Machine learning techniques for the identification of risk factors associated with food insecurity among adults in Arab countries during the COVID-19 pandemic. *BMC Public Health*, *23*, 1805. https://doi.org/10.1186/s12889-023-16694-5

Appendix B

Table B.1. Contingency table calculations, food insecurity to food desert risk ratio.

Rates	2010	2015	2019
National Food Insecurity Rates	14.5%	12.7%	10.5%
Geographic Deserts	1.03	0.96	0.77
Financial Deserts	1.31	1.36	1.67
Combination Deserts	1.54	1.38	1.50

Table B.2. Yearly linear regression model summary statistics.

Factor	Minimum Residual	Maximum Residual	Median Residual	R2	F-Statistic
Food Insecurity	13.51%	13.20%	-2.01	-0.86	0.00071
Poverty	15.81%	13.98%	-0.58	-0.23	0.099
Household Poverty	7.91%	6.70%	-0.61	-0.31	0.17
Income	X	$48,510.50	1.28	-0.47	0.07
Transportation	7.95%	5.43%	0.68	-0.23	0.26
Unemployment	3.97%	3.73%	0.14	-0.039	0.17
Lack of Complete Plumbing	0.48%	0.18%	-0.45	-0.61	0.33
Disability	16.43%	16.10	-1.57	-0.54	0.029
Lack of Health Insurance	9.90%	8.92%	-0.33	-0.24	0.0075
Receiving SNAP	14.27%	12.54%	-0.60	-0.23	0.12
Received SNAP under Poverty Line	49.28%	51.26%	-0.94	-0.20	0.15
Household receives SNAP and is in Poverty	7.79%	6.10%	-0.64	-0.33	0.12
Renters above 30%	38.71%	35.58%	0.84	0.20	0.13
Mortgage above 30%	26.40%	25.52%	-0.0092	-0.30	0.17
Non-Mortgage above 30%	11.63%	10.90%	0.36	-0.25	0.13

Table B.3. Variable Importance on Food Insecurity Ranked.

Factor	Percent Increase Mean Standard Error
Median Income	58.24
Renters above 30%	57.41
Lack Transportation Rate	57.05
Household Poverty Rate	42.16
Percent of Households with SNAP below the Poverty Line	36.43
Percent Total Households with SNAP and below Poverty Line	35.50
Poverty Rate	34.34
Percent of Households with SNAP	33.63
Mortgage above 30%	33.35
Unemployment Rate	29.06
Year	28.98
Non-Mortgage above 30%	25.21
Population	18.44
Percent Lacking Complete Plumbing	14.53

Table B.4. Regression statistics when variable is compared against food insecurity rates.

Factor	Minimum Residual	Maximum Residual	Median Residual	R^2	F-Statistic
Median Household Income	-9.15	18.52	-0.26	0.143	1.37×10^4
Renters above 30%	-8.70	19.42	-0.19	<0.001	0.3228
Lacking Transportation	-8.77	19.37	-0.20	0.002	148.8
Household Poverty Rate	-17.35	18.63	-0.090	0.105	9666

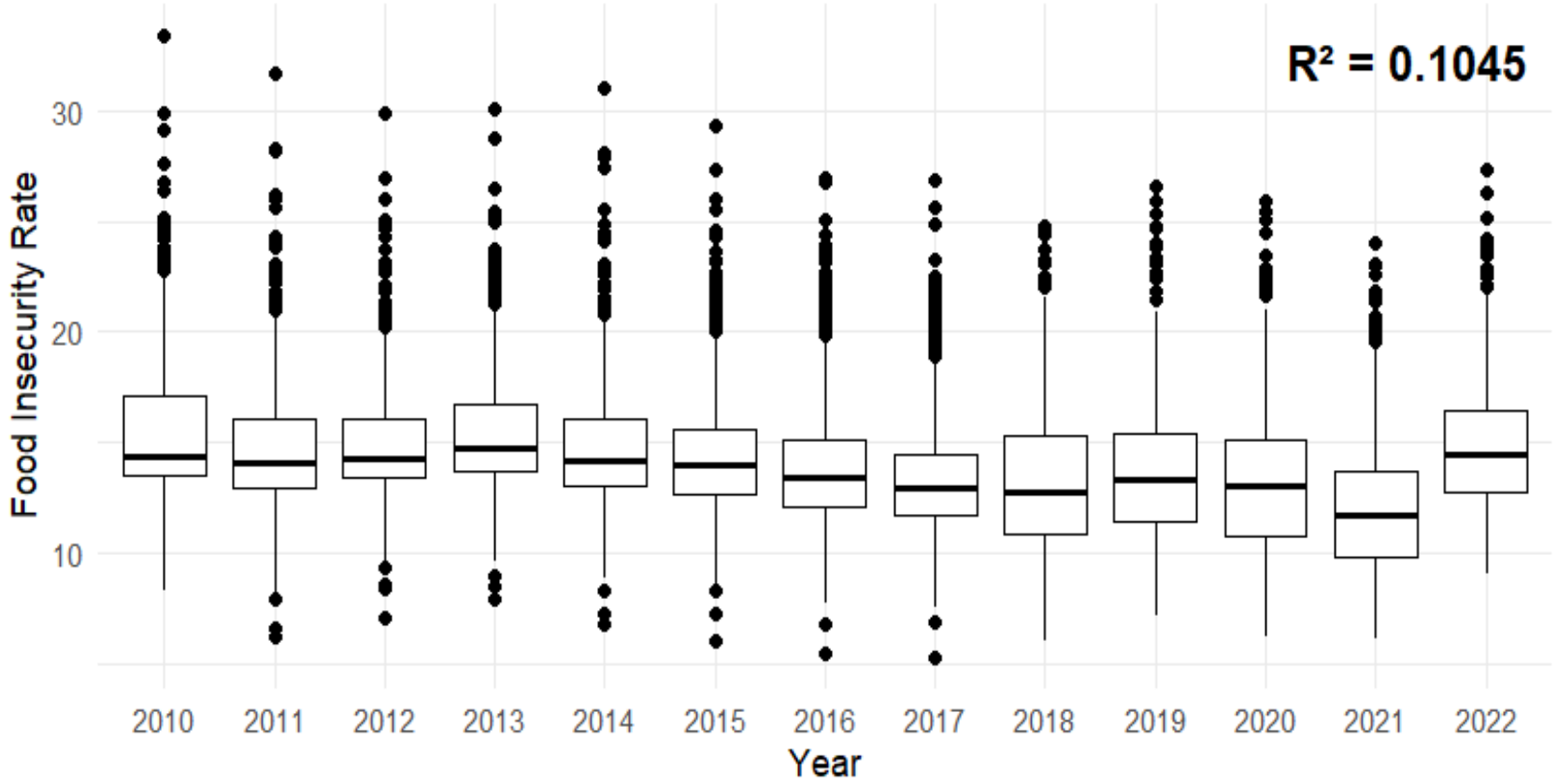


Figure B.1. Food insecurity rates in Appalachia from 2010 to 2022. Data from Feeding America's Mind the Meal Gap Database.

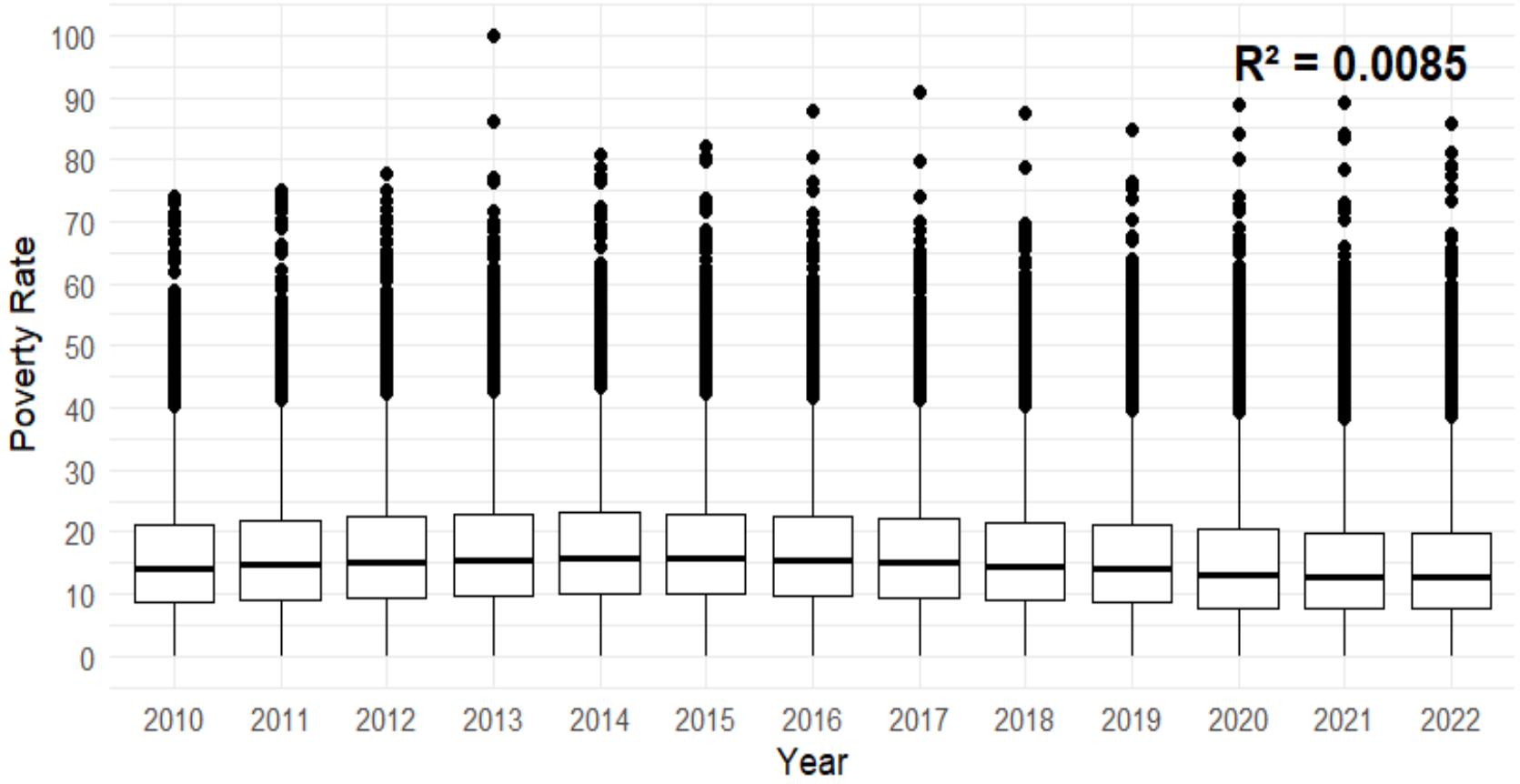


Figure B.2. Overall poverty rates in Appalachia from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022.

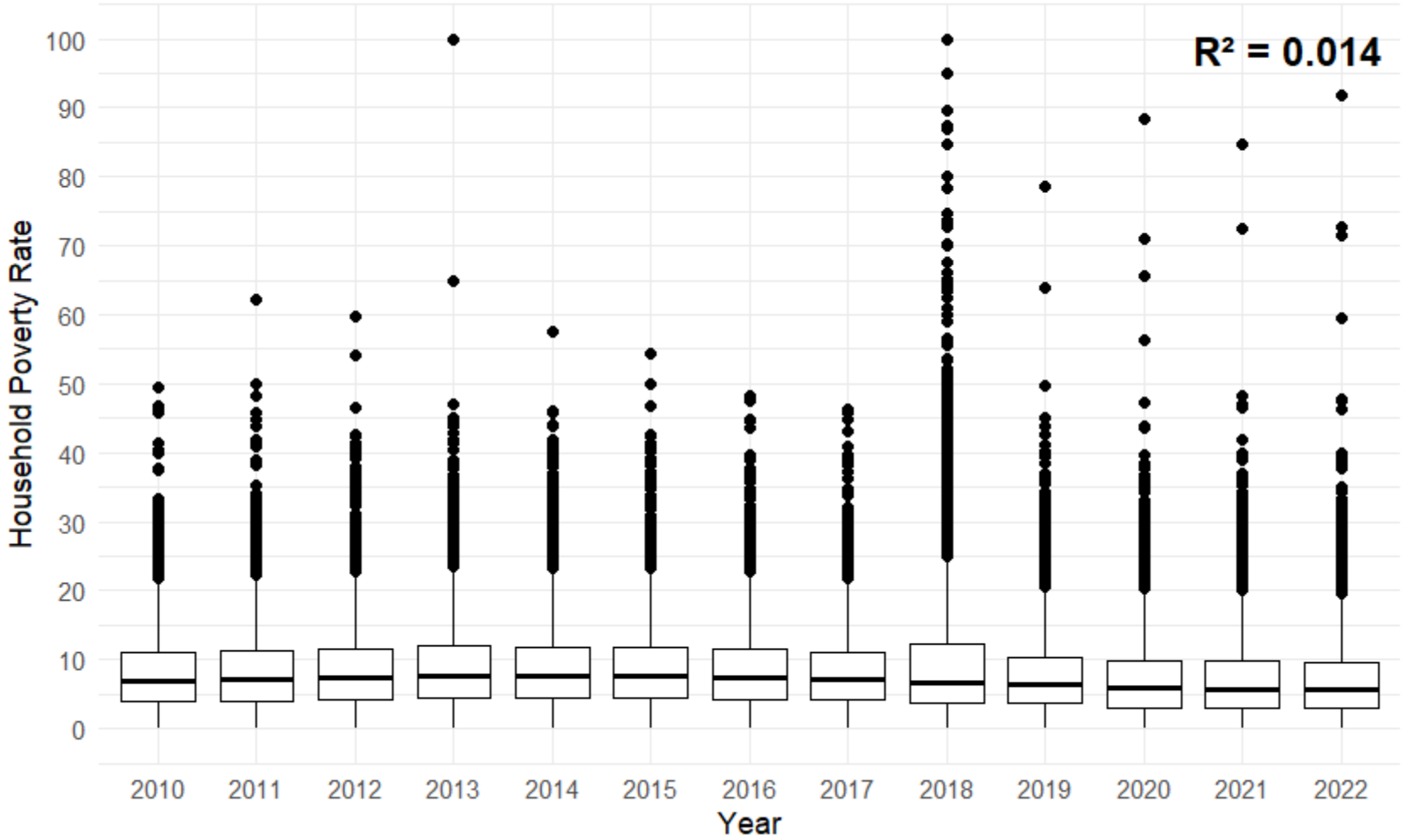


Figure B.3. Household poverty rates in Appalachia from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022.

