Defining Neighborhood for Health Research in Arizona

by

Tiffany 'Monicque' Lee

A Thesis Presented to the
Faculty of the USC Graduate School
University of Southern California
In Partial Fulfillment of the
Requirements for the Degree
Master of Science
(Geographic Information Science and Technology)

May 2019

DEDICATION

For my sons, J.R. and Wyatt; I hope this work will help you both understand that no matter what life throws at you, not to ever give up. To my husband J.R., thank you for your understanding, support and encouragement throughout this journey. To my Blue Dog, my unwavering study companion throughout all the very long nights…

You all are my motivators and my motivation.

Table of Contents

List of Figures

List of Tables

Acknowledgements

A very special thank you to John P. Wilson. It has been a great privilege to learn from you. Thank you for the opportunity.

I would also like to acknowledge all the faculty and staff that I have had the pleasure to work with throughout the GIST program at USC. It has been a challenging, at times intimidating, and all together amazing experience.

Thank you to all my Trojan family.

List of Abbreviations

ACS	American Community Survey
AIANNHAs	American Indian, Alaska Native, and Native Hawaiian Areas
AIRs	American Indian Reservations
ANSI	American National Standards Institute
BAS	Boundary and Annexation Survey
BIA	Bureau of Indian Affairs
DOI	Department of the Interior
EDGE	Education Demographic and Geographic Estimates
FIPS	Federal Information Processing Series
GIS	Geographic Information Science
GIST	Geographic Information Science & Technology
MAUP	Modifiable Areal Unit Problem
MCU	Measured Contextual Unit
MHI	Median Household Income
NCES	National Center for Education Statistics
SDTSAs	State designated tribal statistical areas
TCU	True Contextual Unit
TDSAs	Tribal designated statistical areas
TSAP	Tribal Statistical Areas Program
UGCoP	Uncertain Geographic Context Problem
UPOP	Uncertain Point Observation Problem
U.S.	United States

USC	University of Southern California
USPS	United States Postal Service
ZCTAs	Zip Code Tabulation Areas
ZIP	Zone Improvement Plan

Abstract

Defining place in health studies has been a crux for researchers as the definition of neighborhood is often regarded as adaptable to study needs and/or the preferences of the researcher. Health researchers commonly rely on measures of neighborhood that default to any number of predefined spatial administrative units, providing a relatively quick and cost-effective means to accessing and categorizing population data within a geographic area of interest. This approach to inferring population statistics assumes that median values for variables are relatively evenly disbursed across specific geographic areas of varying sizes.

This thesis explores how research outcomes may be affected by the choice of geographic reporting zones. The primary research goal of this study was to compare geographic reporting zones within the State of Arizona and to determine how the choice of neighborhood would influence the resulting values for three commonly utilized social determinants of health; median household income, numbers of children and the elderly, and the percent Native American population. This study used administrative boundaries at the county, census tract, and census block group levels from the 2000 Decennial Census and examined if and what variation occurred within the resulting outcomes for differing reporting zones within the State of Arizona.

The results of this thesis demonstrate that outcomes cannot be generalized across administrative units, that spatial aggregation will affect final outcomes, and that the choice of spatial reporting zone may produce widely different estimates for the same variable within a given geographic area. This thesis provides the foundation for future work investigating how choice of neighborhood can affect outcomes for small area studies and sets the framework for exploring what effects neighborhood definition might have on estimates of social determinants of health when proximity buffers are applied.

Chapter 1 : Introduction

People are constantly trying to define space. We assign invisible spatial boundaries and lines to the oceans, the Earth, and even the cosmos. Collectively, we tend to hold these boundary lines as quasi-physical representations of belonging. These artificial boundaries represent claim status, power, cultural identification, and give inhabitants a sense of place. Boundary zones are powerful proclamations representative of people and place. Historically, assigning boundary delineations to an area was an authoritative act. We are taught from an early age to acknowledge, generally respect, and not question these unseen lines of boundary delineation that fill our daily routines (Goodchild 2018). Some of these boundaries are often unknowingly assigned: such as census areas, ZIP codes, voting precincts and school districts. Some of these boundary lines are fixed; as in national and state borders. Some we simply accept; such as county boundaries, property lines and street networks. Other boundary lines are more unclear; for example, determining where the exact geographical break occurs along a demographic transition, or where a disease outbreak is likely to occur next. Science and governments are constantly using spatial delineations as measures of process. Possibly nowhere is the issue of boundary delineation more important than in the realm of spatial epidemiology and human health. Disease does not respect these quasi-physical boundaries of place - it is often indiscriminate without regard towards the places or people it affects (Flowerdew et al. 2008).

The merging of GIScience and epidemiology has given rise to a growing branch of health research examining where and how human well-being is affected by the local context which is often delineated by the surrounding spatial patterns and their boundaries. Geography is treated as a potentially direct correlate of human health and well-being in this view of the world by geographers, sociologists and other social scientists (e.g. Matthews and Yang 2013; Jankowska et al. 2014). Epidemiologists now regard GIS as a powerful tool in evaluating disease occurrence and transmission throughout specific areas. This new area of health research merging geography, sociology, and epidemiology is intent on delivering new methods for analyzing the correlations between population, health, and place (i.e. geography).

The emergence of GIS in health research has allowed researchers the ability to reconceptualize boundary delineations and evaluate subjects in localized areas, especially as it applies to examining the health outcomes connecting people and the environment. Health researchers designate small area boundaries encompassing a subject of interest as a given study zone. These spatial zones, known as neighborhoods, are vitally important for understanding population health patterns as a function of location.

The conceptualization and measurement of neighborhood has yielded a dilemma of grand proportions for researchers. An individual's perceived definition of neighborhood often varies significantly from a researcher's delineation of the same vicinity. Within health research, this meaning of locale can vary greatly from one study to another depending on what measures and considerations are selected in defining that neighborhood.

Selecting data derived from differing measures or interpretations of a neighborhood, or points within, can sometimes generate substantially different results yielding uncertainty in the final reporting. The attempt to circumvent reporting uncertainty has led researchers on a quest for the ultimate method in depicting the representativeness of a population. Multiple problems arise in data reporting when spatial linkage inferences between and within boundary zones and point observations are inappropriately applied. Researchers have attempted to use numerous methods including multilevel analysis as a means of dealing with this uncertainty; however, although beneficial in addressing some aspects of uncertainty, all forms of spatial analysis are still subject to classical and emerging spatial problems requiring deliberation in study design and/or data analysis (Robertson and Feick 2018).

As the scale or extent of a study area changes, corresponding details of that area also change. For example, when the scale of a study area gets larger, some details become indiscernible (Openshaw and Alvanides 2001). Also, moving a neighborhood boundary line may change the dynamics of the area under study causing a change in demographics and geography that may affect the relationships being examined and thus the resulting outcomes (Foster and Hipp 2011). The reporting issues caused by utilizing different spatial areas of different sizes and/or scales is commonly recognized as the modifiable areal unit problem (MAUP) (Tatalovitch et al. 2006; Swift et al. 2014; Robertson and Feick 2018).

Sharing some likeness to but separate from the MAUP are the problems associated by using static boundaries for analysis. Constraining the measure of an individual within a health study to a predefined static boundary, often assigned by an administrative unit, does not represent the true measure of a subject's dynamic traverse and exposures through time and space (i.e. potentially spanning multiple administrative boundaries over a given temporal period). This inferential error results in a measurement problem recognized by researchers as the uncertain geographic context problem (UGCoP) (Kwan 2012b; Robertson and Feick 2018).

Further complicating the issue of data uncertainty caused by spatial misreporting is an issue brought to the forefront by technological advances and public accessibility to GIS applications. This problem, the uncertain point observation problem (UPOP), results from misreporting point locations and subsequently incorrectly linking them to areal units enabling users to develop aggregated spatial inferences erroneously (Robertson and Feick 2018).

The conventional assignment of subjects to a predefined measured contextual unit (MCU) (e.g. administrative boundary) is not usually indicative of the true contextual unit (TCU) affecting those individuals and thus potentially creates ecological or atomistic fallacies within the final reporting. Ecological fallacies make incorrect inferences about individuals based off aggregated measures at the 'group-level'; whereas, an atomistic fallacy makes incorrect assumptions about an aggregated population based from measures at the 'individual-level.' Any time an analysis utilizes data between higher and lower level aggregations, that data becomes vulnerable to some form of fallacy (Kwan 2012b; Robertson and Feick 2018). The following diagram (Figure 1) demonstrates the interrelationships that exist within various methods of neighborhood delineation and analysis.

