A Definition of Rurality:

A Sequential Quantitative Analysis Aligning Education to Geography

by
Benjamin J. Anderson

A Dissertation
Submitted in Partial Fulfillment of the Requirements for the
Doctor in Education in Career and Technical Education Degree

Fall 2024

Style Manual Used: American Psychological Association, 7th Edition

Dr. Benjamin Anderson | 08/23/2024

Dr. Michael Selover, Chair | 08/23/2024

Dr. Matthew Simoneau, Committee Member | 08/23/2024

Dr. Debbie Stanislawski Committee Member | 08/23/2024

Students: Please read the statements below and mark each box as “checked” to signify you have read and agree to each item. This page can be deleted from your copy of the dissertation if you choose to publish yet needs to be present for submission to and approval by the Graduate School.

Style Manual Used: American Psychological Association, 7th edition

☒ **I have adhered to the Graduate School Research Guide and have proofread my work.**
☒ **I understand that this research report must be officially approved by the Graduate School. Additionally, by signing and submitting this form, I (the author(s) or copyright owner) grant the University of Wisconsin-Stout the non-exclusive right to reproduce, translate, and/or distribute this submission (including abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video. If my research includes proprietary information, an agreement has been made between myself, the company, and the University to submit a thesis that meets course-specific learning outcomes and CAN be published. There will be no exceptions to this permission.**
☒ **I attest that the research report is my original work (that any copyrightable materials have been used with the permission of the original authors), and as such, it is automatically protected by the laws, rules, and regulations of the U.S. Copyright Office.**
☒ **My dissertation chair has approved the content and quality of this paper.**

Anderson, Benjamin J. ***A Definition of Rurality: A Sequential Quantitative Analysis Aligning Education to Geography***

Abstract

The dissensus on what rural is at the Federal level changes the designation of any given place from rural to urban or vice versa, impacting decisions about allocating resources to communities. Legislative codes conceal heterogeneity and sensitivity to local variations in rural sites, contributing to a growing concern about applying historical approaches to defining rurality and urbanism. This sequential quantitative, descriptive, and correlational study aimed to capture an essence of place as it pertains to rural and evaluates the alignment between the National Center for Education Statistics' (NCES) classification of school districts and the communities they serve. The geospatial analysis aligns the boundaries of the NCES to the United States Census Bureau (USCB; $p > 0.05$), and the agencies disagree on the rural designation of 2,176,937 people within Illinois, Iowa, Minnesota, and Wisconsin. A random forest model of 500 trees is trained on 65 variables in the correlational phase. A sample of 10,000 observations from a test dataset resulted in an accuracy of 98.6% with a No Information Rate of 0.1657 and Cohen's Kappa 0.99. When the classification of the places a school district administrates is compared to the classification of the district, the correlational phase determined that the composition of the places school districts administrate does not align with the classification of the entire district ($p < 0.05$). This research concludes that the conditions of rural will manifest anywhere, regardless of distance from an urban center. The distinguishing factor of rurality is how its circumstances manifest due to mobility, educational and technological access, economic practice, or symbolic ethnicity.

Acknowledgments

I always jested that I did not know what I wanted to be when I grew up. Becoming an adult has always implied that I would be sacrificing my innate curiosity for responsibility. In many ways, this has been true, but I have been fortunate to have had the opportunity to dedicate the last twelve years to exploring my interests. None of this would have been possible without the teachers who encouraged my eccentricities, employers allowing me to devote time to my education, and my peers and colleagues who continue to mentor and share their expertise.

To my dissertation chair, Dr. Michael Selover, and committee, Drs. Matthew Simmoneau and Debbie Stanislawski, I appreciate your input throughout my dissertation journey. Your perspectives and challenges shifted the focus of this research, leading to observations that I would not have found otherwise. None of this would have been possible without your willingness to engage with this multidisciplinary study. It has been a pleasure to learn from and work with you.

I would also like to thank the cohort. Your experiences brought our coursework to life and made this experience like no other. What you have accomplished in your lives and your impacts on others is inspirational. I could not have asked for a finer group of professionals to have walked this path with.

For my wife, Amber, who continues to support my obsession with pursuing anything I find fascinating, I'm fortunate to have you by my side and joyed to take you on the next adventure. To my mostly supportive son, Matthew, who had reservations about me becoming a doctor, don't worry about growing up and picking a career; I haven't. I'm glad you came around once you realized I would be a doctor who does not use needles.

Table of Contents

List of Figures

Chapter I: Introduction

The bonds people have with place have become fragile as globalization, increased mobility, and technological advancements threaten people's connections to space (Scannell & Gifford, 2010). "Mobility" is commonly used to describe one's ability to move upward on the socioeconomic scale (Florida, 2009). In a more literal sense, mobility is freedom of movement across physical space (Rosen, 2011). Within the United States, mobility serves as a means of discovering new prospects, a form of liberation pursued in the spirit of self-reliance (Rosen, 2011). Mobile may also refer to the myriad technologies that allow people to remain in constant contact regardless of location (Rosen, 2011). Unlike historical mobility associated with escaping place, place has minimal impact on technological mobility (Baciu et al., 2017; Rosen, 2011). Technological mobility always informs consumers where they are, what is around them, and where they are going (Rosen, 2011). Even though people are ever more inextricably linked, relationships are not strengthening, and people expect more from technology and less from each other (Matakos & Gionis, 2022; Turkle, 2011).

In a global economy, it can sometimes seem that the world is becoming placeless – mobility and affinity with a physical location have no socioeconomic implications (Florida, 2009; McClay & McAllister, 2014). The advancements in communication and transportation technologies that have made globalization possible have leveled the opportunity gap; "you can innovate without having to emigrate" (Friedman, 2005, para. 11). Since the twentieth century, commentators have been writing about the leveling impacts of trade and technology, making "place" appear unimportant to our economic mobility (Florida, 2009). Technological progress has seemingly eliminated the economic significance of physical location (Florida, 2009).

Succumbing to the idea of a physical place being insignificant is the equivalent of acknowledging that the technologies utilized to communicate are “places” through which people can use digital relationships to substitute friends, family, and community (McClay & McAllister, 2014). To be rooted is a person’s most essential and least recognized need, and a person has roots by virtue of their objective, active, and natural participation in a community, which preserves treasures of the past and expectations of the future (Weil, 2001). People are rooted in their embodiment's particular and finite conditions, which serve as a basis for opportunity (Florida, 2009). Place grounds one’s reality and identity, and people’s roots serve as a basis of place (McClay & McAllister, 2014; Weil, 2001). By substituting place for placeness, people risk losing the basis for a healthy and resilient individual identity (McClay & McAllister, 2014). Rejecting place forfeits the necessary preconditions for cultivating public virtues (McClay & McAllister, 2014; Weil, 2001).

Measuring space is a critical research approach; space alone is a significant factor in a person’s class position, health status, education options, and economic mobility (Dex, Lindley, & Ward, 2007). Where people live impacts how they grow up, spend their free time, and the educational opportunities available (United States Department of Health and Human Services, 2021). A person’s economic status was primarily determined by place of birth in the past (Florida, 2009). People born in disadvantaged locations carry that initial disadvantage across subsequent life stages (Dex, Lindley, & Ward, 2007). In contemporary society, socioeconomic opportunities are affected by one’s ability to move from one space and relocate to another (Florida, 2009). People who are mobile, those who possess the means, resources, and inclination to move to new locations, can seize new economic opportunities (Florida, 2009). In contrast, those tied to place, the rooted, find themselves "stuck" due to limited resources, economic circumstances, and declining financial opportunities (Florida, 2009).

Friedman (2005) emphasizes the unimportance of space, dictating that "the world is flat" and that telecommunication technologies enable everyone to innovate without emigrating to urban centers. The growth of cities and urban places challenges the idea of a "flat world" (Florida, 2009). Between 1800 and 1900, the world's population residing in urban places increased from 3% to 15% (United States Census Bureau, 2023). By 2009, it reached 50%; in the 2020 census, 80% of the United States population resided in urban places (United States Census Bureau, 2023). Bishop and Cushing (2008) describes this pattern as "the big sort" in which communities are further economically stratified by those who are mobile and those who are stuck. Society is segregated and divided across class lines, creating a "spikey" spatial world of economic opportunity (Florida, 2009).

Historically, movement patterns were involuntary and used to find work or associated with the stage of life one is in (Baciu et al., 2017). These spatial patterns are as recent as the 1950s and 1960s in the United States–white-collar and blue-collar workers—moved to find jobs (Florida, 2009). Blue-collar work was concentrated in cities around natural resources and transportation hubs, whereas white-collar workers relocated to where their employers directed them (Florida, 2009). However, according to the 2020 census, work-related mobility only accounts for 23% of all movements in the United States (United States Census Bureau, 2023). Housing accounted for 58.5% of movers; renters want to own, young couples want to upgrade, and retirees want to downsize (U.S. Census Bureau, 2023).

The baby boomers, "boomers,"—those born between 1946 and 1966—usurped traditional spatial patterns associated with previous generations (Florida, 2009). With each new life state the boomers entered, they established new values, creating new habitation, consumption, and location patterns associated with age (Fingerman et al., 2012; Kim et al., 2018). Post-boomer generations continue to expand upon the boomers' values and habitual patterns (Kim et al., 2018). For example, the meaning of “family” is becoming diversified and no longer reflects previous interpretations, and revered securities such as money, employment industry, and education are no longer an accurate indication of how well a place is doing (Florida, 2009). Place is becoming as significant as space for identifying locations needing developmental aid (Neumark & Simpson, 2014).

Statement of the Problem

The socioeconomic class stratification of the United States is due to the human capital and education divide (Florida, 2009). The United States is the world’s most mobile society, and the continued retooling of the United States Census data and shifts in its economy contribute to the growing concern about the application of historical approaches to defining rurality and urbanism (Florida, 2009; Mueller et al., 2020). Administrative, land-use, and economic concepts of populations are the basis of Federal definitions but no longer accurately represent a place's well-being (Florida, 2009; Cromartie & Bucholtz, 2008). Policymakers frequently update definitions to capture the changing essence of place by measuring and mapping population thresholds to determine an area’s political boundaries and resources (Abler, 1992; Mueller et al., 2020; Blaschke et al., 2018). Because definitions do not incorporate experiential data that can be collected and interpreted confidently, the essence of place cannot be captured (Taun, 1977).

Definitions of rural and urban shape public policies, creating legislative action that lessens impediments, builds prospects, or offers other motivations influencing policymaking decisions (Childs et al., 2022; Nanzer, 2004). In the United States, the development of many existing policies and programs is to support rural places and bridge the economic gap between them and urban places (Mueller et al., 2020). However, there is no standard definition of rural or urban (Ratcliffe M., 2022). Using any of the most frequently cited Federal definitions and characteristics of rurality, the percentage of the population classified as rural can vary from 17 to 49% (Cromartie & Bucholtz, 2008).

Many Federal agencies and legislative codes conceal heterogeneity, lacking sensitivity to local variations in rural sites, by dichotomizing urban and rural classifications (Jenkins et al., 2016). The conventional spatial binary classification methods used to distinguish rural places are insufficient, making it impossible to study a well-defined population as these methods do not account for the variation of population density within large areas (Childs et al., 2022). These variations of rurality make it challenging to propose a standard definition (Childs et al., 2022). Arguments suggest that, depending on how the definition is structured, a standardized definition of rural may overlook some rural populations (Childs et al., 2022). Adequately describing the sociocultural aspects of a study's examined population ensures that consumers of research are informed of the people and places examined and allows for the correct interpretation of findings and accurate comparisons between studies (Childs et al., 2022).

In research, policy, and funding initiatives, the definition of rural and urban selected is whichever meets the endeavor's goals, yet rural places remain underserved compared to their nonrural counterparts (Roberts & Green, 2013; Cromartie & Bucholtz, 2008). The dissensus on what rural is at the Federal level changes the designation of any given place from rural to urban or vice versa, impacting decisions about allocating resources to communities, which creates and furthers health, education, and infrastructure disparities for those who are mobile, and those who are rooted or stuck (Bishop & Cushing, 2008; Childs et al., 2022; Florida, 2009). Classifications of rural used by the Federal government justify the planning, application, and monitoring of policy interventions and development efforts to meet the needs of a population, creating an exacerbation of resource issues when different agencies fund communities with department-specific definitions (Cloke & Edwards, 1986). The lax standard for classifying places makes it challenging to justify and set aside resources to aid rural communities across the United States – furthering the economic divide between those who are spatially bound and those who are not (Bishop & Cushing, 2008; Childs et al., 2022).

The growing health, education, and infrastructure disparities between rural and urban areas are due to the lax standard for spatially structuring places (Bishop & Cushing, 2008; Childs et al., 2022). By reducing populations into a binary classification through a single urban lens, the Federal definitions lose more information than they create (Wheatley, 2006). The characteristics that make rural place are lost, leaving unlimited variables to describe rurality, ensuring rural places remain underserved (Childs et al., 2022; Roberts & Green, 2013; Wheatley, 2006).

This research follows the findings of Waldorf (2006) and Selover (2021), where the Midwest and Great Lakes areas as representative subjects capture a large spectrum of rurality and urbanism. The geographical extent includes the states of Illinois, Iowa, Minnesota, and Wisconsin.

Purpose of the Study

The purpose of this sequential quantitative, descriptive, and correlational study is to mathematically capture an essence of place by applying administrative, land-use, economic concepts, and socioeconomic characteristics of each NCES locale classification using census and school district data from the United States Census Bureau and American Community Surveys to describe the existing conditions in which policy derives. Additionally, the results of this research may serve as a means for assessing the progression of policy objectives to inform policymakers and school districts on the impacts of policy or lack thereof.

Research Questions

A coherent understanding of a place's structure is formed by exploring the commonalities across the different permutations of the concepts of space (Scannell & Gifford, 2010). This sequential quantitative, descriptive, and correlational study aims to capture an essence of place as it pertains to rural by using the socio-economic characteristics of school districts and the places they administrate. The following research questions meet this purpose:

Research Question 1: What principles drive the definitions of rural and urban, and how do they align with the locale system for the National Center for Education Statistics?

Research Question 2: What characteristics define rurality and urbanism beyond population thresholds or distances from an urban core?

Research Question 3: How does the National Center of Education Statistics locale classification for school districts align with the classification of the places schools administrate?

Significance of the Study

Any literature on rural places is often from a deficit perspective and for representative sampling rather than for learning something substantial about rural places (Sherwood, 2000). Rural populations are challenging to assess due to being widely dispersed and lacking traditional addresses, leading to undercounting or oversampling fields (United States Census Bureau, 2019a). Sampling techniques are the foundation of the census, and collecting information in rural places tends to produce smaller samples than in urban (United States Department of Health and Human Services, 2020). The result is a more significant margin of error in population estimates and demographics for rural places, furthering the deficiency of conclusive intelligence (Sherwood, 2000; United States Census Bureau, 2019a).

Developing effective research and policies to support the needs of the rural population is impossible without consistent and accurate information (Mulkey, 1988; Stern, 1994). Research is hindered by the lack of information available, preventing it from learning more answers and asking the right questions (Sherwood, 2000). Lack of adequate research with definitional inconsistencies limits policymakers' ability to enact legislation and monitor effectiveness, leading to a continuous cycle of policy implementation with little or no responsibility for their effectiveness on rural places (Sherwood, 2000). No accountability or supporting research ensures that rural issues remain undiscovered, unaddressed, or ignored (Sherwood, 2000).

Ill-defined spatial distinctions between rural and urban are the foundations of rural research and policy (Isserman, 2005). Places are more than points on a map; they exist in many sizes and shapes and can be tangible or symbolic (Low & Altman, 1992). The conventional binary spatial classification methods used to distinguish rural places are insufficient, making it impossible to study a well-defined population as these methods do not account for the variation of population density within large areas (Childs et al., 2022). By explicitly defining rural and quantifying aspects of “place,” researchers better operationalize their construct, enabling future researchers to evaluate and critique the alignment between different conceptualizations of rural policy (Hawley et al., 2016; Marzbani et al., 2020)

Assumptions of the Study

This research operates under the premise proposed by Mueller et al. (2020) that rurality is not the inverse of urban. Vastly different urban or rural places are not the same; there are varying degrees of rurality and urbanism that cannot be reduced to the binary system of Federal agencies.

Further, this research assumes that data from authoritative sources such as the NCES, USCB, Office of Management and Budget (OMB), and Department of Agriculture (USDA) are the highest quality sources for information surrounding the demographics of rural and urban places. The extraction of demographics and classification of places operates under the assumption that population distribution is equal across census blocks, a common assumption in spatial analysis conducted by the U.S. Office of Policy Development and Research (Din & Wilson, 2020). Additionally, this research assumes that variables beyond population and distance from an urban core will accurately describe rurality in further educational research. Lastly, this research assumes that labeling the product a “definition” will attract further attention from education researchers and policymakers.

Limitations of the Study

The data collection methods of authoritative sources impose some limitations on this study. Sourcing data from the American Community Survey, USCB, and the U.S. Department of Labor is only as accurate as their sampling methods, which have varying degrees of error. In addition to historical changes in urban definitions, differing sampling methods, and definitions of rurality, using data from the USCB after 2010 includes added statistical noise due to rising privacy concerns – creating opportunities for error in cross-agency data linking (United States Census Bureau, n.d.). Furthermore, creating a mathematical model on an oversampled population encourages bias and only captures quantifiable aspects of “place.” Lastly, the assumption that population distribution is equal across census blocks is inaccurate (Din & Wilson, 2020). This research uses the smallest partition of statistically significant areas, census blocks, to limit the effects of the assumption of equal distribution.

Definition of Terms

This research uses the following terms:

Education

Pertaining to the physical places primary, secondary, and postsecondary educational institutions administrate.

Geographic Information System

Geographic Information System is a software package or coded hardware device that displays, manages, manipulates, and stores geographic data.

Geographic Object Identifier

Geographic Object Identifier (GEOID) is an acronym for geographic object identifier. A GEOID is a unique identifier assigned to an object to enable linkages between geospatial information in various formats.

JavaScript Object Notation

JavaScript Object Notation (JSON) is a lightweight data-interchange format that is easy for humans to read and write (IBM, 2024).

Kurtosis

In statistics, a measure of the degree to which a distribution has a peak versus to which it is flat. Along with skew, kurtosis is one of the descriptors of the shape of a distribution.

Layer

An individual dataset that is displayed in a geographic information system. A layer may be a single or multiple files within a geodatabase, imagery, or other spatial information.

Mathematical Model

A system that uses mathematical concepts and language to study the effects of different components and predict behavior**.**

Nonrural

Any location defined as urban, suburban, or nonrural by a government agency or non-governmental organization within the context of a funding, economic development, or educational program.

Place

The relationship between physical and social interactions, influenced by the surrounding circumstances and factors of space, creating a sense of belonging that forms one's identity.

Rural

Any location not classified as urban or defined as rural by the United States Census Bureau.

Shapefile

A file that stores, displays, and manipulates geographic data in vector format within a geographic information system. The format is designated by the file extension .shp.

Space

A mechanism to measure geometric areas and volumes to assess the environment's physical impact on people and to determine an area's administrative boundaries, resources, potentials, and wealth.

Urban

A geographical location containing 5,000 people and 2,000 housing units for a given census area as defined by the U.S. Census Bureau (2022).

Vector

Geographic information in the form of a collection of points, lines, and polygons.

Chapter II: Literature Review

This literature review introduces the theoretical framework used to underpin the analysis and mathematical approaches for classifying rural and urban places. It continues with the structuring of place in the educational policy of the United States. Lastly, it highlights how Federal agencies describe rural places in policy and its challenges.

The Essence of Place

Places are for and about people; they are more than collections of buildings and roads augmented by the occasional park (Tibbalds, 1992). Places provide meaning through structure by creating a foundation for necessary, optional, and social activities (Gehl, 2011). Strictly necessary activities are mandatory and occur even if place does not effectively support its inhabitants (Gehl, 2011). Mandatory activities may include commuting to work or waiting for public transportation, and the physical setting of place has minimal impact on these activities (Whyte, 1980). In contrast, optional activities are willingly done behaviors heavily influenced by the environment of place (Marzbani et al., 2020). Through public infrastructure and services, access to open spaces with various social uses, such as meeting spaces, parks, and shops, enables individual activity (Towers, 2005). Social activities form when the environment provides better conditions, such as entertaining amenities, education, and safety, for individual necessary and optional activities (Lang, 1994).

Sense of place is a model for building communication between space and its people by establishing a dialog to understand the activities, icons, and meaning within spaces (Montgomery, 1998). It serves as a foundation to ensure societal well-being in land use planning or to gain a more nuanced understanding of community resilience for conservation and development (Faulkner et al., 2018; García-Díez et al., 2020). Place is a sense of belonging that forms one's identity based on the surrounding circumstances and factors of space; it is a perceptual aspect and the relationship between physical and social interactions that manifest as a home, neighborhood, city, community, state, region, or nation (Carmona, 2021; Canter, 1997). The circumstances and factors of space then impact individual and societal behaviors, values, attitudes, and social activities (Carmona et al., 2003). Some researchers suggest that a sense of place encompasses sub-concepts of place identity, attachment, and dependence (Scannell & Gifford, 2010). However, the emphasis is typically on displacement or the longing for places that are lost and excludes concepts such as mobility (Hirsch & Deutsch, 2005; Gustafson, 2001).

Scannell and Gifford (2010) propose a tripartite organizing framework that extends the "sense of place" to incorporate "place attachment" -- the bonding between individuals and their meaningful environments (Gustafson, 2001). The framework has three dimensions: the person, the psychological "sense of place" process, and the attachment object "space" (Scannell & Gifford, 2010). The person dimension is who is attached (Scannell & Gifford, 2010). The psychological process involves how the affect, cognition, and behavior manifest in place (Scannell & Gifford, 2010). Lastly, the object of attachment is the physical characteristics of place (Scannell & Gifford, 2010).

Place attachment occurs within the person dimension at individual and group levels (Scannell & Gifford, 2010). At the individual level, place attachment involves the personal connections one has to place –settings that evoke personal memories and contribute to a stable sense of self (Twigger-Ross & Uzzell, 1996). Group-framed place attachment is a community process, and attachment forms when cultures, genders, and religions attach to areas, preserving their practices and cultures (Scannell & Gifford, 2010). Culture links people to places through shared historical experiences, values, and symbols (Scannell & Gifford, 2010).

The psychological “sense of place” concerns how individuals and groups relate to place (Scannell & Gifford, 2010). Place attachment as affect is the authentic and emotional bond with an environment that satisfies a fundamental human need (Relph, 1976). As cognition, place manifests itself as “symbolic communities” through social schemas based on historical events (Scannell & Gifford, 2010). Individuals attach to categories of places such as cottages, suburban single-family dwellings, or downtowns (Fieldman, 1990). The circumstances of the categories of place impact individual and societal behaviors, values, attitudes, and social activities (Carmona et al., 2003).

Lastly, the attachment object is the physical representation of place within social and physical dimensions at differing geographic scales (Scannell & Gifford, 2010). Certain physical features, such as density, proximity, and amenities, influence the social representation of place (Fried, 2000). Primarily, social systems of mass society, such as socioeconomic class, race, religion, and life stage, define where people reside within places, forming pockets of relatively homogenous communities (Scannell & Gifford, 2010). People attach themselves to a place because it facilitates “distinctiveness” from other places or affirms the uniqueness of one’s group (Twigger-Ross & Uzzell, 1996).

Rural as Essences of Place

Education is the most essential function of state and local governments, and compulsory education laws demonstrate our recognition of the importance of education, the foundation of good citizenship, and a democratic society (Kauper, 1954). Though the federal government of the United States has limited authority over education, no country has seen a school established so nearly coincident with its first settlements (Denenberg, 1979; United States Department of Education, n.d.). Should any child be denied the opportunity of an education, they are unlikely to succeed, and such an opportunity is a right that must be available to all on equal terms (Kauper, 1954). Furthermore, the educational needs of rural residents are as multifaceted as urban populations (United States Department of Health, Education, and Welfare, 1960). The need and access to good, up-to-date service to the same degree and service standards should apply to urban and rural places (United States Department of Health, Education, and Welfare, 1960).

A standardized definition of rural that incorporates elements of place may help eliminate confusion in research and policymaking (Childs et al., 2022). Any standardized rural and urban taxonomy should measure something explicitly and meaningful, be replicable, be derived from available, high-quality data, be quantifiable and not subjective, and have on-the-ground validity (Hart et al., 2005). Recognizing multiple populations of what constitutes rural in the present delineation of definitions within the physical dimension of place will make future longitudinal rural research studies more reliable (Childs et al, 2022; Dex et al., 2007). Any granular spatial description of rural allows multi-level analytical approaches to rural studies, possibly gaining a better understanding of the psychological and individual nuances in various populations (Meilleur et al., 2013). Additionally, accounting for location and an individual's immutable characteristics will uncover the unique conditions people face, and a more tailored educational response can be formulated (Selover, 2021).

Rural as an Agreement Scale

Which definition of rural should researchers and policymakers use? Long et al. (2021) explore the disparities between definitions by comparing psychological and physical characteristics, including percentage with a bachelor's degree, income below 200% poverty, population density, percentage with health insurance, and whether various health care services within 30 minutes driving time of the census tract centroid. By analyzing 72,506 census tracts across eight Federal rural definitions, Long et al. (2021) created an agreement scale of zero through eight, where zero specifies the number of definitions that agree a place is not rural, and eight is all definitions identify a place as rural. Any place not classified as zero or eight highlights the dissonance between Federal definitions (Long et al., 2021).