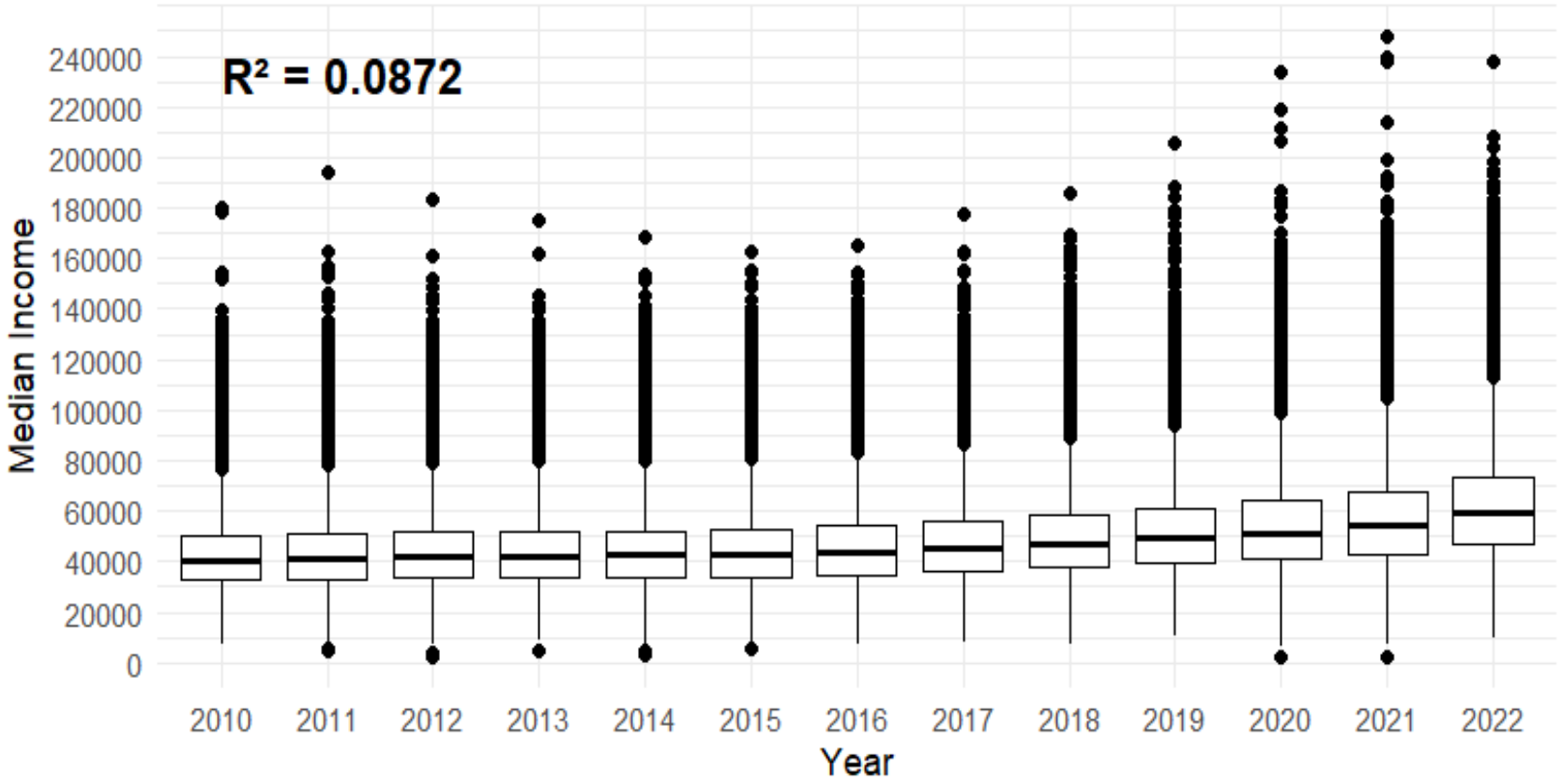


Figure B.4. Median household incomes in Appalachia from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022.

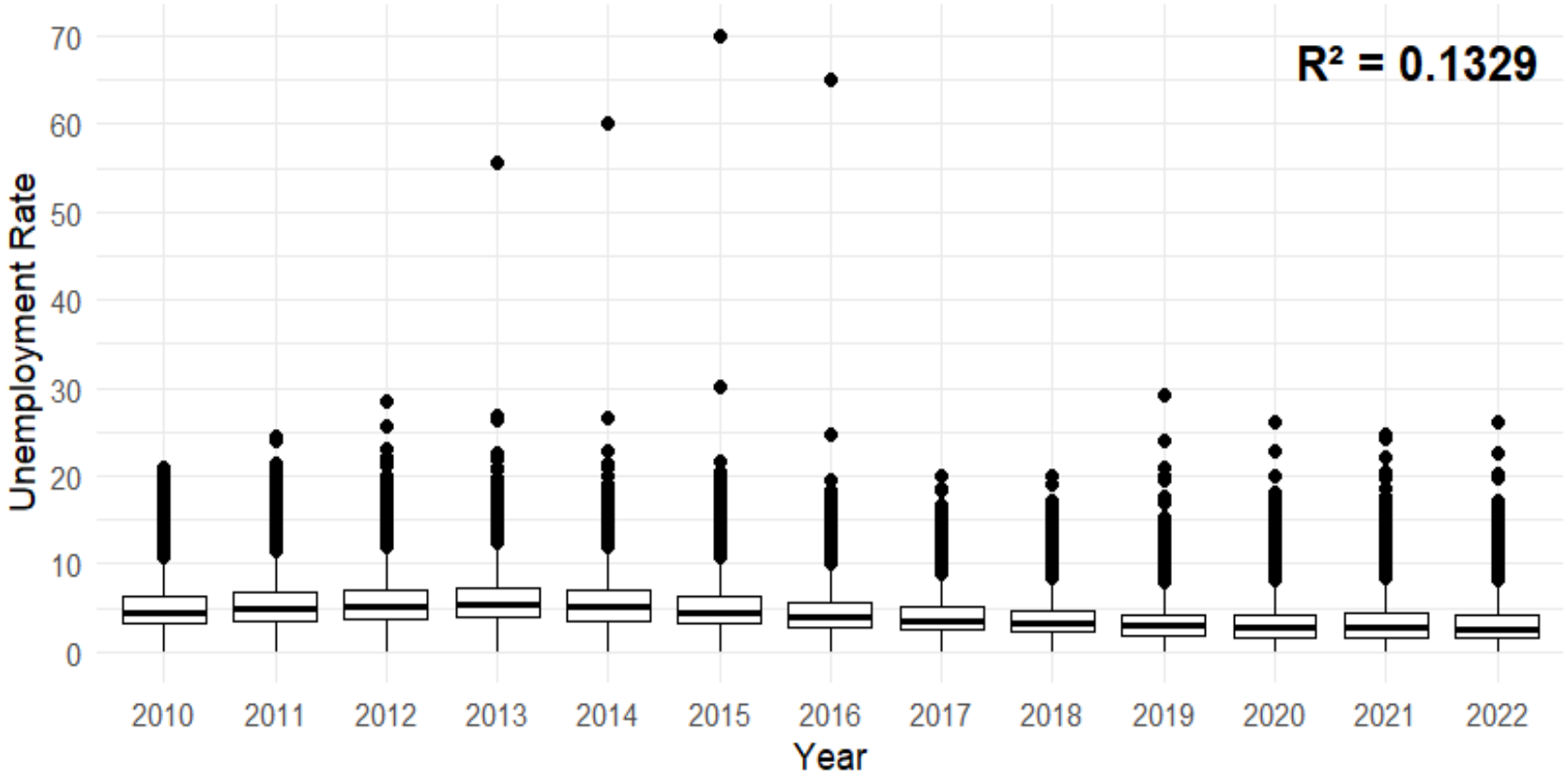


Figure B.5. Unemployment rate in Appalachia from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022.

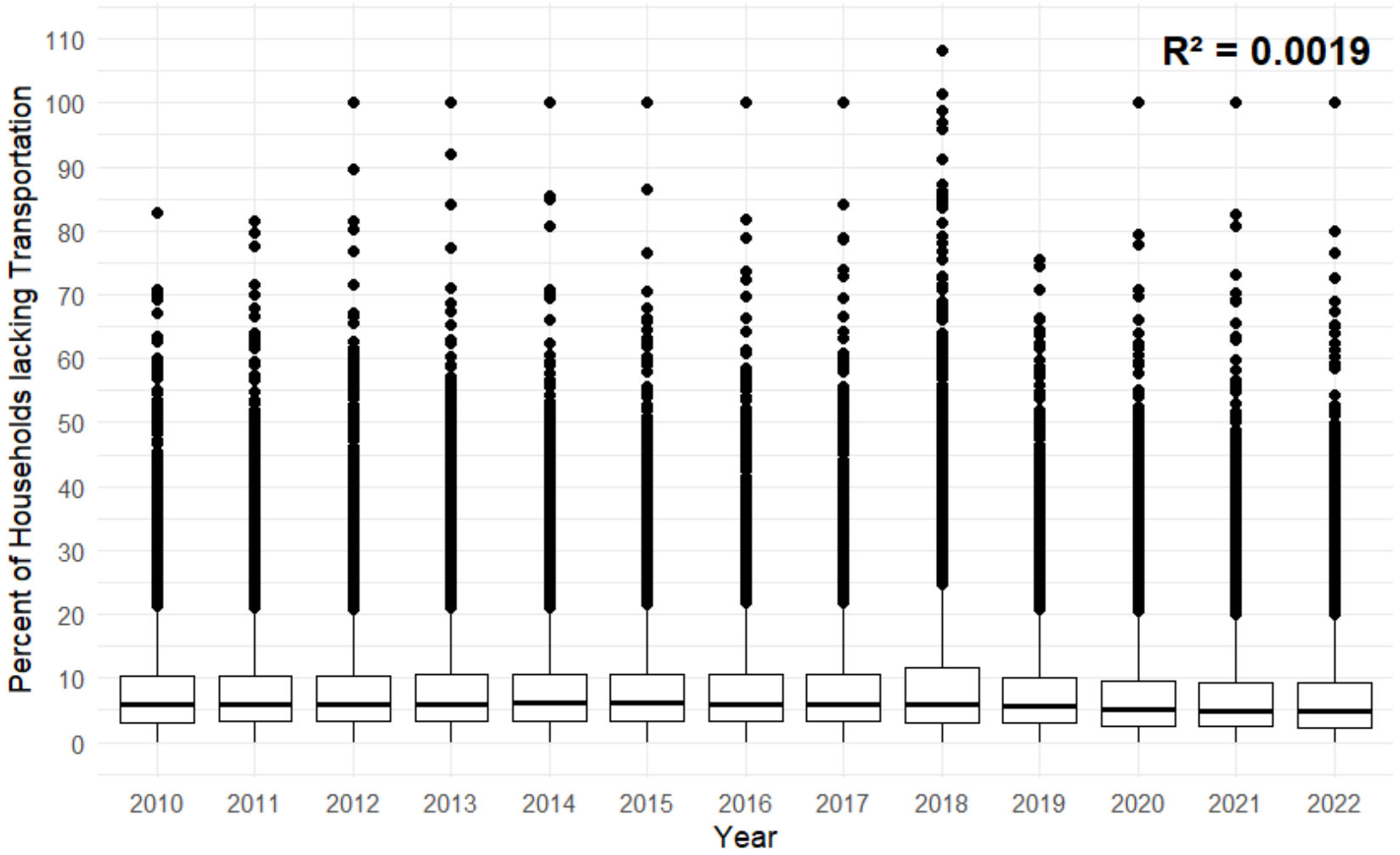


Figure B.6. Number of Appalachian households lacking transportation from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022.