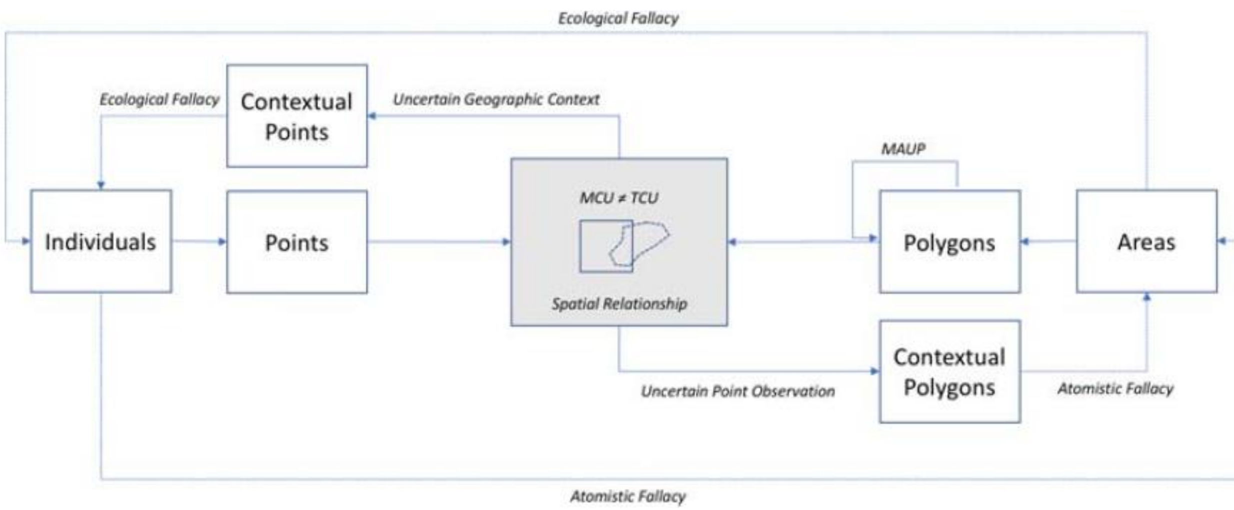

Figure 1. Relationships between issues affecting the spatial analysis of neighborhood (Source: Robertson and Feick 2018)

1.1. **The Ecological Fallacy and Modifiable Areal Unit Problem**

There are misreporting issues to consider when analyzing discrete socioeconomic and demographic data collected at an administrative level yet applied to individuals in small area studies. The resulting uncertainty creates an ecological fallacy which is found in data where assumptions for the microscale (i.e. individual) are collected at the macro scale (i.e. 'group level') but inferred per individual (Swift et al. 2014; Robertson and Feick 2018). This leads to the necessity to account for both micro- and macro-levels with some multilevel analysis to bridge the gap and circumvent or minimize ecological bias (Schule and Bolte 2015; Strominger et al. 2016; Robertson and Feick 2018). The ecological fallacy, however, is also present in multilevel analysis which affects both static and dynamic neighborhood evaluations that use aggregated data for individual inference (Figure 1). This renders both the modifiable areal unit problem and the uncertain geographic context problem vulnerable to the ecological fallacy (Robertson and Feick 2018).

The misreporting issues that are introduced by utilizing data from differing spatial aggregations is commonly recognized as the MAUP (Tatalovich et al. 2006; Swift et al. 2014; Robertson and Feick 2018). This fact and the accompanying complexity render acknowledgment of the MAUP within health studies to often be little more than a footnote or caveat (Root 2012; Swift et al. 2014). The practice and problem of ignoring the MAUP and utilizing aggregated data is that information is lost in the process of aggregation which results in misreporting (Openshaw and Alvanides 2001). The lack of consideration of the MAUP can undermine otherwise sound methodology simply by the uncertainty that might be introduced by not accounting for it.

Some spatial analysis methods are superior in escaping or minimizing the MAUP. The often-cited examples include political administrative units which utilize zone aggregation and 'official zoning systems,' such as the hierarchical census units consisting of blocks, block groups, and tracts used within the U.S. (Openshaw and Alvanides 2001).

The MAUP is additionally complicated by the need to also consider the potential effects of the UGCoP and the UPOP (Robertson and Feick 2018). Although similar in context, both are unique aspects of neighborhood delineation and each requires specific consideration (Kwan 2012b).

1.2. The Uncertain Geographic Context and Uncertain Point Observation Problems

The UGCoP is to be considered when examining neighborhood because the boundary and scale issues addressed by accounting for the MAUP do not account for the contextual uncertainty associated with the dynamic activity space of an individual within a given neighborhood's population. The definition of the neighborhood by administrative boundaries or buffers does not account for the social characteristics within neighborhoods that can influence health outcomes (Kwan 2012b). Accounting solely for the MAUP can still distort results when the potential effects of the UGCoP are ignored because the individual mobility behaviors within a population are rarely limited to conventionally used neighborhood boundaries. People's daily routines often span multiple types of boundary areas within a given spatiotemporal period. Throughout the course of a day spatial behaviors are diverse and dynamic frequently spanning or are affected by multiple locales of different administrative boundary types, often simultaneously (Matthews 2011, 35-54).

Recent research has demonstrated that measures of an individual's actual daily traverse usually cover greater geographic areas than the conventional static neighborhood definition that is used. The result of not accounting for the UGCoP is often misreporting due to mismatched units of analysis, inadvertently omitting important exposure information, and introducing the ecological fallacy. The typical geographic boundaries used to delineate neighborhoods can over- and/or under-estimate the real exposure effects experienced by an individual's activity space. Additionally, an individual's activity space consisting of their daily routine, and less frequent trips, may significantly affect that person's perception of neighborhood. Neighborhood perception is an important consideration for researchers as it can provide insight into potential exposures occuring throughout a person's daily routine that may not be adequately represented by analyzing spatiotemporal traverse patterns alone (Chaix et al. 2009; Kwan 2012b; Jankowska et al. 2017).

Equally important is the spatial uncertainty introduced by the UPOP. The error presented by the UPOP occurs with a mismatch in point observation location data between the MCU and TCU reported at the individual level as a basis for group level inferences, making UPOP subject to the atomistic fallacy (Figure 1). The UPOP can result in incorrect aggregated population characterizations by introducing error and uncertainty in the association between where and how a point observation is created and connected to a neighborhood (Robertson and Feick 2018).

Another consideration in identifying how a neighborhood is to be defined for a study, yet often unacknowledged across health research, is that just as the dynamics of people change over a diurnal period, so do places. The human dynamics influencing an area such as a city street or park may be very different over the course of 24 hours causing a range of potential exposures. The issue of time and space, and the activity of both people and places, can greatly affect exposure outcomes within a given neighborhood. The often overlooked effects of temporal exposures over hours, days, seasons, and/or years should be considered when selecting how a neighborhood is to be determined (Xia et al. 2006; Matthews and Yang 2013). Often the physical neighborhood dynamics of place are not accurately represented using data from administrative sources, leaving the spatiotemporal aspect and effects of place unacknowledged (Yiannakoulias 2011; Delmelle 2016).

1.3. Rurality

Another substantial challenge in designating neighborhood within health studies has to do with the special characteristics of rural areas. Rurality is a challenge for health researchers examining environmental contextual influences. Rural communities often have low population densities, generally isolated within or disbursed across larger geographic regions. Infrastructure measures vary, and frequently street networks are minimal. Administrative boundaries are commonly used as aggregated proxies of rural neighborhoods and do not reflect the geographic actuality of a local area. Small area studies in rural settings are prone to ecological bias caused by aggregation issues, scale effects and inconsistent measures and considerations of neighborhood (Rousseau 1995).

1.4. Congruence

Incompatible and mismatched geographic units present a substantial problem for researchers and complicate the conundrum caused by spatial problems and fallacies. Administrative boundary types vary in size, can be nested within one another, might overlap, and sometimes share little to no commonality with one another – yet all are used to represent place. When data from differing types of boundary systems are overlaid and used together (e.g. spatial interpolation), data is lost or erroneously created resulting in misreporting and uncertainty. These geographic zones have been created for various distinct purposes by differing governmental agencies, sometimes arbitrarily, and facilitate the most convenient and sometimes only means of data acquisition for researchers (Eagleson 2002; Mennis 2003).