While many of the definitions are related, they have essential variations. Census data is the primary derivative for the properties of the eight definitions (Long et al., 2021). Census geography is the foundation of definitions originating from the USDA and the OMB, where the USDA frequently relies on rural-urban commuting areas and the OMB on core-based statistical areas (Office of Management and Budget, 2010; United States Department of Agriculture, n.d.-a). The USCB follows a variable geographical description containing urban areas and clusters (Long et al., 2021). Further definitions under the USCB, such as the USDA's Frontier and Remove Access Codes (FAR), focus on ZIP codes to describe population definitions (Long et al., 2021).

Not all of the eight most cited Federal definitions in medical research use a binary scale to fall within the proposed dichotomous index (Long et al., 2021). Four definitions used a binary classification system, and no reclassification was necessary. Other definitions contain as many as 12 levels of rurality (Long et al., 2021). Thresholds were established by cutoff points frequently used in rural health literature and other indicators of social determinants and rurality (Long et al., 2021).

The 5-year American Community Survey from the USCB from 2014-2018 provided the population characteristics at the census tract scope (United States Census Bureau, 2019b). Population characteristics, such as hospital information, health information, and population estimates, originated from the Integrated Public Use Microdata Series and the National Plan and

Provider Enumeration System files (Long et al., 2021). Furthermore, identifying whether the centroid of each tract was within a 30-minute drive of a health center utilized the ESRI's online application programming interface (API) for network analysis, based on the 2010 census block populations to locate tract population-weighted centroids (Long, Delameter, & Holmes, 2021).

Federal definitions present rural population estimates ranging from 6.9 million to over 75.5 million (Long et al., 2021). Urban Influence codes generated the most significant disparity between definitions for seven dimensions: poverty, population growth, percentage of nonelderly uninsured, and healthcare access measures (Long et al., 2021). Notable discontinuities between zero (not rural) and one (dissonance between rural and urban) occur along the total population, population density, minority population percentages, and the prevalence of depression dimensions (Long et al., 2021). Similar observations occur when definitions agree on a place being rural, seven and eight, including 30-minute travel time to access a hospital and primary care provider (Long et al., 2021). Clear breaks in dimensions may highlight differences between rural and urban populations; however, characteristics such as health outcomes, education, and poverty vary semi-smoothly across the levels of agreement, suggesting that binary classifications may not capture the nuances of the differences between rural and urban populations (Long et al., 2021).

Rural As a Function of Educational Access

Research, policy, and funding initiatives underserve rural secondary and postsecondary career and technical education administering institutions compared to their nonrural counterparts due to inconsistent definitions of what constitutes rural and the social and economic realities in these locations (Selover, 2021). By correctly defining the boundaries of rural in the United States, even on a regional basis, a more careful analysis will uncover the unique conditions facing the people who live and work there, and a more tailored educational response can be formulated (Selover, 2021). To begin a conversation about educational equity based on physical location rather than on traditional designators of diversity, Selover (2021) seeks to determine the misalignment between locations federally defined as rural under one of the many definitions of rural in use. Lastly, Selover (2021) provides recommendations to frame future conversations in terms that do not include views of rural and peripheral educational needs in deficit terms but instead use their inherent resiliencies.

Selover (2021) follows a similar quantitative process to Long et al. (2021) by incorporating high-quality, available data from federal definitions of rurality and presenting the classification disagreement. However, the context is with educational institutions and campuses instead of exclusively census tracts (Selover, 2021). The geographic information system software calculates the maximum differences between the institutional rurality and the surrounding population using a nearest-neighbor analysis, resulting in a nearest-neighbor index value for each institution (Selover, 2021). Maximum index values indicate the most substantial disagreement, or inverse correlation, between the rurality of the census tract housing the institution and the census tracts of the population surrounding it, and the significance of these disagreements use Tukey's test statistic (Selover, 2021).

Like Long et al. (2021), Selover (2021) proposes using smaller administrative units to analyze populations. Altering the geographic unit from census tracts to smaller features, such as census block groups or blocks, may lead to a better understanding of the disagreement between institutional locations and the populations they serve (Selover, 2021). Additionally, there is a differentiation between the experiences of rural and nonrural individuals from which policy may resolve (Selover, 2021). Discussions surrounding educational equity should include an accounting of not just a student's immutable characteristics but their point of origin as well (Selover, 2021).

Rural as a Continuous Multi-Dimensional Measure

Ill-defined spatial distinctions between rural and urban are the foundations of rural research and policy (Isserman, 2005). The metro/non-metro perception is a standard proxy for, or is synonymous with, a rural/urban distinction (Waldorf, 2006). Similarly, the Rural-Urban Continuum Code distinguishes metropolitan counties by the population size of their metro area and non-metropolitan counties by the degree of urbanization and adjacency to a metro area (United States Department of Agriculture, n.d.-b). Although the FARs and USDA provide coding and suggest a continuous and monotonic increase of rurality on a scale, this suggestion may be an illusion as it hides the initial distinction between metro and non-metro areas (Long et al., 2021; Waldorf, 2006).

To remedy the shortcomings of coding, Isserman (2005) suggests a rural-urban density topology. Isserman's topology assigns counties into one of four categories based on the percentage of urban residents, the total number of urban residents, population density, and the population size of the county's largest urban area. Though this approach is excellent at identifying extremes, it "pigeonholes" counties, potentially separating similar counties and joining dissimilar counties (Waldorf, 2006). Furthermore, threshold-based categorization is not independent of a spatial scale, requiring new sets of thresholds when using a different spatial scale. Waldorf (2006) asserts that the question should not be "Is a county rural or urban?" but "What is a county's degree of rurality?"

To develop a continuous, multi-dimensional measure of rurality, Waldorf follows the approach used by the Human Development Index to measure a country's average achievement on a continuous scale from zero to one hundred (Waldorf, 2006). This process involves four phases: identifying the dimensions of rurality, selecting measurables to represent each dimension, re-scaling the variables onto a comparable scale, and selecting a function that links the re-scaled variables in a function that reduces multidimensionality into one-dimensionality (Waldorf, 2006). Each step involves a series of subjective decisions that will impact the outcomes – defendable justifications for each step must be part of the approach. The benefit of this approach is that a continuous measure does not confine the resulting index to a particular physical scale, such as counties (Waldorf, 2006). However, due to the elusive nature of rurality, it will ultimately be impossible to assess the precision of the continuous measure (Waldorf, 2006).

Data availability limits the selection of variables to represent each dimension (Sherwood, 2000). Waldorf (2006) uses “simple measures” that can be replicated and updated. Due to the skewed population distributions, abundance of small populations, and low population densities, Waldorf (2006) takes the logarithm of the population size and population density. Further dimensions include the distance to the closest metropolitan area and the percentage of the population living in an urban area (Waldorf, 2006).

When developing a continuous model, variables must be re-scaled and linked when developing a continuous model (Milne & Cohen, 1999). The four measures need to be on compatible, bounded scales to ensure the resulting index is independent of the units of measurement (Waldorf, 2006). Linking the variables determines how the measures jointly determine the rurality of a place (Waldorf, 2006). For example, is population size more significant than population density? Waldorf (2006) chose an unweighted average re-scaled on a zero-to-one scale, with zero being the most urban and one being the most rural.

Using the USCB 2000 census data, the resulting index reflects the lowest rurality scores along the Great Lakes and most urban counties in the U.S. – the New York Metro area, Cook County, Illinois, and Los Angeles, California (Waldorf, 2006). The most rural counties include Daniels County, Montana, and four counties in Nebraska: McPherson, Blaine, Logan, and Thomas (Waldorf, 2006). In non-metro counties, the percentage of adults without a high school degree is higher, and the percentage of adults with at least a bachelor's degree is lower than in metro counties (Waldorf, 2006). Though there was little change in the rurality pattern between the 1990 and 2000 censuses, the percentage of adults with low academic attainment decreased substantially (Waldorf, 2006).

Structure of Place in United States Education Policy

The policy process in the United States includes the fundamental concept of representation, the notion that legislation should represent the region's attitudes and the individuals living there (Nanzer, 2004). Those who enact policy represent the interests of the public (National Conference of State Legislatures, 2021). Effective representation requires officials to keep constituent beliefs and values in mind while developing legislation that will impact the public (National Democratic Institute, 2008). When considering courses of action that can impact the roots of identity and community, policymakers must understand the attachments individuals have formed with the regions in which they live (Nanzer, 2004). Understanding place will aid policymakers in considering issues that create burdens or benefits within space (Baekgaard, Moynihan, & Thomsen, 2021).

The Attachment Object as Foundations of Education Policy

United States Federal and State agencies provide fundamental descriptions for categorizing rural and nonrural populations in educational research and policy through visible and perceptible phenomena (Baumgartner & Zielowski, 2007). These descriptions generally fall along population thresholds (Long et al., 2021). The foundation of using population thresholds for policy originates from one of the most significant legislative acts of the Confederation Congress, the Northwest Ordinance of 1787 (United States Library of Congress, 2020). The Northwest Ordinance of 1787 developed a public land survey system and provisions three steps for admitting new states into the Union, requiring population thresholds through each phase (National Archives, n.d.). In its initial phase, the Act requires a congressionally appointed governor, secretary, and three judges for a territory (National Archives, n.d.). Once the population of a territory reaches "five thousand free male inhabitants of full age," the territory will elect an assembly and one non-voting delegate to Congress (National Archives, n.d.). At 60,000 inhabitants, the territory will draft a state constitution and request membership into the Union (National Archives, n.d.).

The 1787 Land Ordinance developed a public land survey system, the New England survey system, partitioning townships into square, six-mile parcels on each side with thirty-six inner sub-sections and a parcel, Section 16, being reserved for maintaining public education (U.S. Department of the Interior Bureau of Land Management, 1991; Carleton, 2001). Initially, the population requirement for statehood in the Ordinance was the State's population with the least numerous inhabitants (United States Library of Congress, 1776). Using the new survey system, proposals for the population threshold would incorporate density (Library of Congress, n.d.). Conceptualizations for the Northwest Territory ranged from introducing states with a minimum size of 10,000 square miles to 29,000, 8 to 25 states, and up to 200,000 inhabitants (Library of Congress, n.d.).

Congress continued to land grant provisions from the Northwest Ordinance of 1787 to support common public schools throughout the nineteenth century (Tyack et al., 1987). Under The Act Donating public lands to the several States and [Territories] which may provide colleges for the benefit of agriculture and the Mechanic arts, the Morrill Act of 1862, the U.S. government ceded land from Native Americans into six-mile square townships with a one-square-mile section reserved for public schools' support (National Archives, n.d.). In the Smith-Hughes Act, the Federal board partitioned the United States into five regions for administration and inspection purposes (Houston, 1919). Similarly, the Library Services Act of 1956 partitioned the U.S. into five regions: Northeast, South, North Central, West, and Outlying Parts due to the vast differences in population disbursement (United States Department of Health, Education, and Welfare, 1960). The Library Services Act authorized funding to improve public library services to rural places with a population of 10,000 or less, where rural areas are less than 5 miles from an urbanized place or less than 2.5 miles from an urban cluster (Office of Educational Technology, n.d.).

The Individual as Education Policy

Throughout the mid-20th century, United States policymakers categorized federal policies and programs into three standards that indicated equity levels to further advance them: access, participation, and outcomes (Ashford-Hanserd et al., 2020). The access standard mandates equal availability of facilities and services for all students (Ashford-Hanserd et al., 2020). The participation principle focuses on ensuring the equal engagement of students within the educational system (Ashford-Hanserd et al., 2020). Lastly, the equal outcomes standard achieves post-graduation parity in educational attainment (Ashford-Hanserd et al., 2020).

The *Plessy v. Ferguson* ruling sanctioned the "separate but equal" doctrine under the equal protection clause of the Fourteenth Amendment, enforcing the framework as the basis for measuring educational equity (Ashford-Hanserd et al., 2020; Kauper, 1954). Lawsuits in Kansas, South Carolina, Virginia, Delaware, and the District of Columbia challenged the equity framework under *Plessy v. Ferguson,* citing the validity of segregation in public education (Kauper, 1954). Rulings by the Supreme Court, however, suggested that the segregation of educational facilities was either equal or in the process of being equalized and that segregation as a social pattern would be present; that even with standard geographical districting of school lines, a routine program of districting for school purposes will have the same effect as a segregation program (Kauper, 1954).

Brown v. Board of Education rules that separation based on race violates the 14^{th} Amendment and, therefore, is unconstitutional (Brown v. Board of Education, 1954). Prior rulings leading up to *Brown v. Board of Education* imply a spatial mismatch between the education policy of the United States and those impacted by it –residentially and economically segregated zones are separate but not equal and limit career opportunities (Kain, 1968; Wilson, 2012). In *Sipuel v. Board of Regents,* the Federal access standard was struck down in Oklahoma because it refused to admit minority students to the State law school when no separate law school was maintained (Kauper, 1954). Participation fell shortly after in the *Sweatt v. Painter* ruling of 1950, challenging admittance into the University of Texas Law School even though a separate law school existed for minority students. The ruling in *Sweatt v. Painter* states that the separation of schools deprives minority students of the opportunity for the interplay of ideas, exchange of views, and association with students who would eventually constitute a large part of the legal profession (Kauper, 1954).

Furthermore, *Sweatt v. Painter* and *McLaurin v. Oklahoma State Regents* contest equal outcomes, with the ruling supporting that prestige associated with a university limits employment opportunities and that segregation inhibits a student's ability to exchange views to learn their profession (Ashford-Hanserd et al., 2020; Kauper, 1954). One of the most significant policies to arise from spatial mismatch, providing Federal funding directly to school districts, appeared in 1965 with the enactment of Title I of the Elementary and Secondary Education Act (ESEA; Archambault & Pierre, 1980). The eligibility of Title I schools, both public and private, requires that at least 40% of students originate from low-income families, as defined by the U.S. Census (Congressional Research Service, 2024).

Psychological Attachment as Education Policy

The United States public education system is responsible for educating ninety percent of students (United States Congress Joint Economic Committee, 2019). Traditionally, public education is tied directly to housing through spatial residential assignment policies that assign homes to schools via attendance zones (United States Congress Joint Economic Committee, 2019). Officials design attendance zones so that students attend schools near home, and despite growing opportunities to opt-out or choose another via charters, magnets, and public-school open enrollment policies, seventy-one percent of students attend their assigned public school (National Center for Education Statistics, 2019). Policymakers treat housing and education policies separately, even though they are practically enmeshed (United States Congress Joint Economic Committee, 2019). Students will naturally attend schools within commuting distance of their homes, associating school quality with home prices (Chiodo et al., 2010). However, the relationship between schools and housing is a policy choice (United States Congress Joint Economic Committee, 2019). Districts tend to draw school boundaries by neighborhood and assign students to school by home address, even when other options are more convenient or appealing to families (United States Congress Joint Economic Committee, 2019).

Residential zoning regulations control the type, size, and number of homes built in school zones, segregating cities by income within and across metropolitan places (Lens & Monkkonen, 2016). The housing characteristics systematically vary across school attendance zones, with larger houses and single-family homes more common with high-performing school boundaries, and more significant housing cost gaps exist across high and low-quality schools in places with more restrictive residential zoning (Kane et al., 2005; Rothwell, 2012). Strategies, such as charter schools, magnet schools, and districtwide lotteries, decouple housing from schools within specified places (OCEDiLibrary, 2012). Though these strategies reduce the link between housing and education, the traditional relationship persists: nine percent of K-12 students attend a private school, twenty percent attend a chosen public school, and seventy-one attend their assigned public school (Fischel, 2015; United States Congress Joint Economic Committee, 2019).

A study linking housing data to ZIP code and neighborhood found that public elementary schools tend to be associated with smaller attendance zones than middle and high schools (United States Congress Joint Economic Committee, 2019). The median home value varies widely across ZIP codes in the U.S., and the average ZIP code associated with a high-quality public school has a significantly higher median home value than the average ZIP code associated with a low-quality public school, even when regional differences are incorporated (United States Congress Joint Economic Committee, 2019). The Social Capital Project report of 2019 uses a rating scale of D- through A+, with D- being the lowest rating and A+ the highest, in a D-, D, D+, C-, C, C+, …. sequence (United States Congress Joint Economic Committee, 2019). Schools ranked “A+” in the Social Capital Project report have a median home value of $486,104 – four times higher than any public school ranked a D or lower ($122,061; United States Congress Joint Economic Committee, 2019). According to the U.S. Census Bureau (2019), the median home value in the United States is $200,000. The findings of the Social Capital Project report suggest that many families would struggle to find schools, according to the report's scale, with a B+ or higher rating due to housing costs (United States Census Bureau, 2018; United States Congress Joint Economic Committee, 2019).

Academic and teacher quality follows the same trend as school quality, where the median home value is $106,900 for schools with the lowest rating in the Social Capital Project report and $508,485 for schools with the highest (United States Congress Joint Economic Committee, 2019). For each increase in teacher quality, the median home value doubles, whereas a school with the highest rating has a median home value of $334,626 (United States Congress Joint Economic Committee, 2019). Lastly, the median home value is flatter across health and safety ratings, with the lowest health and safety rating having a median home value of $167,989 and the highest having a median home value of $425,021 (United States Congress Joint Economic Committee, 2019).

Arguments against the association of school and teacher quality suggest that quality simply reflects individual choices about where to live (United States Congress Joint Economic Committee, 2019). How households value educational amenities and invest in their school district will improve educational quality and increase home value (OCEDiLibrary, 2012). Furthermore, home value does not describe what type of households' children excel in academics (OCEDiLibrary, 2012). However, past and present school and comprehensive residential zoning suggest that the relationship between homes and schools is unlikely to reflect a natural product of free choice (United States Congress Joint Economic Committee, 2019).

Ambiguity in Defining Rural

The phrase “rural” creates an amalgam of images (Childs, Boyas, & Blackburn, 2022). Many liken the word “rural” to socially isolated and sparsely populated places with farmland, dirt roads, and a distinctive sociocultural context with unique demographic configurations (Childs, Boyas, & Blackburn, 2022). Many people claim to identify rural when they see it, but most cannot define it and often overlook its complexity (Kettler et al., 2016). Though depicting what rural looks like in ordinary life seems natural, arriving at a precise and meaningful definition is much more challenging (De Gennaro & Fantini, 2009). In 2019, the former U.S. Secretary of the USDA, Sonny Perdue, testified before the U.S. Senate Appropriations Committee that “we would love to have a comprehensive definition of rural” (The State of the Rural Economy, 2019).

In a study comprising 929 documents from bibliographic databases, Childs et al., (2022) found 33 definitions of rural used by the federal government in 49 sources. These definitions classified places as rural based on population size, population density, and integration into urban places (Childs et al., 2022). Of the 33 definitions, 30 incorporated population thresholds, 10 included population density requirements, and nine integrated with urban places, such as adjacency conditions (Childs et al., 2022). Some definitions, such as the FAR codes, include all three requirements (Childs et al., 2022). FAR codes consider population, population density, and commute time to an urban when classifying a region as urban or rural (Childs et al., 2022). Additionally, the classification schemes follow country and sub-county partitions, where the OMB distinguishes counties as metropolitan, micropolitan, and rural (Childs et al., 2022). In contrast, the USCB considers the population at the census tract or census block level (Hawley et al., 2016). Before the 2010 census, a town with a population of 3,000 in a remote place was considered urban under the USCB, but the same town is often nonmetropolitan under the OMB definition (Hart et al., 2005). On several occasions, a single agency utilized multiple definitions of rural in research and policy (Mueller et al., 2020). For example, the U.S. Department of Agriculture offers 11 definitions of rural dispersed across multiple departments, and the Farm Security and Investment Act of 2002 includes four definitions of rural (Farm Security and Rural Investment Act of 2002, 2002).

Definitions for rural populations range from 7 to 49% of the U.S. population, with classification thresholds of an area requiring as few as 1,500 people to as many as 1,000,000 (Childs, et al., 2022; Cromartie & Bucholtz, 2008). Childs et al. (2022) describe nine thresholds in the 30 definitions that use population to classify a given area: 3% of definitions set a threshold of 1,500 people or less as rural, 2,500 (3%), 4,000 (3%), 10,000 (13%), 20,000 (19%), 25,000 (7%), 35,000 (3%), 50,000 (45%), and 1,000,000 (3%). Further complications arise when comparing definitions as descriptions of what constitutes an “area” vary (Childs, Boyas, & Blackburn, 2022). Areas defined by the USCB, OMB, USDA, and FAR codes may consist of a county, territory, geographical area, community, place, town, city, open country, or village (Vanderboom & Madigan, 2007). Definitions may also extend across counties and states (Vanderboom & Madigan, 2007). The subtle nuances in the distinction of an area make it difficult for areas to obtain resources that otherwise may be available (Childs et al., 2022).

The USCB definition is one of the most commonly used rural taxonomies in education and health policy (Childs et al., 2022; Isserman, 2005; Ratcliffe et al., 2016). The USCB appropriates rural by the absence of urban, where urban thresholds consider the population of an area and population density (Ratcliffe M., 2022). Before the 2020 census, urban areas required a population of over 50,000 people or an urban cluster with a population between 5,000 and

49,999 people (United States Census Bureau, 2023). In addition to population thresholds, core census block groups or blocks with a population density of at least 1,000 people per square mile and surrounding census blocks with an overall density of at least 500 people per square mile are urban (Ratcliffe M., 2022). Findings in the 2020 census shifted the population requirement to the number of housing units (Ratcliffe M., 2022).

The incorporation of urban clusters in 2000 and the shift from population to housing units make longitudinal assessments of census data not easily comparable (Childs, Boyas, & Blackburn, 2022). The USCB modifies its definition decennially, complicating the ability to assess data over time and between rural studies (Childs et al., 2022). This process is not unique to the USCB as the OMB, USDA, National Center for Education Statistics, and Rural-Urban Commuting Codes are frequently updated (Childs et al., 2022). In contrast, some definitions, such as the Library Services and Construction Act, have not been updated in nearly 60 years (Farrell, 2012). Despite challenges with fluctuating classifications, the wide adoption of the USCB's definition in research permits comparing data between studies (Vanderboom & Madigan, 2007). Additionally, the USCB taxonomy allows researchers to conduct studies that span geographical regions to aggregate data at the county level (Vanderboom & Madigan, 2007).

To address the complexity of rurality, researchers and Federal agencies use definitions that best support their goals, introducing a new set of measurement challenges (Cromartie & Bucholtz, 2008; United States Department of Health and Human Services, 2020). The U.S. Census statistical measure classifies many places as rural and overrepresents the rural population (Rural Health Information Hub, n.d.). In contrast, the OMB classifications include rural places in metropolitan counties, leading to the underrepresentation of the rural population (United States Department of Health and Human Services, 2020). The Federal Office of Rural Health Policy

(FORHP) addresses the underrepresentation of rural places by classifying all non-metro counties as rural and incorporating Rural-Urban Commuting Area (RUCA) codes (United States Department of Agriculture, n.d.-a). RUCA codes follow Census Tracts and have allowed the identification of rural places in Metro counties (United States Department of Health and Human Services, 2020). RUCA codes, however, do not account for distance to services and sparse population, resulting in the manual classification of 132 large census tracts as rural (United States Department of Health and Human Services, 2020).

In 2020, the U.S. Census Bureau redefined urban and rural classification, reducing urbanized places and clusters to one urban category (Ratcliffe M., 2022). The minimum population threshold to be classified as urban was increased from 2,500 people to 5,000 for a given census area (United States Census Bureau, 2019a). Urban qualification no longer depends on population density but on a housing unit qualification of at least 2,000 (Ratcliffe M., 2022). The new classification offers an "either/or" approach for urban classification (Ratcliffe M., 2022). The change considers areas of higher population density but a lower number of housing units or areas of lower population density but a higher number of housing units to be urban (Mueller et al., 2020). This classification changed approximately 1,000 areas from urban to rural status, containing an estimated 3.5 million people (Ratcliffe M., 2022).

In comparison, the OMB provides the designation of counties as Metropolitan ("Metro"), Micropolitan ("Micro"), or neither (Office of Management and Budget, 2010). A metropolitan area contains a core urban area of 50,000 or more people, and a micropolitan area contains an urban core of at least 10,000 but less than 50,000. A Metropolitan Statistical Area onsists of a core area containing a substantial population nucleus with adjacent counties having a high degree of economic and social integration with the core (Office of Management and Budget, 2010).

Rural areas are all counties not part of a metropolitan or micropolitan area or part of a Metropolitan Statistical Area (Office of Management and Budget, 2010).

Inconsistencies arise with the definitions presented by the USCB and the OMB, two of the most widely accepted rural designations, as the development of their taxonomies were for different reasons and geographical areas (Childs et al., 2022). Depending on the definition, individuals living in low-population density areas may be classified urban by the USCB and rural by the OMB if they are adjacent to or near an urban area (Vanderboom & Madigan, 2007). Even more significant disparities arise when comparing eight of the most commonly used Federal rural taxonomies (Long et al., 2021).

Summary

In research, policy, and funding initiatives, the definition of rural and urban selected is whichever meets the endeavor's goals, yet rural areas remain underserved compared to their nonrural counterparts (Roberts & Green, 2013; Cromartie & Bucholtz, 2008). A granular description of rural allows multi-level analytical approaches to rural studies, possibly gaining a better understanding of the nuances in various populations (Meilleur et al., 2013; Selover, 2021). Many Federal agencies and spatial legislative codes conceal heterogeneity, lacking sensitivity to local variations in rural sites, by dichotomizing urban and rural classifications (Jenkins et al., 2016). Any coding schemas create imperfect representations of neighborhood and school boundary geography, leading to the assumption that the population characteristics represent Federal administrative structures (Saporito et al., 2013). The benefits of using a standardized definition across multiple agencies and research studies outweigh the risks (Childs et al., 2022).