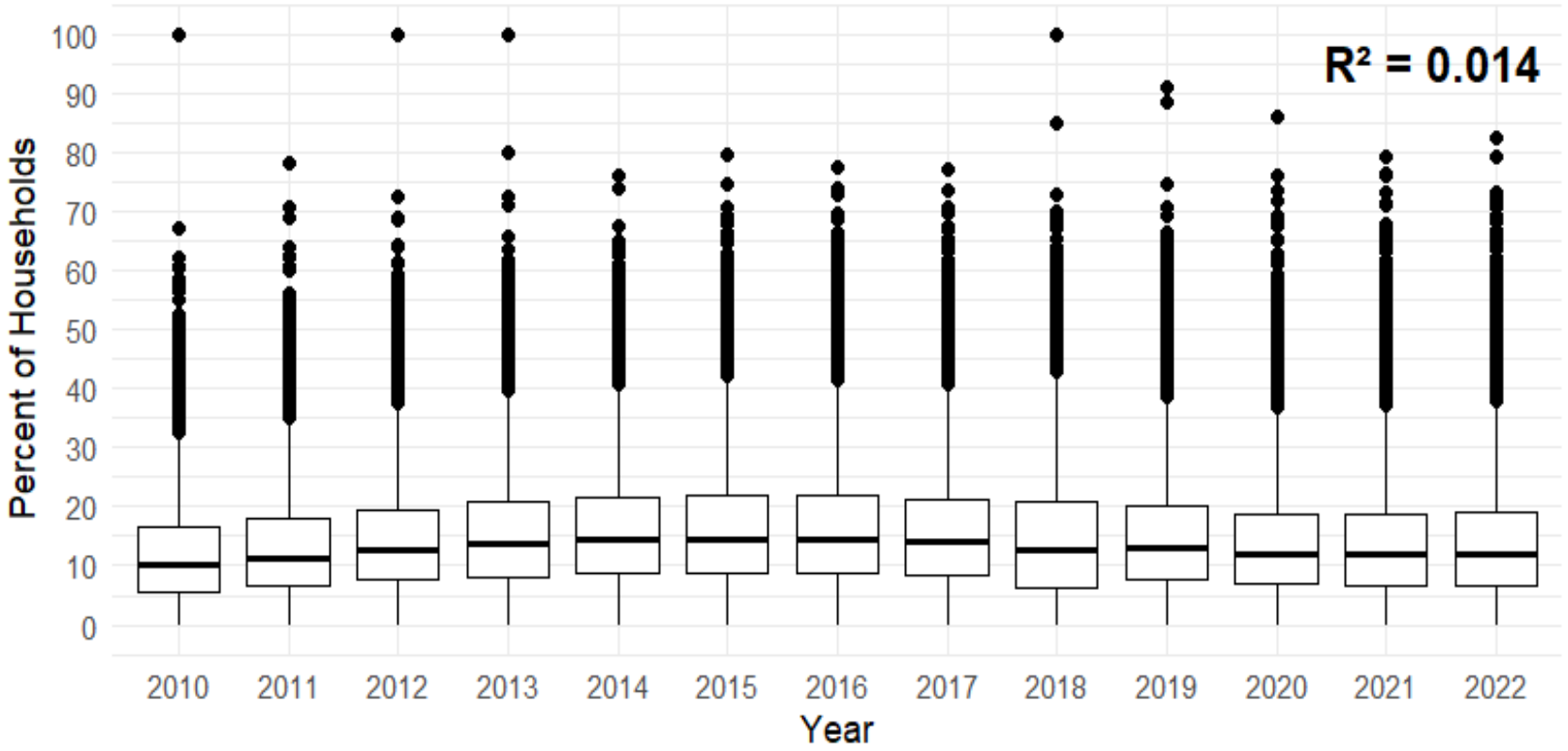


Figure B.7. Percentage of Appalachian households that receive SNAP from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. Supplemental Nutrition Assistance Program (SNAP).

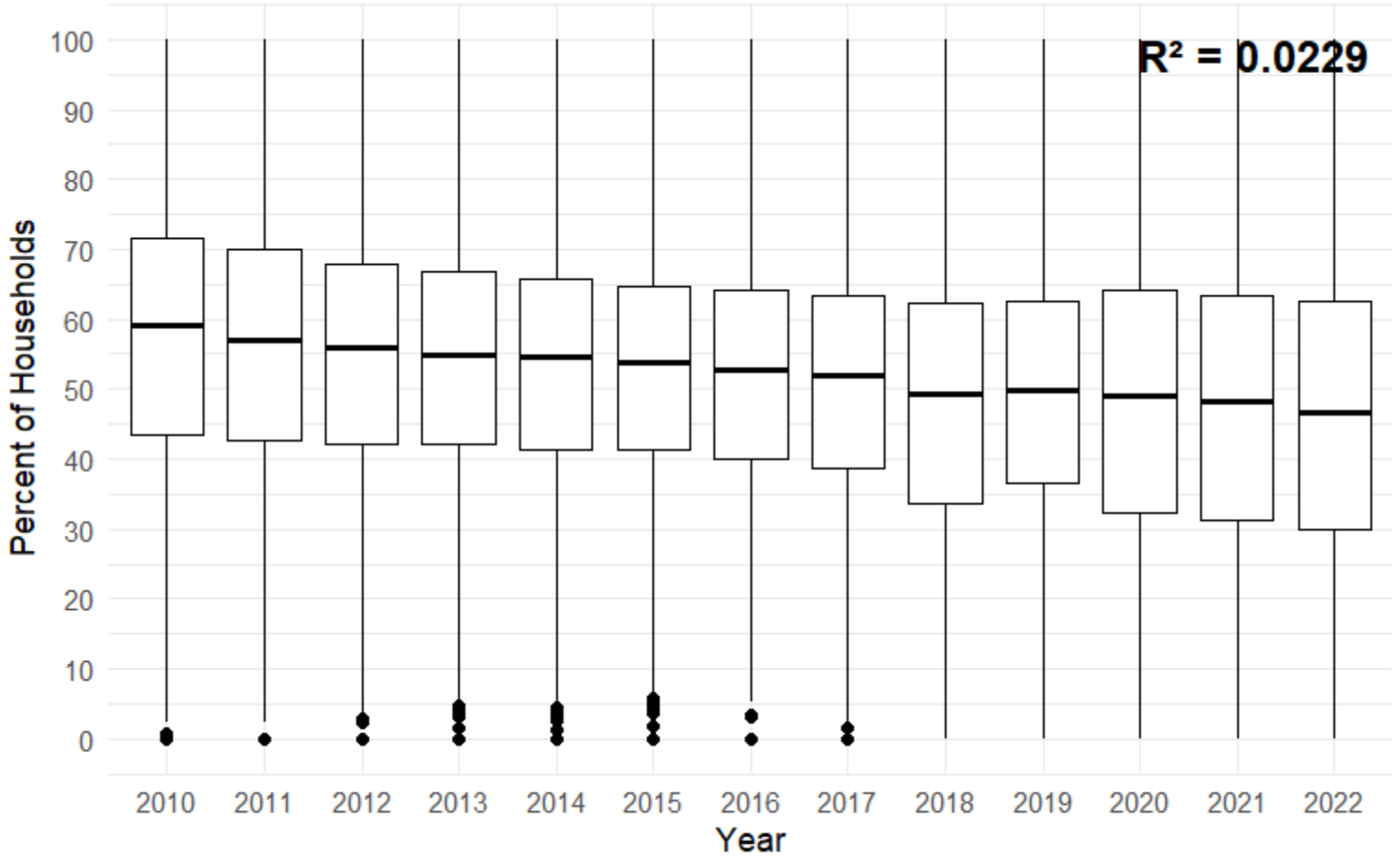


Figure B.8. Percentage of Appalachian households that receive SNAP that are below the poverty line from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. Supplemental Nutrition Assistance Program (SNAP).

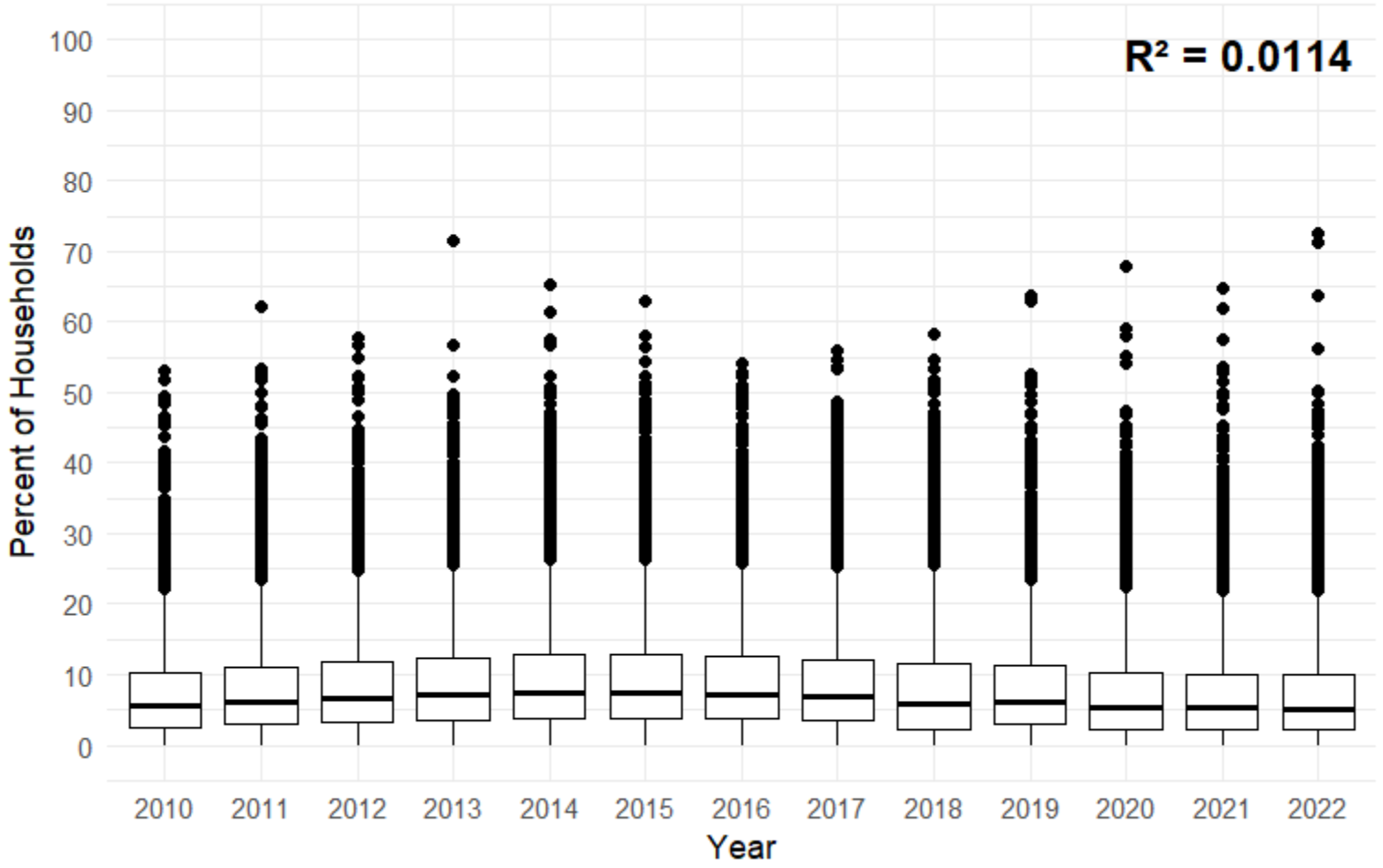


Figure B.9. Percentage of total Appalachian households that receive SNAP and are below the poverty line from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022. Supplemental Nutrition Assistance Program (SNAP).

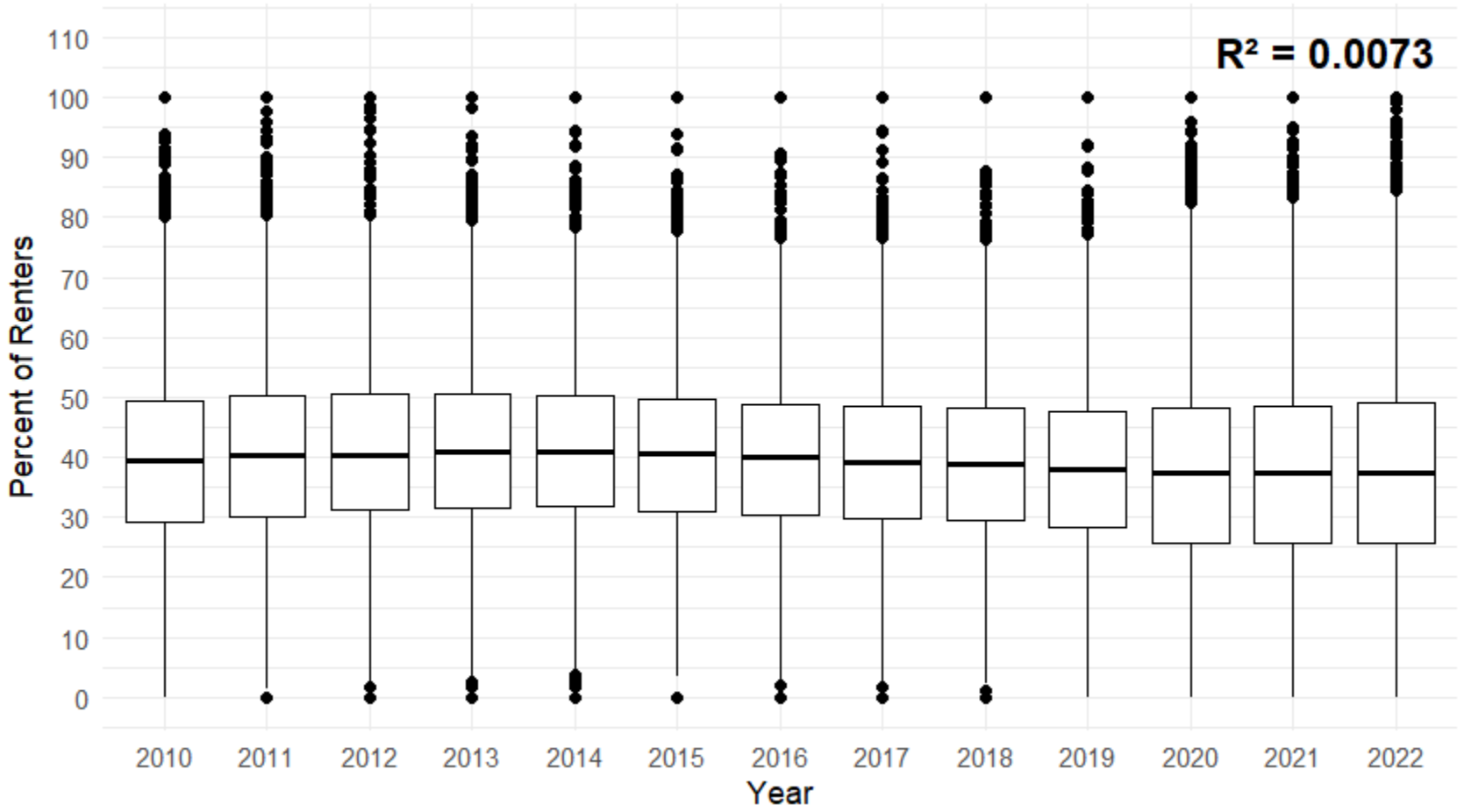


Figure B.10. Percentage of rental units with excessive housing costs from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022.