To complicate matters, the selection of these units for health research is often equally arbitrary. Furthermore, researchers may be limited in data availability for a given area, requiring the use of a geographic unit that might not be the most representative of a subject (Diez Roux 2001; Eagleson 2002; Chaix et al. 2005; Santos et al. 2010; Robertson and Feick 2018).

1.5. Thesis Objective

This thesis set out to examine and compare the descriptive statistics which result from using aggregated data for various neighborhood contexts, similar to what is commonly done in current health research. The best available geographic units were used to generate measures of neighborhood and the potential pitfalls that are likely to follow from using this approach to measure geographic variability for a large region.

The State of Arizona was selected because of its size, socioeconomic and demographic diversity, and the various population reporting methodologies used for different sub-populations in the state. Arizona's population dynamics with its urban and rural distribution and unique racial and ethnic diversity offer numerous sampling scenarios. Data were derived from Federal and State reporting for the general population as well as American Indian reporting areas.

1.6. Thesis Organization

The next chapter, Chapter Two, details the most common methods used to define neighborhood in health research within the U.S. This related work sets the framework for the neighborhood differentiation methods that were used in this thesis. Chapter Three details the methodology for how this study was conducted. This chapter describes the data sources and spatial analysis that was performed. Chapter Four details the results and how the variables of interest varied across the state as a function of both the geographic variability and/or the ways in which geographic units for collecting and representing people in various parts of the state varied from one another. Finally, Chapter Five discusses the significance of these results and the implications for planning neighborhood selection strategies in future work.

Chapter 2 : Related Work

The integration of GIS into epidemiology and health research has become a powerful instrument for analysis and demonstration of the spatial distribution and movement of populations, disease, health services and their infrastructure. GIS platforms have allowed health researchers the accessibility to spatial tools that readily enable the measurement of various dimensions examining the relationships between health and place (Rushton 2003). The area attributes of a place, in health research, are often categorized within the boundaries of a neighborhood defined as some measurable assignment to a geographic locale (Diez Roux 2007). The method of neighborhood delineation has created a quandary for researchers as the measure and use of a neighborhood, and its relevance to a given topic, is arbitrarily determined within a study design and non-specific within any given scientific field (Boscoe and Pickle 2003; Duncan et al. 2014; Perchoux et al. 2016). Researchers have struggled to determine what the best means of delineating neighborhood are, as different definitions are known to present varying outcomes even within the same vicinity and which method to use for a given study is customarily left to the discretion of the researcher (Diez Roux 2001; Swift et al. 2014). The majority of studies that have examined the role of boundary selection for neighborhood have focused on pre-determined administrative units which are not necessarily representative of the population and/or area of interest (Diez Roux 2007; Santos et al. 2010, Foster and Hipp 2011). Recent research into neighborhood definition has also examined the role of the perceived environment in defining neighborhood. Such studies are subjective, examining the activity spaces of an individual and utilizing self-reporting from subjects and ground truthing within an area of interest (Perchoux et al. 2016; Kirby et al. 2017). The validity of the classical fixed boundary study is now challenged against sliding boundary studies, and the quest and argument continue as to how a neighborhood should be defined (Chaix et al. 2009). Few studies, however, have examined the variation and problems caused by this fluid definition of neighborhood. This chapter sets out to review the methodologies used for neighborhood delineation and explores the various issues associated with the selected definitions.

2.1. Administrative Boundaries

Administrative boundaries are government-defined and universally recognized jurisdictional divisions of geography (Chang 2010). How a type of administrative boundary zone is defined varies from one country to another, but their purposes are the same and serve as a way for political administrations to organize infrastructure, political elections, community and emergency services, and often by happenstance, the contextual data residing within geographical units (Sabel et al. 2013). Administrative boundaries have been predominantly used as a measure for and proxy of neighborhood for health studies (Diez Roux 2007; Santos et al. 2010). In the U.S., administrative boundaries are categorized by State, County (or Parish, the geographical equivalent to a county, within the State of Louisiana), District (school district, voting precinct, and separate judicial district), ZIP code and Census (Mu et al. 2015). For this thesis, administrative neighborhoods are examined at the county, census, and American Indian Area levels (Figure 2).

Health researchers are often limited to using data collected at the administrative level for small area studies such that neighborhood effects are aggregated using differing geographic scales and extents (Leite et al. 2015). The use of data collected and analyzed at different scales within a study may become subject to both the ecological fallacy and the MAUP by introducing significant potential bias into a given study (Tatalovich et al. 2006; Swift et al. 2014).

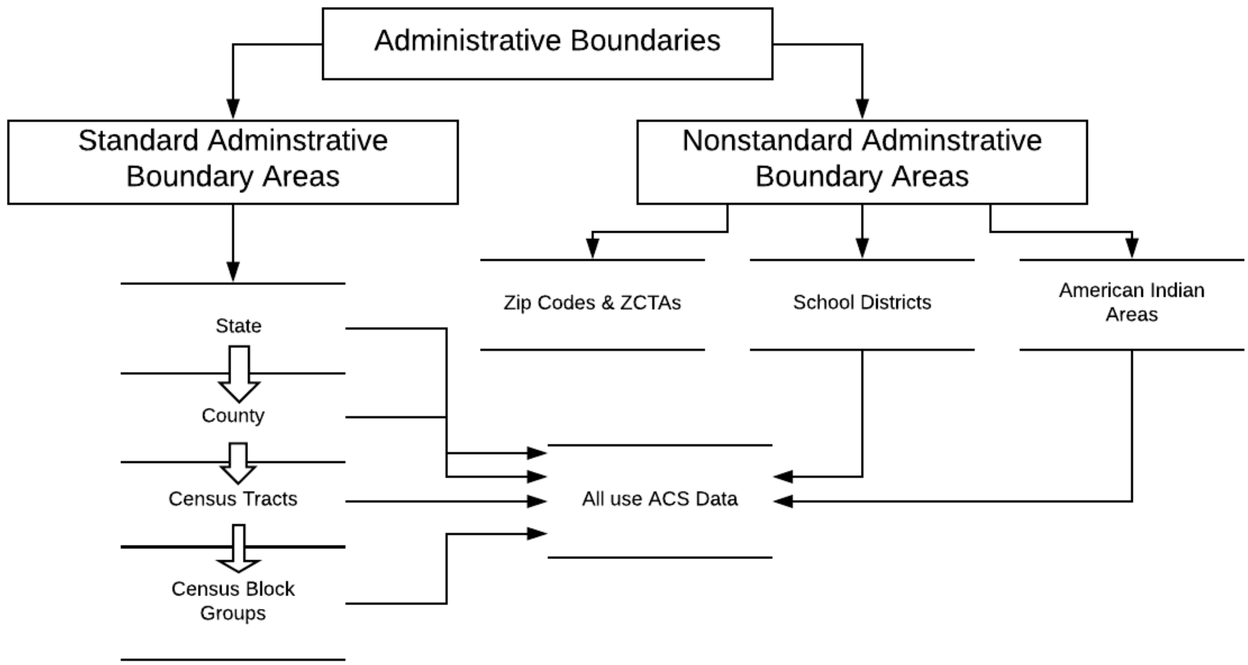

Figure 2. Administrative boundary types discussed in this thesis

These fixed boundary zones can be additionally problematic as they regularly are not representative of the true contextual environment and often do not correlate to the actual study area of interest (Chaix et al. 2005; Tatalovich et al. 2006; Root 2012; Perchoux et al. 2013). Population distribution and geographic obstacles, such as water and terrain or barriers created by the built environment, may have significant effects on data reported for a given administrative boundary (Santos et al. 2010).

Administrative boundary areas have further limitations as these zones are generalizations of the populations they encompass at each level and they may not account for the within zone variability. Furthermore, scale plays a critical role in the final reporting, especially in small area studies. Leite et al. (2015) states that it has been speculated that the scale used in a study often dictates the population inferences determined by researchers, where health is reflected as being a function of social and economic factors in larger reporting areas, and a result of the individual in small areas. Small areas are susceptible to small, limited sample size which can misrepresent and misreport the actual variance in a population by potentially leaving out people within that population that fall outside the sampling such as minorities, children, and the elderly (Institute of Education Sciences, National Center for Education Statistics 2018b). Administrative boundary units may not convey the actual geographic and human variability of a neighborhood (Santos et al. 2010).