Chapter III: Method and Procedures

The purpose of this sequential quantitative, descriptive, and correlational study is to mathematically capture an essence of place by applying administrative, land-use, economic concepts, and socioeconomic characteristics of each NCES locale classification using data from the United States Census Bureau and American Community Surveys to describe the existing conditions in which policy derives. A granular description of rural allows multi-level analytical approaches to rural studies, possibly gaining a better understanding of the nuances in various populations (Meilleur et al., 2013; Selover, 2021). Identifying these nuances may help eliminate confusion in educational research and policymaking (Childs et al., 2022). The following research questions support these objectives:

1. Research Question 1: What principles drive the definitions of rural and urban, and how do they align with the locale system for the National Center for Education Statistics?
2. Research Question 2: What characteristics define rurality and urbanism beyond population thresholds or distances from an urban core?
3. Research Question 3: How does the National Center of Education Statistics locale classification for school districts align with the classification of the places schools administrate?

This research employs a sequential quantitative, descriptive, and correlational design to answer these research questions. Quantitative research is a paradigm that tests objective theories by examining the relationship among variables or groups (Creswell & Creswell, 2018). Descriptive research explores the aspects of a specific variable without proposing relationships between variables, correlation, or cause and effect -- it focuses on describing or measuring a specific phenomenon or group (Newhart & Patten, 2023). With the reduction of geographical descriptions in research and policy down to a binary system, the unique characteristics of populations are not well known, and concepts vary greatly (Cromartie & Bucholtz, 2008; Wheatley, 2006). Describing the characteristics of populations as they apply to the concept of place represents a post-positivistic view of research, that there are limitations to objective knowledge, recognizing the importance of subjectivity, and that truths are imperfect and influenced by our values and experiences (Legg & Hookway, 2021). This research considers a quantitative methodology appropriate as synthesizing the vast population demographics and school districting information from each classification at the census block partition supplements the missing geographical attributes; statistically significant data collection can yield a good understanding of a phenomenon (Ramsay-Jordan et al., 2022).

To address the research questions, phase 1 models the conditions of the NCES classification codes at the census block scope, creating a replicable description of rural and urban derived from available, high-quality data (Hart et al., 2005). Phase 2 synthesizes demographic and school district information from the USCB, USDA, and NCES to capture an essence of place using the permutation feature importance technique as a foundation to present the statistically significant characteristics and differences within the NCES coding. Lastly, phase 2 applies the model to the United States Census Bureau's classification of census blocks within school districts' administrative areas to assess alignment.

Phase 1 Research Methodology

Arriving at a precise and meaningful description of rural is challenging, and its complexities are often overlooked (De Gennaro & Fantini, 2009; Kettler et al., 2016). Phase 1 of this research employs a descriptive, nomothetic methodology, deriving data from the NCES and

USCB for spatial analysis. This phase does not seek the relationship between the variables; it only seeks to understand the specific phenomena with controlled variables through a post-positivist perspective. Unlike True and Quasi-Experimental designs, where variables require treatment and interventions on the observed population, descriptive and correlational designs control their variables and do not have interventions or treatments (Creswell & Creswell, 2018). Other research designs are impractical as the condition of rurality is not universally defined, readily observable, and does not respond to experimental treatment (Selover, 2021). Furthermore, the analysis in this phase is geospatial; it does not fit neatly into a correlational or comparative paradigm as both constructs have defined outcomes that spatial analysis does not satisfy (de Smith et al., 2023). A quantitative methodology and descriptive design phase are necessary to develop a more comprehensive understanding of the characteristics of rural and urban areas.

Phase 1 Subject Selection and Description

The geographical extent of this phase includes the states of Wisconsin, Minnesota, Illinois, and Iowa and their entire population. This study does not conduct any sampling. The rationale supporting this geographical extent rests in the researcher's accessibility to this region and the diverse consensus, or lack thereof, on which areas are rural, as presented by Long et al., (2021). The locale definitions of the NCES serve as the classification system for census blocks. The NCES locales categorize U.S. territory into four types of areas: City, Suburban, Town, and Rural (National Center for Education Statistics, 2019). Each type of area contains three subtypes that can be fully collapsed into a fundamental urban-rural dichotomy or expanded into a more detailed collection of 12 distinct categories. Size differentiates a subtype in the case of city and suburban assignments, and proximity differentiates in the case of town and rural assignments

(National Center for Education Statistics, 2019). Appendix A presents the classification requirements for each locale and its categories.

Phase 1 Instrumentation

The data used in phase 1 is from publicly available secondary data from the National Center for Education Statistics and the United States Census Bureau. The National Center for Education Statistics collects information, in collaboration with the United States Census Bureau, through surveys using complex assessments, administrative sources, and samples of schools, institutions, and households (National Center for Education Statistics, 2023). The United States Census Bureau partitions statistical areas into census blocks bounded by visible features such as roads, streams, and railroad tracks and by nonvisible boundaries such as selected property lines and city, township, school district, and county limits and short line-of-sight extensions of streets and roads (United States Census Bureau, 2022). The National Center for Education Statistics divides areas based on population size or proximity to populated areas (National Center for Education Statistics, 2023).

The rural and urban structures of the National Center for Education Statistics and the United States Census Bureau differ. The Python programming language (version 3.7), an interpreted, high-level, general-purpose programming language, comes packaged in a free geographic information system that supports geospatial data analysis, QGIS version 3.3.0 “Hertogenbosch.” Python is utilized to align observations of the two geospatial systems (Python Software Foundation, n.d.).

The ArcGIS for Student Use Web API for Python is a web service offering maps, features, imagery, and analysis tools that can be retrieved by any device that can make network requests (Esri, n.d.). The ArcGIS API NCES Open Data “identify” service accepts a geometry “x” and “y” point object and returns the NCES locale information of the queried point. ArcGIS for Student Use is a proprietary GIS analysis software published by Esri and requires a yearly subscription (Esri, n.d.). This software was paid for and licensed by the researcher.

R, version 4.2.3, a programming language and environment for statistical computation and graphics, is utilized for statistical analysis and graphic generation (The R Foundation, n.d.). R is licensed as Free Software under the Free Software Foundation’s GNU General Public License terms in source code form (The R Foundation, n.d.). R is available to download through the Comprehensive R Archive Network as a precompiled binary distribution (The R Foundation, n.d.).

Phase 1 Data Collection Procedures

Phase 1 derives its data from the NCES and USCB. The NCES Education Demographic and Geographic Estimates (EDGE) program develops geographic data to help policymakers, program administrators, and the public understand the relationships between educational institutions and their communities – aligning with the objectives of this research (National Center for Education Statistics, 2019). Furthermore, data from the USCB is the most frequently cited information in geospatial research (Long et al., 2021). This research considers the data from NCES and USCB best suited for this phase due to information availability, institutional authority, and completeness.

For each State's 118th Congressional District administrative divisions, all legal boundaries and names are current as of January 1, 2020, except for the 118th Congressional Districts and State Legislative Districts, which are current as of January 1, 2023. The data for the administrative divisions was obtained through the USCB Census Mapping Files file transfer protocol archive on the bureau’s website in a Topologically Integrated Geographic Encoding and

Referencing (TIGER) and TIGER/Line in shapefile format. A shapefile, denoted with the extension .shp, is a nontopological format for storing geographic features' geometric location and attribute information (Esri Developer, n.d.). Points, lines, or polygons (areas) represent the geographic features in a shapefile.

A project containing shapefiles may also contain dBASE tables, denoted as .dbf, which contain additional attributes for shapefiles. The index file, denoted as .shx, links the attributes in the dBASE tables to the geographical features in a shapefile. For each census block, the geographical package includes GEOIDs linked to census demographic data: area (land), area (water), block name, number of houses, population, and the centroid of the census block polygon in longitude-latitude format.

This phase uses the Demographic Data download utility on the IES-NCES website to download each State's NCES locale boundaries shapefiles. The geographic codes for the NCES dataset come from TIGER 2021. Congressional districts in TIGER 2021 represent the 116th Congress. TIGER 2021 uses the OMB updates to metropolitan and micropolitan delineations as of March 2020. The population estimates represent the 2020 apportionment figures. Urban Areas represent the 2010 Census definitions in TIGER 2021. With all data sources for phase 1 either in a spatially native format or prepared for use in a spatial environment, all quantitative data was ready for analysis and required no modification.

Phase 1 Data Analysis

The spatial analysis in Phase 1 classifies a given census block according to the NCES codes using the desktop version of QGIS, version 3.3.0 “Hertogenbosch” (QGIS, 2023). QGIS operates in both desktop and cloud-based computing environments. QGIS was developed by the Open-Source Geospatial Foundation and licensed under GNU General Public License version 2 or any future revisions of the GNU license (QGIS, n.d.). Within a GIS, data layers can be displayed and built within an architecture that produces manipulable and stackable objects, which can then be analyzed (de Smith et al., 2023). In spatial analysis, a layer is a spatial dataset, usually a map, with data associated with the spatial area in an attribute table (de Smith et al., 2023).

Following the data validation process outlined by Selover (2021), QGIS opens each shapefile as a layer and confirms that all use the same datum and spheroid. A spheroid is a mathematical model approximating the earth's shape, and a datum defines the spheroid’s location relative to the earth's center (de Smith et al., 2023). This process ensured that all datasets aligned spatially, ensuring an accurate analysis (Selover, 2021). The datum used was North American Datum 1983, and the spheroid used was Geodetic Reference System 1980.

Using the Python programming language, QGIS’ API, and QGIS’ Software Development Kit (SDK), a script (see Appendix B) loads the NCES and USCB shapefiles as Vector Layers containing the census block and NCES features (polygons). The script then stacks the layers into a Map Layer, with the NCES layer being the first layer and the USCB layer being the second layer – allowing spatial boundary intersection calculations. For each polygon in the USCB shapefile, the script loads each polygon into a thread-safe queue. Thread-safe queues enable multiple instances of the script to read data from it, allowing the geospatial analysis to execute more efficiently and retry a task should one fail. Multiple instances of the script, “threads,” then pull the USCB polygons from a queue.

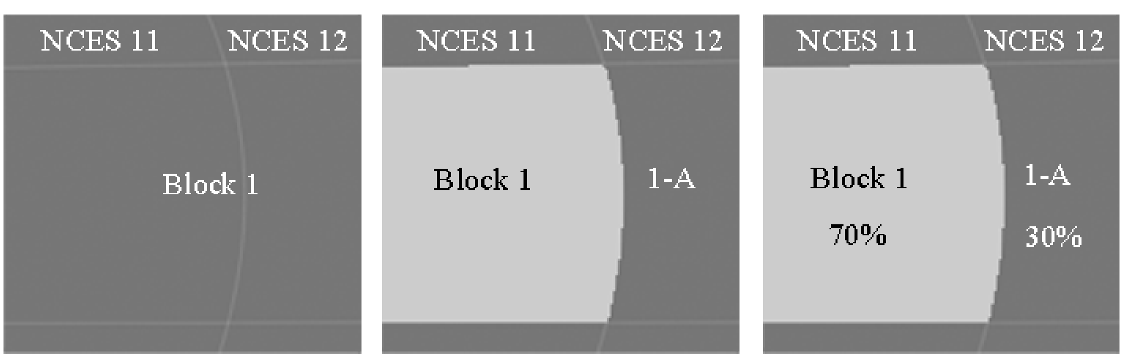
Each thread computes the centroid of the polygons by partitioning them into triangles. The weighted sum of each triangle's x and y coordinates determines the centroid of the entire polygon (QGIS, 2023). The script checks if the geometry of the centroid of each USCB polygon area is within any of the NCES polygon areas. A second script detects if the geometry of the USCB polygon crosses multiple NCES polygons (see Figure 1).

All code and data processing executes on eight Mac mini M1 computers, each running eight-core computer processing units and 64 gigabytes of memory. The Mac minis are clustered over an offline network and communicate over ethernet. One Mac mini generates a queue of all census blocks to be processed, and the remaining seven operate as consumers, computing the NCES locale intersections. The output of each intersection computation is written to a shared network drive and aggregated by the Python script once the queue is empty.

Lastly, a third script computes the percent area where the polygons intersect and creates a separate vector layer containing each intersection area's GEOID and LOCALE attributes. Three new attributes are added and set for each USCB feature. The "LOCALE" attribute from the NCES polygon in which the centroid of the USCB polygon lies "*centroid_nces_code*." For each NCES polygon the USCB polygon crosses, a comma-delimited list of each NCES "LOCALE" attribute "*nces_cross*." Lastly, the area percent for each polygon crosses "*nces_cross_pct*." Appendix B contains a sample of the Python code used for the geospatial analysis.

Figure 1

A Polygon "Block 1" Partitioned into NCES Geometries



The NCES's ArcGIS Identify API validates each census block's classification per its geometry centroid. The ArcGIS Identify API allows the Python programming language to query data via an object's geometry through an *x* and *y* geometry parameter, the longitude and latitude of the polygon's centroid, and retrieve the NCES code for the area. Due to the ArcGIS API limiting the number of requests per hour ("rate limiting"), the validation phase uses a random sample of 33% of the census blocks for each State with a random seed value of 42. A set seed value ensures that results are reproducible (Benureau & Rougier, 2018). The Python script queues the features, retrieves the classification code from the ArcGIS API response, and compares the results. Non-successful responses from the API due to request limiting were re-queued. The script then saves the resulting records' geographical unique identifier, "GEOID," and the NCES response code.

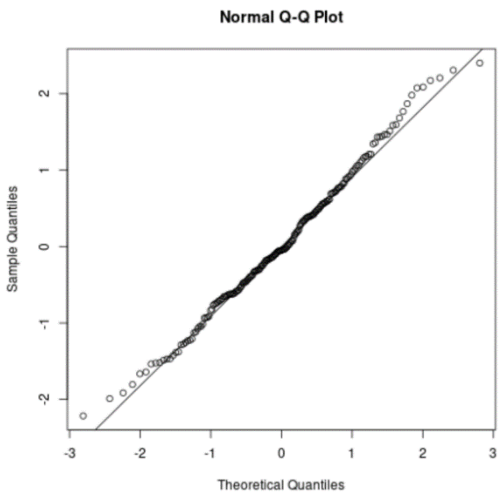
QGIS then loads the results from the geospatial analysis and exports the canvas as an image with the "canvas size" option selected. The feature tables are then exported to a Comma Separated Value (.csv) file to conduct a descriptive analysis of the data with the R programming language. The attributes in the census block layers include the state code "STATEFP20", county code "COUNTYFP20", the census tract "TRACTCE20", the census block "BLOCKCE20", the geographical unique identifier "GEOID20", urban-rural code as per the USCB "UR20", land area in square meters "ALAND20", water area in square meters "AWATER20", number of houses "HOUSING20", population "POP20", NCES locale boundary crossing list "nces_cross," and the census block centroid NCES code "tabblock_centroids_centroid_nces_code." For census blocks with multiple NCES codes, each layer export contains the geographical identifier "GEOID20", the NCES code "nces_code," and the percent overlap "overlap."

An R script parses and aggregates each .csv file into a single data frame from each shapefile. All rows in the block layers without an NCES code were reclassified as “NA” as the functional status codes of a census block, such as military, nonfunctional legal entities, and active state-recognized entities, are not evaluated by the NCES (National Center for Education Statistics, 2019). It then filters out Census blocks with an NCES boundary intersection and creates a new record containing the NCES code for every boundary crossing. The script modifies each new record’s land area, water area, housing, USCB urban-rural identification, and population attributes. The resulting observations are saved to a .csv for future processing in phase 2.

A Quartile-Quartile (Q-Q) scatter plot is generated from the resulting data frame to identify the data distribution. If both sets of quartiles came from the same distribution, the points would roughly follow the reference line. The further the points on the plot lie away from the reference line, the more substantial the evidence is against the data originating from a normal distribution. Figure 2 presents a normally distributed Q-Q. A Sharpio-Wilk test is conducted on a random sample of 5,000 observations with a significance (alpha) threshold of 0.05 and a random seed of 42 to determine what assumptions are present in the distribution of the NCES codes.

Figure 2

Q-Q Plot of a Normal Distribution



A Person’s test for kurtosis was conducted using R to describe the distribution's characteristics further. Kurtosis measures whether data is heavy-tailed or light-tailed relative to a normal distribution (National Institute of Standards and Technology, n.d.). A distribution with over-dispersed data relative to a normal distribution will have high kurtosis and tend to have heavy tails or outliers – a leptokurtic distribution (National Institute of Standards and Technology, n.d.). Under-dispersed data will have low kurtosis and tend to have a light tail or lack of outliers – a platykurtic distribution (National Institute of Standards and Technology, n.d.).

Additionally, the R script conducts a Hartigan Dip Test of Unimodality to determine if multimodality is present in the distribution. The dip test measures multimodality in a sample by the maximum difference, over all sample points, between the empirical distribution function and the unimodal distribution function that minimizes that maximum difference (Hartigan & Hartigan, 1985). The uniform distribution is the asymptotically least favorable unimodal distribution, and the distribution of the test statistic is determined asymptotically and empirically when sampling from the uniform. "[D] values less than 0.05 indicate significant bimodality and values greater than 0.05 but less than 0.10 suggest bimodality with marginal significance" (Hartigan & Hartigan, 1985). The null hypothesis (H_0) is that the distribution is unimodal. The alternative hypothesis (H_a) is that the distribution is non-unimodal and at least bimodal.

Lastly, a Mann-Whitney *U* test identifies whether the census blocks and responses from the NCES API originate from the same population. The Mann-Whitney *U* is also applied to the USCB classification of census blocks and the NCES classification to identify geospatial alignment. The null hypothesis (H_0) of the Mann-Whitney *U* test is that the two populations are equal. The alternative hypothesis (H_a) is that the two populations are different. These hypotheses represent a two-way test, as there is no assumption about one value being higher or lower than the other. The Mann-Whitney *U* test was selected as there is no dependence on normality and its ability to test ordinal data.

This phase models the conditions of the National Center for Education Statistics locale codes at the census block scope, creating a replicable description of rural and urban derived from available, high-quality data. The result is a verifiable geospatial alignment process that enables supplemental information to enhance the descriptions of Federal administrative structures. In the context of this phase, the NCES locale code is aligned with census blocks, initiating a comparative analysis of the rural and urban designation of places between the NCES and USCB. The geospatial alignment process, NCES locale designation, and the USCB rural-urban dichotomy are carried into phase 2.

Phase 2 Research Methodology

Ill-defined distinctions between rural and urban are the foundations of rural research and rural policy (Isserman, 2005). This phase applies a correlational quantitative design, seeking to identify the relationship between the variables and the NCES locale code for a given area and model the population using the information collected from the observed population. The quantitative methodology is appropriate as the initial statistical data for the NCES locales, when applied to census blocks and school districts, has not been collected.

Phase 2 Subject Selection and Description

The geographical extent for phase 2 maintains the same scope as phase 1: the states of Wisconsin, Minnesota, Illinois, and Iowa and all school districts they contain. School districts are geographic entities and single-purpose governmental units that operate schools and provide public educational services at the local level (National Center for Education Statistics, 2023). The NCES uses the school district boundaries to develop a broad collection of district-level demographic estimates from the Census Bureau’s American Community Survey. The USCB annually updates school district boundaries, names, local education agency codes, grade ranges, and school district levels based on information provided by state education officials (National Center for Education Statistics, 2023). The school district boundaries in this spatial analysis include unified public, secondary, and elementary school districts.

School demographic data originates from the American Community Survey Education Tabulation (ACS-ED). The ACS-ED is a custom tabulation of the ACS produced for the National Center of Education Statistics (NCES) by the U.S. Census Bureau (National Center for Education Statistics, 2023). The surveys provide a collection of social, economic, demographic, and housing characteristics for school systems, school-age children, and the parents of school-age children. In addition to focusing on school-age children, the ACS-ED provides enrollment iterations for children enrolled in public schools. The data profiles include percentages (along with associated margins of error) that allow for comparison of school district-level conditions across the U.S. All information the data profiles contain is in the public domain.

For feature selection in the classification model, classification and regression tree (CART) models effectively explain any variance in a response variable (Rigatti, 2017). CART models are non-parametric and output a decision tree for classification or regression. A CART algorithm selects the optimal decision threshold for a variable based on recursive partitioning, testing the potential value of different thresholds for each predictor, and implementing the most valuable until further splitting no longer improves discrimination (Rigatti, 2017). Combined with ensemble techniques, creating multiple models and combining them to improve accuracy, CART models discover meaningful interactions and non-linear effects of the predictors (Rigatti, 2017).

Phase 2 Instrumentation

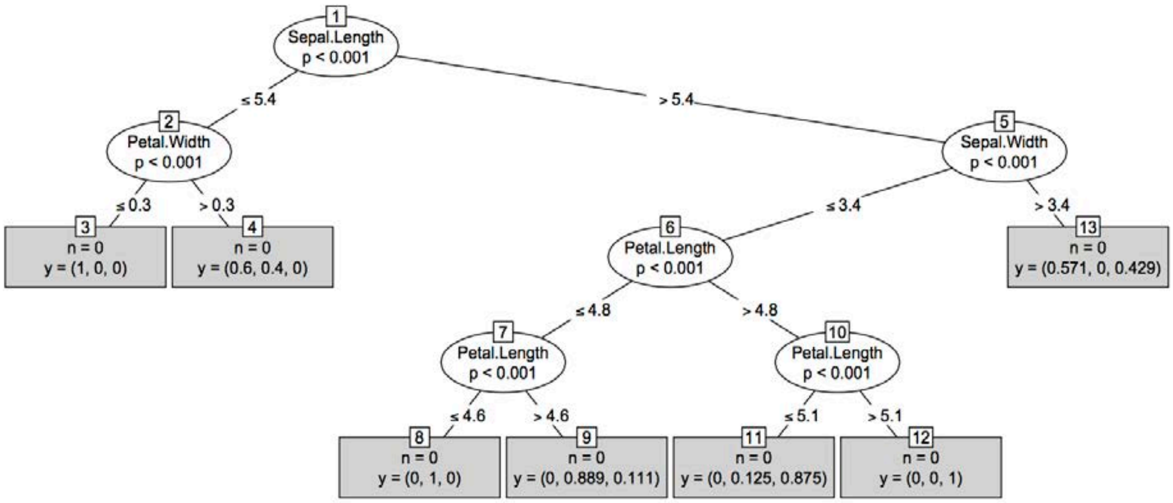
Phase 2 retrieves all secondary data from phase 1 and publicly available datasets through the NCES EDGE Open Data. For statistical analysis and modeling, R, version 4.2.3, a programming language and environment for statistical computation and graphics, loads the output from phase 1 (R Foundation, 2023). The Python programming language (version 3.7) aggregates information across the USCB, ACS-ED, and NCES for descriptive analysis and classification. A Python script uses the QGIS API and SDK to align census blocks to school districts and NCES locales to classify their boundaries. All code and subsequent data processing follow the same architecture as phase 1.

CART models are not generally up to the task of analyzing complex data with many interacting predictors (Rigatti, 2017). However, an assemblage of many such trees using different sets of predictors may, when aggregated, produce excellent prediction results. The R script uses Breiman and Cutler's Random Forests for Classification and Regression “randomForest” R library, version 4.7-1.1, configured to output conditional inference trees, and the Classification and Regression Training library, version 6.0-94, to select the features for the statistical model and address the limitations of single-tree CART models. CART models have two fundamental problems: overfitting and a selection bias towards covariates with many possible splits, leading to biased importance ranking (Hothorn & Zeileis, 2015). In non-conditional inference trees, to avoid overfitting, the trees created are pruned; however, maximizing a splitting criterion over all possible splits does not permanently eliminate bias (Hothorn et al., 2015). Conditional inference trees use a statistical approach that considers the distributional properties, measuring the association between responses and covariates (Hothorn et al., 2015).

Random forest is a tree-based classification and regression method comprising an ensemble of individual decision trees. The trees used as base learners in random forest can be of different types. A random forest model combines the output of multiple decision trees for classification and regression problems into a single result (IBM, 2023). Using the desktop version of QGIS, version 3.3.0 “Hertogenbosch,” a visual representation of the model's classifications is generated.

Figure 3

Random Forest Decision Tree Process



Random Forest uses randomization to create many decision tree predictors (Rigatti, 2017). The output of the trees is aggregated into a single output using voting for classification problems or averaging for regression problems. Random Forest applies randomization in two ways. First, it samples data with replacement "bagging," where there is no guarantee that every subject will appear in the new sample, and some may appear more than once. The second randomization occurs at the decision nodes. The algorithm chooses several predictors from the spatial analysis in phase 1 at each node. The decision tree then tests all possible thresholds for all selected variables and selects the variable-threshold combination that results in the best split. The random selection of variables and threshold testing continues until the algorithm encounters nodes containing only cases, controls, or other pre-defined endpoints to determine the classification of a census block (see Figure 3).