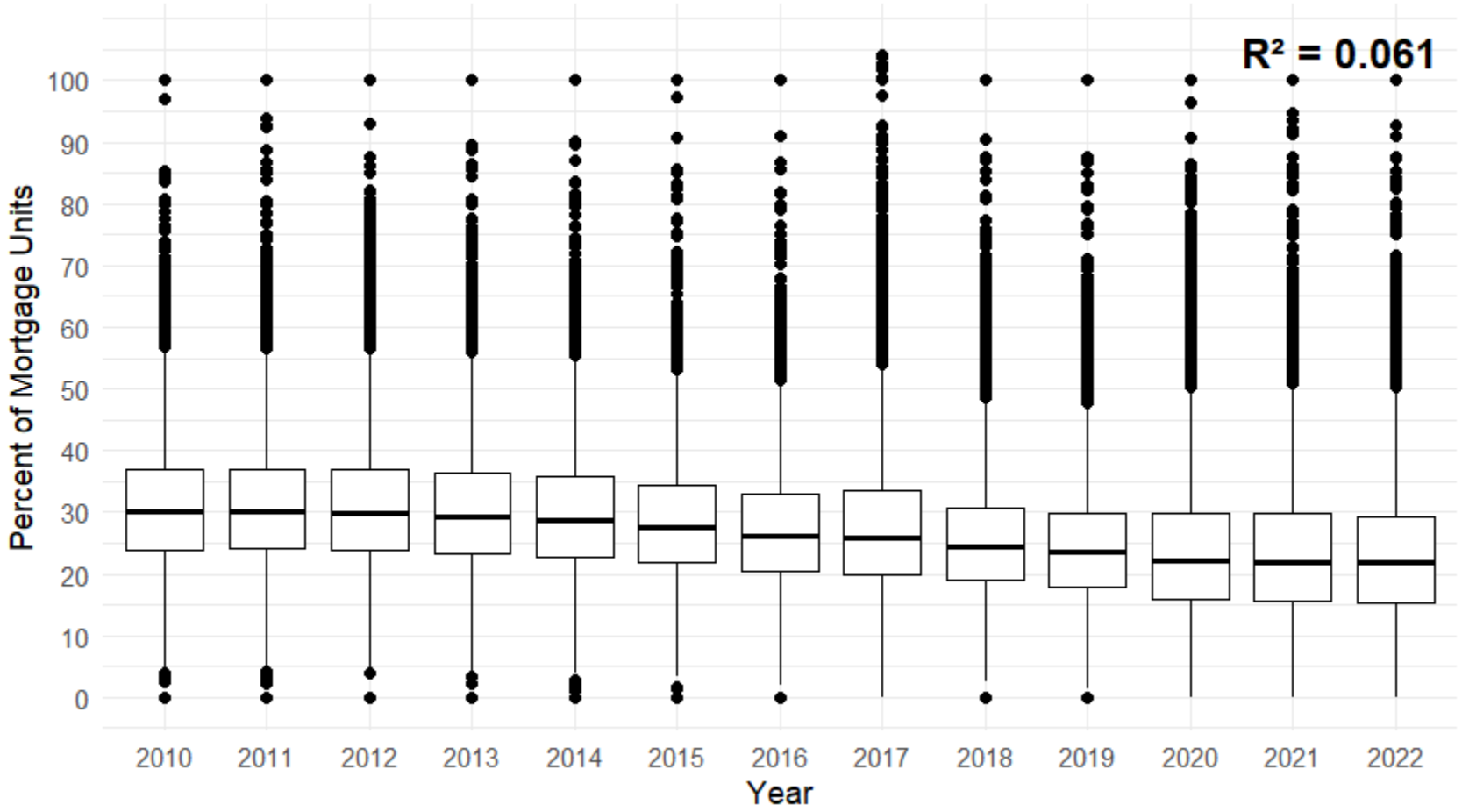


Figure B.11. Percentage of housing units with a mortgage that have excessive housing costs from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022.

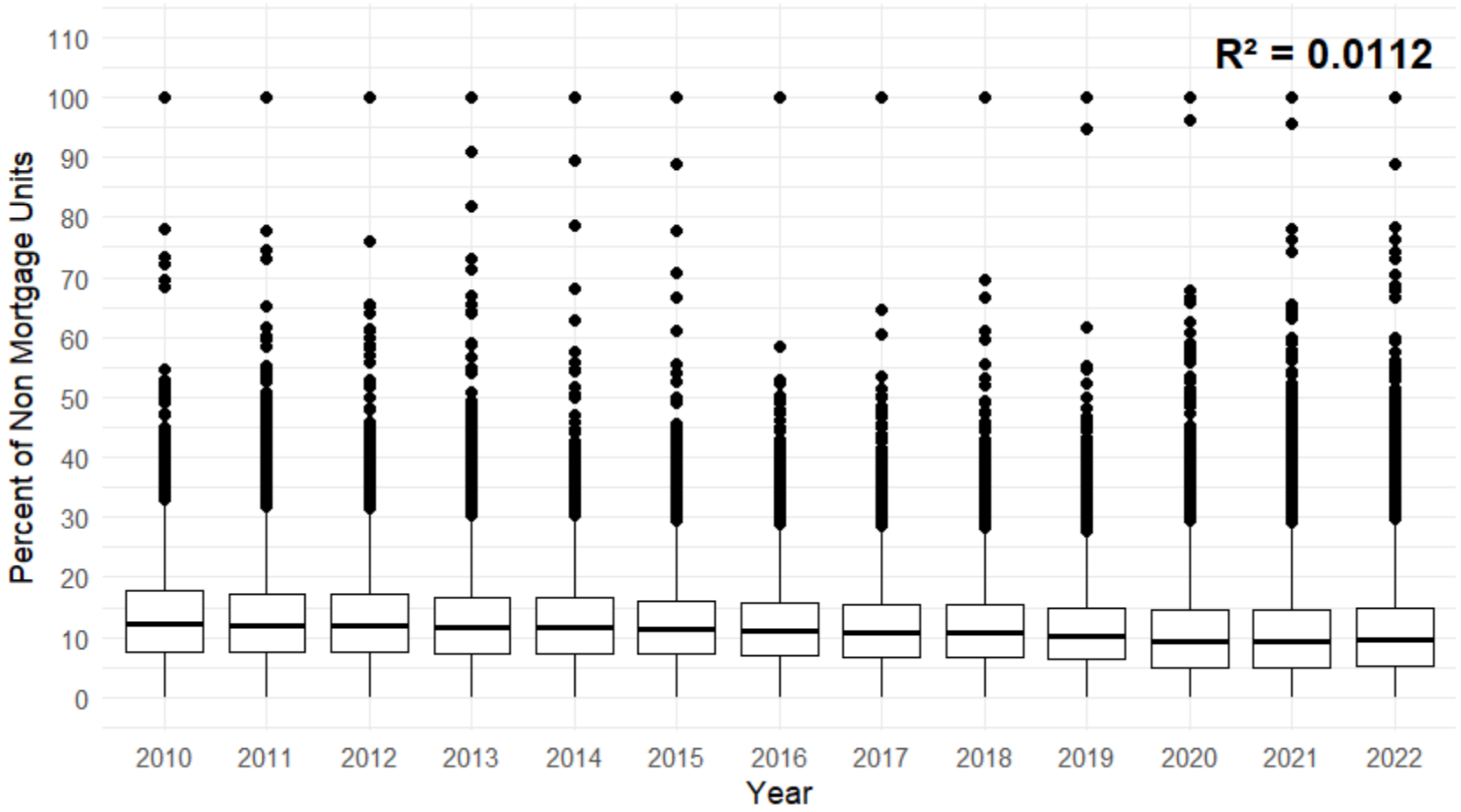


Figure B.12. Percentage of housing units without a mortgage that have excessive housing costs from 2010 to 2022. Data from the U.S. Census Bureau 5-Year American Community Survey 2010 to 2022.

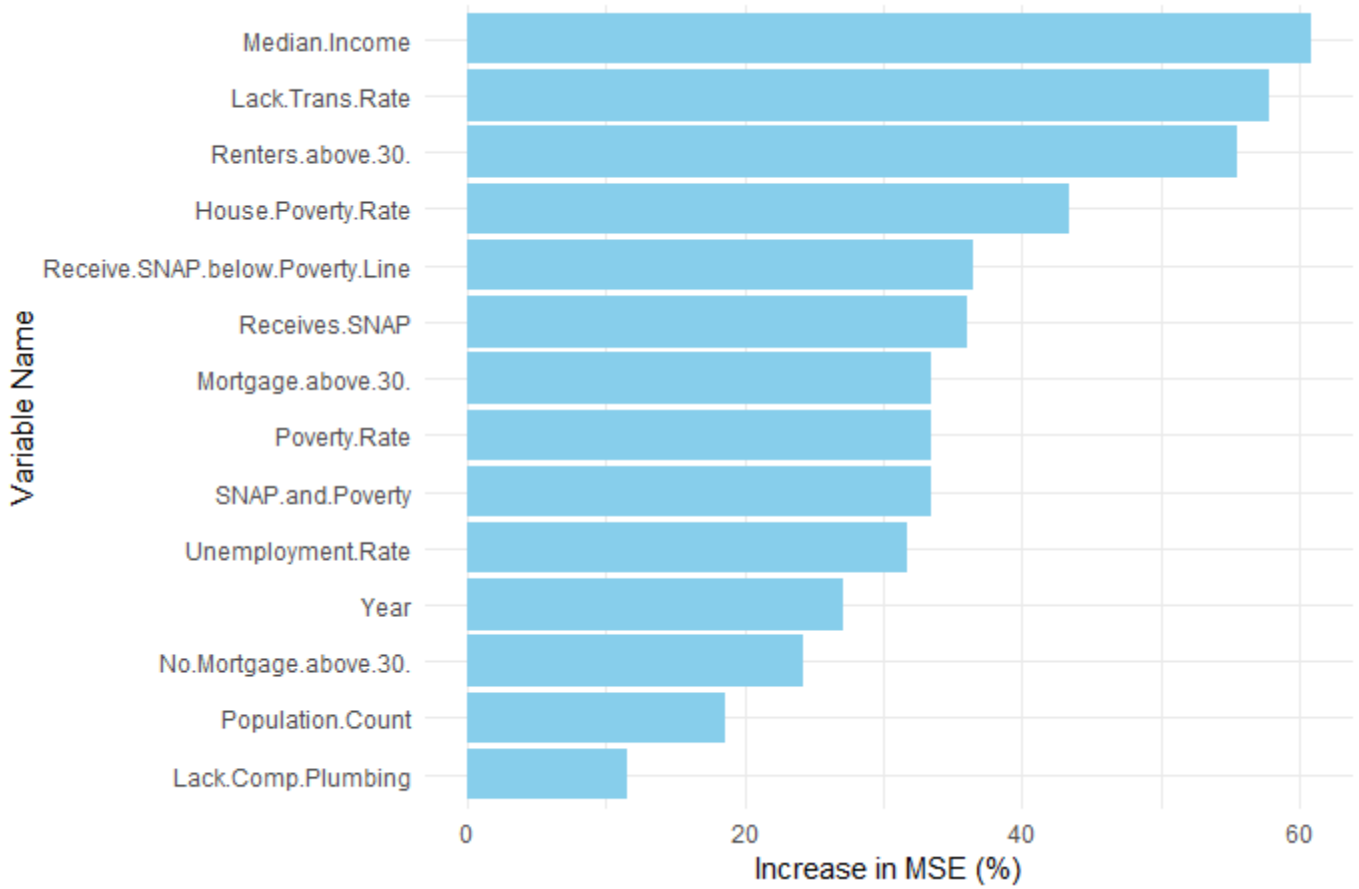


Figure B.13. Socioeconomic variables ranked by influence on food insecurity. Calculated using a Random Forest model in R.