2.1.1. State and County Level

Within the administrative boundary system of the U.S., state and county territories are the main geographic building blocks for data reporting from the U.S. Census (US Census Bureau 1990). State borders are temporally static boundaries. These lines of delineation do not move and are not adjusted over time to account for population and/or urban expansion. County lines, however, are not always permanent and their boundaries can change over time. Counties can experience ‘substantial changes’ where borderlines are moved, also new counties can be created, or existing counties ‘deleted’ (U.S. Census Bureau 2016). Data collected and standardized at state and county levels are often not practical for small area studies because of the introduced ecological fallacy (Swift et al. 2014; Robertson and Feick 2018).

2.1.2. Census Level

The U.S. Census is a decennial survey that counts every resident of the country by constitutional mandate. Counts and demographic information are collected by a combination of mailed questionnaire and ground truthing by census workers to ensure complete (100%) coverage of all areas. Census information is compiled nationally at block, block group, and tract levels (U.S. Census Bureau 2010c).

Census data is used for statistical purposes by both the government and the public, private, and not-for-profit sectors. The government uses census data to identify geographical areas in need of infrastructure and community services and to allocate federal support programs and funding for those communities deemed as worthy recipients by the reported population statistics. The Census is also used to determine how many seats within the House of Representatives each State can hold. Census data is available publicly; however, personal information reported per household address is not available to the public. Only count data per administrative unit is available for “public” use (U.S. Census Bureau 2010c).

In 2010, the U.S. Census Bureau revamped their questionnaire procedure changing the way population data is gathered and reported. Previously, the decennial census utilized both short and long forms of questionnaires to collect population information. The short form contains basic demographic questions consisting of ‘name, sex, age, date of birth, race, ethnicity, relationship and housing tenure.’ The long form questionnaire was sent by random sampling to an estimated one of every six households and contained additional detailed socioeconomic questions (U.S. Census Bureau 2010c, 2018a).

In 2010, the U.S. Census switched to using only the short form and eliminated the long form. In place of the long form the U.S. Census implemented the American Community Survey (ACS), enacted in 2005, that collects detailed demographic and socioeconomic data monthly, compiled and reported annually, with the goal of collecting continuously updated population information for all areas across the U.S. The ACS questionnaire uses random sampling throughout the U.S. to collect detailed population information from a ‘small percentage’ of households and does not collect data from the same household more than once every five years (U.S. Census Bureau 2014, 2018a).

The ACS has become an important instrument for the U.S. Census Bureau's efforts to maintain up-to-date, continuous, population estimates and is a valued resource for providing small area statistics to data users. Census statistics are ever-changing however, and data from one reporting decennial census to another can produce substantially different estimates. The temporal challenges with the decennial census means many have come to regard the ACS as the solution because the data presented provides a snapshot of dynamic population information available for small area evaluations during the time between census reports. The issue of sample size is not often considered when ACS data is reported or utilized. ACS data is the result of a small percentage sampling versus the decennial census complete population sampling. The data collection methodology for the ACS creates an inherent atomistic fallacy within the data reported by how it uses non-majority data and creates statistical assumptions for the whole based on those values (Sabel et al. 2013; U.S. Census Bureau 2014; Roberson and Feick 2018).

The ACS provides researchers a convenient option to review annual population estimates for small areas at the 'census tract and block group' levels. These small area statistics were previously only available from the decennial census report. The adoption of the continuously collected ACS, in place of the traditionally used long-form collected once every ten years, is often used by researchers as an interim resource for representing changing population conditions across the country in between decennial reporting's, notwithstanding potential sampling bias (U.S. Census Bureau 2014).

Census tracks are nested within counties and are the larger of the two divisions. Census tracts generally follow observable physical characteristics and consist of populations ranging from 2,500 to 8,000 residents (US Census Bureau 1990). Census tracts were initially designed to represent areas with similar socioeconomic and demographic characteristics. Since census tract inception, however, population dynamics throughout the country have changed resulting in census tracts that are no longer homologous relative to one another.

Census block groups are nested within census tracts and represent the smallest published spatial division of the Census, with populations under 2,500 residents. These areas often have physical geographic boundaries and have been considered valuable for their use in small area studies (US Census Bureau 1990). Census block groups are often thought to be representative of neighborhood; however, their boundaries are administratively defined and may have no relevance to the local culture or functionality of a specific zone.

Matthews (2018) has noted that 'there is a centralized thought throughout the health sciences that census data defines neighborhood.' This generalization has led researchers to use census data as representations of place in health studies with little regard to the actual functional neighborhood. Census areas are relatively arbitrary with little regard to the actual practicality, contextual effects, or social structure within a neighborhood creating potential misreporting and error when used in health research (Sharkey and Faber 2014).

The ability to access a wide array of population data at these predefined administrative units makes census and ACS information the most easily accessible and cost-effective data resource for health researchers (Foster and Hipp 2011; U.S. Census Bureau 2018a). Census and ACS data are often used as a rudimentary proxy for neighborhood. Although Census data is considered 100% coverage, it is assigned to administrative census units that are subject to MAUP effects, miscount, and other non-sampling errors. The ACS is also assigned to administrative census units; however, in contrast to decennial census reporting, the ACS uses a relatively small sampling of the entire population introducing potential misreporting error. and although the ACS is commonly used for population insight between census reports it may be only somewhat representative of current population dynamics. The ACS sampling methodology utilizing small sample size also means some measures of population characteristics may not be reliable making the resulting data subject to the atomistic fallacy (Spielman and Logan 2012; Robertson and Feick 2018). Additionally, administrative census units are subject to boundary adjustments as their populations change between census periods creating incongruent boundary zones. Utilizing aggregated census and ACS data can lead to misreporting caused by the bias introduced from zonal effects, the MAUP, the atomistic fallacy, and by using estimated and continuously changing population information (Openshaw and Alvanides 2001; Diez Roux 2007; Kwan 2012b).

2.1.3. 'Nonstandard' Administrative Unit Areas

The U.S. Census Bureau also tabulates data for administrative units that fall outside of the conventional census grouping levels. These special localities have been created for different reasons than standard census areas and consequently share no geographic connection to administrative census boundaries. These special zones were created either independently from or in joint effort with the U.S. Census Bureau; however, the U.S. Census Bureau does compile data within these areas (U.S. Census Bureau 1990).

2.1.3.1. ZIP Code

Zone Improvement Plan (ZIP) code divisions are completely independent of other types of administrative boundaries. ZIP codes do not correspond with census tracts but instead were devised by the U.S. Postal Service (USPS) as a means of identifying the primary mail delivery system area associated with a municipality or postal office. The system was devised by the Postmaster General to accommodate increasing mail flow to growing populations across the U.S. and follow a network of mail delivery routes throughout a neighborhood (US Census Bureau

2018c). ZIP codes evolve with a neighborhood as it grows and are an independent zoning system that are frequently spatially incompatible with other types of traditional administrative boundaries.

ZIP Code Tabulation Areas (ZCTAs) are reported by the U.S. Census Bureau as spatial delineations of USPS routing zones. These reporting areas are designated by determining the most recurrent ZIP code inside of a census block and then delegating it to the entire census block. ZCTAs are often different than the actual ZIP code they represent. Additionally, very rural areas may not have a ZIP code assignment, leaving a void in ZCTA coverage (US Census Bureau 2018c).

2.1.3.2. School Districts

School districts are administrative units created and determined by local governments and maintained in joint effort between the U.S. Census Bureau and the National Center for Education Statistics (NCES), a part of the Institute of Education Sciences within the U.S. Department of Education (U.S. Census Bureau 2018b; Institute of Education Sciences, NCES 2018c). School districts are independent from standard census areas and serve as a demographic and economic measure of the school-aged population inside the geographic area within district boundaries. School district boundaries are updated biennially by the U.S. Census Bureau's Geography Division for socio-demographic reporting and are used to create spatial layers for census TIGER/Line files (U.S. Census Bureau 2018b).

School districts are boundary 'catchment' areas that dictate which area schools residents can attend (Institute of Education Sciences, NCES 2018b). The NCES uses school district data compiled from both the ACS and the decennial census for their Education Demographic and Geographic Estimates (EDGE) program (Institute of Education Sciences, NCES 2015).