Phase 2 Data Collection Procedures

Phase 2 of this research derives data from the results of phase 1, the NCES elementary and unified school district boundary shapefiles, NCES school district characteristics, and the

ACS-ED datasets. The data for the school district boundaries was obtained through the USCB Census Mapping Files file transfer protocol archive on the USCB's website in a TIGER and TIGER/Line in shapefile format. The NCES EDGE, Open Data utility, hosts the ACES-ED and School District Characteristic datasets in .csv format. The ACES-ED datasets include housing, demographic, economic, and social characteristics for the entire school district population. All ACES-ED datasets are current as of February 18th, 2022. For each state's school district administrative division, all boundaries and names are as of January 1, 2020.

Phase 2 Data Analysis

The geospatial analysis of phase 2 uses the Python programming language and the QGIS SDK. A Python script loads the USCB school district and the resulting phase 1 shapefiles as Vector Layers containing the district boundaries and census block classification features (polygons). Following the same process as phase 1, the script stacks the layers into a single Map Layer, with the phase 1 layer being the first layer and the USCB school district boundary layer being the second. For each polygon in the phase 1 shapefiles, the Python script loads each polygon into a thread-safe queue. Threads pull from a queue and check if the geometry of the centroid of each phase 1 polygon area is within any of the school district polygon areas. The script also detects if the geometry of the polygon crosses multiple school district polygons; Figure 1 highlights this process. Any crossing computes the percent area where the polygons intersect and creates new feature attributes containing the percent intersection area "*d_pct*" and the school district GEOID comma-delimited list "*d_geoid.*" QGIS loads the updated phase 1 shapefiles and exports the feature tables to a Comma Separated Value (.csv) file to conduct a correlational data analysis with the R programming language. The attributes include those in the initial phase 1 export, the percent intersection area, and the school district GEOID comma-delimited list.

Table 1

Spatial Analysis Attributes

Attribute	Description	Example
GEOID20	Census block identifier	550899900000001
UR20	USCB urban/rural code	R
ALAND20	Area – land (meters)	32,767
AWATER20	Area – water (meters)	10,250
HOUSING20	Number of housing units	58
POP20	Population (2020 census)	159
nces_cross	NCES locales a census block intersects	41,42
d_geoid	School district GEOID	5511220
d_pct	Percent area of a census block within a district	0.33
centroid_nces_code	NCES code for the centroid of a census block	41
nces_cross_pct	Percent area of a census block with locales	0.6, 0.4

The ACES-ED social, economic, demographic, and housing characteristics for school systems, school-age children, and the parents of school-age children surveys account for all Wisconsin, Minnesota, Iowa, and Illinois public and private school districts. Features from the economic characteristics survey include employment status by sex, unemployment rate, method of transportation to commute to work, population working from home, occupation industry, income and benefits, public assistance income, health insurance coverage, families below the poverty level, and the number of families with children. Housing characteristics include property value, number of housing units with a mortgage, number of units paying rent, rent thresholds, average number of children in a household, number of vehicles, and number of housing units lacking telecommunications infrastructure. Social characteristics incorporate the number of married households, married households with children under 18, marriage status by sex, nursery school and preschool enrollment, kindergarten enrollment, elementary school enrollment, high school enrollment, college or graduate school enrollment, educational attainment, veteran status, residence status, computer access, and internet access. Lastly, demographic characteristics include the total population by sex, age, and race. Table 1 presents all the attributes used in the spatial analysis with a description and example.

An R script filters ACES-ED and NCES School Characteristics data by the State abbreviation column “STATEABBREV” for Wisconsin “WI,” Minnesota “MN,” Illinois “IL,” and Iowa “IA.” Additionally, the script filters out districts with no functional design and those labeled “School District Undefined.” The script then aggregates all datasets into a single data frame, including the school ranking data, linking records by their GEOID field – except for the NCES School Characteristics data. The script joins the NCES School Characteristics with the ACES-ED data on the “School district identifier” columns into a single data frame. Lastly, it joins the school district rankings on the GEOID field. The script then saves the data to a file in a comma-separated values .csv format.

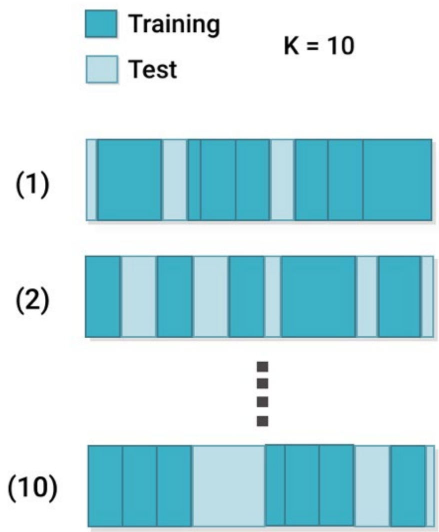
The R script imported the observations from phase 1, filtered out census blocks with a zero population, and removed census blocks without an NCES classification. It then links census blocks from the USCB to the characteristics of the school districts from the ACES-ED surveys. When linked to census blocks, the ACES-ED surveys' characteristics were unmodified if the school district contained the entire block. If the block resides in multiple school districts, the weighted average of the ACES-ED metrics was applied.

The script then conducts a descriptive assessment of the classification of school districts by the NCES and the locale classification of the census blocks they contain. A table in Appendix C presents the classification of each school district against the number of census block-locale observations. This table gives a high-level overview of the distribution of school districts, the areas they represent, and the alignment of the NCES classification. Appendix D lists the unmodified descriptive variables to preserve and present each variable's margin of error.

A Random Forest model is generated against all variables in the dataset, excluding population, to predict the school district locale classification of a given census block through the "randomForest" and "carat" libraries. The script improves and assesses the prediction power of the model by employing a K-fold cross-validation procedure. K-fold cross-validation splits the dataset into training and testing sets, the "holdout" method (Stone, 1976). In the holdout method, the model is trained on 70% of the training data and uses the test data to assess prediction power. K-fold validation splits the testing and training data into five *k*-subsets and repeats the holdout method *k*-times (see Figure 4). A sample of the R script is presented in Appendix E.

Figure 4

K-Fold Cross Validation Process



The script uses each *k*-subset as a test set, and the other *k*-1 subsets are combined to form a training set – computing the average error across all trials. The advantage of this method is the flexibility of how the data gets divided, ensuring every data subject gets tested at least once and is present in the training data *k*-1 times. This advantage comes at the cost of having to re-run random forest algorithm *k*-times, resulting in the computation taking *k*-times longer to make an evaluation.

The script then replaces the school district classification on the test data with the USCB classification. The new test data is then run through the model with the most significant NCES locale disagreements by subtracting the predicted locale code from the actual, identifying the "pockets" of homogenous communities described by Scannel and Gifford (2009). A confusion matrix, a table that presents the performance of an algorithm where each row of the matrix represents the instances in class while each column represents the instances in a predicted class, or vice versa, is generated to present potential misalignment between the United States Census Bureau classification of places and the NCES.

The Kappa statistic, a statistic that is used to measure inter-rater reliability, serves as a mechanism to assess the agreement of the classification of places according to the NCES and the model. The Kappa statistic is more robust than a percent agreement as it accounts for the possibility of agreement occurring by chance (Cohen, 1960). Cohen (1960) suggests the Kappa result be interpreted as follows: values ≤ 0 as indicating no agreement, 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement.

Lastly, a Mann-Whitney *U* test determines if the results of the school district classifications are significantly different from the predicted census block classifications. The null hypothesis (H_0) of the Mann-Whitney *U* test is that the two populations are equal. The alternative hypothesis (H_a) is that the two populations are different. These hypotheses represent a two-way test, as there is no assumption about one value being higher or lower than the other. The Mann-Whitney *U* test was selected as there is no dependence on normality and its ability to test ordinal data.

Limitations

The NCES derives its locale definition population thresholds from the USCB, where a locale is either urban or rural. Sourcing data from the USCB and the NCES is only as accurate as their sampling methods, with varying degrees of error. In addition to historical changes in urban definitions, differing sampling methods, and definitions of rurality, using data from the USCB includes added statistical noise due to rising privacy concerns – creating opportunities for error in cross-agency data linking (United States Census Bureau, n.d.). Despite these limitations, this research considers data from the USCB to be the best available information on rural and urban areas and sufficient to declare a complete population.

"Aligning Education to Geography" suggests a deficit perspective and possible researcher bias toward the spatial designations of school districts. At the time of this dissertation, the researcher lived in a rural (distant) area, according to the NCES, while employed in higher education and working remotely for a corporation in a large metropolitan area. Through their employment, the researcher has observed how corporate policy, initiatives, perceptions, and benefits underserve remote workers and rural employees. This concept expands to smaller educational institutions and educational policies intended to support their operations. Should these observations be invalidated in future research, it would suggest that the perceptions of rurality and urbanism no longer vary.

The geographical extent of this research is a small region of the United States and not representative of the nation. Though the findings of Childs et al. (2022) and Waldorf (2006) present some form of literature supporting the selection of the Midwest and Great Lakes areas as representative subjects, the reasoning behind the selection is one of the researcher's familiarities with the geographic bounds – not one of supporting literature. Additional forms of sub-regionality may be present and not represented in this research.

Lastly, as presented in the literature review, there are many definitions of rural and urban in policy development. In research, policy, and funding initiatives, the definition of rural and urban selected is whichever meets the endeavor's goals (Cromartie & Bucholtz, 2008). The definitions discussed in this research all serve or have served a purpose to meet the diverse nature of U.S. geography. The topic of this research is education, and the NCES is a Congressional mandate to collect, collate, analyze, and report complete statistics on the condition of education in the U.S. (National Center for Education Statistics, 2023). The context of education presents a limited perspective of the conditions of rural and urban areas and may not be applicable outside of it.

Chapter IV: Presentation of the Findings

The purpose of this sequential quantitative, descriptive, and correlational study is to mathematically capture an essence of place by applying administrative, land-use, economic concepts, and socioeconomic characteristics of each NCES locale classification using census and school district data from the United States Census Bureau and American Community Surveys to describe the existing conditions in which policy derives. The following research questions fulfill this purpose:

1. Research Question 1: What principles drive the definitions of rural and urban, and how do they align with the locale system for the National Center for Education Statistics?
2. Research Question 2: What characteristics define rurality and urbanism beyond population thresholds or distances from an urban core?
3. Research Question 3: How does the National Center of Education Statistics locale classification for school districts align with the classification of the places schools administrate?

This chapter presents the data generated by analyzing each phase and concludes with a summary of the findings.

Phase 1 Findings

Creating a reproducible and verifiable description of the National Center for Education Statistics locale system through high-quality and available data for analysis is the goal for phase 1 of this research. The geographic area of interest is constrained to Wisconsin, Minnesota, Iowa, and Illinois. Within the geographical information system, groups for each state contain the census block, NCES locale boundaries, centroid, and intersecting boundary layers. Of the

946,941 census blocks across the four States, 831,541 reside within a given NCES locale with no intersecting boundaries – 125,400 census blocks fell within multiple locales. With the census blocks falling into locales, they were broken into smaller blocks to fit within each locale and merged into the blocks list, resulting in 1,071,958 observations. The artifact in Appendix F presents a map of this intermediate phase.

For the State of Wisconsin, the script processed 203,059 features. Exactly 700 features, identified as census blocks with no land area and no population, had no intersections with any NCES locales and were left blank. Iowa had zero unclassified census blocks in its 175,199 features. Minnesota contains 198,705 features, with 247 falling outside NCES bounds. Of these blocks, 192 had no land area with no population. Illinois has 369,979 features, with 161 unclassified and 155 without land area or population.

The NCES is the United States' leading independent education research, evaluation, and statistics source (National Center for Education Statistics, 2019). It provides a foundation for understanding characteristics that define rurality and urbanism and how they apply to education (National Center for Education Statistics, 2019). Computing the spatial intersections of the NCES boundaries over census blocks and classifying blocks within their boundaries enables phase 2 to link attributes from other Federal agencies and further investigate the characteristics that define rurality and urbanism using the NCES locale codes. The Python script for the geospatial analysis made 312,671 requests to the ArcGIS API to assess the accuracy of the locale classification of census blocks. The sample sizes included 67,009 features for Wisconsin, 65,753 for Minnesota, 122,093 for Illinois, and 57,816 for Iowa. Fourteen features in Wisconsin mismatched the API response, seven in Illinois, and five in Minnesota.

After computing the spatial analysis and comparing a random sample with the NCES API response locales, the Mann-Whitney *U* test ($U = 4.88e10$, $p > 0.05$) suggests insufficient evidence to reject the null hypothesis. The spatial analysis recreates the boundaries of the NCES locales to use for a correlational analysis in Phase 2. Table 2 summarizes the locales, attributes available for each locale, and the percentage of census blocks within each locale with a zero population. Of the places classified as rural or urban by the USCB, 284,704 census block classifications are misaligned. The Mann-Whitney *U* test ($U = 7.82e11$, $p < 2.2e-16$) suggests sufficient evidence to reject the null hypothesis and that the binary classification of places by the USCB does not align with the binary classification of places by the NCES.

Table 2

Demographics for Each NCES Classification Code

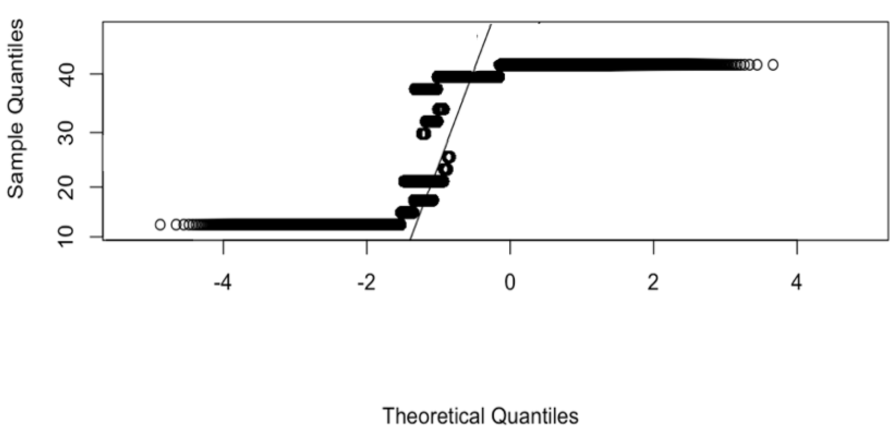
Code	*N*	μ Land area	μ Water area	μ Population	μ Housing	*0* Population %
11	62,235	20,644.58	748.23	69.59	31.68	16
12	27,293	44,743.55	955.27	47.73	20.91	23
13	55,988	44,854.81	1,403.11	47.60	20.79	23
21	163,111	59,614.97	2,496.86	53.24	21.24	22
22	22,777	56,830.72	1,352.41	38.08	16.32	26
23	16,029	52,409.03	1,104.73	29.06	12.86	25
31	22,144	52,049.03	2,101.18	32.67	13.82	17
32	70,440	37,737.32	860.11	25.59	11.54	18
33	35,747	36,328.17	1,931.17	22.83	10.90	18
41	169,709	533,436.55	20,277.28	13.36	5.66	35
42	282,280	1,085,437.19	19,605.63	9.69	4.53	31

Code	*N*	*μ* Land area	*μ* Water area	*μ* Population	*μ* Housing	*0* Population %
43	143,958	1,499,780.31	74,917.32	7.10	4.45	39
NA	247	154,084.81	1,083,222,343.24	0.49	0.75	96

The distribution of NCES observations exhibits a negative kurtosis of -1.094129, suggesting under-dispersed data following a platykurtic distribution. The Sharpio-Wilk test ($p < 0.05$) and the visual representation in Figure 5 further suggest that the NCES classification codes do not follow a normal distribution. Hartigan's Dip Test of Unimodality ($D = 0.079, p < 0.05$) rejects the null hypothesis that the distribution is unimodal and suggests it is either unimodal or bimodal. There is statistically significant evidence to suggest that two populations are present in the NCES observations. The dip statistic *D* falls between values greater than 0.05 but less than 0.10, suggesting bimodality with marginal significance.

Figure 5

The NCES Coding Distribution Q-Q Plot



Phase 2 Findings

Treating rural as a place and conducting a descriptive and correlational analysis to present the statistically significant characteristics and differences within the NCES locales are the primary goals of phase 2 in this research. Defining place in rurality is achieved by incorporating school district information from the American Community Survey Education Tabulation and selecting features presented in Scannell and Gifford's (2010) sense of place and place attachment framework and applying them to a Random Forest statistical model. The model highlights the statistically significant features that make up rural and urban areas according to the NCES locales, capturing an essence of place and offering insights into the geographical locations and socioeconomic characteristics of each NCES locale. A dialog can be initiated from the model's output for future policy implementation recommendations.

Locale Profiles

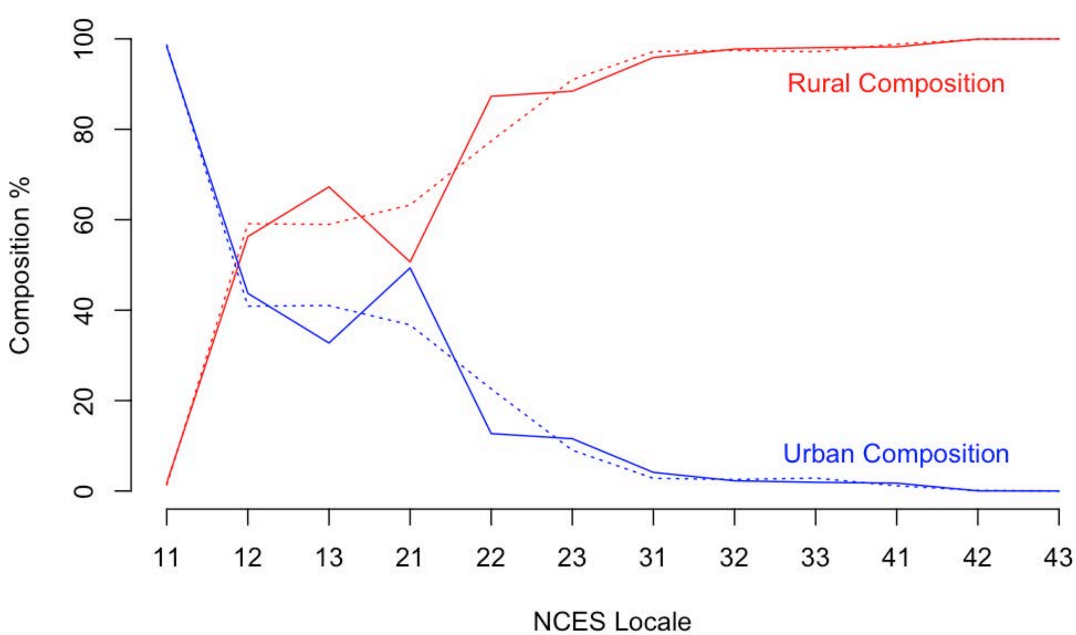
The geographic area of interest is constrained to Wisconsin, Minnesota, Iowa, and Illinois in the descriptive analysis. Within the geographical information system, groups for each state contain the census block, NCES locale boundaries, intersecting boundaries, and school district boundary layers. Approximately 10.52 gigabytes of data containing 1,852 variables across the American Community Surveys and United States Census Bureau were analyzed. The American Community Surveys consist of a 0.9% oversample rate, 27,748,760 people, compared to the 27,515,031 people surveyed in the 2020 United States Census data. The geospatial processing aligning school districts to census blocks took 13 hours with zero errors.

Of the 1,926 school districts, five fall within NCES locale 11, 11 in locale 12, 55 in locale 13, 406 in locale 21, 28 in locale 22, 34 in locale 23, 92 in locale 31, 220 in locale 32, locale in locale 33, 185 in locale 41, 519 in locale 42, and 276 in locale 43. After filtering out census blocks with no intersections with school district boundaries, 945,196 observations were recorded. Figure 5 presents the distribution of school district classifications and the classification of the places they administrate, reflecting a form of multimodality as per the results of Hartigan's Dip Test of Unimodality ($D = 0.079, p < 0.05$) in phase 1.

The descriptive analysis of NCES Locales identified that 98.23% of the places administrated by school districts classified as locale 11 are considered urban by the United States Census Bureau—the urban composition in locales 12 through 21 ranges between 32.73% and 49.34%. At locale 22, the composition drops to 12.7%, 4% in locale 31, and less than 2% in locale 33. Figure 6 depicts the rural-urban composition trend of school districts by NCES locale classification. Appendix C describes the composition of school districts by locale classification. Appendix D presents the unmodified variables and each variable's margin of error.

Figure 6

School District Rural-Urban Composition by NCES Locale Classification



Large Cities. School districts within NCES locale 11 consist of 95.6% of places classified as locale 11 by the United States Census Bureau, 4.34% as locale 21, and 0.0656% as locale 13. Large Cities operate 1,114 education centers, servicing 518,842 students, and maintain a 14-to-1 student-to-teacher ratio. All school district records for locale 11 are complete with no missing data.

Large Cities encompass 3,475,872 people residing within 1,551,301 homes with a median value of $208,013.00. 2.56% of homes lack telecommunication services, 0.98% lack adequate kitchen facilities, and 0.20% do not have complete plumbing services. Households consist of an average of 2.61 residents per home for homeowners and 2.02 for renters. 5.93% of people relocated to their current homes between 2015 and 2018. 15.36% moved between 2010 and 2014, 32.89% between 2000 and 2009, 22.62% from 1990 to 1999, 11.06% from 1980 to 1989, and 12.13% before 1979.

Residents of school districts in Large Cities earn a median income of $74,084.15. 54.12% of the population is in the civilian workforce, and 7.36% are unemployed. 59.76% of the workforce drives to work, 20.63% utilizes public transportation, 6.28% walks, 8.59% carpools, and 4.74% works from home. The average commute is 24.94 minutes. 20.87% of residents are employed in Management and Business occupations. Education employs 12.62% of residents; sales, 11.27%; Service, 10.05%; Production, Retail, and Manufacturing, 6.37%; Arts, 5.71%; Waste Management, 7.47%; Finance, 3.94%; 1.98% in Public Services, and 0.14% in Agriculture. A high school diploma is the highest academic attainment for 26.22% of the population. 21.26% have attended some college. 7.54% obtained an associate degree, 26.88% an undergraduate, and 18.10% have a graduate degree.

The median age of those over 18 in Large Cities is 33. 9.52% of the population is married, and 23.43% of families live in poverty with children. 1.50% of families receive public cash assistance, and 7.29% of the population receives SNAP benefits. 9.59% of families speak a language other than English at home, and 88.05% have healthcare coverage. 49.19% of residents identify as White; 23.89% African American or Black; 0.43% Native American or Alaska Native; 6.30% Asian; and 20.20% Hispanic or Latino.

Mid-Size Cities. School districts within NCES locale 12 consist of 80.6% of places classified as locale 12 by the United States Census Bureau, 11.5% as locale 21, 3.45% as locale 22, 3.2% as locale 41, 0.827% as locale 31, 0.27% as locale 42, and 0.167% as locale 13. Mid-sized Cities operate 598 education centers, servicing 292,345 students, and maintain a 15-to-1 student-to-teacher ratio. All school district records for locale 12 are complete with no missing data.

Mid-sized Cities encompass 2,100,576 people residing within 875,930 homes with a median value of $169,630.00. 2.26% of homes lack telecommunication services, 0.70% lack adequate kitchen facilities, and 0.033% do not have complete plumbing services. Households consist of an average of 2.66 residents per home for homeowners and 2.06 for renters. 4.77% of people relocated to their current homes between 2015 and 2018. 12.14% moved between 2010 and 2014, 28.04% between 2000 and 2009, 26.74% from 1990 to 1999, 14.33% from 1980 to 1989, and 13.97% before 1979.

Residents of school districts in Mid-sized Cities earn a median income of $78,270.32. 53.22% of the population is in the civilian workforce, and 6.80% are unemployed. 81.48% of the workforce drives to work, 3.70% utilizes public transportation, 2.00% walks, 8.48% carpools, and 4.34% works from home. The average commute is 20.94 minutes. 19.73% of residents are employed in Management and Business occupations. Education employs 12.10% of residents; sales, 12.44%; Service, 8.55%; Production, Retail, and Manufacturing, 6.85%; Arts, 4.50%; Waste Management, 5.53%; Finance, 3.99%; 1.77% in Public Services, and 0.35% in Agriculture. A high school diploma is the highest academic attainment for 27.94% of the population. 23.04% have attended some college. 10.20% obtained an associate degree, 24.51% an undergraduate, and 14.31% have a graduate degree.

The median age of those over 18 in Mid-sized Cities is 38. 11.25% of the population is married, and 16.61% of families live in poverty with children. 1.01% of families receive public cash assistance, and 4.81% of the population receives SNAP benefits. 10.57% of families speak a language other than English at home, and 92.07% have healthcare coverage. 73.70% of residents identify as White; 10.07% African American or Black; 0.44% Native American or Alaska Native; 4.88% Asian; and 10.90% Hispanic or Latino.

Small Cities. School districts within NCES locale 13 consist of 88% of places classified as locale 13 by the United States Census Bureau, 7.27% as locale 21, 2.52% as locale 22, 0.898% as locale 41, 0.754% as locale 23, 0.434% as locale 42, 0.118% as locale 12, 0.00255% as locale 11, and 0.00031% as locale 31. Small Cities operate 905 education centers, servicing 442,470 students, and maintain a 14-to-1 student-to-teacher ratio. All school district records for locale 13 are complete with no missing data.