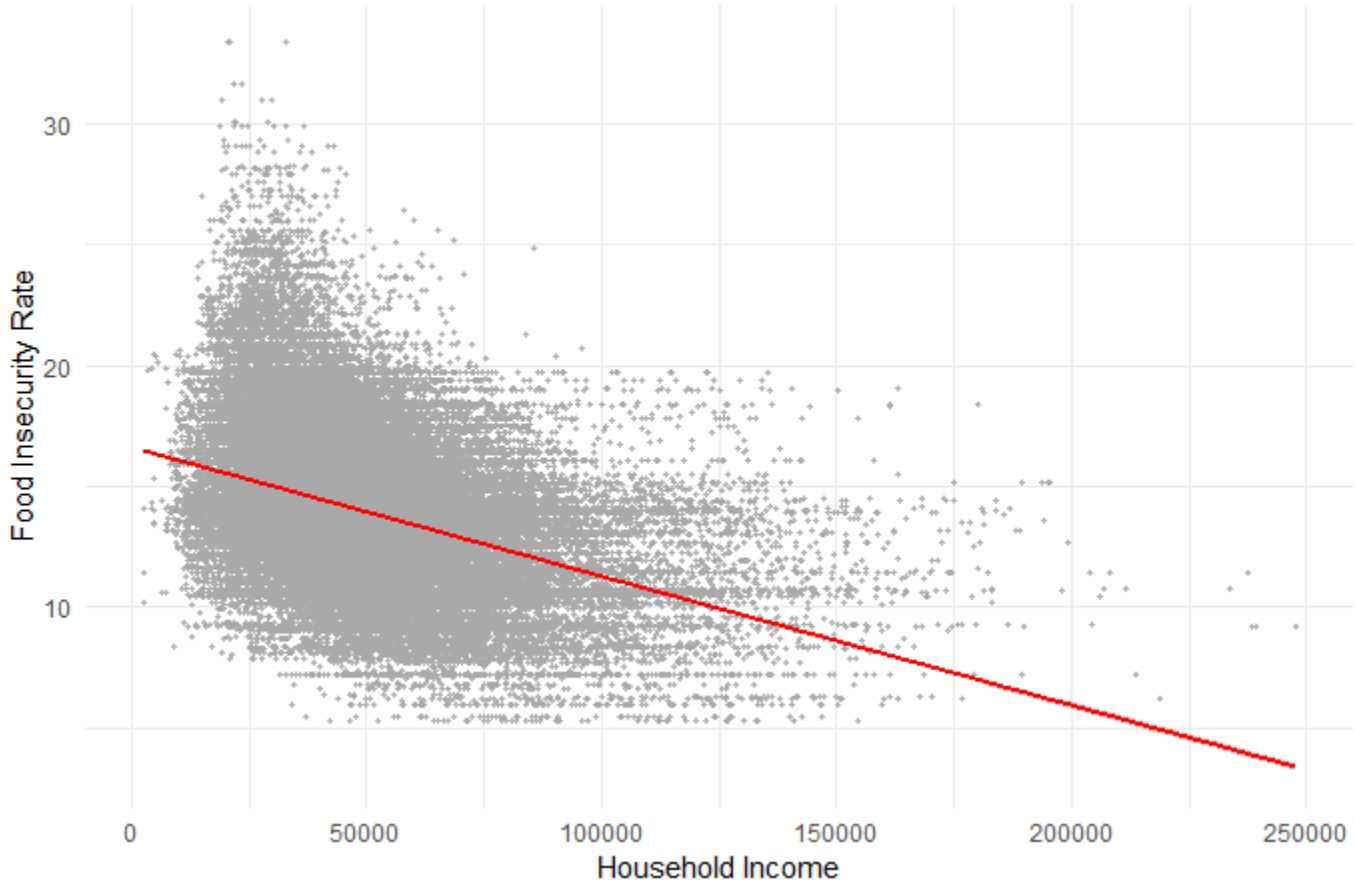


Figure B.14. Linear regression between median household income and food insecurity in Appalachia.

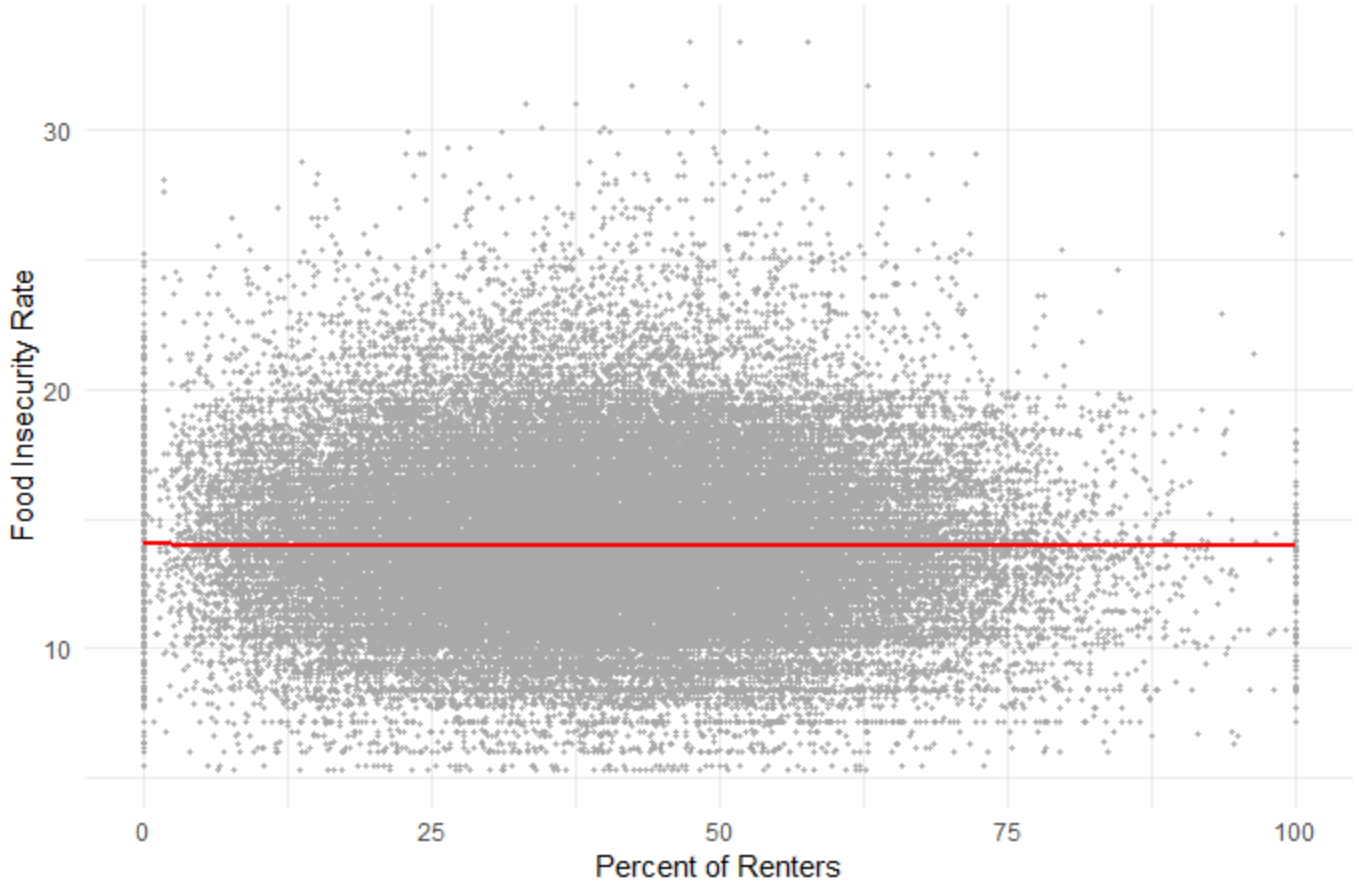


Figure B.15. Linear regression between rental units with excessive housing costs and food insecurity in Appalachia.

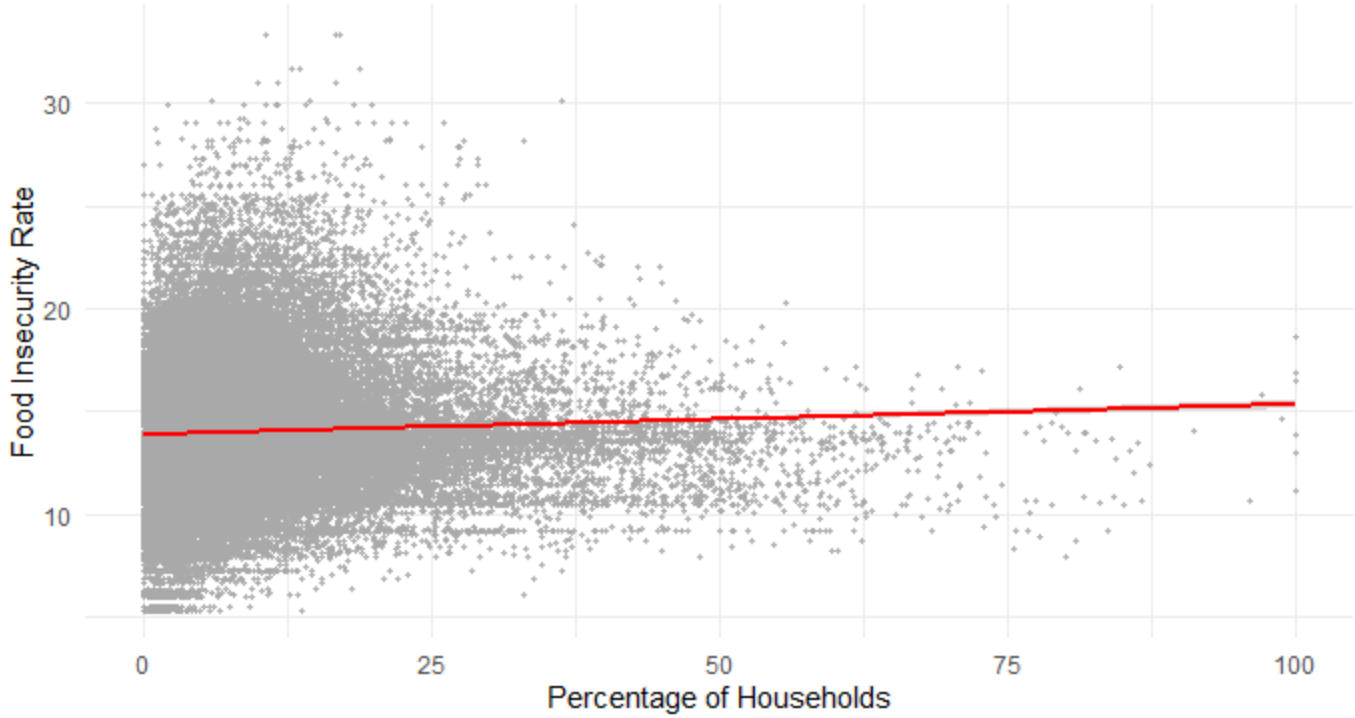


Figure B.16. Linear regression between households that lack transportation and food insecurity in Appalachia.

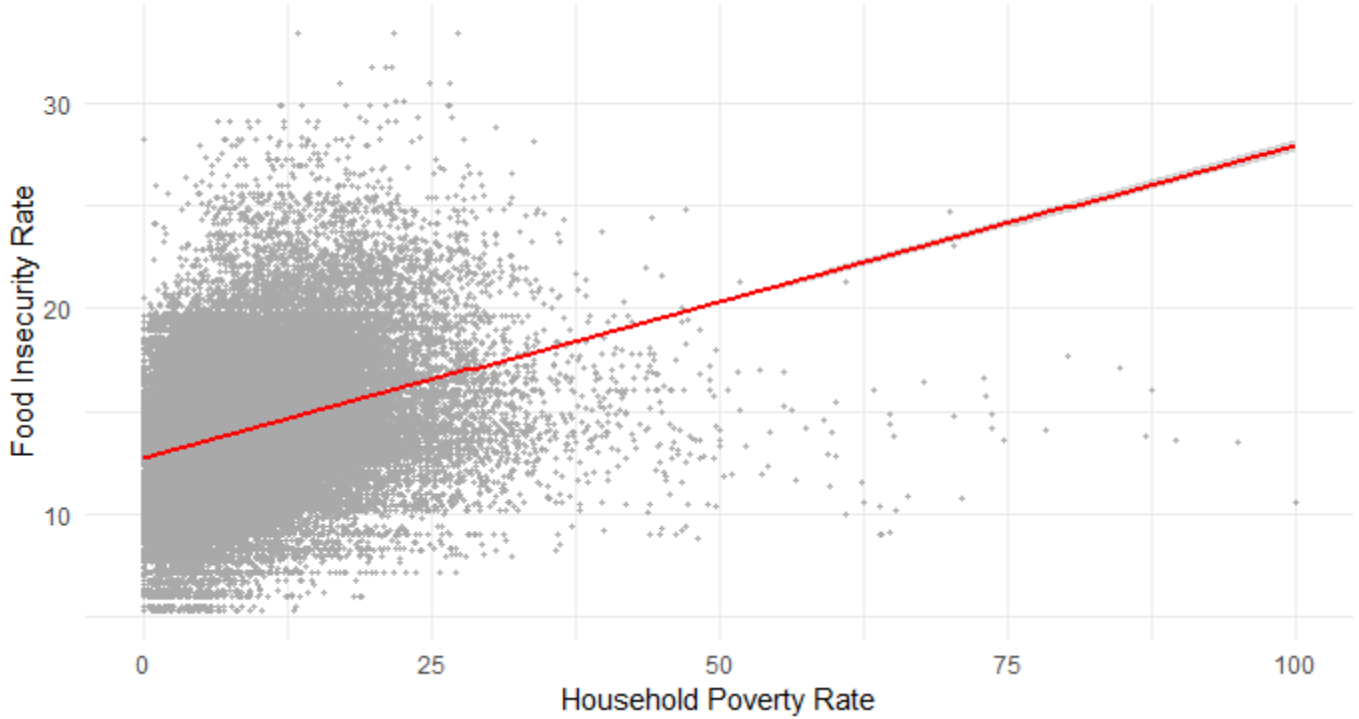


Figure B.17. Linear regression between household poverty rate and food insecurity in Appalachia.