The EDGE program uses geographic information to inform policy makers and the public about the correlation between schools and the people and area they serve. The EDGE program compiles data to create school district locales, used to classify school districts into four categories; urban, suburban, town, or rural and three subsequent subtypes. The function of the EDGE locale is to provide spatial data for research and analysis that gives researchers the ability to customize detailed investigations of the dynamics occurring between the social and physical properties of a locale area (Institute of Education Sciences, NCES 2015).

The EDGE locale boundary function also offers a ZCTA locale file where researchers can assign 'NCES indicators' to ZCTAs for use with TIGER/Line files. The ZCTAs are determined by the boundaries of the NCES locales (Institute of Education Sciences, NCES 2018a).

School districts are not static and are relatively arbitrarily defined. They often conflict with other types of administrative units such as counties, census tracts, and ZIP codes. The potential incompatibility of school districts with other administrative unit types causes school districts to sometimes have multiple, simultaneous, spatial unit assignments, i.e. spanning multiple counties and/or having varying ZIP codes within the same school district (Institute of Education Sciences, NCES 2018c). The use of mismatched spatial units may automatically introduce bias by potentially creating or eliminating important data (Eagleson 2002). Combining potentially mismatched spatial units with aggregated ACS data adds to uncertainty and creates significant possibilities for misreporting and error.

2.1.3.3. American Indian, Alaska Native, and Native Hawaiian Areas (AIANNHAs)

The U.S. Census Bureau collects and provides population data for Native American Indian lands situated throughout the U.S. as well. The Bureau of Indian Affairs (BIA), a branch of the U.S.

Department of the Interior (DOI), keeps a record of American Indian tribes and maintains governance over tribal affairs. The U.S. Census Bureau compiles population data from Indian area census reporting zones for the collection and quantification of populations within tribal geographic boundaries and off-reservation trust lands (U.S. Environmental Protection Agency 2016; U.S. Census Bureau 2010a).

Tribal areas known as American Indian Reservations (AIRs) and off-reservation trust areas are Federally and/or State recognized Indian use territories independent from any other form of administrative boundary. The geographic boundaries of AIRs reside within the borders of the U.S. and were determined on an individual tribal basis by final tribal treaties or judicial orders agreed upon with the appropriate State Government(s) and the Federal Government. Each tribe has governing authority over their people which consists of their independent tribal governments, local laws, and boundary zones. The U.S. Federal Government maintains ultimate federal jurisdiction over AIRs (U.S. Census Bureau 2010a).

Federal AIR borders can span all forms of classical administrative boundaries including state and county borders (U.S. Environmental Protection Agency 2016). State AIRs have a government-appointed intermediary that determines and reports State recognized tribal boundaries to the U.S. Census Bureau. State AIRs cannot cross State boundaries but can cross county lines (U.S. Census Bureau 2010a)

The U.S. Census Bureau works with tribal government agencies to annually identify and update reservation boundary lines and features through the Boundary and Annexation Survey (BAS). The main function of the BAS is to make certain that legal tribal boundaries are documented correctly so that population data is accurately recorded and reported to local, tribal, county, and federal agencies (US Census Bureau 2017).

The Tribal Statistical Areas Program (TSAP) is a decennial survey that provides tribes the option of delineating physical boundaries within their specific AIR to create tribal Census tabulation areas. TSAP data is utilized by the U.S. Census Bureau and the American Community Survey (ACS) to provide statistical socioeconomic and demographic data for tribal, federal and state agencies. Tribes can designate boundary lines creating the reporting areas for each State identified tribal statistical area (SDTSA), Tribal designated statistical area (TDSA), Tribal census tract, and Tribal block group (U.S. Census Bureau 2010b; U.S. Department of Homeland Security 2017).

The U.S. Census Bureau designates a unique numeric tribal census code, alphabetically by tribe, for each federal AIR across the U.S. independent from any non-reservation coding system. Additionally, Federal Information Processing Series (FIPS) and American National Standards Institute (ANSI) codes are also assigned, and are unique, to each tribe by State. Although appointed alphabetically by State, federal AIRs for the same tribe can have completely different FIPS codes if they cross a State line (U.S. Census Bureau 2010a).

2.2. **Buffer Zones**

GIS serves as a valuable tool for facilitating the ability to measure proximity around subjects of interest (Perchoux et al. 2013). Proximity zones measure the extent of a neighborhood's environment and commonly use buffers as a measure of locality (Faber and Sharkey 2015). Buffers allow researchers flexibility to examine neighborhood exposures at an individual level in contrast to being constrained by data aggregated at the scale of traditionally used administrative boundaries (Spielman and Logan 2012; Matthews and Parker 2013).

There are several types of buffers used in health research. Three of the more common buffer methods are circular, street network, and activity space. Figure 3 compares circular and network buffers (Chaix et al. 2009). Ego-centric buffers are often considered the best approach at representing a subject's immediate neighborhood relying on aggregated proxies combining contextual characteristics of place with an individual's geocoded location as a representation of physical address (Spielman and Logan 2012; Matthews and Parker 2013).

Within these different types of buffer zones, boundaries may be sharp or fuzzy depending on methodology (Chaix et al. 2009; Spielman and Logan 2012). Buffer sizes are subjective and are often determined by following those used in prior research. Measurement within a buffer zone may be linear or not and may not always account for obstacles created by physical surroundings (Leite et al. 2015).

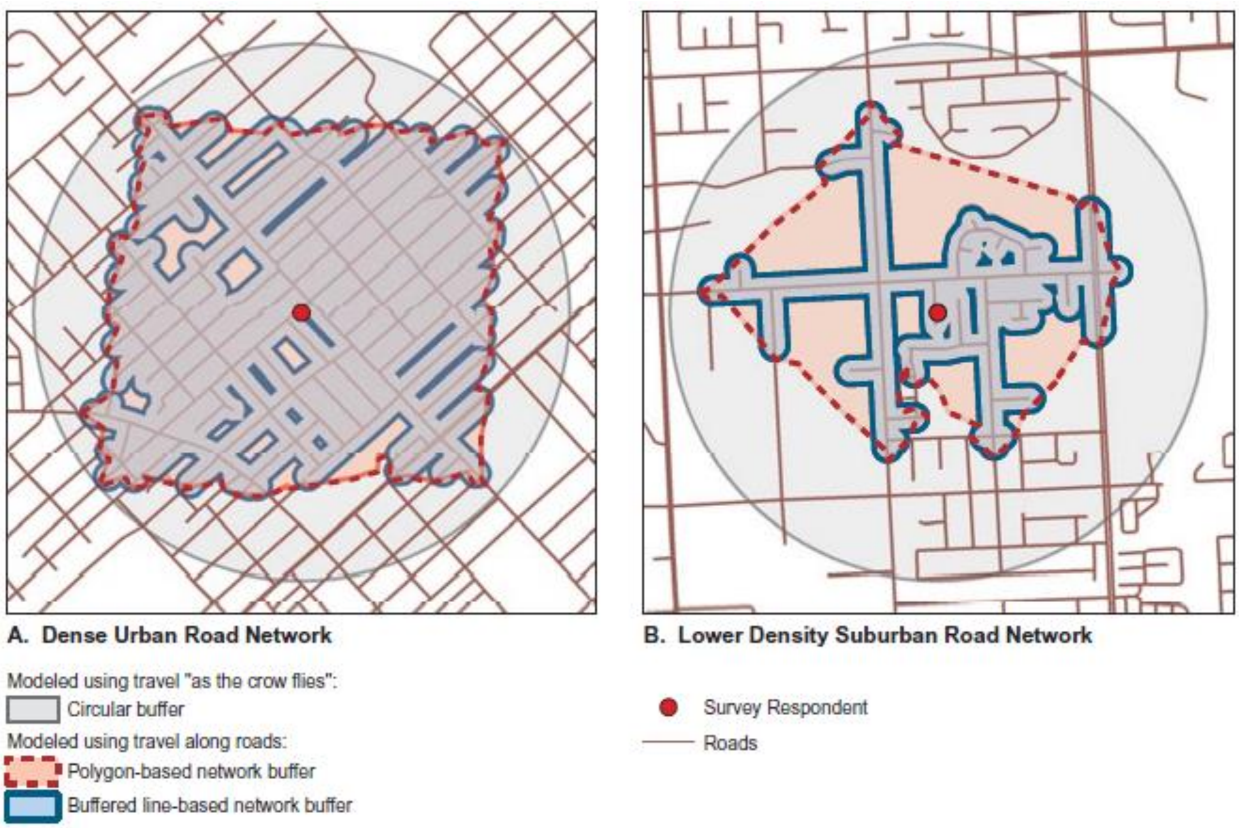

Figure 3. Circular and network buffers around a point of interest (Source: Hall et al. 2007)

2.2.1. Circular

Circular buffers create an isotropic neighborhood defined by a specified distance extending out equally in every direction from the center (Chaix et al. 2009). Frequently, circular study zones are used in health research as a means of quantifying factors within a specified boundary distance. These ego-centric zones are often a preferable, all-inclusive, measure of the relevant neighborhood and are used as an alternative to fixed administrative delineations representative of a locale (Perchoux et al. 2013).