Small Cities encompass 3,368,293 people residing within 1,418,512 homes with a median value of $198,269.00. 1.83% of homes lack telecommunication services, 0.74% lack adequate kitchen facilities, and 0.12% do not have complete plumbing services. Households consist of an average of 2.57 residents per home for homeowners and 2.16 for renters. 5.49% of people relocated to their current homes between 2015 and 2018. 13.50% moved between 2010 and 2014, 28.32% between 2000 and 2009, 25.47% from 1990 to 1999, 13.35% from 1980 to 1989, and 13.87% before 1979.

Residents of school districts in Small Cities earn a median income of $81,649.78. 54.00% of the population is in the civilian workforce, and 5.48% are unemployed. 80.43% of the workforce drives to work, 3.63% utilizes public transportation, 3.42% walks, 8.19% carpools, and 4.33% works from home. The average commute is 21.01 minutes. 19.97% of residents are employed in Management and Business occupations. Education employs 12.85% of residents; sales, 12.17%; Service, 8.69%; Production, Retail, and Manufacturing, 7.09%; Arts, 4.76%; Waste Management, 5.13%; Finance, 3.76%; 1.42% in Public Services, and 0.36% in Agriculture. A high school diploma is the highest academic attainment for 26.73% of the population. 22.25% have attended some college. 10.28% obtained an associate degree, 25.20% an undergraduate, and 15.54% have a graduate degree.

The median age of those over 18 in Small Cities is 38. 10.61% of the population is married, and 14.37% of families live in poverty with children. 0.99% of families receive public cash assistance, and 4.28% of the population receives SNAP benefits. 9.64% of families speak a language other than English at home, and 92.33% have healthcare coverage. 76.66% of residents identify as White; 7.89% African American or Black; 0.41% Native American or Alaska Native; 6.07% Asian; and 8.97% Hispanic or Latino.

Large Suburbs. School districts within NCES locale 21 consist of 90.7% of places classified as locale 21 by the United States Census Bureau, 3.49% as locale 41, 3.06% as locale 13, 1.45% as locale 12, 0.853% as locale 31, 0.262% as locale 11, 0.139% as locale 23, and 0.0339% as locale 42. Large Suburbs operate 2,211 education centers, servicing 1,112,360 students, and maintain a 14-to-1 student-to-teacher ratio. The descriptive analysis excludes two records (*NA*) missing teacher information.

Large Suburbs encompass 6,169,333 people residing within 2,210,323 homes with a median value of $241,527.00. 1.79% of homes lack telecommunication services, 0.61% lack adequate kitchen facilities, and 0.15% do not have complete plumbing services. Households consist of an average of 2.81 residents per home for homeowners and 2.39 for renters. 4.17% of people relocated to their current homes between 2015 and 2018. 10.98% moved between 2010 and 2014, 26.48% between 2000 and 2009, 28.50% from 1990 to 1999, 15.61% from 1980 to 1989, and 14.26% before 1979.

Residents of school districts in Large Suburbs earn a median income of $96,165.58. 53.84% of the population is in the civilian workforce, and 6.37% are unemployed. 81.25% of the workforce drives to work, 4.63% utilizes public transportation, 1.51% walks, 7.75% carpools, and 4.86% works from home. The average commute is 27.53 minutes. 20.78% of residents are employed in Management and Business occupations. Education employs 11.29% of residents; sales, 12.67%; Service, 7.81%; Production, Retail, and Manufacturing, 6.42%; Arts, 4.23%; Waste Management, 6.11%; Finance, 4.21%; 1.66% in Public Services, and 0.26% in Agriculture. A high school diploma is the highest academic attainment for 25.43% of the population. 22.44% have attended some college. 9.72% obtained an associate degree, 26.71% an undergraduate, and 15.69% have a graduate degree.

The median age of those over 18 in Large Suburbs is 41. 11.55% of the population is married, and 11.29% of families live in poverty with children. 0.79% of families receive public cash assistance, and 3.30% of the population receives SNAP benefits. 10.85% of families speak a language other than English at home, and 92.13% have healthcare coverage. 72.05% of residents identify as White; 8.99% African American or Black; 0.28% Native American or Alaska Native; 5.48% Asian; and 13.20% Hispanic or Latino.

Mid-size Suburbs. School districts within NCES locale 22 consist of 58.4% of places classified as locale 22 by the United States Census Bureau, 24.6% as locale 13, 9.33% as locale 12, 6.29% as locale 41, 1.09% as locale 31, 0.125% as locale 42, 0.105% as locale 21, and 0.0133% as locale 32. Mid-size Suburbs operate 419 education centers, servicing 201,542 students, and maintain a 14-to-1 student-to-teacher ratio. All school district records for locale 22 are complete with no missing data.

Mid-size Suburbs encompass 1,409,110 people residing within 600,551 homes with a median value of $180,191.00. 1.97% of homes lack telecommunication services, 0.56% lack adequate kitchen facilities, and 0.11% do not have complete plumbing services. Households consist of an average of 2.61 residents per home for homeowners and 2.09 for renters. 5.38% of people relocated to their current homes between 2015 and 2018. 13.52% moved between 2010 and 2014, 28.55% between 2000 and 2009, 25.88% from 1990 to 1999, 13.15% from 1980 to 1989, and 13.53% before 1979.

Residents of school districts in Mid-size Suburbs earn a median income of $80,418.88. 53.93% of the population is in the civilian workforce, and 4.65% are unemployed. 82.12% of the workforce drives to work, 2.40% utilizes public transportation, 3.15% walks, 8.12% carpools, and 4.21% works from home. The average commute is 20.14 minutes. 19.48% of residents are employed in Management and Business occupations. Education employs 12.97% of residents; sales, 12.19%; Service, 8.71%; Production, Retail, and Manufacturing, 7.17%; Arts, 4.60%; Waste Management, 4.57%; Finance, 3.90%; 1.65% in Public Services, and 0.51% in Agriculture. A high school diploma is the highest academic attainment for 28.51% of the population. 22.44% have attended some college. 11.38% obtained an associate degree, 24.02% an undergraduate, and 13.66% have a graduate degree.

The median age of those over 18 in Mid-size Suburbs is 39. 10.96% of the population is married, and 11.96% of families live in poverty with children. 0.94% of families receive public cash assistance, and 4.41% of the population receives SNAP benefits. 9.97% of families speak a language other than English at home, and 92.71% have healthcare coverage. 81.43% of residents identify as White; 6.44% African American or Black; 0.55% Native American or Alaska Native; 4.66% Asian; and 6.92% Hispanic or Latino.

Small Suburbs. School districts within NCES locale 23 consist of 66.4% of places classified as locale 23 by the United States Census Bureau, 24.5% as locale 13, 5.28% as locale 41, 3.13% as locale 42, 0.561% as locale 21, and 0.159% as locale 31. Small Suburbs operate 381 education centers, servicing 168,328 students, and maintain a 14-to-1 student-to-teacher ratio. The descriptive analysis excludes one record (*NA*) missing student and teacher information.

Small Suburbs encompass 1,238,843 people residing within 532,796 homes with a median value of $163,760.00. 1.88% of homes lack telecommunication services, 0.80% lack adequate kitchen facilities, and 0.15% do not have complete plumbing services. Households consist of an average of 2.67 residents per home for homeowners and 2.17 for renters. 5.21% of people relocated to their current homes between 2015 and 2018. 12.86% moved between 2010 and 2014, 27.34% between 2000 and 2009, 26.28% from 1990 to 1999, 13.62% from 1980 to 1989, and 14.69% before 1979.

Residents of school districts in small suburban places earn a median income of $77,321.34. 53.52% of the population is in the civilian workforce, and 5.58% are unemployed. 81.85% of the workforce drives to work, 2.59% utilizes public transportation, 3.27% walks, 8.11% carpools, and 4.19% works from home. The average commute is 21.83 minutes. 18.93% of residents are employed in Management and Business occupations. Education employs 12.76% of residents; sales, 12.08%; Service, 8.78%; Production, Retail, and Manufacturing, 7.59%; Arts, 4.61%; Waste Management, 4.46%; Finance, 3.39%; 1.57% in Public Services, and 0.71% in Agriculture. A high school diploma is the highest academic attainment for 29.78% of the population. 23.29% have attended some college. 11.05% obtained an associate degree, 22.62% an undergraduate, and 13.25% have a graduate degree.

The median age of those over 18 in Small Suburbs is 40. 10.61% of the population is married, and 12.97% of families live in poverty with children. 0.94% of families receive public cash assistance, and 4.38% of the population receives SNAP benefits. 9.83% of families speak a language other than English at home, and 92.67% have healthcare coverage. 81.81% of residents identify as White; 6.40% African American or Black; 0.41% Native American or Alaska Native; 4.47% Asian; and 6.92% Hispanic or Latino.

Fringe Towns. School districts within NCES locale 31 consist of 84.1% of places classified as locale 31 by the United States Census Bureau, 10.7% as locale 41, 4.35% as locale 21, 0.593% as locale 42, 0.178% as locale 32, 0.11% as locale 13. Fringe Towns operate 515 education centers that serve 220,657 students while maintaining a 15-to-1 student-to-teacher ratio. The descriptive analysis excludes three records (*NA*) missing student and teacher information.

Fringe Towns encompass 1,441,264 people residing within 589,585 homes with a median value of $191,823.00. 1.31% of homes lack telecommunication services, 0.59% lack adequate kitchen facilities, and 0.22% do not have complete plumbing services. Households consist of an average of 2.76 residents per home for homeowners and 2.24 for renters. 4.01% of people relocated to their current homes between 2015 and 2018. 10.48% moved between 2010 and 2014, 25.25% between 2000 and 2009, 29.06% from 1990 to 1999, 15.82% from 1980 to 1989, and 15.38% before 1979.

Residents of school districts in Fringe Towns earn a median income of $84,246.01. 53.60% of the population is in the civilian workforce, and 4.61% are unemployed. 83.47% of the workforce drives to work, 1.78% utilizes public transportation, 1.86% walks, 7.83% carpools, and 5.05% works from home. The average commute is 25.40 minutes. 18.84% of residents are employed in Management and Business occupations. Education employs 11.80% of residents; sales, 12.09%; Service, 8.01%; Production, Retail, and Manufacturing, 7.32%; Arts, 3.89%; Waste Management, 4.60%; Finance, 3.51%; 1.93% in Public Services, and 0.98% in Agriculture. A high school diploma is the highest academic attainment for 31.13% of the population. 23.76% have attended some college. 11.77% obtained an associate degree, 22.01% an undergraduate, and 11.33% have a graduate degree.

The median age of those over 18 in Fringe Towns is 41. 11.65% of the population is married, and 9.78% of families live in poverty with children. 0.74% of families receive public cash assistance, and 3.14% of the population receives SNAP benefits. 10.94% of families speak a language other than English at home, and 93.43% have healthcare coverage. 86.77% of residents identify as White; 3.70% African American or Black; 0.50% Native American or Alaska Native; 2.29% Asian; and 6.74% Hispanic or Latino.

Distant Towns. School districts within NCES locale 32 consist of 87% of places classified as locale 32 by the United States Census Bureau, 11.5% as locale 41, 1.29% as locale 42, and 0.192% as locale 31. Distant Towns operate 915 education centers, servicing 323,904 students while maintaining a 14-to-1 student-to-teacher ratio. All school district records for locale 32 are complete with no missing data.

Distant Towns encompass 2,298,704 people residing within 1,042,656 homes with a median value of 139,809.00. 2.00% of homes lack telecommunication services, 0.87% lack adequate kitchen facilities, and 0.25% do not have complete plumbing services. Households consist of an average of 2.40 residents per home for homeowners and 2.10 for renters. 4.33% of people relocated to their current homes between 2015 and 2018. 10.85% moved between 2010 and 2014, 25.80% between 2000 and 2009, 26.41% from 1990 to 1999, 14.93% from 1980 to 1989, and 17.68% before 1979.

Residents of school districts in Distant Towns earn a median income of $69,539.86. 51.55% of the population is in the civilian workforce, and 4.82% are unemployed. 82.58% of the workforce drives to work, 0.53% utilizes public transportation, 3.30% walks, 9.03% carpools, and 4.56% works from home. The average commute is 21.84 minutes. 15.83% of residents are employed in Management and Business occupations. Education employs 12.16% of residents; sales, 11.21%; Service, 9.02%; Production, Retail, and Manufacturing, 9.77%; Arts, 4.03%; Waste Management, 3.14%; Finance, 2.56%; 1.92% in Public Services, and 2.02% in Agriculture. A high school diploma is the highest academic attainment for 38.85% of the population. 24.60% have attended some college. 12.41% obtained an associate degree, 16.28% an undergraduate, and 7.86% have a graduate degree.

The median age of those over 18 in Distant Towns is 42. 10.63% of the population is married, and 14.11% of families live in poverty with children. 0.90% of families receive public cash assistance, and 4.75% of the population receives SNAP benefits. 9.79% of families speak a language other than English at home, and 91.37% have healthcare coverage. 90.43% of residents identify as White; 2.46% African American or Black; 0.48% Native American or Alaska Native; 1.15% Asian; and 5.48% Hispanic or Latino.

Remote Towns. School districts within NCES locale 33 consist of 79.1% of places classified as locale 33 by the United States Census Bureau, 20.8% as locale 41, 0.0884% as locale 42, and 0.075% as locale 43. Remote Towns operate 433 education centers, servicing 159,661 students while maintaining a 14-to-1 student-to-teacher ratio. All school district records for locale 33 are complete with no missing data.

Remote Towns encompass 1,080,729 people residing within 519,959 homes with a median value of $129,863.00. 2.11% of homes lack telecommunication services, 0.85% lack adequate kitchen facilities, and 0.31% do not have complete plumbing services. Households consist of an average of 2.30 residents per home for homeowners and 2.11 for renters. 4.33% of people relocated to their current homes between 2015 and 2018. 10.81% moved between 2010 and 2014, 26.07% between 2000 and 2009, 26.70% from 1990 to 1999, 14.58% from 1980 to 1989, and 17.52% before 1979.

Residents of school districts in Remote Towns earn a median income of $67,623.65. 50.98% of the population is in the civilian workforce, and 4.91% are unemployed. 82.18% of the workforce drives to work, 0.68% utilizes public transportation, 3.31% walks, 9.11% carpools, and 4.73% works from home. The average commute is 18.67 minutes. 16.24% of residents are employed in Management and Business occupations. Education employs 12.46% of residents; sales, 11.34%; Service, 8.97%; Production, Retail, and Manufacturing, 9.26%; Arts, 3.92%; Waste Management, 3.06%; Finance, 2.67%; 1.88% in Public Services, and 2.07% in Agriculture. A high school diploma is the highest academic attainment for 36.03% of the population. 24.87% have attended some college. 13.42% obtained an associate degree, 17.45% an undergraduate, and 8.23% have a graduate degree.

The median age of those over 18 in Remote Towns is 43. 10.66% of the population is married, and 16.27% of families live in poverty with children. 1.11% of families receive public cash assistance, and 4.83% of the population receives SNAP benefits. 9.96% of families speak a language other than English at home, and 91.87% have healthcare coverage. 90.16% of residents identify as White; 2.46% African American or Black; 0.96% Native American or Alaska Native; 1.48% Asian; and 4.94% Hispanic or Latino.

Fringe - Rural. School districts within NCES locale 41 consist of 24.4% of places classified as locale 32 by the United States Census Bureau, 18.4% as locale 41, 12.6% as locale 31, 10.3% as locale 33, 9.79% as locale 42, 9.44% as locale 21, 5.28% as locale 13, 4.27% as locale 22, 3.7% as locale 23, 1.52% as locale 12, 0.199% as locale 43, and 0.111% as locale 11. Fringe-Rural places operate 852 education centers, servicing 331,745 students, and maintain a 14-to-1 student-to-teacher ratio. The descriptive analysis excludes two records (*NA*) missing student and teacher information.

Fringe-Rural places encompass 2,329,537 people residing within 1,001,362 homes with a median value of $164,270.00. 1.86% of homes lack telecommunication services, 0.73% lack adequate kitchen facilities, and 0.25% do not have complete plumbing services. Households consist of an average of 2.54 residents per home for homeowners and 2.18 for renters. 4.24% of people relocated to their current homes between 2015 and 2018. 10.93% moved between 2010 and 2014, 25.89% between 2000 and 2009, 27.50% from 1990 to 1999, 15.22% from 1980 to 1989, and 16.23% before 1979.

Residents of school districts in Fringe-Rural places earn a median income of $76,723.80. 52.75% of the population is in the civilian workforce, and 4.97% are unemployed. 82.12% of the workforce drives to work, 2.32% utilizes public transportation, 2.48% walks, 8.29% carpools, and 4.80% works from home. The average commute is 22.81 minutes. 18.34% of residents are employed in Management and Business occupations. Education employs 11.87% of residents; sales, 11.89%; Service, 8.36%; Production, Retail, and Manufacturing, 7.99%; Arts, 4.05%; Waste Management, 4.46%; Finance, 3.38%; 1.80% in Public Services, and 1.28% in Agriculture. A high school diploma is the highest academic attainment for 32.28% of the population. 23.56% have attended some college. 11.44% obtained an associate degree, 21.28% an undergraduate, and 11.44% have a graduate degree.

The median age of those over 18 in Fringe-Rural places is 42. 11.09% of the population is married, and 12.81% of families live in poverty with children. 0.85% of families receive public cash assistance, and 3.94% of the population receives SNAP benefits. 10.36% of families speak a language other than English at home, and 92.17% have healthcare coverage. 83.06% of residents identify as White; 5.16% African American or Black; 0.47% Native American or Alaska Native; 3.15% Asian; and 8.15% Hispanic or Latino.

Distant - Rural. School districts within NCES locale 42 consist of 45.5% of places classified as locale 42 by the United States Census Bureau, 19.7% as locale 32, 11% as locale 41, 10.6% as locale 33, 8.98% as locale 43, 2.46% as locale 31, 0.674% as locale 22, 0.527% as locale 13, 0.326% as locale 23, 0.172% as locale 21, and 0.0591% as locale 12. Distant-Rural places operate 972 education centers, servicing 254,675 students while maintaining a 13-to-1 student-to-teacher ratio. All school district records for locale 42 are complete with no missing data.

Distant-Rural places encompass 1,705,433 people residing within 810,194 homes with a median value of $138,641.00. 1.95% of homes lack telecommunication services, 0.75% lack adequate kitchen facilities, and 0.39% do not have complete plumbing services. Households consist of an average of 2.43 residents per home for homeowners and 2.18 for renters. 3.62% of people relocated to their current homes between 2015 and 2018. 9.58% moved between 2010 and 2014, 23.57% between 2000 and 2009, 27.46% from 1990 to 1999, 16.03% from 1980 to 1989, and 19.74% before 1979.

Residents of school districts in Distant-Rural places earn a median income of $70,835.89. 51.60% of the population is in the civilian workforce, and 4.72% are unemployed. 82.51% of the workforce drives to work, 0.51% utilizes public transportation, 2.99% walks, 8.70% carpools, and 5.29% works from home. The average commute is 23.37 minutes. 16.09% of residents are employed in Management and Business occupations. Education employs 11.74% of residents; sales, 11.01%; Service, 8.45%; Production, Retail, and Manufacturing, 9.51%; Arts, 3.48%; Waste Management, 3.01%; Finance, 2.63%; 1.94% in Public Services, and 2.91% in Agriculture. A high school diploma is the highest academic attainment for 39.19% of the population. 24.51% have attended some college. 12.96% obtained an associate degree, 16.10% an undergraduate, and 7.24% have a graduate degree.

The median age of those over 18 in Distant-Rural places is 43. 10.72% of the population is married, and 13.11% of families live in poverty with children. 0.84% of families receive public cash assistance, and 4.08% of the population receives SNAP benefits. 10.22% of families speak a language other than English at home, and 92.26% have healthcare coverage. 92.67% of residents identify as White; 1.67% African American or Black; 0.73% Native American or Alaska Native; 0.92% Asian; and 4.01% Hispanic or Latino.

Remote - Rural. NCES locale 43 consists of 70.5% of places classified as locale 43 by the United States Census Bureau, 18% as locale 42, 6.4% as locale 41, 3.75% as locale 33, and 1.3% as locale 32. Remote-Rural places operate 814 education centers, servicing 170,121 students while maintaining a 12-to-1 student-to-teacher ratio. The descriptive analysis excludes two records (*NA*) missing teacher information.

Remote-Rural places encompass 1,131,066 people residing within 643,178 homes with a median value of $124,774.00. 2.27% of homes lack telecommunication services, 0.93% lack adequate kitchen facilities, and 0.68% do not have complete plumbing services. Households consist of an average of 2.34 residents per home for homeowners and 2.14 for renters. 3.38% of people relocated to their current homes between 2015 and 2018. 8.81% moved between 2010 and 2014, 22.49% between 2000 and 2009, 27.88% from 1990 to 1999, 16.73% from 1980 to 1989, and 20.72% before 1979.

Residents of school districts in Remote-Rural places earn a median income of $65,825.09. 50.05% of the population is in the civilian workforce, and 4.86% are unemployed. 80.89% of the workforce drives to work, 0.44% utilizes public transportation, 3.38% walks, 9.01% carpools, and 6.28% works from home. The average commute is 23.52 minutes. 15.77% of residents are employed in Management and Business occupations. Education employs 11.39% of residents; sales, 10.59%; Service, 8.55%; Production, Retail, and Manufacturing, 9.49%; Arts, 3.59%; Waste Management, 2.59%; Finance, 2.41%; 1.96% in Public Services, and 4.05% in Agriculture. A high school diploma is the highest academic attainment for 40.81% of the population. 24.78% have attended some college. 13.06% obtained an associate degree, 15.04% an undergraduate, and 6.30% have a graduate degree.

The median age of those over 18 in Remote-Rural places is 45. 10.31% of the population is married, and 14.34% of families live in poverty with children. 0.97% of families receive public cash assistance, and 4.22% of the population receives SNAP benefits. 10.11% of families speak a language other than English at home, and 91.71% have healthcare coverage. 93.44% of residents identify as White; 0.93% African American or Black; 1.96% Native American or Alaska Native; 0.65% Asian; and 3.03% Hispanic or Latino.

A Model of Place

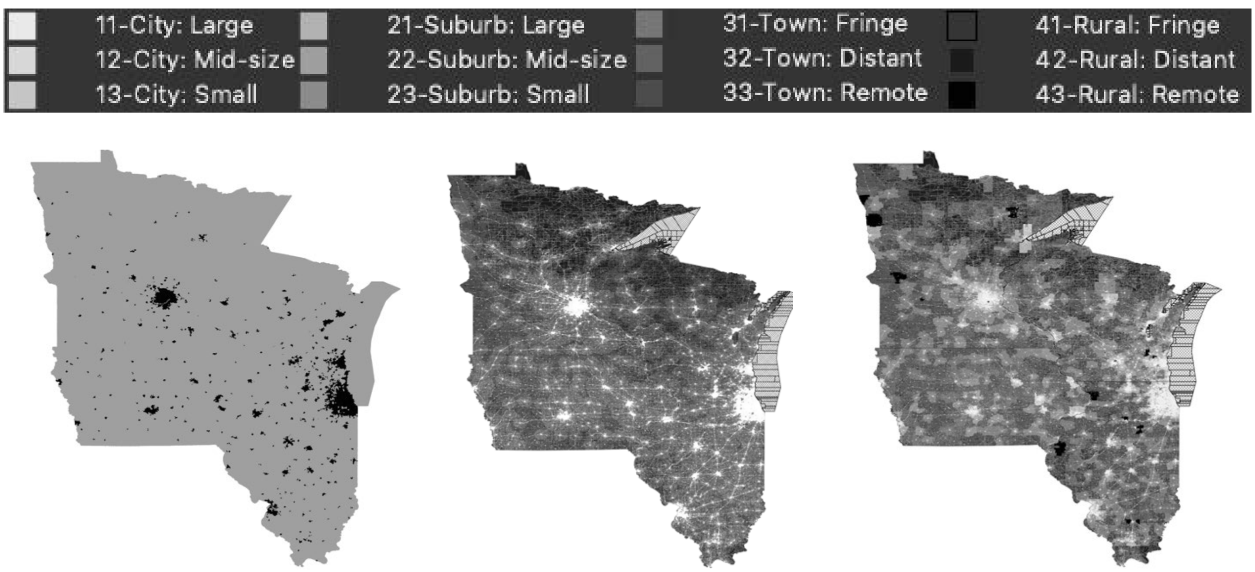
The computational phase involves 500 trees, 398,279 observations, each *mtry = 2*, α = 0.05, and out-of-the-bag (OOB) predictions to evaluate variable importance. The model includes 64 variables and applies five *k*-fold cross-validation steps with one repeat. The processing took 4 hours to complete. On the training data, the resulting OOB error is 0.00%. Applying a random sample of 10,000 observations, using a random seed of 42, from the test set resulted in an accuracy of 98.6% with a No Information Rate of 0.1657 and Kappa 0.99 – an almost perfect agreement according to Cohen (1960).

The most significant feature when determining the classification of a place is the percentage of the workforce over the age of 16 employed in professional, scientific, management, administrative, and waste management services with an importance score "*I*" of 100. The second most significant is the percentage of households with children under 18 (*I* = 95.37). The years in which a percentage of the population moved to an area, between 2000 and 2009, resulted in an importance score of *I* = 91.19. Employment in management, business, science, and arts occupations, *I* = 84.98; educational attainment (bachelor's), *I* = 83.8; broadband access, *I* = 81.56; health insurance coverage, *I* = 81.28; the mean value of homes, *I* = 79.54; transportation, *I* = 79.24; race – Asian, *I* = 78.69; educational attainment (graduate), *I* =

77.07; commute time, 76.57; race – Hispanic or Latino, *I* = 75.91; percentage of the population – SNAP benefits, *I* = 75.85; percentage of households with a computer, *I* = 75.83; percentage of the population – living with non-family, *I*=75.04; and employment in financial services, *I* = 74.74. A graphical representation of the classification of places by the model is presented in Appendix G. Figure 7 presents the classification of places by the United States Census Bureau, the National Center for Education Statistics, and the model.