Chapter 4

Predictive Modeling of Food Insecurity Values

Abstract

Using data collected from 2010 to 2022 and the information collected from the hotspot and trend analysis, two separate Long Short-Term Memory Models (LSTM) were created using R and the Keras, TensorFlow, and Caret packages to predict food insecurity rates. The first LSTM was a single-year model, using the previous year to predict the next, and was used to predict from 2023 to 2027. The second was a rolling 5-year timestep model, using a 5-year range to predict the next year. The single-year model was made for comparison, providing confidence in the ability of the rolling model to predict. Both of the models saw a steady decline in the mean absolute and mean squared errors, indicating that the LSTM network was learning the trends found within the data before they stabilized. The single-timestep model produced an R^2 value of 0.9929, and the model covered 99.29% of the variation within the training dataset. Between 2021 and 2022, the food insecurity rate in the Appalachian Region jumped from 13.3% to 16.9%. The single model predicted that from 2023 to 2027, the food insecurity rates would stay elevated, only decreasing from 14.48% to 14.40%. This was expected as the model is basing its predictions on a single year. The rolling timestep model produced rates that were more realistic. The R^2 value produced was 0.39, which was on par with the literature for similar projects. The model took into account the spike in food insecurity between 2021 and 2022, with an increase seen between the median in 2021 and 2023, but was not dramatically influenced by the spike. The rolling model predicted that food insecurity will increase from 13.9% in 2023 to 15.5% in 2027.

4.1 Introduction

Understanding the concerns within Appalachia, what's causing them, and what the trends are over the years is necessary to create solutions that work best for the people in those areas. As

the food insecurity rates and the number of food deserts continue to rise, planning for future action will prevent continued struggle or keep these rates from growing. Using the information gathered proactively would enable solutions to be developed comprehensively. To do this, it is reasonable to take the gathered data and use the analyzed relationships to predict the trends in the coming years. Predicting data based on previous years has its caveats. It can be influenced heavily and quickly by unforeseen circumstances or legislation passed. Influences on food insecurity can be either negative or positive and sometimes deviate from model predictions. For these predictions, a Long Short-Term Memory (LSTM) network was chosen because of its ability to handle time series and sequential data that have complex relationships among factors (Multivariate Time Series Forecasting with LSTMs in Keras, 2024). Additionally, LSTMs are capable of handling outliers, missing observations, and inconsistencies in the data, which were present in this dataset.

Using the LSTM network, the food insecurity rates for the region were predicted over the next 5 years (2023 - 2027). These predictions were then used to estimate the coming year's food deserts using the relationship between food deserts and food insecurity. The predictions were then added into the yearly regressions to see the updated trends, as well as mapping out the changes in ArcGIS Pro to visualize the changes seen in the region compared to the recorded years. This information can be beneficial for future policy and decision-making at a local level to combat rising food insecurity and expanding food deserts.

Research questions addressed in this chapter include:

1. What are the future trends of food insecurity in Appalachia from 2023 to 2027?
2. What is the effect of future food insecurity on food deserts and the loss of farmland in the region, particularly in the geographic distribution covered in Chapter 1?

3. How can a predictive LSTM model be used to make decisions regarding food accessibility in the region to proactively address food insecurity and food deserts?

4.2 Methods

The data used in this section was the same as the data that was used in the linear regression models. Compiled into a singular comma-separated-values (CSV) spreadsheet, any null values were filtered and dropped from the dataset.

The data was separated into normalized training and testing subsets based on years, with training including 2010 to 2017 and testing using 2018 to 2022. Using the Caret package, the preProcess function was used to normalize the data values to a scale of 0 to 1. This was done to have a standardized data format across both the training and testing subsets. The values for the year and census tract identification numbers were excluded from the normalization as they were the identifiers for the model. Due to the changing number of census tracts following the decennial census conducted in 2020, the added census tracts (increased from 6,189 to 7,241) were aggregated into the FIPS code associated with the pre-2020 census tract from which it was split. This was done to ensure that the model had adequate amounts of data to make predictions on.

The Long Short-Term Memory (LSTM) network was built using both the TensorFlow and Keras packages in R, with TensorFlow's virtual environment connecting with Python. The network was built using an LSTM layer with 60 units, with a single timestep. The number of units was chosen using a trial-and-error method to reduce the risk of overfitting, given the size of the dataset. Two LSTM models were trained for a one-year lag and a five-year lag, meaning that one model will use the previous time step to predict the next time step, and the other model will use the previous five time steps to predict the next one. Similarly to the LSTM layer units, during

the validation process, a dropout layer needed to be added to prevent overfitting. The dropout rate was 0.2, which tells the model to drop 20% of the neurons in a layer, preventing overfitting by reducing the amount of memorization and prioritizing trend learning. To finish building the model, a dense layer with a single unit was added. This layer transforms the LSTM layer's output into numeric prediction values for, in this case, food insecurity. The rolling five-timestep model was built with more units (100) as well as a regularizer (using the L2 regularizer set at 0.02) to help with the overfitting that was seen in the building and learning of the model. The dropout rate in the rolling model remained at 0.2.

After both of the models were built, they were trained at a learning rate of 0.0005 using the Adam optimizer. A slower learning rate (the standard being 0.001) was used to increase the stability of the model, and was decided on through grid searching due to overfitting. Adam optimizer, standing for Adaptive Moment Estimation, is a function in the keras package that reduces loss during model training by adjusting learning rates based on past gradients. This is done by combining the momentum and root mean squared propagation into a first moment and second moment estimation, leading to a stable and effective optimization. As this model is based on regression, and not classification, the evaluation metrics used were the mean absolute and mean squared error (MAE and MSE). The mean absolute error indicates how the model predicts the values by calculating the difference between the predictions and the actual points used in the training set. The mean squared error was selected as a metric because it is more affected by outliers, as it squares the differences in the predicted versus actual values. As seen in the previous chapter, outliers are present in this data. The MSE was used as a metric to determine if they are influential enough to be detrimental to the model's performance.

Lastly, the final product of the single timestep was trained using 50 epochs and a batch size of 16 for the years 2010 to 2017. The number of epochs and batch sizes was determined using a grid search to find the best combination of hyperparameters to reduce the risk of overfitting. The model training using 50 epochs means that the model will run through the training dataset 50 times to learn and identify trends. It was found that 50 epochs was the highest value that could be used before the model began to memorize the data and overfit. The size of data batches is the number of data points that the model learns from before updating. If the batches are too large, the model would not update frequently, leading to poor generalization. It was found that the smallest batch, which could be used before the model began to see too much noise and overfitting, caused high amounts of variance. The testing data was used as the validation dataset, and the plots generated by the training were used to adjust the model as needed and ultimately finalize the model. The final product of the rolling model was based on 100 epochs and a batch size of 32 to aid with the overfitting that was seen.

The predictions of food insecurity rates for each census tract for the next year were made using the model for the years 2023 to 2027. The predicted values were then incorporated into the prediction calculation of the following years. For example, in the single-timestep model, 2022 was used to predict 2023, and then 2023 was used to predict 2024. Whereas in the five-timestep model, 2018 to 2022 was used to predict 2023, and 2019 to 2023 was used to predict 2024. This is generally seen as more reliable, given that each predicted year has a year grounded in the dataset. Once predicted, the values were compiled into a CSV and mapped out using a boxplot to compare with previous years.

4.3 Results

4.3.1 Single Timestep

The top graph in **Figure C.1**, Appendix C, is the loss seen during the training of the model using mean absolute error (MAE). The model sees a significant drop following the first datapoint, which is expected as it passes through the first dataset for the first time. The training set's MAE decreased from 0.092 to 0.0061, and the testing set's MAE decreased from 0.058 to 0.0026. The decrease in loss and plateau seen in **Figure C.1** signifies that the model has learned the trends within the data and is achieving stability. The second graph is the mean squared error (MSE) calculated for the dataset. The training set's MSE decreases from 0.019 to 0.0001, and the testing set's MSE decreases from 0.0051 to 0.0001. The decrease in MSE is ideal as it shows that the model is learning the data and adjusting to the present outliers. The final MAE and MSE values dictate that the model is not majorly influenced by present outliers.

The model was evaluated by comparing the actual values from the testing dataset (2018 to 2022) against the values that were generated by the model after training, shown in **Figure C.2**. The predicted values are closely related to the actual values as seen by the line of best fit, with an R^2 value of 0.993. This high R^2 suggests that 99.3% of the variation in the food insecurity dataset is covered by this model, demonstrating the model's strong predictive capabilities against the actual dataset. The model is capable and stable when predicting, but as a precautionary measure, the predictions that were generated were limited to 2023 to 2027. While the predictions could be extended past 2027, the further the predictions get from the actual dataset, the more variable the predictions can become.

The predicted values were compiled together and then visualized in the yearly regression model described in Chapter 2 (**Figure C.3**). It was found that after the sudden spike in food

insecurity from 2021 to 2022, the median increased from 13.3% to 16.9%, the predicted food insecurity rates from 2023 to 2027 remained elevated but began to gradually decrease (beginning at 14.48% in 2023 and ending at 14.40% in 2027). This increase was most dramatically seen in the central and southern subregions. Exploring the relationships found and explained previously leads not only to being able to predict the future of food insecurity, but also to predicting the factor response to food insecurity within the region. With the increase in food insecurity, it's assumed that the poverty rates will also remain elevated, and the median income will remain fairly stagnant. This captures the negative feedback loop that Appalachia has caught itself in. As food insecurity rates rise, household poverty and median incomes will remain stagnant, contributing to the rise in food insecurity, which contributes to the stagnation of poverty and income. This cycle is hard to break, but it becomes even more difficult as it remains unconsidered.

4.3.2 Rolling Five Timesteps

The model was adjusted to perform a rolling prediction using five timesteps instead of a single one to compare the prediction capability. Seen in **Figure C.4**, the adjusted model had fairly similar training performance to the single-step model, gradually improving its prediction performance. The training set's MAE value decreased from 0.61 to 0.0045, and the testing set's MAE decreased from 0.043 to 0.036. The decrease and stability seen in the MAE values mean that the model has learned the patterns within the dataset to the point of stability and prediction. The training set's MSE values decreased from 0.043 to 0.0018, and the testing set's MSE decreased from 0.0088 to 0.0014. The decrease in MSE signifies that the model is adjusting to outliers within the data and can learn the patterns without major complications from outliers.

The adjusted model was evaluated by comparing the actual values from the testing dataset (2020 to 2022) against the values that were generated by the model after training, shown in **Figure C.5**. The evaluation was based on the R value, which was 0.3896. The high R demonstrates the model's strong predictive capability. The predictions made by this model are based on the previous three years, using 2020 to 2022 to predict 2023, and so on.

The predictions made by the rolling model were visualized in **Figure C.6**. The predictions made by the adjusted model are similar to those of the trends seen before the spike in 2022, which is a good sign that the model is predicting based on the trends that were identified throughout the entirety of the model. The predicted values showed the median dropping back down to 13.9% in 2023 and then gradually increasing to 15.5% in 2027. This trend is similar to that seen from 2017 to 2021, where the median gradually increased from 13.4% to 15.1% before decreasing

4.4 Discussion

The predicted food insecurity rates for 2023 to 2027 were based on single and rolling timesteps. For both models, the mean absolute and mean squared errors were decreasing, with the rolling timestep model seeing more of a decrease with MSE. This is to be expected as the rolling model uses more historical data to learn and predict, which reduces the number of errors made. The fitness of the data was tested by graphing the actual versus predicted values, which yielded a high R^2 value of 0.9927 for the single time step model, which is expected as it's basing its predictions on a single year. The rolling model had a lower and modest R^2 value, 0.3856, which is due to the complexity and noise associated with incorporating large amounts of socioeconomic data. Based on the literature covered in the introduction, a modest R^2 value is common in research and models such as these. The confidence in the model is built on the fitness

of the single time step model, the predicted trend of the rolling model, and the correlations that were determined in previous chapters.

The regression trend was updated with the model values. Both models' trend displays show that the food insecurity rates had a gradual decline in values following an increase between 2021 and 2022. The decline was slightly larger using the rolling model, but only marginally. Based on Chapter 2's hotspot analysis, food insecurity rates were highest in the central and southern subregions of Appalachia, with the significant increases being in the central region. This follows the trends found in the hotspot analysis for food insecurity. The correlation between food insecurity and food deserts, calculated in Chapter 3, suggests that as the food insecurity rates stay elevated through the next several years, the number of food deserts will continue to climb, particularly the number of combination deserts. This is paired with the expected continued decline in farmland acreage across both the region and the country. Thus, it is reasonable to assume that food inaccessibility will become increasingly common.

4.5 Conclusion

The hotspot analysis and linear regressions were used to create two predictive Long Short-Term Memory networks to predict the rates of food insecurity of the Appalachian Region over the 5 years following the most recent data release (2023 to 2027). This was done through building both a single and a rolling 5-year timestep prediction model. The single-step model predicted that food insecurity rates would stay elevated and only slightly decline from 2023 to 2027. However, the more stable and reliable five-timestep model predicted that the food insecurity rates would follow the trend seen from 2017 to 2021, and decrease in 2023 after the spike in 2022, before gradually starting to increase again. Given the relationship between food

insecurity and food deserts, the number of food deserts, particularly the number of combined geographic and financial food deserts, will also continue to gradually increase.

References

(2024, February 17). *Multivariate Time Series Forecasting with LSTMs in Keras*. GeeksforGeeks.https://www.geeksforgeeks.org/multivariate-time-series-forecasting-with-lstms-in-keras/

Deléglise, H., Interdonato, R., Bégué, A., Maître d'Hôtel, E., Teisseire, M., & Roche, M. (2022). *Food security prediction from heterogeneous data combining machine and deep learning methods. Expert Systems With Applications, 190, 116189.* https://doi.org/10.1016/j.eswa.2021.116189

The author would like to acknowledge the use of Grammarly, a language model developed by Grammarly Inc., in the preparation of this publication. Grammarly was used for grammar and spell-checking in this thesis.