Circular buffer sizes are not standard or fixed, allowing researchers to adjust the sizing to whatever threshold deemed most appropriate for a study (Hall et al. 2007). There is no consensus as to what the buffer size should be; however, recent health studies have used a radius of half a mile around a specific point of interest (Perchoux et al. 2013). Research has shown however that the buffer size selected directly affects the reported functionality of a buffer zone for land use (Strominger et al. 2016).

Circular buffers around a point of interest can introduce bias into a study depending on what type of contextual variable is being measured. Isotropic neighborhoods can be valuable for examining environmental and epidemiological exposures around a point of interest; however, they may have limited efficacy in assessing health issues influenced by an individual's daily traverse.

Often circular buffers utilize Euclidian distance as a measure of contextual variables within a neighborhood. This straight-line distance is measured from point A to point B and commonly referred to 'as the crow flies' which generally does not account for barriers introduced by physical geography or the built environment. Many of the aspects relating to health research do not pertain to exposures occurring 'as the crow flies' (Chaix et al. 2009; Hall et al. 2007). Euclidian distance does not account for the interaction between population and physical environment; however, its usefulness and relevance depend on what proxy is being calculated. Notwithstanding these potential problems, health research frequently uses circular buffer areas and straight-line distance as a measure of proximity (Foster and Hipp 2011). The built environment, street networks, and physical geography can directly affect the walkability and otherwise traversable extent of a neighborhood (Boruff et al. 2012; Perchoux et al. 2016).

2.2.2. Network

Street network buffers are line-based and defined by some measure of the road system from a feature of interest (Hall et al. 2007). Street network buffer sizes, like circular buffers, are not standardized and threshold distances vary depending on study purpose. Areas experiencing higher population densities have demonstrated the need for an increase in buffer size compared to lesser populated areas (Perchoux et al. 2016).

In contrast to circular buffers, street network buffers are theorized to be a more accurate representation of neighborhood since they provide a course of navigable passage around and through the physical and built environment (Hall et al. 2007). Street network buffer zones are often considered a more human-oriented means of measuring neighborhood as they reflect routes of regular commute and use, translating to routine environmental exposures at an individual level (Perchoux et al. 2013).

Past research has attempted to operationalize network buffers by street pattern, viewing major roadways as buffer zone borders and minor road systems as potential routes of traverse within a neighborhood. Use of this methodology, although centrally important to the concept of neighborhood and the use of street networks as buffer zones for analysis, is considered a building block for delineating neighborhood units (Cutchin et al. 2011). It is theorized that using minor street network connectivity as a measure of social interactions representative of the neighborhood will reduce measurement bias by minimizing intra-neighborhood variability and amplifying neighborhood differentiation (Foster and Hipp 2011).

Network distances in urban settings often use a grid-based, Manhattan method, of distance measurement where the distance between points A and B is determined using right angles along an axis or on a grid. Manhattan distance is different from Euclidean and network distance because Manhattan distance is the measure of the two right-angle sides of a triangle (Apparicio et al. 2017).

This method is commonly used to measure two points on a municipality map and often represents 'city block distance' (Charreire et al. 2010). Grid networks can be a useful measure of traverse and walkability in certain environments; however, their usefulness is limited to regular and relatively close urban street networks. When used in environments with sparse or irregular street networks, grid networks can introduce substantial measurement error (Hall et al. 2007).

Street-based buffer zones are often identified as walkability zones, usually determined from preceding studies to be within a 15-minute walk from a residence or feature of interest, translating to roughly 1,000 m from the point of interest (Hall et al. 2007; Perchoux et al. 2016).

Street network buffer distances are commonly determined by a researcher and subsequently, they do not provide accurate representations of the actual physical area, or direction, most frequented by a resident. Street network buffers have a multitude of applications; however, they are not always realistic of an individual's daily traverse exposures or the anisotropic nature of human routine (Chaix et al. 2009; Perchoux et al. 2016).

2.2.3. Activity Space & Perceived Environment

One of the primary limitations to the classical measures of neighborhood is their inability to represent the true dynamic nature of the human element. Traditional measures of neighborhood focus on place of residence and are defined by either fixed administrative boundaries or set buffers theorized as representative of a resident's daily exposures. This conventional view of neighborhood neglects that the human routine is not static but more often variable and changing (Macintyre et al. 2002; Matthews and Yang 2013).

Twenty-first century society has created an increasingly mobile culture (Chaix et al. 2009; Matthews and Yang 2013). Economy and opportunity have added to the transient nature of residence. The incorporation of and advancements in GIS and related geospatial technologies now provide health researchers with new methods of capturing the complex dimensions of life outside an individual's dwelling (Perchoux et al. 2013).

The concept of activity space follows a person's daily traverse, spatially and temporally, assessing direct and potential exposures relative to that individual (Kwan 2012b; Perchoux et al. 2013). The range of actual exposures an individual experiences throughout their average daily routine and traverse are often far greater than what has been deemed by classical measures of neighborhood (Chen and Kwan 2015).

Activity spaces tend to demonstrate a directionally oriented neighborhood composed of residential, public, and institutional places. Activity spaces vary by age group and residential type based on socioeconomic status (Perchoux et al. 2013). Current methodologies commonly utilized in recording an individual's routine activity space include GPS tracking, reporting through questionnaires, self-reporting, or other means of volunteered geographical information (Matthews and Parker 2013; Perchoux et al. 2013). Geoprocessing tools are then employed to evaluate data and create representative activity spaces (Perchoux et al. 2013).

The perceived neighborhood can be highly subjective often differing from an actual traversed activity space by encompassing areas thought of as preferable even if outside of a person’s true realm of daily use (Flowerdew et al. 2008; Perchoux et al. 2016). The issue of differentiated geographic delineations between individual activity spaces is further complicated by the temporal aspect affecting a perceived or actual neighborhood (Kwan 2012b). The variability of this form of spatial evaluation creates scale and zonal effects which can introduce congruence problems and aggregation bias (Swift et al. 2014).

2.2.4. Dasymetric Mapping

Dasymetric mapping is an area-based mapping technique that constructs population information from multiple aspatial, areal, and linear datasets (Holt et al. 2004; Swift et al. 2014). Dasymetric maps use ancillary data to create dimensional zones of measure disaggregated from the confines of administrative boundaries. Dasymetric mapping shares similarities with choropleth mapping techniques with some substantial differences (Mennis 2003; Mennis and Hultgren 2006; Nelson 2010).

The use of ancillary data in dasymetric mapping creates a more realistic depiction of real-world attributes compared to choropleth maps. Dasymetric maps follow existing spatial patterns and are not subject to the contrasting population differences often depicted in the arbitrarily defined administrative boundaries used in choropleth mapping (Figures 4 and 5) (Mennis 2003; Nelson 2010).

Figures 4 and 5 demonstrate the contrast between choropleth and dasymetric mapping techniques. Figure 4 shows how choropleth maps depict population density throughout a county using traditional administrative boundaries. The choropleth maps show distinct, sharp variations in area population density between administrative boundary zones and are likely not representative of the actual population distribution over those areas. Figure 5 depicts the same county using dasymetric mapping. The use of ancillary data creates a more probable representation of that county's actual population distribution throughout the area (Nelson 2010).