Figure 7

The United States Census Bureau Classification of Rural and Urban in Comparison to the National Center for Education Statistics and the Model



United States Census Bureau | National Center for Education Statistics | Model

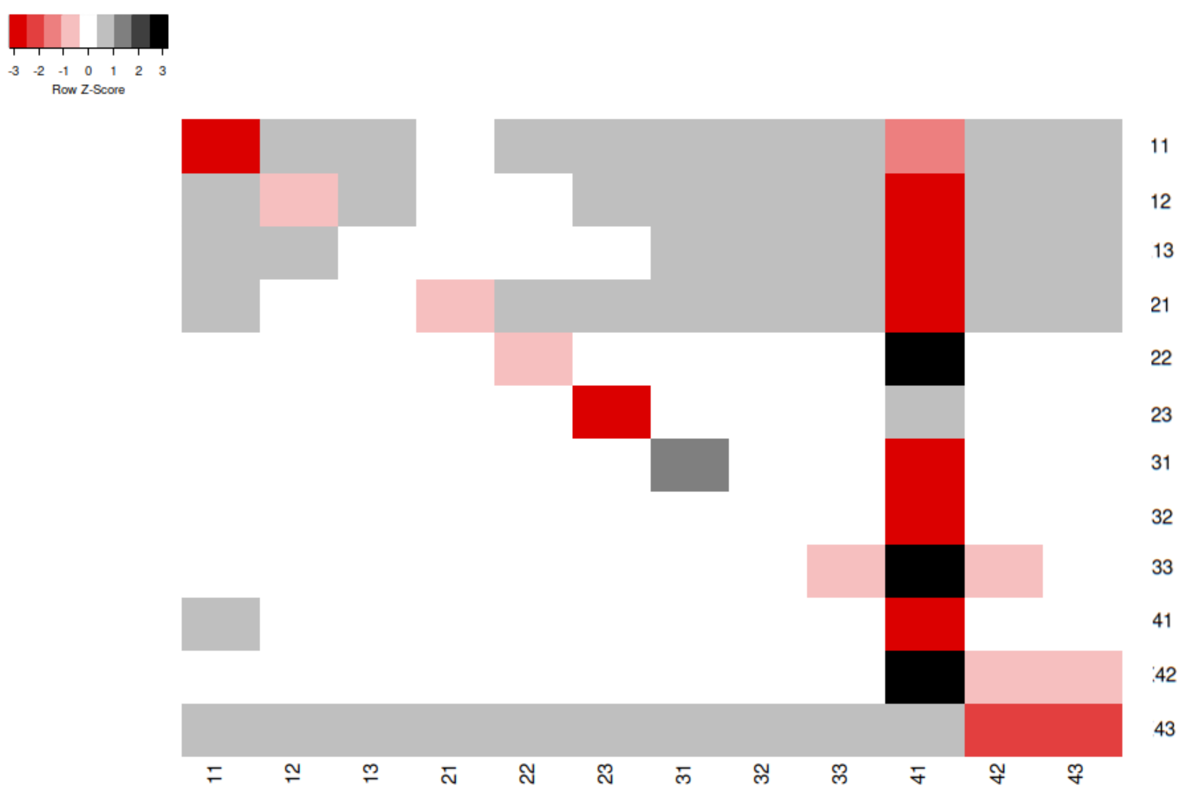
Note. *The United States Census Bureau classification color scheme is inverted.*

The reclassification of census blocks from the NCES school district classification to the NCES definition contained 1,136,588 complete observations. The model characterized 64% of census blocks following the NCES classification with a No Information Rate of 0.2571 and a

Kappa of 0.5918 – a moderate agreement (Cohen, 1960). The Mann-Whitney *U* test (U = 7.24e11, p < 2.2e-16) suggests sufficient evidence to reject the null hypothesis – the results are not from the same population. The composition of the places school districts administrate does not align with the classification of the entire district. Additionally, the urban-rural classification in which the NCES is derived disagrees with the binary classification of 26.55% of census blocks. A confusion matrix is tabulated in Appendix H, and a z-score heatmap plots out the differences between the model's classification and the NCES in Figure 8.

Figure 8

Z-Score Heatmap of Model Classification Predictions in Comparison to NCES Classification

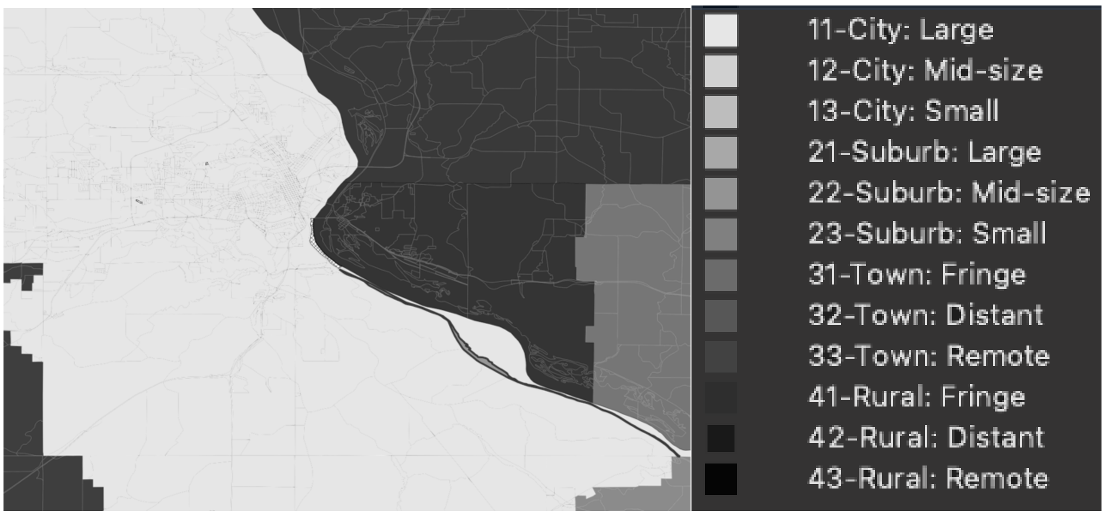


Model Alignment. The most significant disparity between the NCES locale classification of a place and the model involves census block 1002 in Iowa (see Figure 9). The United States

Census Bureau classifies census block 1002 in Iowa as rural. In contrast, the National Center for Education Statistics considers the place part of the locale Fringe (Rural), Distant (Rural), and Small Suburb Locales. The model considers the block part of a small city.

Figure 9

Census Block 1002 in Iowa

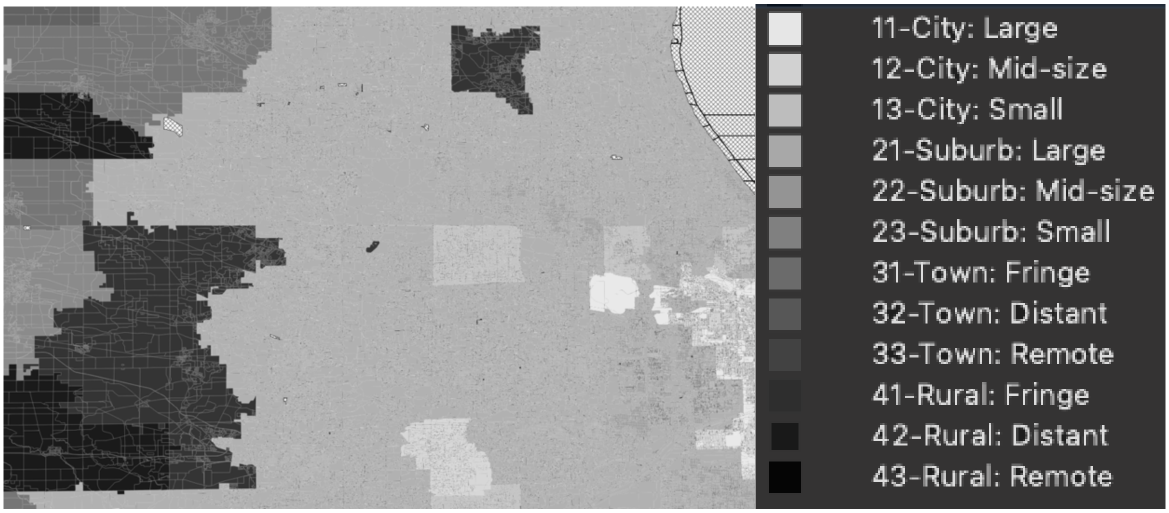


Census block 1002 in Iowa employs 11% of its inhabitants in professional, scientific, management, administrative, and waste management services. The median age is 37, 10.9% of the population is married, 93.7% have healthcare, and 24.5% hold a bachelor's degree. 35.7% of residents moved to the area between 2000 and 2009. 32.3% have access to broadband internet. 92.1% of the population identifies as White, 4% African American or Black, 1.4% Asian, and 2.3% Hispanic or Latino.

The United States Census Bureau classifies census block 1008 in Illinois as urban (see Figure 10). In contrast, the National Center for Education Statistics considers the place part of the Locale as a Mid-sized City, Small City, Large Suburb, and Fringe (Rural). The model considers the block part of a large suburb.

Figure 10

Census Block 1008 in Illinois

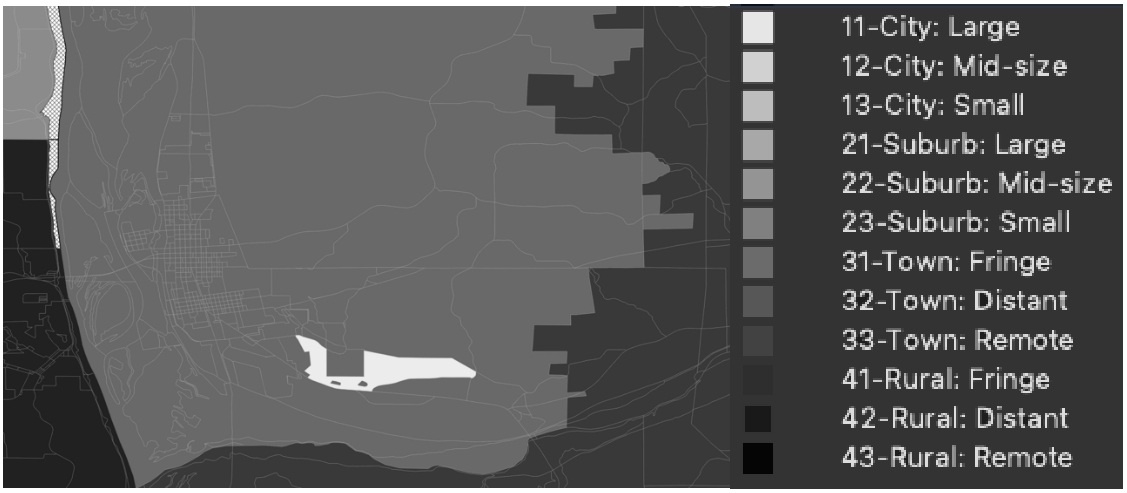


Census block 1008 in Illinois employs 7.8% of its inhabitants in professional, scientific, management, administrative, and waste management services. The median age is 38, 11% of the population is married, 88.5% have healthcare, and 13% hold a bachelor's degree. 13.1% of residents moved to the area between 2000 and 2009. 28.3% have access to broadband internet. 56.7% of the population identifies as White, 3.6% African American or Black, 8.6% Asian, and 30.4% Hispanic or Latino.

Lastly, The United States Census Bureau classifies census block 2046 in Wisconsin as rural (see Figure 11). The National Center for Education Statistics agrees with the classification, labeling it Locale Fringe (Rural). The model considers the block part of a Remote Town.

Figure 11

Census Block 2046 in Wisconsin



Census block 2046 in Wisconsin employs 2% of its inhabitants in professional, scientific, management, administrative, and waste management services. The median age is 46, 8.9% of the population is married, 87.1% have healthcare, and 14.2% hold a bachelor's degree. 25.6% of residents moved to the area between 2000 and 2009. 31% have access to broadband internet. 94.7% of the population identifies as White, 3% African American or Black, and 1.2 % Hispanic or Latino.

Summary

This chapter begins by describing a reproducible and verifiable description of the National Center for Education Statistics locale system through high-quality and available data for analysis. The analysis identified 946,941 census blocks across Wisconsin, Minnesota, Iowa, and Illinois, of which 125,400 were within multiple NCES locales. All census blocks within multiple locales were broken into smaller blocks to fit within each locale and merged into the blocks list, resulting in 1,071,958 observations. A random sample of 312,671 census blocks and their classification were made to the ArcGIS API to assess the accuracy of the locale classification of census blocks. Fourteen census blocks in Wisconsin mismatched the API response, seven in Illinois, and five in Minnesota. The Mann-Whitney *U* test (U = 4.88e10, $p > 0.05$) suggests insufficient evidence to reject the null hypothesis that the computed NCES locale is not significantly different from the classification within ArcGIS. The spatial analysis recreates the boundaries of the NCES locales.

The chapter continues with a descriptive analysis by spatially aligning and classifying 1,071,958 census blocks according to the school district's NCES locale code. After filtering out census blocks with no intersections with school district boundaries, 945,196 observations were recorded. Data from the American Community Surveys and United States Census Bureau containing 1,852 variables are mapped to each census block. Using features presented in Scannell and Gifford's (2010) sense of place and place attachment framework, 65 variables are selected to generate a profile for each NCES locale, concluding the descriptive analysis.

In correlational analysis, a random forest model of 500 trees, 398,279 training observations, $mtry = 2$, $\alpha = 0.05$, and OOB predictions is generated to evaluate variable importance. On the training data, the resulting OOB error is 0.00%. Applying a random sample of 10,000 observations, using a random seed of 42, from the test set resulted in an accuracy of 98.6% with a No Information Rate of 0.1657 and Kappa 0.99. The most significant feature when determining the classification of a place is the percentage of the workforce over the age of 16 employed in professional, scientific, management, administrative, and waste management services with an importance score "*I*" of 100. The second most significant is the percentage of households with children under 18 (I = 95.37). Lastly, the years in which a percentage of the population moved to an area, between 2000 and 2009, resulted in an importance score of I = 91.19.

The classification of 1,136,588 census blocks by the NCES was run through the model, characterizing 64% of census blocks following the NCES classification with a No Information Rate of 0.2571 and a Kappa of 0.5918. The Mann-Whitney U test (U = 7.24e11, p < 2.2e-16) suggests sufficient evidence to reject the null hypothesis – the results are not from the same population. The composition of the places school districts administrate does not align with the classification of the entire district.

The following chapter presents the results of the research questions being asked. The final chapter also discusses how this analysis's results interact with each other and the existing literature on place and application to policy, as presented in chapter 2.

Chapter V: Discussion, Conclusion and Recommendations

The purpose of this sequential quantitative, descriptive, and correlational study is to mathematically capture an essence of place by applying administrative, land-use, economic concepts, and socioeconomic characteristics of each NCES locale classification using census and school district data from the United States Census Bureau and American Community Surveys to describe the existing conditions in which policy derives. If, according to Florida (2009), the socioeconomic class stratification of the United States is due to the human capital and education divide, then policymakers need to know more about what features of place-based policies make them more effective or less effective and how to reconcile the findings (Neumark & Simpson, 2014). The NCES EDGE program develops geographic data to help policymakers, program administrators, and the public understand the relationships between educational institutions and their communities (National Center for Education Statistics, 2019). Assessing the characteristics of the NCES locale system, the alignment of the classification of school districts and the classification of the places they administrate, and identifying the significant characteristics of each locale when applying concepts of place will better inform policymakers on how to address the growing economic disparities between rural and urban places.

Chapter 4 presents the results of the descriptive and correlational analyses of the NCES locale system. This chapter takes the results of chapter 4 and applies them in relation to the research questions and literature in chapters 1 and 2. In doing so, this chapter intends to serve as an interpretation of the realities of the NCES locale system beyond the ill-defined spatial distinctions between rural and urban. This chapter also discusses the conclusions reached because of this research and makes recommendations for future research and individual action.

Discussion

This section discusses research results in alignment with the research questions proposed in chapter one. Where appropriate, this section also aligns the findings with those of the research discussed in chapter two. The intent of this section is to expand the knowledge of the realities of the relationships between school districts and the communities by challenging the alignment of the NCES classification of a school district and the places they administrate. This section continues to apply Scannell and Gifford's (2010) tripartite organizing framework as a guide to secure the interpretation within a consistent theoretical lens.

Research Question 1

Research question 1 asked, “What principles drive the definition of rural and urban, and how do they align with the locale system for the National Center for Education Statistics?” Principles are fundamental truths that serve as a foundation for a system of belief, behavior, or a chain of reasoning. Under Schein's model of organizational culture, people formulate beliefs when implementing values that repeatedly achieve a desired result (Schein, 1992). Beliefs include underlying assumptions about how members of an organization should work with one another and what will lead to success (Schein, 1992). Members of an organization, such as policymakers, then develop a high degree of consensus toward those assumptions and values.

Creating competing descriptions of urban and rural affirms that populations are underserved or misrepresented in policy. United States Federal agencies provide fundamental descriptions for categorizing rural and nonrural populations in educational research and policy through visible and perceptible phenomena (Baumgartner & Zielowski, 2007). Scannell and Gifford's (2010) concept of space and the attachment object, the physical representation of place within social and physical dimensions at differing geographic scales, are the foundation of rural and urban policy. From the public land survey system introduced in the Northwest Ordinance of 1787 up to the 2010 United States Census Bureau's definition of urban, policymakers have formulated the belief that developing space and population density policy will achieve the desired result. Measuring space is a critical approach; space alone is a significant factor in a person's class position, health status, education options, and economic mobility (Dex et al., 2007). Where people live impacts how they grow up, spend their free time, and the educational opportunities available (United States Department of Health and Human Services, 2021). In the correlational analysis of this research, space is significant and aligns with 64% of the NCES locale classifications of school districts and the places they administrate. When applying policy incorporating the attachment object alone, policymakers can only assess "where" policy impacts, not "who," ensuring rural communities remain misrepresented and underserved.

The spatial definitions of the NCES that derive the requirements from the USCB for classifying a place as rural or urban only moderately agree, according to Cohen's Kappa statistic. The NCES relies on standard urban and rural definitions developed by the 2010 U.S. Census Bureau, and each type of locale is either urban or rural. Within the NCES Locale system and the geospatial context of this study, 13% of census blocks fell within multiple locales, and 4,160,392 out of 27,515,031 people, 15.12%, lived in rural places. Following the U.S. Census Bureau's rural-urban dichotomy, 26.55% of census blocks within the NCES locale system do not spatially align, and 6,337,329, 23.03%, of people live in rural places. The misalignment is partially due to the NCES Locales not neatly following census block structures, and the location of the centroid of a census block determines its classification. Additionally, the NCES and U.S. Census Bureau manually classify places that do not align with their definitions.

Research Question 2

Research question 2 asked, “What characteristics define rurality and urbanism beyond population thresholds or distances from an urban core?” The NCES is a binary description of places where everything that is not urban is rural. Long et al. (2021) suggest that binary classifications may not capture the nuances of the differences between rural and urban populations as there are no apparent breaks in the dimensions of their study. One of the products of this research is a profile of each NCES Locale using school district information from the American Community Surveys. The 64 dimensions of these profiles are geospatially aligned and weighted to census blocks – an inquiry recommended by Selover (2021). These profiles do not serve as a definite mechanism to identify what is rural and what is not. The profiles offer a brief perspective other than population thresholds and distances from an urban center to identify “breaks” in dimensions for further inquiries.

The profiles are limited in scope. There is an ethereal component to place that cannot be characterized quantitatively. Place is a sense of belonging that forms one’s identity based on the surrounding circumstances and factors of space; it is a perceptual aspect and the relationship between physical and social interactions that manifest as a home, neighborhood, city, community, state, region, or nation (Carmona, 2021; Canter, 1997). Scannell and Gifford (2010) capture the essence of place through people (cultural and individual aspects), psychological (i.e., affects, cognition, and behavioral components), and attachment object categories (i.e., physical and social spaces). Of these categories, each profile encompasses cultural, social, physical, and behavioral dimensions under the guise of density and proportion in contrast to counts.

The attachment object serves as a physical location for educational and employment opportunities. Skilled workers are attracted to places where the amenities for their economic success are available. Concentrations of skilled individuals in places with adequate public infrastructure attract employers (Florida, 2009). Educational institutions manifest in locations where employers require their services. The result is more availability and easier access to educational opportunities in places of more significant population concentration. Those with the means, resources, and inclination to move to new locations can seize the opportunities suburban and city places offer. In contrast, those tied to place, the rooted, find themselves “stuck” due to limited resources, economic circumstances, and declining financial opportunities – a process Bishop and Cushing (2008) labeled as “the big sort.”

Florida (2009) attributes the socioeconomic class stratification in the United States to the human capital and education divide. The educational attainment and employment dimensions provide insights into places where people with similar educational backgrounds assimilate for career opportunities. Overall, the distribution of the highest educational attainment decreases as the remoteness of a place increases. Those with graduate degrees flock around large cities and undergraduates in large suburban places, where 6 to 8% of the population works in an administrative, professional, or waste management environment. More significant proportions of people with associate degrees reside outside the suburbs, whereas those residing in towns start to work more in agricultural industries – 1.1% to 4.9%. People with some college education remain consistent across all locales. However, those with a high school diploma as their highest educational attainment in the most remote places have the highest population ratio.

Though the traditional habitual patterns of previous generations may no longer accurately reflect the stage of life a person is in, there are signs of behavioral patterns within the locale profiles. Pre-1999, 65.4% of residents of remote rural places moved to their homes. In contrast, 45.8% moved to their residences in large cities and 57.9% in the suburbs. The general movement pattern consisted of lower growth in large cities and increasing rates further away from city cores. Between 2000 and 2009, the trend inverted. City locales experienced a 30% growth, and suburbs experienced 26.6%. Post-2009, the profiles continue to reflect declining growth in places outside of city and urban cores. The youngest population now inhabits large cities; the oldest is in the most remote rural places. Those who are married with children under 18 reside in the suburbs. Suburbs also house the largest concentration of people with undergraduate degrees and the highest income earners. In contrast, the most negligible concentrations of families married with children under 18 reside in large cities and remote rural places.

People Self-Segregate. Social systems of mass society, such as socioeconomic class, race, religion, and life stage, define where people reside within places, forming pockets of relatively homogenous communities (Scannell & Gifford, 2010). People attach themselves to a place because it facilitates “distinctiveness” from other places or affirms the uniqueness of one’s group (Twigger-Ross & Uzzell, 1996). Though race alone is not a significant indicator of culture, it is significant when looking at places where people attach themselves (Ford, 2004).

In rural places, 93.7% of the population is White, 1.0% African American or Black, 2.8% Hispanic or Latino, 0.8% Asian, and 1.5% Native American or Alaska Native. Urban places are more racially diverse, consisting of 48.3% White, 23.5% African American, 22.3% Hispanic or Latino, 5.5% Asian, and 0.3% Native American or Alaska Native. The most significant African American population resides in large and medium-sized cities, with the population dropping as the distance increases from urban centers. Similarly, the Hispanic or Latino population resides in large cities, with additional peaks in large suburban places. As distances from urban centers increase, the White population generally increases, with concentrations in small cities, mid-sized-to-small suburbs, fringe towns, and remote rural places. The Asian population is concentrated in large suburban places, with other pockets in large cities. The Native American or Alaska Native population is negligible until fringe towns, where they consist of 1.2% of the population, increasing to 1.99% in remote rural places.

Districts tend to draw school boundaries by neighborhood and assign students to school by home address, even when other options are more convenient or appealing to families (United States Congress Joint Economic Committee, 2019). One of the goals of phase 2 of this research was to identify “pocket” communities by assessing the most significantly different classifications of a census block by the model compared to the NCES. If people self-stratify, then places that differ in prediction versus classification will exhibit behaviors and be composed of populations like school districts in other locales. The limitation of this approach is that the significance of the changes across the 64 dimensions in the profiles is not explicit, and a pocket community’s heterogeneity is concealed along the boundaries of locales with similar composition.

Census block 1002 in Iowa highlights how a place intersecting and bordering multiple locale groups will express the extremes of its composition where neither the NCES, the United States Census Bureau, nor the model agree. There are places, such as census block 2046 in Wisconsin, where the NCES and USCB agree it is rural. The model distinguishes census block 2046 as part of a remote town (urban) locale, reflecting the blurry line of where urban ends and rural begins (Appendix H). Where the NCES, USCB, and the model disagree, unique behaviors are revealed about the places where communities attach themselves. Census block 1008 in Illinois, where a significant Hispanic or Latino community resides compared to the average city and suburban locales, exhibits Twigger-Ross and Uzzell’s (1996) notion of people conglomerating to affirm the uniqueness of one’s group.

Research Question 3

Research question 3 asked, “How does the National Center of Education Statistics locale classification for school districts align with the classification of the places schools administrate?” Places are more than points on a map; they exist in many sizes and shapes and can be tangible or symbolic (Low & Altman, 1992). Within the NCES locale system, a school district’s physical building’s location determines its classification. However, a school building is a relic of where communities were once located at their construction. Rural communities were abandoned by policy and research once “modernity” was believed to be achieved in the second half of the 20th century (Biddle & Azano, 2016). The location of a school building continues to serve as the function of a Federal administrative schema for identifying locations needing developmental aid even though people are more mobile than ever and are leaving rural for urban places.