Appendix C

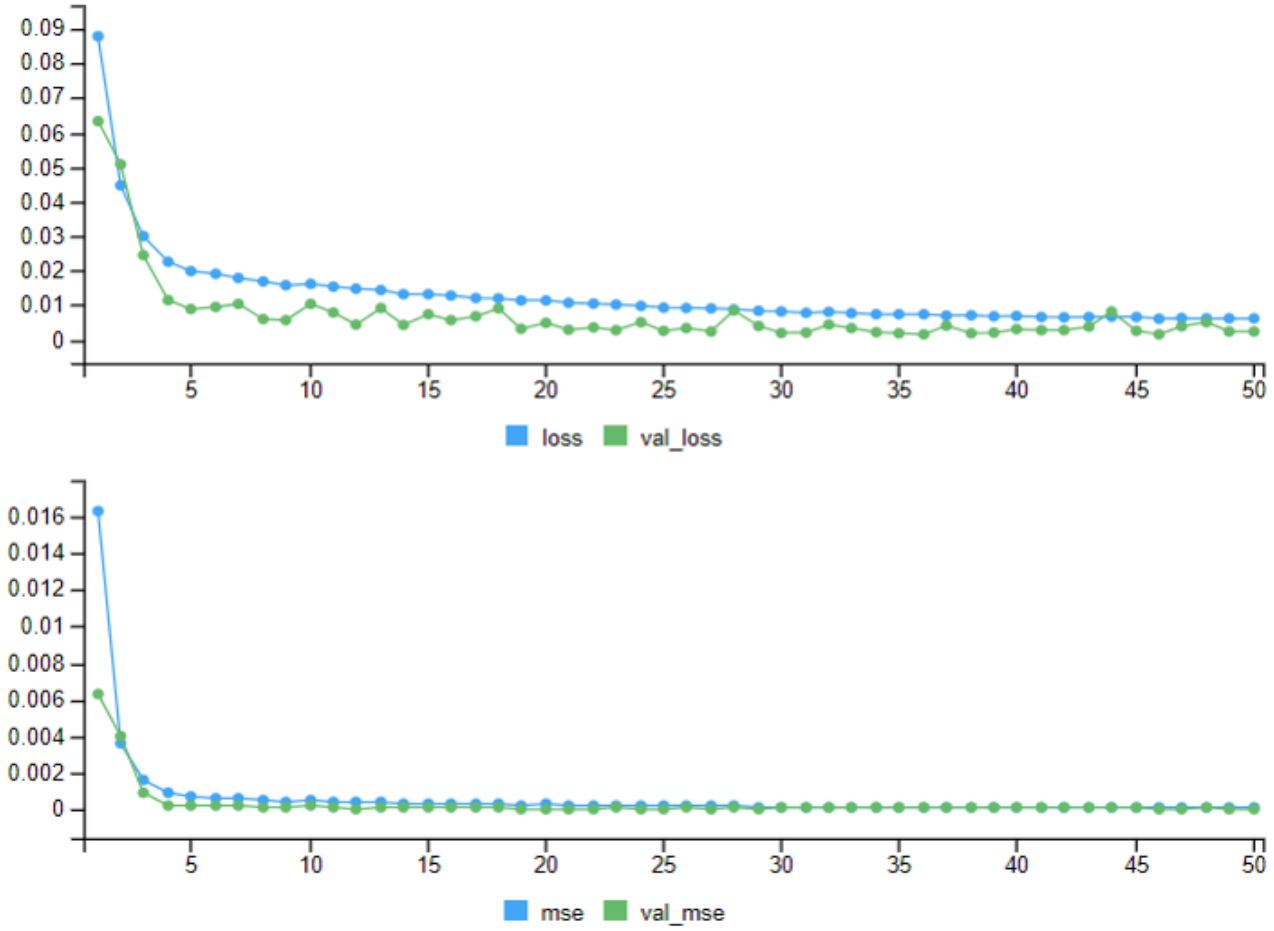


Figure C.1. Loss and validation loss plots for the Long Short-Term Memory (LSTM) model built for food insecurity. Loss metrics used were mean absolute error (first graph) and mean standard error (second graph).

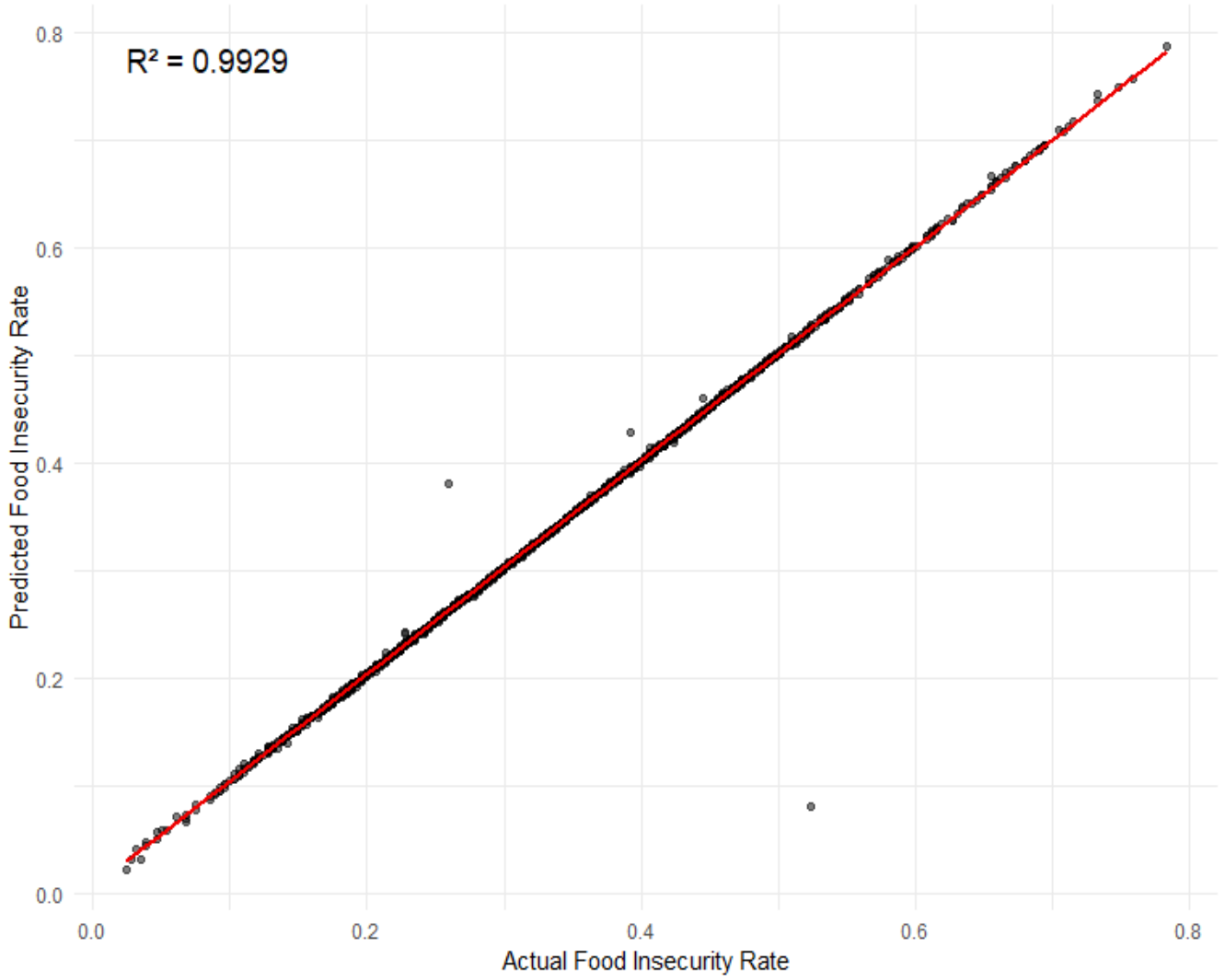


Figure C.2. Predicted versus actual food insecurity rates for single time step LSTM validation.

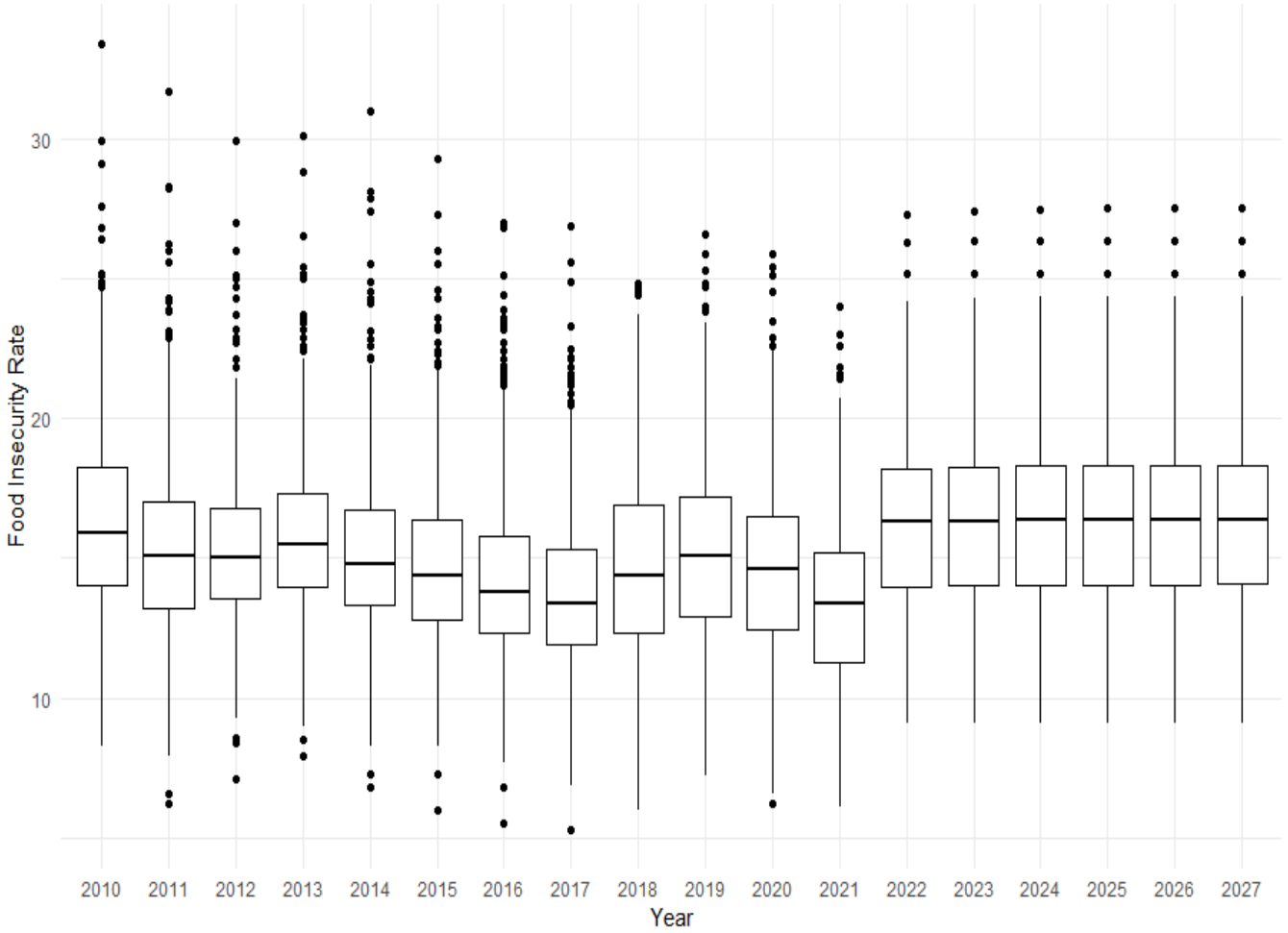


Figure C.3. Food insecurity rates in Appalachia from 2010 to 2027. Data from Feeding America and the U.S. Census. Years 2023 to 2027 were predicted using a single time step long short-term memory model.

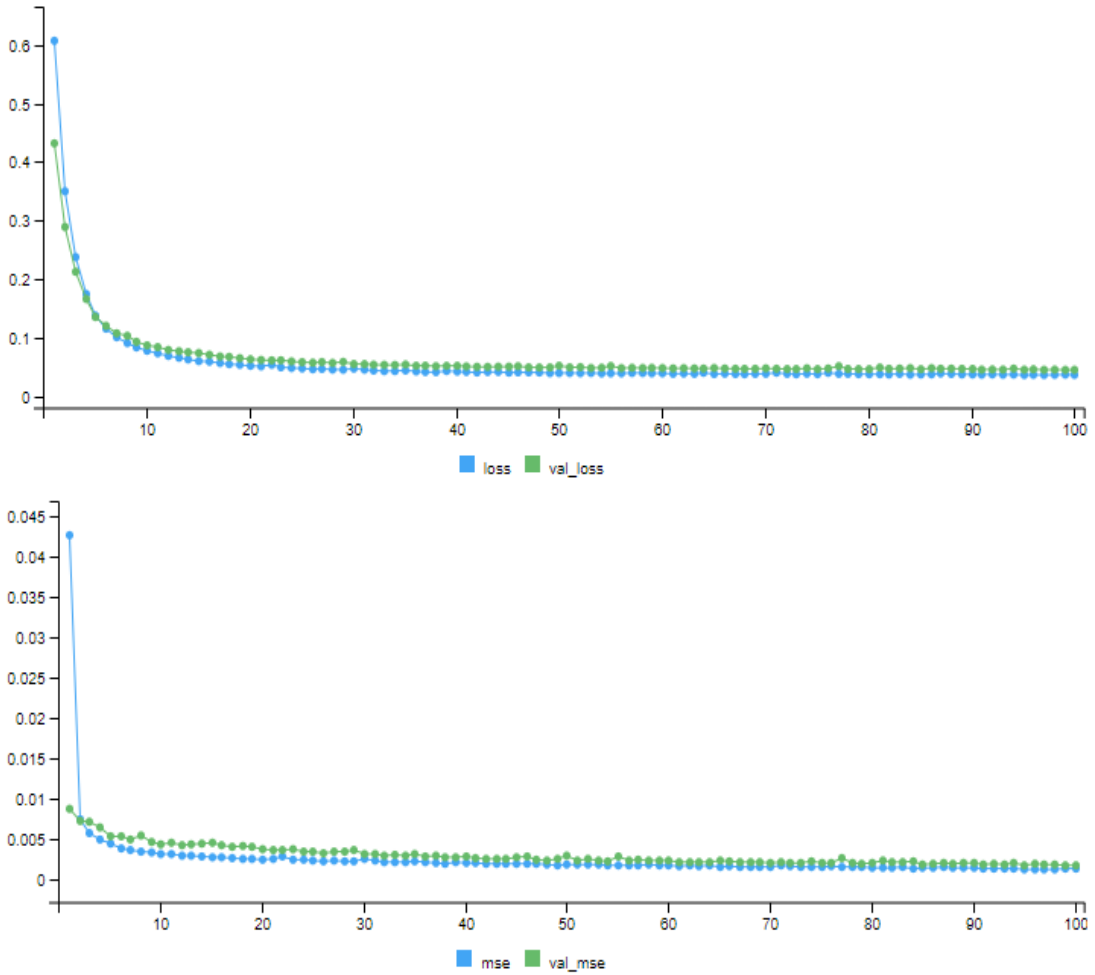


Figure C.4. Loss and validation loss plots for the Long Short-Term Memory (LSTM) model built for food insecurity. Loss metrics used mean absolute error (first graph) and mean standard error (second graph).