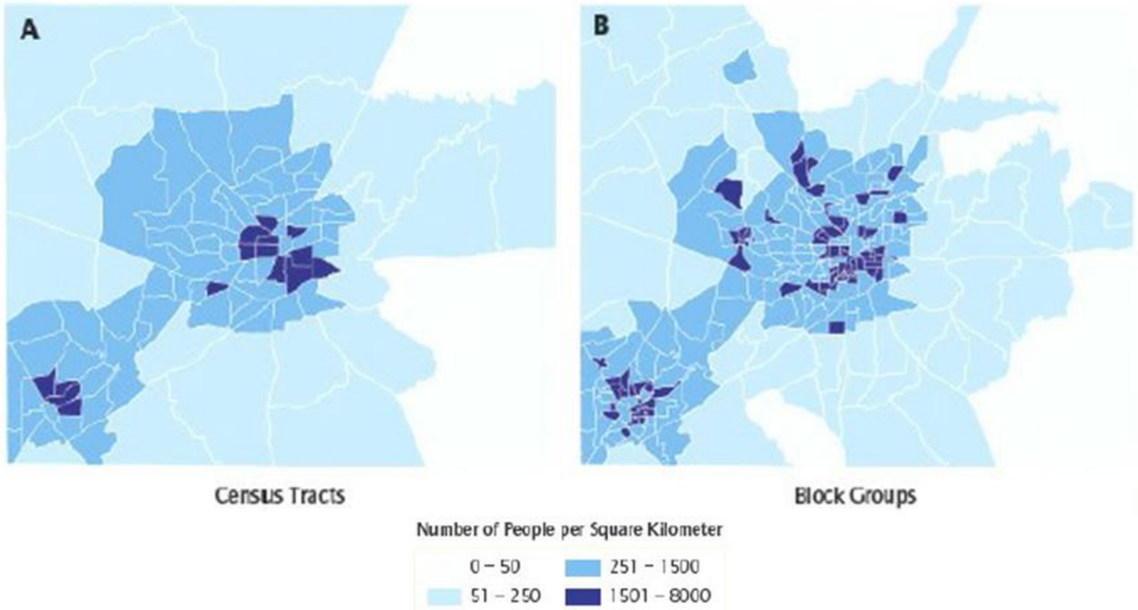

Figure 4. Guilfrod County, North Carolina census tract and block group population density (Source: Nelson 2010)

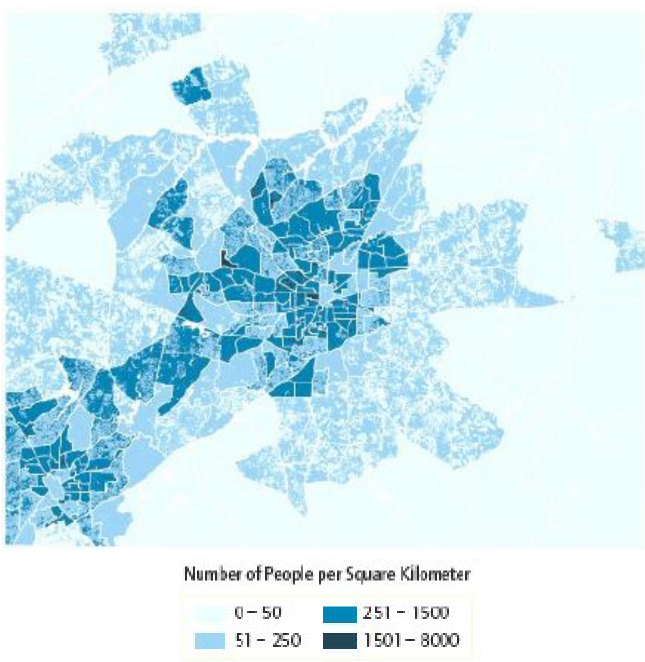

Figure 5. Dasymetric map of Guilford County, North Carolina population density (Source: Nelson 2010)

Dasymetric mapping also shares close similarities with areal interpolation as its function is to transform geographic data from one boundary system to another (Mennis 2003; Mennis and Hultgren 2006). The inherent problem of utilizing different boundary systems together is the inability to analyze and combine data of differing boundary types together without losing and or biasing the data where population counts are neither fabricated nor eliminated in the final results (Eagleson 2002; Mennis 2003). A spatial discrepancy, as is found in trying to use incompatible boundary systems, can also be a result of temporal effects as is often experienced when comparing census data from different reporting periods (Zandbergen and Ignizio 2010).

An additional complication is that ancillary data often contain geographic and/or attribute errors, further biasing results and can be difficult to account for because of contextual effects. Associating population data with area attributes can be problematic and can introduce uncertainty that is usually not included in reporting (Nagle et al. 2013).

Dasymetric mapping attempts to alleviate these situations by utilizing aggregated data and transforming those values by combining it with ancillary data by reaggregation into smaller spatial zones that more closely depict the actual population distribution throughout a given area, as was illustrated on Figures 4 and 5 (Mennis and Hultgren 2006).

Dasymetric maps are helpful in illustrating heterogeneity within a specific population and are a valuable means of visualizing cluster events throughout a geographic area (Barrozo et al. 2016). Dasymetric mapping is also theorized as providing a truer representation of populations dispersed throughout small areas, such as block group population distribution in both urban and rural environments, making this technique a valuable tool for researchers conducting small area studies (Mennis 2003; Mennis and Hultgren 2006).

Chapter 3 : Methods and Data Sources

The state of Arizona was chosen as the study area for this thesis because of the population distribution and diversity found throughout the state. This type of population dispersion was an important factor in being able to show how choosing varying forms of neighborhood for a study can affect the reporting outcomes. Administrative boundaries at the county, census tract, and census block group levels were compared to evaluate what reporting differences existed between reporting types for specified variables. Within these variables, specific populations of concern were identified and focused on to determine if and how these neighborhood variations lead to poor representation of specific population groups. The temporal aspects of this study used U.S. Census Bureau data from 2000 and 2015. The 2000 decennial U.S. Census was utilized for the majority of the analysis as it was the last census using 100 percent reporting. The 2010-2015 ACS estimates published in 2015 were used in the latter part of the study for further comparison as it was the last ACS to include data for areas within the state that only report once every five years (as is common for AIRs). Census tracts and census block group boundaries were also adjusted for the appropriate temporal period, as boundary changes affected the delineation of reporting units used by the ACS. The population variables examined were median household income, age, and ethnicity. Data were downloaded from the U.S. Census Bureau. Microsoft Excel 2016 was utilized for data conversion from .csv and file format preparation. ArcGIS Desktop (versions 10.5.1 and 10.6) were used for data analysis.

3.1. Study Area

The state of Arizona is the sixth largest state in the U.S., covering a total area of 113,990.30 square miles. There are 15 counties and by 2010 the U.S. Census Bureau reported there being 1,526 census tracts and 4,178 census block groups dispersed throughout the state (U.S. Census

Bureau 2015). The state is also home to 22 federally recognized American Indian Nations. Figure 6 shows the largest of the aforementioned geographic units in the study area. The maps in Figure 6 also show the study area by (a) Arizona counties; (b) census tracts; (c) census block groups; and (d) AIR’s with their associated area’s minimum, mean, and maximum geographic extent.

The U.S. Census Bureau estimated the Arizona population in July of 2017 to be at 7,016,270 (U.S. Census Bureau 2018d). Approximately 68% of the state’s population is located within the Phoenix Metro area (Maricopa County), and 15% in the Tucson Metro area (Pima County). That leaves the remaining 17% dispersed throughout the predominantly rural, remaining 13 counties. The variability found throughout the state provides important insight into how neighborhood delineations might affect outcomes in both traditionally examined metropolitan zones as well as in less frequently studied rural areas where aggregation issues may have a greater effect.

3.2. **Hypothesis**

Data acquired from the U.S. Census Bureau is a common resource for health researchers. These data are reported per administrative unit. The null hypothesis presented and assumed in many research studies is that median values for variables are relatively unchanging, and equivalent, regardless of the administrative unit utilized. The null hypothesis used for the current study offers that aggregation has little statistical relevance throughout a geographic neighborhood, regardless of population dispersion. The alternative hypothesis presented in this thesis is that actual values vary across administrative units and that aggregation can affect resulting outcomes. This variability dictates that that consideration of how a neighborhood is selected is important and affects how outcomes are reported.