The locale profiles reflect a rapid decline in rural growth and an influx in suburban locales since 1999. As the suburbs grow, the boundaries expand and envelop rural communities as part of a homogenous urban locale. The result is that within the rural-urban dichotomy, the rural-urban distinction within the NCES locale system falls at locale 41, where the rural composition of school districts increases from 98.1% in Locale 33 to 98.3% in Locale 41 (Figure 6). Any variations in rural places disappear as the NCES codes conceal heterogeneity, as Jenkins et al. (2016) suggest.

In agreement with Saporito et al. (2013), the NCES schema is an imperfect representation of school boundary geography, and the realities of a place's rural-urban designation can vary widely. The spatial distinction of distance from an urban center misrepresents the rural-urban designation of 9,905,411 people. Rural exhibits more unique behaviors than what the NCES presents. According to the model, 2,176,937 fewer people live in rural places than in the NCES.

Populations in City and Large Suburban places administrated by rural schools are more like those in Mid-sized Suburbs (see Figure 8). Fringe and Distant towns split into peaks in Small Suburban and Remote Towns. 49.4% of rural–fringe places are more like remote towns and distant–rural places than the NCES believes is the threshold of rural and urban places (see Appendix I).

The classification of the places schools administrate partially aligns with a school district's classification. Cohen's (1960) Kappa statistic suggests a moderate agreement. School districts classified as Large Cities administrate 96.5% of places within the Large City locale. Similar alignment is found in Mid-sized Cities, 80.6%; Small Cities, 88%; Large Suburbs, 90.7%; Fringe Towns, 84.1%; and Distant Towns, 87%. Moderate alignment is found in Mid-sized Suburbs, 58.4%; Small Suburbs, 66.4%; Remote Towns, 79.1%; and Remote – Rural, 70.5%. There is minimal alignment in Fringe – Rural, 18.4%, and Distant – Rural, 45.5%, places.

It is effortless to perceive communities only as percentages and space. Proportions shroud the magnitude in which a dimension signifies a community. Each tabulation in Appendix I is a place -- people's realities and identities that do not align with the NCES designation. The delta of the confusion matrix and the composition of school district locale classifications offers some insights into where communities' socioeconomic activities and behaviors misalign with their designation by the NCES. The most prominent misalignment is at the rural-urban boundary at locale 41, where a place in a large city can exhibit the same realities as the most remote rural places. Uniquely, Mid-sized suburban places express minimal variation, and all places within the locale designated as a Mid-sized suburb align. However, most locales convey communities that manifest characteristics of other locales within their boundaries.

Conclusions

This sequential quantitative, descriptive, and correlational study set out to capture an essence of place in rural and urban places by synthesizing information from the ACS with the National Center for Education Statistics (NCES) locale classification system in response to the growing economic disparities between rural and urban. To understand the historical approaches, phase 1 of this research successfully aligns 946,941 census blocks to 12 NCES Locales, 1,926 school districts, and 1,852 variables from the American Community Surveys and United States Census Bureau according to the Mann-Whitney *U* test (U = 4.88e10, $p > 0.05$). From a pure spatial alignment, the rural-urban designation of 284,704 census blocks misaligns between the NCES and the United States Census Bureau (USCB) due to census blocks not fitting neatly into the NCES boundaries. The Mann-Whitney *U* test (U = 7.82e11, $p <$ 2.2e-16) of this spatial alignment suggests sufficient evidence to reject the null hypothesis and that the binary classification of places by the USCB does not align with the binary classification of places by the NCES.

Phase 2 applies a correlational quantitative design treating space as place by implementing a Random Forest statistical model. The model incorporates 500 trees, 64 variables, 398,279 observations, each *mtry = 2*, α=0.05, and OOB predictions to evaluate variable importance. On the training data, the resulting OOB error is 0.00%. A random sample of 10,000 observations, using a random seed of 42, from a test set not included in the training data results in an accuracy of 98.6% with a No Information Rate of 0.1657 and Kappa 0.99 – a near-perfect alignment. When distinguishing between rural and urban, the model characterized 64% of census blocks following the NCES classification of rural and urban with a No Information Rate of 0.2571 and a Kappa of 0.5918 – a moderate alignment.

This research results in an analysis that suggests a disagreement on what rural and urban are spatially and socioeconomically within the NCES locale system. Conglomerating the school districting policies and practices of each state's department of public instruction may be a contributing factor. It must be said that the physical location of a school building classified by the NCES is not representative of the populations it contains when population counts and distance from an urban center are excluded from the equation. Most locales exhibit communities with characteristics of all other locales within their boundaries. The temporal dimensions capture an evolution of rural and urban places, describing a condition in which suburban expansion has not resulted in the reclassification of school districts. In a rural-urban dichotomy, the locale system lacks sensitivity to local variations, and up to 98.1% of places an urban school district administrates may be considered rural.

What remains to be said in this research, though not a research question, is "What is the definition of rurality?" From the perspective of the socioeconomic classes in the United States, the conditions of what the United States Census Bureau denotes as rural are prevalent across all NCES locales. The essence of rural will manifest anywhere, regardless of distance from an urban center. The distinguishing factor of rurality is how its circumstances manifest due to mobility, educational and technological access, economic practice, or symbolic ethnicity. Within the context of this research, the spatial genesis of rural places is in mid-sized to small suburban places. However, any spatial distinction of rural as a distance from an urban core obstructs future research, and assessing the present is only an immutable lens into the dynamic realities of rural places. Without historical context, the essence of rural places will not be captured.

Recommendations

This research identified an egregious proportion in which the spatial classifications of places, according to the National Center for Education Statistics' classification of school districts, can distort the realities of the populations they encompass. Geospatially aligning census blocks to larger entities – school districts – was necessary as the socioeconomic characteristics used for the analysis are not publicly available at the smallest administrative unit surveyed by the United States Census Bureau. Though the locale codes of the National Center for Education Statistics for school districts are misrepresentative of the corporeality of places in the states of Wisconsin, Minnesota, Iowa, and Illinois, it is within this project that a method exists to enhance Federal spatial descriptions with elements of place. It would be futile to assert that the behaviors of rural and urban populations in this study apply to other regions. From this conclusion, an opportunity in rural and urban theory arises for future research to repeat the procedures of this research at the State level or in other regions.

There is an ethereal component to place that cannot be captured quantitatively. Scannell and Gifford's (2010) tripartite organizing framework is a new concept melding sense of place and place attachment. Though intended for urban development, applying concepts from Scannell and Gifford's framework to what is typically treated as "space" unveils rural communities as multifaceted as their urban counterparts. The raw data in this research identifies pockets of unique behaviors that misalign with their rural and urban classifications. A possible direction for future review is to explore the raw data, identify new communities, and analyze these places using the individual components of sense of place and place attachment quantitatively and qualitatively. Through such a hyper-focused lens, future research can determine whether places needing aid can be mathematically identified.

Another opportunity exists within the modeling process. The timeframes in which people moved to their current residence are significant, and locale profiles present an evolving composition of the places school districts administrate. Developing an ecological time series state-space model will capture the transformation of the places school districts administrate, furthering the understanding of the environment in which the NCES locale system was developed and how it came to be in its current state. Such a model must also incorporate the changes in school district boundaries; only the end state is reflected without information such as school redistricting, closure, or consolidation. Additionally, a state-space model offers insights into the future and a potential for policy and spatial descriptions to recognize places needing aid proactively.

It is within the findings of this research that the act of stratifying dynamic behaviors of populations with relatively static school district boundaries will eventually cause the classification of the school district to misrepresent the places they contain. During school integration in the United States, legal comments on the composition of school districts argue that even with standard geographical districting of school lines, a routine districting program for school purposes will have the same effect as a segregation program (Kauper, 1954). People will self-segregate to affirm their group's uniqueness (Twigger-Ross & Uzzell, 1996). The act of self-stratification inhibits students' ability to learn and exchange unique views, limiting their perspective and impacting future outcomes (Ashford-Hanserd et al., 2020; Kauper, 1954). Keeping static school boundaries in a system that reassesses and redraws its classification system every ten years is prejudicial. From here, there is an opportunity for a change in policy. Depending on a State's educational goals, a school district's boundaries can be redrawn to reflect the building’s classification, to encompass a diverse set of locales, or to affirm a community’s identity.

Finally, investigating the qualification criteria for policy, aligning it to places eligible for aid within policy, and those that receive it will further contribute to the growing understanding of the socioeconomic divide in the United States. The context should not be what is rural and what is urban; it appears this is not the correct question as the rural condition does not map into geographic areas delineated in existing data sources. Instead, any future inquiry should be into the barriers that prevent places needing aid from receiving it. Only then will research ask the right questions and contribute actionable findings in policymaking.

References

The state of the rural economy: Hearing before the committee on agriculture house of representatives, 116th Cong. (2019). https://democrats-agriculture.house.gov /uploadedfiles/116-01_-_35465.pdf

Archambault, F. X., & St. Pierre, R. G. (1980). The effect of federal policy on services delivered through ESEA Title I. *Educational Evaluation and Policy Analysis, 2*(3), 33-46. https://doi.org/10.3102/01623737002003033

Ashford-Hanserd, S., Springer, S. B., Hayton, M.-P., & Williams, K. E. (2020). Shadows of Plessy v. Ferguson: The dichotomy of progress toward educational equity since 1954. *The Journal of Negro Education 89*(4), 410-422. https://www.jstor.org/stable/10.7709/jnegroeducation.89.4.0410

Baciu, A., Negussie, Y., & Geller, A. (2017). Communities in action: Pathways to health equity. *Journal of Nutrition Education and Behavior 54*(1), 94-95. https://doi.org/10.17226/ 24624

Baekgaard, M., Moynihan, D. P., & Thomsen, M. K. (2021). Why do policymakers support administrative burdens? The roles of deservingness, political ideology, and personal experience. *Journal of Public Administration Research and Theory 31*(1), 184-200. https://doi.org/10.1093/jopart/muaa033

Barton, H., Grant, M., & Guise, R. (2021). *Shaping neighbourhoods* (3rd ed.). Routledge.

Baumgartner, R. J., & Zielowski, C. (2007). Analyzing zero emission strategies regarding impact on organizational culture and contribution to sustainable development. *Journal of Cleaner Production 15*(13), 1321-1327. https://doi.org/10.1016/j.jclepro.2006.07.016

Benureau, F. C. Y., & Rougier, N. P. (2018, January 3). Re-run, repeat, reproduce, reuse, replicate: Transforming code into scientific contributions. *Front Neuroinform, 11*(69). https://doi.org/10.3389/fninf.2017.00069

Biddle, C., & Azano, A. (2016). Constructing and reconstructing the “rural school problem”: A century of rural education research. *Review of Research in Education, 40*(1), 298-325. https://doi.org/10.3102/0091732X16667700

Bishop, B., & Cushing, R. (2008). The big sort: Migration, community, and politics in the United States of “Those people.” In R. Teixeira (Ed.), *Red, blue, and purple America* (pp. 50-). Brookings Institution Press.

Brown v. Board of Education, 347 U.S. 483 (1954). https://tile.loc.gov /storage-services/service/ll/usrep/usrep347/usrep347483/usrep347483.pdf

U.S. Department of the Interior Bureau of Land Management. (1991). *A history of the rectangular survey system*. https://www.blm.gov/sites/default/files/histrect.pdf

California Department of Housing and Community Development. (n.d.). Enterprise zone program. https://www.hcd.ca.gov/grants-and-funding/programs-archived/enterprise-zone-program

Canter, D. (1997). The facets of place. In: G. T. Moore & R.W. Marans (eds.) *Toward the integration of theory, methods, research, and utilization. Advances in environment, behavior and design*, (pp. 109-147). https://doi.org/10.1007/978-1-4757-4425-5_4

Carleton, D. (2001). *Landmark congressional laws on education: Student’s guide to landmark* *congressional laws.* Greenwood Press.

Carmona, M. (2021). *Public places, urban spaces: The dimensions of urban design* (3rd ed). Routledge. https://doi.org/10.4324/9781315158457

Carmona, M., Heath, T., OC, T., & Tiesdell, S. (2003). *Public places - Urban spaces*. Routledge.

Childs, E. M., Boyas, J. F., & Blackburn, J. R. (2022). Off the beaten path: A scoping review of how ‘rural’ is defined by the U.S. government for rural health promotion. *Health Promotion Perspectives*, *12*(1), 10-21. https://doi.org/10.34172/hpp.2022.02

Chiodo, A. J., Hernández-Murillo, R., & Owyang, M. T. (2010, May). Nonlinear effects of school quality on house prices. *Federal Reserve Bank of St. Louis Review*, *92*(3), 185-204. https://doi.org/10.20955/r.92.185-204

Cloke, P., & Edwards, G. (1986). Rurality in England and Wales 1981: A replication of the 1971 index. *Regional Studies*, *20*(4), 289-306. https://doi.org/10.1080/09595238600185271

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, *20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Congressional Research Service. (2024). *The elementary and secondary education act (ESEA), as amended by the every student succeeds act (ESSA): A primer*. https://crsreports.congress.gov/product/pdf/R/R45977

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed). SAGE Publications, Inc.

De Gennaro, B., & Fantini, A. (2009). *The concept of rurality and the rural-urban relationship as perceived by young people* [Paper, Università degli Studi di Bari and Università degli Studi di Teramo]. https://ifsa.boku.ac.at/cms/fileadmin/Proceeding2002 /2002_WS03_01_Degennaro.pdf

de Smith, M. J., Goodchild, M. F., & Longley, P. A. (2023). *Geospatial analysis: A comprehensive guide to principles, techniques, and software tools* (6th ed.) https://www.spatialanalysisonline.com/extractv6.pdf

Denenberg, D. (1979, June). The missing link: New England's influence on early national educational policies. *The New England Quarterly*, *52*(2), 219-233. https://doi.org/ 10.2307/364840

U.S. Census Bureau. (2023). *Housing Characteristics: 2020*. https://www2.census.gov/library/ publications/decennial/2020/census-briefs/c2020br-09.pdf

Hirsch, B.J., Deutsch, N. (2005). A second home. In B. J. Hirsch, *A place to call home: After-school programs for urban youth* (pp. 41-56). Teachers College Press; American Psychological Association. https://doi.org/10.1037/11087-003

Dex, S., Lindley, J., & Ward, K. (2007, March). *Vertical occupational mobility and its measurement* (Sheffield Economic Research Paper Series 2007006). Department of Economics: University of Sheffield/ https://eprints.whiterose.ac.uk/9963/1/ SERP2007006.pdf

Din, A., & Wilson, R. (2020). Crosswalking ZIP codes to census geographies: Geoprocessing the U.S. Department of Housing & Urban Development’s ZIP code crosswalk files. *Cityscape: A Journal of Policy Development and Research*, *22*(1), 293-314. https://www.huduser.gov/portal/periodicals/cityscpe/vol22num1/ch12.pdf

OCEDiLibrary. (2012). *Equity and quality in education: Supporting disadvantaged students and schools*. https://doi.org/10.1787/9789264130852-en

Esri. (n.d.). *What is a shapefile?*. https://desktop.arcgis.com/en/arcmap/latest/manage-data/shapefiles/what-is-a-shapefile.htm

Esri Developer. (n.d.). *Get started*. https://developers.arcgis.com/documentation/mapping-apis-and-services/get-started/

Farm Security and Rural Investment Act of 2002, Pub. L. 107-171, 116 Stat. 134. (2002). https://www.congress.gov/107/plaws/publ171/PLAW-107publ171.pdf

Farrell, M. (2012). *A Brief History of National Support for Libraries in the United States*. International Federation of Library Associations and Institutions. https://www.ifla.org/past-wlic/2012/140-farrell-en.pdf

Faulkner, L., Brown, K., & Quinn, T. (2018). Analyzing community resilience as an emergent property of dynamic social-ecological systems. *Ecology and Society*, *23*(1), 24-. https://doi.org/10.5751/ES-09784-230124

Rural Health Information Hub. (n.d.). *What is rural?* Retrieved January 1, 2023, from https://www.ruralhealthinfo.org/topics/what-is-rural

Fieldman, R. M. (1990). Settlement identity: Psychological bonds with home places in a mobile society. *Environment and Behavior*, *22*(2), 183-229. https://doi.org/10.1177/0013916590222002

Fingerman, K., Pillemer, K., Silverstein, M., & Suitor, J. (2012). The baby boomers' intergenerational relationships. *Educational Resources Information Center*, *52*(2), 199-209. https://doi.org/10.1093/geront/gnr139

Fischel, W. A. (2015, July). *Zoning rules! The economics of land use regulation*. Lincoln Institute of Land Policy.

Florida, R. (2009). *Who's Your City?: How the Creative Economy Is Making Where to Live the Most Important Decision of Your Life*. Basic Books; Illustrated edition.

Ford, R. T. (2004). *Racial Culture – A Critique*. Princeton University Press. https://doi.org/10.1515/9781400826308

Fried, M. (2000, September). Continuities and discontinuities of place. *Journal of Environmental Psychology*, *20*(3), 193-205. https://doi.org/10.1006/jevp.1999.0154

Friedman, T. L. (2005). *The world is flat: A brief history of the twenty-first century* (1st ed). Farrar, Straus & Giroux.

García-Díez, V., García-Llorente, M., & González, J. A. (2020). Participatory mapping of cultural ecosystem services in Madrid: Insights for landscape planning. *Land*, *9*(8), 244-. https://doi.org/10.3390/land9080244

Gehl, J. (2011). *Life between buildings: Using public space* (6th ed). Island Press.

Gustafson, P. (2001). Roots and routes: Exploring the relationship between place attachment and mobility. *Environment and Behavior*, *33*(5), 667–686. https://doi.org/10.1177/00139160121973188

Hart, L. G., Larson, E. H., & Lishner, D. M. (2005, July). Rural definitions for health policy and research. *American Journal of Public Health*, *95*(7), 1149-1155. https://doi.org/10.2105/AJPH.2004.042432

Hartigan, J. A., & Hartigan, P. M. (1985, March). The dip test of unimodality. *The Annals of Statistics*, *13*(1), 70-84. https://doi.org/10.1214/aos/1176346577

Hawley, L. R., Koziol, N. A., Bovaird, J. A., McCormick, C. M., Welch, G. W., Arthur, A. M., & Bash, K. (2016). Defining and describing rural: Implications for rural special education research and policy. *Rural Special Education Quarterly*, *35*(3), 3-11. https://doi.org/10.1177/8756870516035003027

Abler, R. F. (1992). *Geography's inner worlds: Pervasive themes in contemporary American geography* (1st ed). Rutgers University Press.

Hothorn, T., & Zeileis, A. (2015). Partykit: A modular toolkit for recursive partitioning in R. *Journal of Machine Learning Research*, *16*(118), 3905-3909. http://jmlr.org/papers/v16 /hothorn15a.html

Hothorn, T., Hornik, K., & Zeileis, A. (2015). *ctree: Conditional inference trees*. Comprehensive R Archive Network. https://cran.r-project.org/web/packages/partykit/ vignettes/ctree.pdf

Houston, D. (1919). *Annual Report of the Federal Board for Vocational Education.* Washington Government Printing Office. https://www.va.gov/vetdata/docs/FY1919.pdf

IBM. (2023). *What is random forest?* https://www.ibm.com/topics/random-forest

IBM. (2024). *JavaScript Object Notation (JSON) format*. https://www.ibm.com/docs/en/baw/ 22.x?topic=formats-javascript-object-notation-json-format

Isserman, A. M. (2005, October). In the national interest: Defining rural and urban correctly in research and public policy. *International Regional Science Review*, *28*(4), 465-499. https://doi.org/10.1177/0160017605279000

Jenkins, W. D., Lipka, A. E., Fogleman, A. J., Delfino, K. R., Malhi, R. S., & Hendricks, B. (2016, July). Variance in disease risk: Rural populations and genetic diversity. *Genome*, *59*(7), 519-525. https://doi.org/10.1139/gen-2016-0077

Kain, J. F. (1968, May 1). Housing segregation, Negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, *82*(2), 175–197. https://doi.org/10.2307/1885893

Kane, T. J., Staiger, D. O., & Reigg, S. K. (2005, May). School quality, neighborhoods and housing prices: The impacts of school desegregation. *National Bureau of Economic Research Working Paper Series, 11347*. https://doi.org/10.3386/w11347

Kauper, P. G. (1954, June). Segregation in public education: The decline of Plessy v. Ferguson. *Michigan Law Review*, *52*(8), 1137-1158. https://doi.org/10.2307/1285295

Kim, H., Lee, S., Cheon, J., Hong, S., & Chang, M. (2018, May 02). A comparative study to identify factors of caregiver burden between baby boomers and post baby boomers: A secondary analysis of a US online caregiver survey. *BMC Public Health*, *18*(1), 579. https://doi.org/10.1186/s12889-018-5488-4

Lang, J. (1994). *Urban design: The American experience* (1st ed). Wiley.

Legg, C., & Hookway, C. (2021). *Pragmatism*. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/pragmatism/

Lens, M. C., & Monkkonen, P. (2016). Do strict land use regulations make metropolitan areas more segregated by income? *Journal of the American Planning Association*, *82*(1), 6-21. https://doi.org/10.1080/01944363.2015.1111163

Library of Congress. (n.d.). *From Thomas Jefferson to James Monroe*, *9 July 1786.* https://tile.loc.gov/storage-services/service/mss/mtj//mtj1/005/005_1125_1130.pdf

Library of Congress. (n.d.). *To Thomas Jefferson from James Monroe 16 July 1786*. https://tile.loc.gov/storage-services/service/mss/mtj//mtj1/005/005_1188_1190.pdf

Long, J. C., Delameter, P. L., & Holmes, G. M. (2021, October). Which definition of rurality should I use? The Relative Performance of 8 Federal Rural Definitions in Identifying Rural-Urban Disparities. *Medical Care*, *59*(5), S413-S419. https://doi.org/10.1097/MLR.0000000000001612

Low, S. M., & Altman, I. (1992). Place attachment: A conceptual inquiry. In I. Altman & S. M. Low, (Eds), *Place attachment: Human behavior and environment* (Vol. 12, pp. 1-12). Springer. https://doi.org/10.1007/978-1-4684-8753-4_1

Marzbani, M., Awad, J., & Rezaei, M. (2020). The sense of place: Components and walkability. old and new developments in Dubai, UAE. *The Journal of Public Space*, *5*(1), 21-36. https://doi.org/10.32891/jps.v5i1.1249

Matakos, G., Gionis, A. (2022). Strengthening ties towards a highly connected world. *Data Mining and Knowledge Discovery*, *36*, 448–476. https://doi.org/10.1007/s10618-021-00812-1

McClay, W. M., & McAllister, T. V. (2014). *Why place matters: Geography, identity, and civic life in modern America*. Encounter Books.

Meilleur, A., Subramanian, S. V., Plascak, J. J., Fisher, J. L., Paskett, E. D., & Lamont, E. B. (2013, October 1). Rural residence and cancer outcomes in the United States: Issues and challenges. *Cancer Epidemiology, Biomarkers, & Prevention*, *22*(10), 1657-1667. https://doi.org/10.1158/1055-9965.epi-13-0404

Milne, B. T., & Cohen, W. B. (1999, October). Multiscale assessment of binary and continuous landcover variables for MODIS validation, mapping, and modeling applications. *Remote Sensing of Environment*, *70*(1), 82-98. https://doi.org/10.1016/S0034-4257(99)00059-0

Montgomery, J. (1998). Making a city: Urbanity, vitality and urban design. *Journal of Urban Design*, *3*(1), 93-116. https://doi.org/10.1080/13574809808724418

Mueller, K. J., Coburn, A. F., Knudson, A., Lundblad, J. P., MacKinney, A. C., & McBride, T. D. (2020). *Considerations for defining rural places in health policies and programs*. Rural Policy Research Institute. https://rupri.org/2020/05/07/considerations-for-defining-rural-places-in-health-policies-and-programs

Mulkey, D. (1988, October 3-5*). Education policy and rural development: A perspective from the southern region* [paper presentation]. Southern Region Rural Development Policy Workshop, Birmingham, Alabama. https://files.eric.ed.gov/fulltext/ED302359.pdf

Nanzer, B. (2004). Measuring sense of place: A scale for Michigan. *Administrative Theory & Praxis*, *26*(3), 362–382. http://dx.doi.org/10.1080/10841806.2004.11029457

National Archives. (n.d.). *Northwest ordinance (1787)*. https://www.archives.gov/milestone-documents/northwest-ordinance

National Archives. (n.d.). *Morrill act (1862)*. https://www.archives.gov/milestone-documents/morrill-act

National Center for Education Statistics. (2019, January). *Digest of education statistics 2017* (53rd ed.). U.S. Department of Education. https://nces.ed.gov/pubs2018/2018070.pdf

National Conference of State Legislatures. (2021). *How states define lobbying and lobbyist*. https://www.ncsl.org/ethics/how-states-define-lobbying-and-lobbyist

National Democratic Institute. (2008). *Constituent relations*. United States Agency for International Development [USAID]. https://www.ndi.org/sites/default/files/2407_gov_constituentsmall_engpdf_10082008.pdf

National Information Center for Children and Youth with Disabilities. (1997, July). Interventions for students with learning disabilities. *NICHCY News Digest*, *25*. https://files.eric.ed.gov/fulltext/ED415607.pdf

National Institute of Food and Agriculture. (2022, August 30). *Celebrating the second Morrill Act of 1890*. U.S. Department of Agriculture. https://www.nifa.usda.gov/about-nifa/blogs/celebrating-second-morrill-act-1890

National Institute of Standards and Technology. (n.d.). *Engineering statistics handbook*. U.S. Department of Commerce. https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm

Neumark, D., & Kolko, J. (2010, July). Do enterprise zones create jobs? Evidence from California's enterprise zone program. *Journal of Urban Economics*, *68*(1), 1-19. https://doi.org/10.1016/j.jue.2010.01.002

Neumark, D., & Simpson, H. (2014). *Place-based policies* [NBER Working Paper 20049]. National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w20049/w20049.pdf

Office of Educational Technology. (n.d.). *Funding digital learning*. U.S. Department of Education. https://tech.ed.gov/funding/

Office of Management and Budget. (2010, June 28). 2010 *Standards for delineating metropolitan and micropolitan statistical areas*. Office of Information and Regulatory Affairs, Federal Register, *75*(123) 1-7. https://www.govinfo.gov/content/pkg/FR-2010-06-28/pdf/2010-15605.pdf

Newhart, M., & Patten, M. (2023). *Understanding research methods* (11th ed). Routledge.