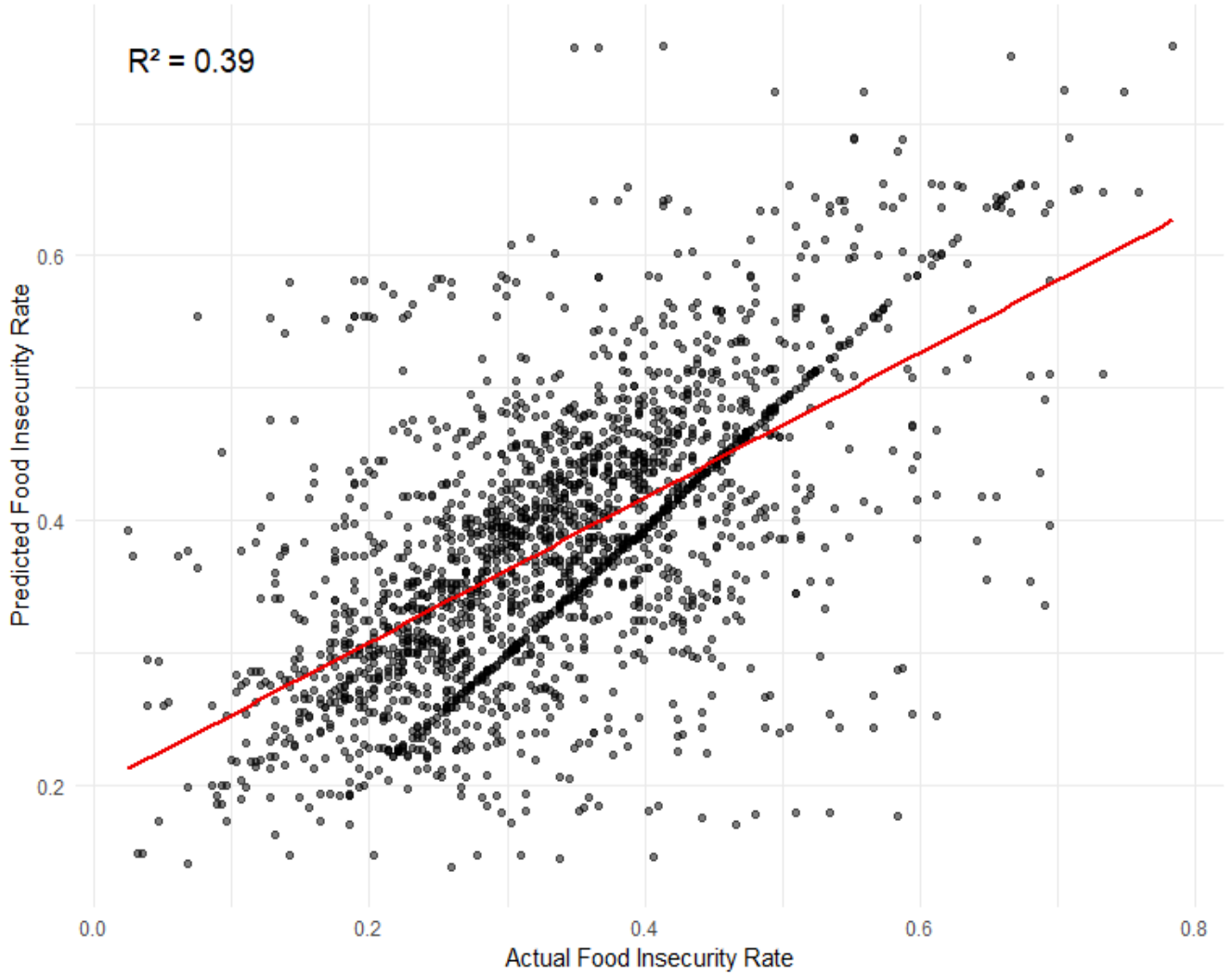


Figure C.5. Predicted versus actual food insecurity rate for rolling five time step long short-term memory model validation.

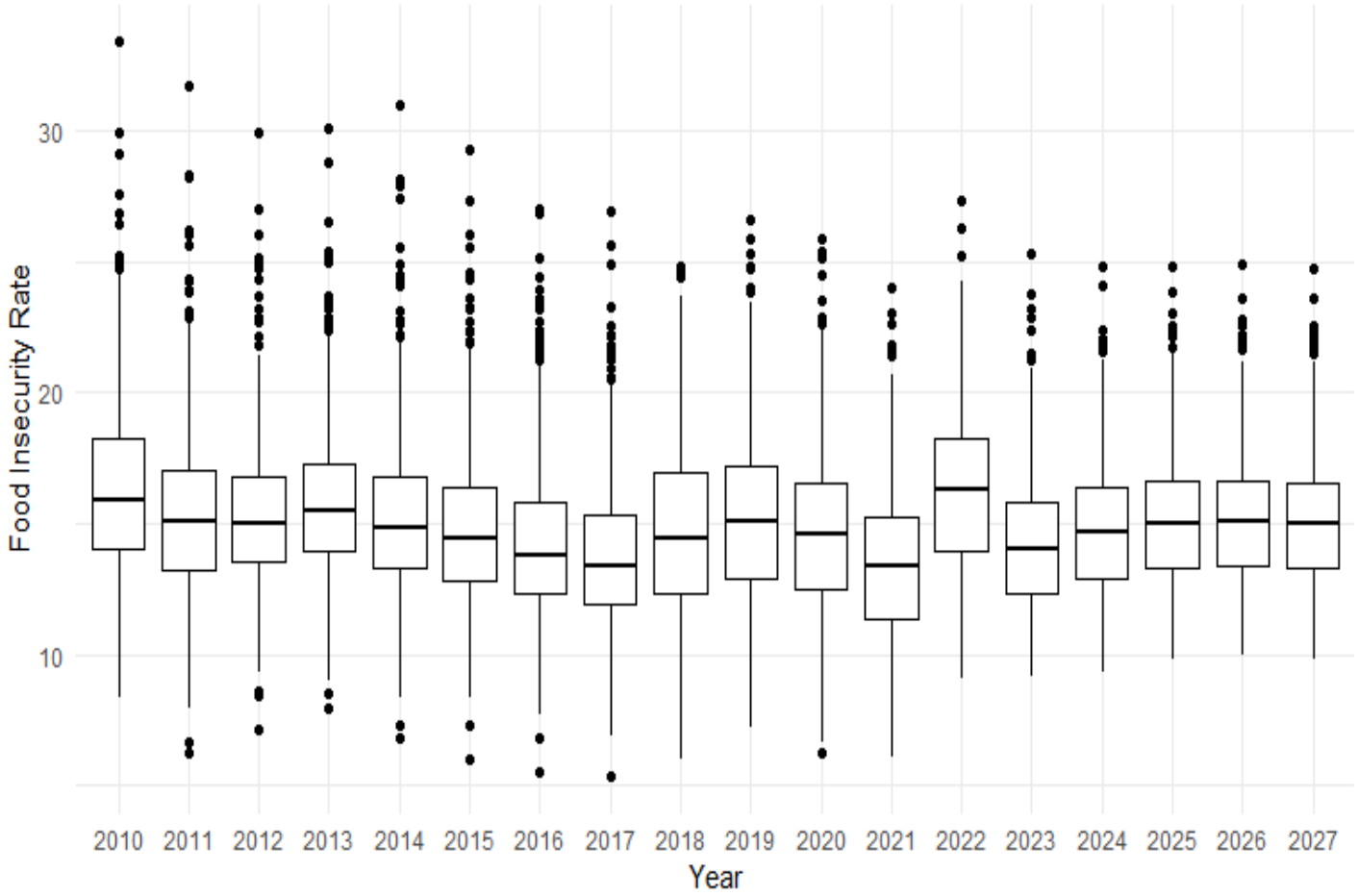


Figure C.6. Food insecurity rates in Appalachia from 2010 to 2027. Data from Feeding America and the U.S. Census. Years 2023 to 2027 were predicted using a rolling five time step Long Short-Term Memory model.

Chapter 5

Conclusions

5.1 Key Findings

The Appalachian Region has historically been underserved regarding research and practical applications. As most of the research in the region is limited or outdated, this project aimed at providing updated and useful information and analyses for current suggestions or future researchers. The predictive LSTM model's ability to simulate how factors such as poverty, income, unemployment, etc., as well as the relationships between food insecurity, food deserts, and farmland loss, enables communities and organizations to target their efforts into successful tactics. The model predicts that food insecurity will continue to gradually rise, contributing to a rise in food desert growth. Coupled with the rapid loss of farmland, the region is currently playing defense against food inaccessibility. The combination of temporal and spatial modeling allows for prediction but also analysis of the geographic distribution of these predictions, allowing decision makers to shift from reactive planning to preventative in their work to reduce food insecurity. The detailed explanations of the relationships examined throughout this research, leading to the final product of the predictive model, can be of great use for future research or policy work. This model can be altered and interacted with to identify ways to reduce the future of food insecurity. For example, changing the percent change in median household income over a period of time to see its effect across the region. This can aid policymakers in selecting a focus on the factors with the most impact, as well as setting a goal to reach over the next five years (i.e., increase the median household income by 5% over the next ten years). The other models and analyses completed provide future researchers and projects with updated figures and calculations to aid in their predictions and solutions.

5.2 Future Research Directions

The Food-Energy-Water nexus can lead solutions to be more effective and proactive through utilizing it's interdisciplinary framework. Food access cannot alone be improved, but rather, it can be improved by addressing the contributing factors through creating comprehensive policies and plans. Intertwining the other two factors, energy burdens and water accessibility, with the farmland loss and socioeconomic hardships of the Appalachian Region will yield those comprehensive plans. At the forefront, future research should focus on the rapid and continued farmland loss in the region. The current Farmland Preservation Policy Act, as a part of the 1981 Farm Bill, protects agricultural land from federal acquisition but should be updated to include protections against large land acquisitions for land leasing agreements. Updates in the 2018 Farm Bill include initiatives to support smaller and lower-income farms, however, farmland continues to decline at alarming rate. Future policy focus on farmland preservation efforts and aid to farmers is crucial to prevent irreparable loss. To address food insecurity and food deserts, alternative food production systems could offer ways to increase regional food production and create a circular economy within the region. Systems such as controlled environment agriculture (CEA) can increase the amount of food produced in a local community and cut down on the food foodprint, making the food cheaper and more geographically accessible. However, examining this through the FEW nexus lens, CEA systems can have a heavy demand on energy and water. Future research and decision-making will need to determine if the Appalachian Region can shoulder the weight of the demand in exchange for the increase in local economy and food systems. Current policy is offering promising directions for future research. The Bipartisan Infrastructure Law allocated $1 billion, the largest grant thus far, to the Appalachian Regional Commission to support their efforts in boosting the socioeconomic standing of the region. The

roadmap for Appalachian food sovereignty is complex, but focusing and working on key aspects, such as farmland preservation, alternative food systems, and a regional economy, is the first step in a proactive Food-Energy-Water nexus approach.

Vita

Angela Price was raised in Polk County, North Carolina, in the small town of Columbus. She attended Appalachian State University and graduated with her Bachelor of Science in Environmental Science with a concentration in Life Sciences and minors in Geology and Chemistry in May of 2023. Angela worked on an undergraduate research project working to quantify Phosphorus exports on Christmas tree farms in the High Country of North Carolina, where she discovered her passion for outreach with farmers. She also partook in a semester research study in Sampson and Duplin County, North Carolina, working with an environmental justice group (EJCAN) regarding the concentrated animal feeding operations in the community. Through this passion for outreach and environmental justice, she decided to pursue graduate school. Angela was offered a research assistantship with Dr. Jie Zhuang at the University of Tennessee in August of 2023. She graduated with her Master of Science in Environmental and Soil Science in May of 2025. Angela's research interests involve rural food access, agricultural and natural land conservation, and outreach. Post-graduation, Angela will pursue a career in geospatial environmental modeling with a focus on environmental justice and food accessibility. She is grateful for the extensive support of the department, as well as her family, friends, and cat named Bug.