Kettler, T., Puryear, J. S., & Mullet, D. R. (2016, July 1). Defining rural in gifted education research: methodological challenges and paths forward. *Journal of Advance Academics*, *27*(4), 245-265. https://doi.org/10.1177/1932202X16656896

Python Software Foundation. (n.d.). About Python. https://www.python.org

QGIS. (2023). *Cheat sheet for PyQGIS*. Retrieved February 19, 2023, from https://docs.qgis.org/3.28/en/docs/pyqgis_developer_cookbook/cheat_sheet.html

QGIS. (n.d.). *Spatial without compromise*. Retrieved December 18, 2023, from https://www.qgis.org/en/site

The R Foundation. (n.d.). *What is R?* Retrieved February 12, 2024, from https://www.r-project.org/about.html

Ramsay-Jordan, N., Crenshaw, A. & Chestnutt C. (2022). Examining purposeful researchable questions in mathematics education. *Journal of Honai Math*, *5*(2). https://doi.org/10.30862/jhm.v5i2.260

Ratcliffe, M. (2022, December 22). *Redefining urban areas following the 2020 census*. United States Census Bureau. https://www.census.gov/newsroom/blogs/random-samplings/2022/12/redefining-urban-areas-following-2020-census.html

Ratcliffe, M., Burd, C., Holder, K., & Fields, A. (2016, December). *Defining rural at the U.S. Census Bureau: American community survey and geography brief*. U.S. Department of

Commerce, Economics and Statistics Administration: U.S. Census Bureau. http://doi.org/10.13140/RG.2.2.16410.64969

Regnault, A., Willgoss, T., & Barbic, S. (2018, April). Towards the use of mixed methods inquiry as best practice in health outcomes research. *Journal of Patient-Reported Outcomes*, *2*(19). https://doi.org/10.1186/s41687-018-0043-8

Relph, E. (1976). *Place and placelessness: Research in planning and design* (1st ed). London: Pion Limited.

Rigatti, S. J. (2017). Random forest. *Journal of Insurance Medicine*, *47*(1), 31-39. https://doi.org/10.17849/insm-47-01-31-39.1

Roberts, P., & Green, B. (2013). Researching rural places: On social justice and rural education. *Qualitative Inquiry*, *19*(10), 765–774. http://doi.org/10.1177/1077800413503795

Rosen, C. (2011). The new meaning of mobility. *The New Atlantis*, *31*, 40-46. https://www.thenewatlantis.com/publications/the-new-meaning-of-mobility

Rothwell, J. (2012, April). *Housing costs, zoning, and access to high-scoring schools*. Brookings Institution. https://eric.ed.gov/?id=ED531228

Saporito, S., Van Riper, D., & Wakchaure, A. (2013). Building the school attendance boundary information system (SABINS): Collecting, processing, and modeling K to 12 educational geography. *Journal of the Urban and Regional Information Systems Association*, *25*(2), 49–62. https://pubmed.ncbi.nlm.nih.gov/29151773

Scannell, L., & Gifford, R. (2010, March). Defining place attachment: A tripartite organizing framework. *Journal of Environmental Psychology*, *30*(1), 1-10. https://doi.org/10.1016/j.jenvp.2009.09.006

Schein, E. H. (1992, November). *Organizational culture and leadership* (2nd ed). Jossey-Bass.

Selover, M. (2021). *The marginalization of rural communities: A mixed-methods study of equity within rural postsecondary career and technical education* [Doctoral dissertation, University of Wisconsin – Stout]. Minds @ UW. https://minds.wisconsin.edu/bitstream/handle/1793/83105/2021seloverm.pdf

Sherwood, T. (2000). Where has all the "rural" gone? Rural education research and current federal reform. *Journal of Research in Rural Education*, *16*(3), 159-167. https://jrre.psu.edu/sites/default/files/2019-08/16-3_1.pdf

Smith, N. B. (1999). A tribute to the visionaries, prime movers and pioneers of vocational education, 1892 to 1917. *Journal of Vocational and Technical Education, 16*(1). https://files.eric.ed.gov/fulltext/EJ598592.pdf

Stern, J. D. (1994). *The condition of education in rural schools*. U.S. Department of Education, Office of Educational Research and Improvement. https://files.eric.ed.gov/fulltext/ED371935.pdf

Stone, M. (1976, September). Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, *38*(1) 111-133. https://doi.org/10.1111/j.2517-6161.1976.tb01573.x

Taun, Y-F. (1977). *Space and place: The perspective of experience*. University of Minnesota Press.

Taylor, H. C. (2008). *The educational significance of the early federal land ordinances*. Kessinger.

Blaschke, T., Merschdorf, H. Cabrera-Barona, P, Gao, S., Papadakis, E., Kovacs-Györi, A. (2018). Place versus space: From points, lines and polygons in GIS to place-based representations reflecting language and culture. *International Journal of Geo-Information, 7*(11), 452. https://doi.org/10.3390/ijgi7110452

Tibbalds, F. (1992). *Making people friendly towns. Improving the public environment in towns and cities* (1st ed). Longman.

Towers, G. (2005). *An introduction to urban housing design: At home in the city.* Routledge.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

Twigger-Ross, C. L., & Uzzell, D. L. (1996, September). Place and identity processes. *Journal of Environmental Psychology*, *16*(3), 205–220. https://doi.org/10.1006/jevp.1996.0017

Tyack, D., James, T., & Benavot, A. (1987). *Law and the shaping of public education, 1785 – 1954*. University of Wisconsin Press.

United States Census Bureau. (2018). *2017 American housing survey data*. https://www.census .gov/newsroom/press-releases/2018/ahs.html

United States Census Bureau. (2019a, July). *Counting the hard to count in a census: Select topics in international censuses*. U.S. Department of Commerce Economics and Statistics Administration. https://www.census.gov/content/dam/Census/library/working-papers/2019/demo/Hard-to-Count-Populations-Brief.pdf

United States Census Bureau. (n.d.). *How is a differentially private system different from the census bureau’s traditional disclosure avoidance techniques*. https://ask.census.gov/ prweb/PRServletCustom?pyActivity=pyMobileSnapStart&ArticleID=KCP-5434

United States Census Bureau. (2019b). *Understanding and using American community survey data: What users of data for rural areas need to know*. U.S. Department of Commerce Economics and Statistics Administration. https://www.census.gov/content/dam/Census/library/publications/2019/acs/ACS_rural_handbook_2019.pdf

United States Census Bureau. (2022, April 11). *Glossary*. https://www.census.gov/programs-surveys/geography/about/glossary.html

United States Census Bureau. (2023, June). *2020 census urban areas facts*. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2020-ua-facts.html

United States Congress Joint Economic Committee. (2019, November 2019). *Zoned out: How school and residential zoning limit educational opportunity*. https://www.jec.senate.gov/public/index.cfm/republicans/2019/11/zoned-out-how-school-and-residential-zoning-limit-educational-opportunity

United States Library of Congress. (1776). *Journals of the continental congress, 1774-1789* (Vol. 4). Washington Government Printing Office. https://tile.loc.gov/storage-services/service/ll/llscd/lljc004/lljc004.pdf

Cromartie, J & Bucholtz, S. (2008, June 1). Defining the “rural” in rural America. U.S. Department of Agriculture. https://www.ers.usda.gov/amber-waves/2008/june/defining-the-rural-in-rural-america/

United States Department of Agriculture. (n.d.-a). *Rural classifications: Overview*. Retrieved November 29, 2022, from https://www.ers.usda.gov/topics/rural-economy-population/rural-classifications/

United States Department of Agriculture. (n.d.-b). *Rural-urban continuum codes*. Economic Research Service. Retrieved January 18, 2023, from https://www.ers.usda.gov/data-products/rural-urban-continuum-codes

United States Department of Education. (n.d.). *The federal role in education*. Retrieved January 18, 2023, from https://www2.ed.gov/about/overview/fed/role.html

United States Department of Health and Human Services. (2021). *Community health and economic prosperity: Engaging businesses as stewards and stakeholders—A report of the surgeon general*. Center for Disease Control and Prevention. https://www.hhs.gov/sites/default/files/chep-sgr-full-report.pdf

United States Department of Health and Human Services. (2020, June 25). *Defining rural population*. https://www.hhs.gov/guidance/document/defining-rural-population

United States Department of Health, Education, and Welfare. (1960). *State plans under the library Services Act: A progress report the first three years: Fiscal years 1957, 1958, 1959* (OE-15012, Bulletin 1960, No. 27). U.S. Government Printing Office. https://files.eric.ed.gov/fulltext/ED543968.pdf

United States Library of Congress. (2020). *Northwest ordinance: Primary documents in American history*. Retrieved March 16, 2022, from https://guides.loc.gov/northwest-ordinance

Vanderboom, C.P., & Madigan, E.A. (2007, March 22). Federal definitions of rurality and the impact on nursing research. *Research in Nursing and Health*, *30*(2), 175-184. https://doi.org/10.1002/nur.20194

Waldorf, B. S. (2006). *A continuous measure of rurality: Moving beyond threshold measures*. [Paper presentation]. Agricultural and Applied Economics Association (AAEA) conferences, 2006 Annual Meeting, July 23-26, Long Beach, CA.. http://dx.doi.org/10.22004/ag.econ.21383

Weil, S. (2001, November 9). *The need for roots: Prelude to a declaration of duties towards mankind* (1st ed). Routledge.

Wheatley, M. J. (2006). *Leadership and the new science: Discovering order in a chaotic world*. Berrett-Koehler.

Whyte, W. H. (1980). *The social life of small urban spaces*. Conservative Foundation.

Wilson**,** W. J. (2012, July 10). *The truly disadvantaged: The inner city, the underclass, and public policy* (2nd ed). University of Chicago Press.

Appendix A

NCES Locale Classification System

Code	Locale	Description
11	Large City	Territory inside an urbanized area and a principal city with a population of 250,000 or more.
12	Midsize City	Territory inside an urbanized area and a principal city with a population less than 250,000 and greater than or equal to 100,000.
13	Small City	Territory inside an urbanized area and a principal city with a population of less than 100,000.
21	Large Suburb	Territory outside a Principal City and inside an Urbanized Area with a population of 250,000 or more.
22	Midsize Suburb	Territory outside a Principal City and inside an Urbanized Area with a population less than 250,000 and greater than or equal to 100,000.
23	Small Suburb	Territory outside a Principal City and inside an Urbanized Area with a population of less than 100,000.
31	Fringe Town	Territory inside an Urban Cluster less than or equal to 10 miles from an Urbanized Area.
32	Distant Town	Territory inside an Urban Cluster that is more than 10 miles and less than or equal to 35 miles from an Urbanized Area.
33	Remote Town	Territory inside an Urban Cluster more than 35 miles from an Urbanized Area.

41	Fringe (Rural)	Census-defined rural territory less than or equal to 5 miles from an Urbanized Area and rural territory less than or equal to 2.5 miles from an Urban Cluster.
42	Distant (Rural)	Census-defined rural territory is more than 5 miles but less than or equal to 25 miles from an Urbanized Area, and rural territory is more than 2.5 miles but less than or equal to 10 miles from an Urban Cluster.
43	Remote (Rural)	Census-defined rural territory is more than 25 miles from an urbanized area and more than 10 miles from an urban cluster.

Appendix B

Python Code Sample for Spatial Analysis

```
def block_district_classifier(self, block):
  if self.add_layer:
   for district in self.add_layer.getFeatures():
    self.block_district_pct(block, district)

  for district in self.vectors.getFeatures():
   self.block_district_pct(block, district)

 def block_nces_pct(self, block):
  for ncesf in self.vectors.getFeatures():
   if block.geometry().overlaps(ncesf.geometry()):
    intsct = block.geometry().intersection(ncesf.geometry())
    feature = QgsFeature()
    feature.setGeometry(intsct)
    feature.setAttributes(
     [
      block.attribute("GEOID20"),
      int(ncesf.attribute("LOCALE")),
      ((intsct.area() / block.geometry().area()) * 100),
     ]
    )
    self.vector_layer.addFeatures([feature])

def block_nces_crosses(self, feature):
  codes = []
  for ncesf in self.vectors.getFeatures():
   if feature.geometry().within(ncesf.geometry()):
    break
   if feature.geometry().overlaps(ncesf.geometry()):
    codes.append(ncesf.attribute("LOCALE"))
  if len(codes) > 0:
   self.vector_layer.changeAttributeValue(
    feature.id(), feature.fieldNameIndex(NCES_CROSS_CODE), ",".join(codes)
   )

def classify_blocks(self, feature):
  for ncesf in self.vectors.getFeatures():
   if feature.geometry().within(ncesf.geometry()):
    self.vector_layer.changeAttributeValue(
     feature.id(),
     feature.fieldNameIndex(CENTROID_NCES_CODE),
     int(ncesf.attribute("LOCALE")),
    ); break
```

Appendix C

Composition of School District Locale Classifications and NCES Census Block Classifications

	11	12	13	21	22	23	31	32	33	41	42	43
11	**63674**	-	3	1005	-	-	-	-	-	314	-	-
12	-	**24800**	60	3795	3478	-	-	-	-	3612	277	-
13	28	53	**57027**	5666	6579	5838	26	-	-	12407	1705	-
21	825	3462	2760	**161125**	90	148	841	-	-	26335	935	-
22	-	1206	1294	-	**15971**	-	-	-	-	7652	1606	-
23	-	-	628	228	-	**11888**	-	-	-	6967	1000	-
31	-	64	6	1287	152	8	**22896**	94	-	22755	8916	-
32	-	-	-	-	15	-	71	**71888**	-	46934	55849	2825
33	-	-	-	-	-	-	-	-	**33225**	18937	26814	5943
41	-	838	667	5840	1150	1180	3191	11107	9260	**34335**	30854	5509
42	-	76	327	103	78	608	303	1233	77	14926	**145253**	28836
43	-	-	-	-	-	-	-	-	69	340	26174	**107834**

Appendix D

Unmodified Locale Variables and Statistics

Var	*ME*	*I*	*Var*	*ME*	*I*
Ed. High School	500.89	55.72	Employment - Wholesale	144.84	54.62
Ed. Some College	469.72	57.96	Employment - Retail	287.03	59.86
Ed. Bachelor’s	461.61	84.27	Employment – Transport.	202.46	59.02
Ed. Graduate	337.58	77.07	Employment – Information	111.39	56.60
Ed. Associate	294.91	60.53	Employment - Finance	228.75	74.73
Commute - Drive	596.75	79.23	Employment – Waste Mgt.	282.69	100
Commute - Public	237.60	72.91	Employment - Education	396.77	65.66
Commute - Walk	194.52	54.70	Employment - Arts	301.84	71.99
Work From Home	185.38	48.13	Employment – Public	168.09	47.55
Commute - Minutes	1.54	76.56	Cash Assistance	117.66	59.68
Commute - Carpool	298.59	66.46	Food / SNAP Benefits	253.12	75.85
Employment – Mgt.	518.33	84.98	Health Insurance	691.38	81.27
Employment - Service	383.08	72.36	Poverty - All	4.82%	43.86
Employment - Sales	413.38	83.80	Poverty – With Children	2.46%	57.14
Employment – Maint.	234.49	45.99	Race - White	760.83	72.15
Employment – Product.	327.54	60.13	Race – African American	374.70	69.69
Employment – Agricult.	84.64	58.16	Race – Native / Alaska	99.40	40.91
Employment – Construct.	199.74	46.15	Race - Asian	293.04	78.68
Employment – Manufact.	319.20	51.84	Race – Hisp. or Latino	407.35	75.91

Var	*ME*	*I*	*Var*	*ME*	*I*
Median Age	1.87	31.90	Mean Household Size	0.0005	29.29
Families with Children	266.66	95.37	Moved – 2015 or Later	176.52	66.04
Households - Non-Family	410.45	75.03	Moved – 2010-2014	273.60	74.48
Population in school	326.91	79.26	Moved – 2000-2009	364.20	91.19
Median House Rooms	-0.001	12.06	Moved – 1990-1999	337.13	68.73
Housing Units	346.65	78.90	Moved – 1980-1989	230.81	55.93
Median Home Value	7787.44	79.54	Moved – Pre-1979	217.58	63.72
Mean Income	4864.74	68.24	No Plumbing Facilities	0.50%	0
Computer Access	401.82	75.83	No Kitchen Facilities	0.63%	5.94
Internet Access	441.73	81.56	No Telephone Service	0.83%	7.09
Housing – Rent Rooms	0.049	7.12	Unemployment Rate	1.52%	15.19

Appendix E

R Sample Code for Data Processing and Modeling

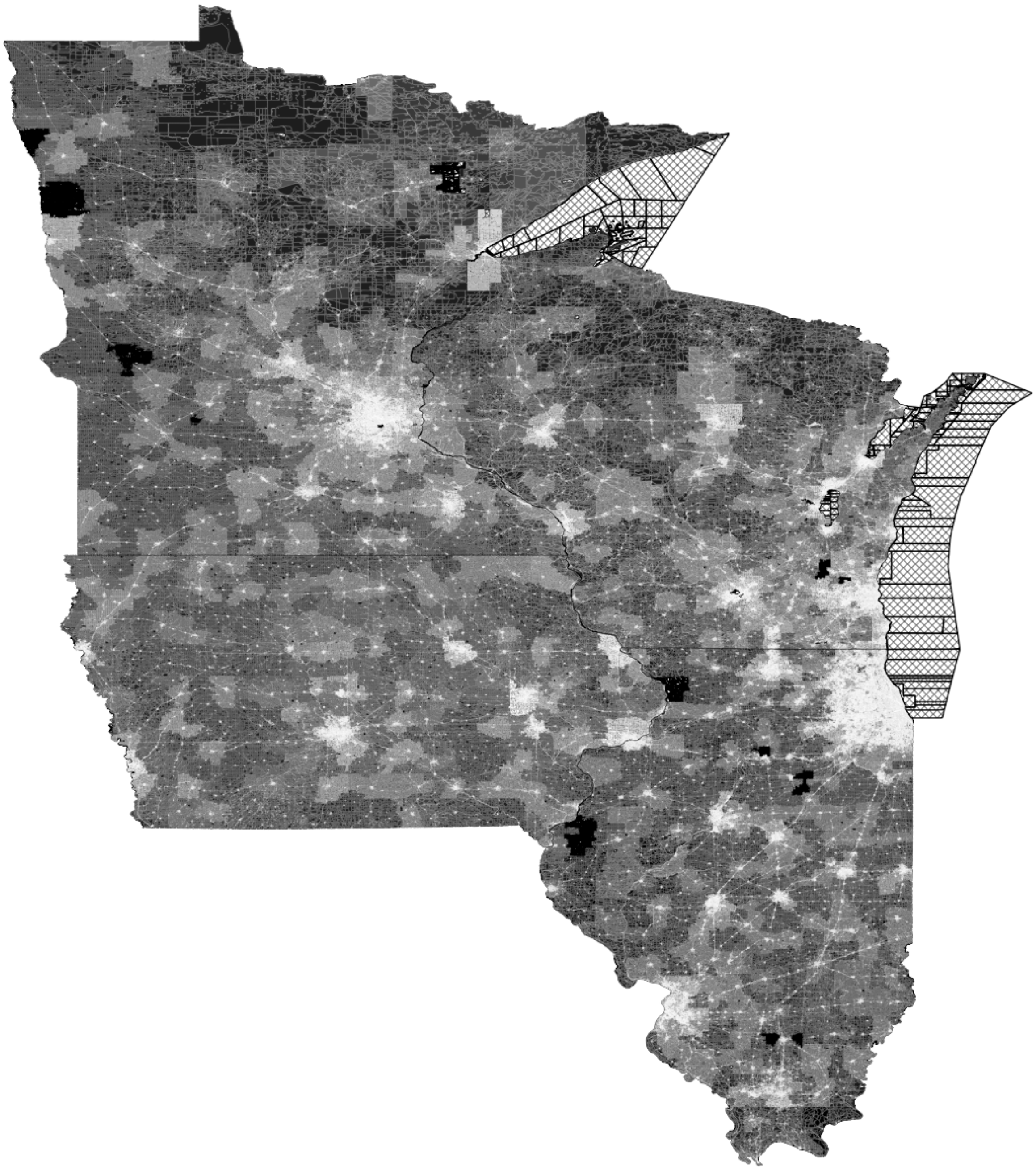
```
block_layers <- list.files(
  path=paste(getwd(), "/data/blocks", sep="")
) %>% paste(paste(getwd(), "data/blocks", sep="/"), ., sep="/") %>% read_csv %>% bind_rows
#...
blchunks <- split(data.frame(block_layers_crosses), seq(nrow(block_layers_crosses)) %/%
detectCores())
process_layer_crosses <- function(data_chunk) {
  library(dplyr)
  results <- list()
  for (i in 1:nrow(data_chunk)) {
    row <- data_chunk[i,]
    matches <- sigints %>% filter(GEOID20 == row$GEOID20)
    rows <- by(matches, seq_len(nrow(matches)), function(m) {
      # omitted for brevity
      row$HOUSING20 <- ceiling(row$HOUSING20 * m$overlap / 100)
      row$POP20 <- ceiling(row$POP20 * m$overlap / 100)
      return(row)
    })
    results <- append(results, rows)
  }
  return(results)
}
cl <- makeCluster(detectCores())
clusterExport(cl,c('process_layer_crosses', 'sigints', 'block_layers_crosses',
'block_layers_intersects'))
results <- parLapply(blchunks, process_layer_crosses, cl = cl)
stopCluster(cl)

set.seed(42)
library(doParallel)
registerDoParallel(cores=3)
model.sample <- model.data[createDataPartition(model.data$locale_description, p = 0.50,
list=F),]
model.cv <- trainControl(method='repeatedcv', number=5, repeats=2)
model.partitions <- createDataPartition(y=model.sample$locale_description, p=0.7, list=FALSE)
model.trainset <- model.sample[model.partitions, ]
model.testset <- model.sample[-model.partitions, ]

model.rf <- train(locale_description ~ ., data=model.trainset, method='cforest',
trControl=model.cv,
metric='Accuracy', na.action=na.exclude, importance = "permutation")
```

Appendix F

Phase 1 Artifact

Appendix G

Phase 2 Artifact



Appendix H

Confusion Matrix for NCES Model Classifying Census Blocks by Locale Code

	11	12	13	21	22	23	31	32	33	41	42	43
11	**62988**	-	4	988	-	-	-	-	-	-	-	-
12	-	**23676**	56	3234	3049	-	-	-	-	306	273	-
13	19	44	**55495**	5240	6283	5434	26	-	-	3346	1676	-
21	711	3144	2563	**156328**	90	131	752	-	-	11714	888	-
22	-	999	1170	-	**15299**	-	-	-	-	23827	1542	-
23	-	-	591	185	-	**10928**	-	-	-	7187	831	-
31	-	58	6	1161	146	5	**31391**	79	-	6266	8530	-
32	-	-	-	-	15	-	54	**68220**	-	20978	54268	2697
33	-	-	-	-	-	-	-	-	**31583**	43163	25991	5677
41	-	701	560	5350	1057	1004	3008	10499	8485	**17376**	29743	5298
42	-	60	267	89	65	526	262	1120	61	32147	**143043**	27863
43	-	-	-	-	-	-	-	-	68	329	25387	**106876**

Appendix I

Delta of the Confusion Matrix and the Composition of School District Locale Classifications and NCES Census Block Classifications

	11	12	13	21	22	23	31	32	33	41	42	43
11	**-686**		1	-17						-314		
12		**-114**	-4	-561	-429					-3306	-4	
13	-9	-9	**-1532**	-426	-296	-404	26			-9061	-29	
21	-114	-318	-197	**-4797**		-17	-89			-14621	-47	
22		-207	-124			-672				16175	-64	
23			-37	-43		**-960**				220	-169	
31		-6		-126	-6	-3	**8495**	-15		-16489	-386	
32							-17	**-3668**		-25956	-1581	-128
33									**-1642**	24226	-823	-266
41		-137	107	-490	-93	-176	-183	-608	-775	**-16959**	-1111	-211
42		-16	-60	-14	-13	-82	-41	-113	-16	17221	**-2210**	-973
43									-1	-11	-787	**-958**