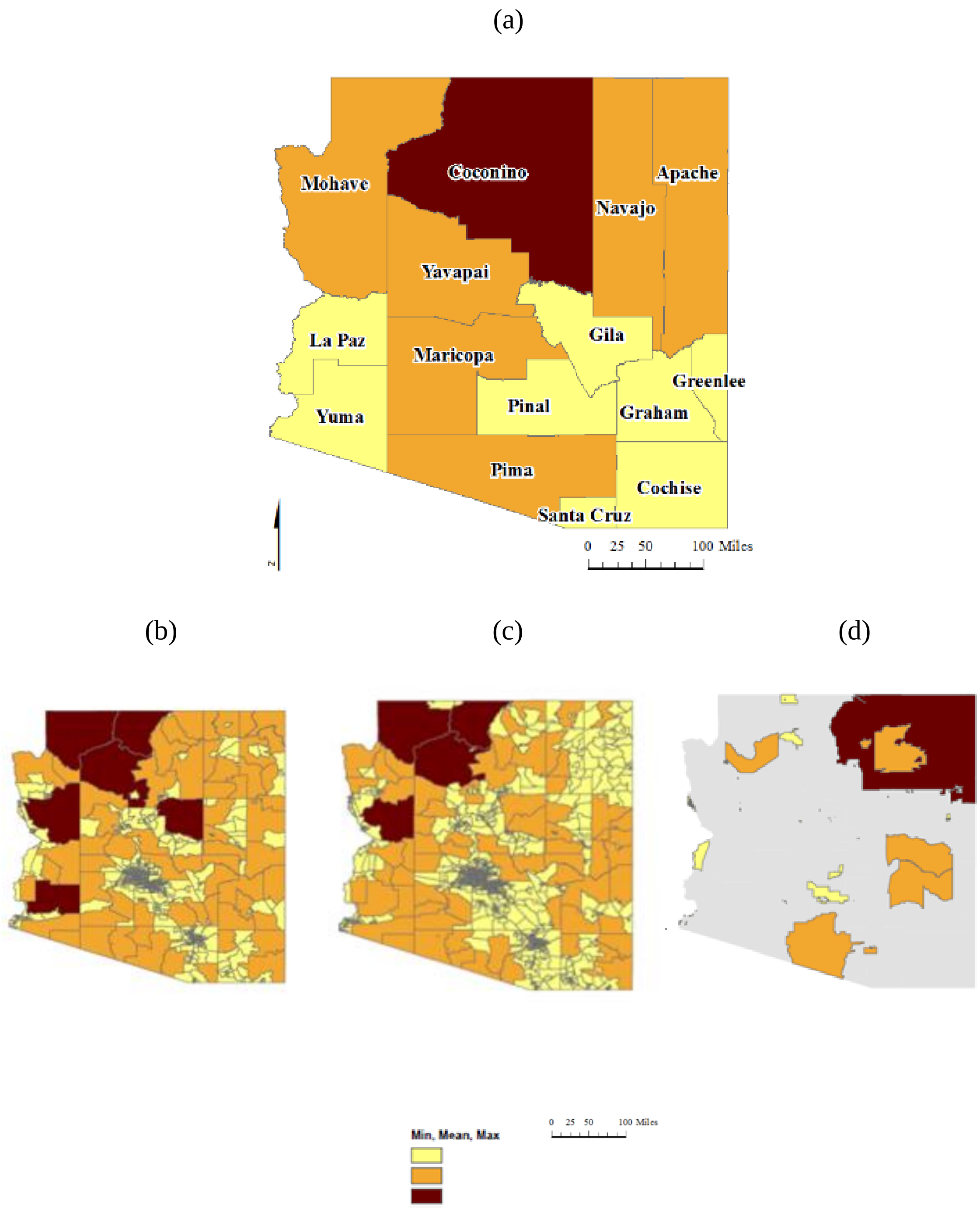

Figure 6. Area maps by: (a) Arizona counties; (b) census tracts; (c) census block groups; and (d) AIR’s. The colors show the geographic extent by minimum, mean, and maximum areas for each administrative unit.

The hypothesis is tested using three variables that are commonly used in research as social health determinants: Median household income, vulnerable population represented by the percentage of children and the elderly, and ethnicity represented by percentage of Native American population. These variables were then examined at three different administrative units to evaluate how and if there are significant variations in terms of their estimation depending on the size of the geographic reporting unit.

For this thesis the administrative units used were county, census tract, and census block group. The results presented in Chapter 4 examine decennial census data from the last 100 percent reporting period (i.e. 2000 - 2010) and investigates what variation occurs for selected variables within differing administrative units. The second phase of this analysis that is briefly referred to in Chapter 5 examined the variability of selected variables using ACS data for three of Arizona's 15 counties. The first ACS reporting that fulfills the needs of this study is from the period 2010-2015.

3.3. Data Sources

Shapefiles and attribute data was downloaded from the U.S. Census Bureau and imported into ArcGIS (Table 1). Initially, Esri Business Analyst was used; however, it was determined that Esri utilized numerous sources for their reports and values, potentially confounding results. This required sourcing data exclusively from the U.S. Census Bureau's American Factfinder for data consistency. This was crucial to ensure that the appropriate reporting survey type (decennial vs. ACS), administrative unit, and temporal period was used for each of the variables examined in this thesis.

Table 1. Census data sets and sources.

Dataset	**Description**	**File Type**	**Data Type**	**Temporal Reporting Period**	**Source**
Counties	Arizona county boundaries	.shp files	Vector data (polygon)	Vintage: 2007 (for 2000 Census) 2015	U.S. Census Bureau TIGER/Line files U.S. Decennial Census ACS 2015
Census Tracts	Boundary lines for all census tracts in Arizona	.shp files	Vector data (polygon)	Vintage: 2007 (for 2000 Census) 2015	U.S. Census Bureau TIGER/Line files U.S. Decennial Census ACS 2015
Census Block Groups	Boundary lines for all census block groups within Arizona	.shp files	Vector data (polygon)	Vintage: 2007 (for 2000 Census) 2015	U.S. Census Bureau TIGER/Line files U.S. Decennial Census ACS 2015
Demographic Profile	Reporting for population by age group per administrative unit	.csv converted to .xlsx and .dbf	Vector data (discrete, point)	2000 2015	U.S. Census Bureau 2000 Census; Demographic Summary file 1 ACS 2015
Ethnicity	Separate datasets reporting Non-Hispanic & Hispanic ethnicity counts per administrative unit	.csv converted to .xlsx and .dbf	Vector data (discrete, point)	2000 2015	U.S. Census Bureau 2000 Census; Ethnicity non-hispanic; Hispanic Ethnicity Summary file 1 ACS 2015
Median household income	Median household income per administrative unit	.csv converted to .xlsx and .dbf	Vector data (discrete, point)	2000 2015	U.S. Census Bureau 2000 Census; Summary file 3 ACS 2015

Data from the 2000 decennial census was determined to be the last decennial census to use 100 percent reporting and that would accurately represent population dynamics for this analysis. The boundaries for census reporting units were also different in 2000 as compared to the 2010 decennial Census and later ACS tabulations. The U.S. Census Bureau converted all 2000 boundary files to shapefile compatible format in 2007 requiring usage of a 2007 vintage file for acquisition of the 2000 boundary shapefiles (Table 1).

The 2010-2015 ACS data was utilized for the second phase of analysis in this thesis. The appropriate temporally corresponding vintage shapefiles were also downloaded to support this part of the analysis (see Table 1 for additional details).

3.4. **Methodology**

Data were downloaded from the U.S. Census Bureau website. Variable data were selected per administrative unit within Arizona, downloaded as a .csv file and converted to Excel files. Spreadsheets were then designed in Excel and imported into ArcGIS Desktop. In ArcGIS, the Excel files were converted into database files that were then joined with their corresponding administrative units (see Figure 7 for workflow).

Administrative boundaries for three commonly used census reporting units were utilized for comparison. Arizona's 15 counties were the basis for examining health determinant variables and if and how results varied between the census reporting unit that were used. Three census reporting units were used in this study: County level, Census Tract level, and Census Block Group level. All of the maps displayed throughout the thesis use the UTM Zone 12N map projection; however, transformations were conducted using ArcGIS, ArcToolbox Data Management toolset to convert to the most appropriate visualization projection from the U.S. Census Bureau source data which utilized GCS North American 1983.

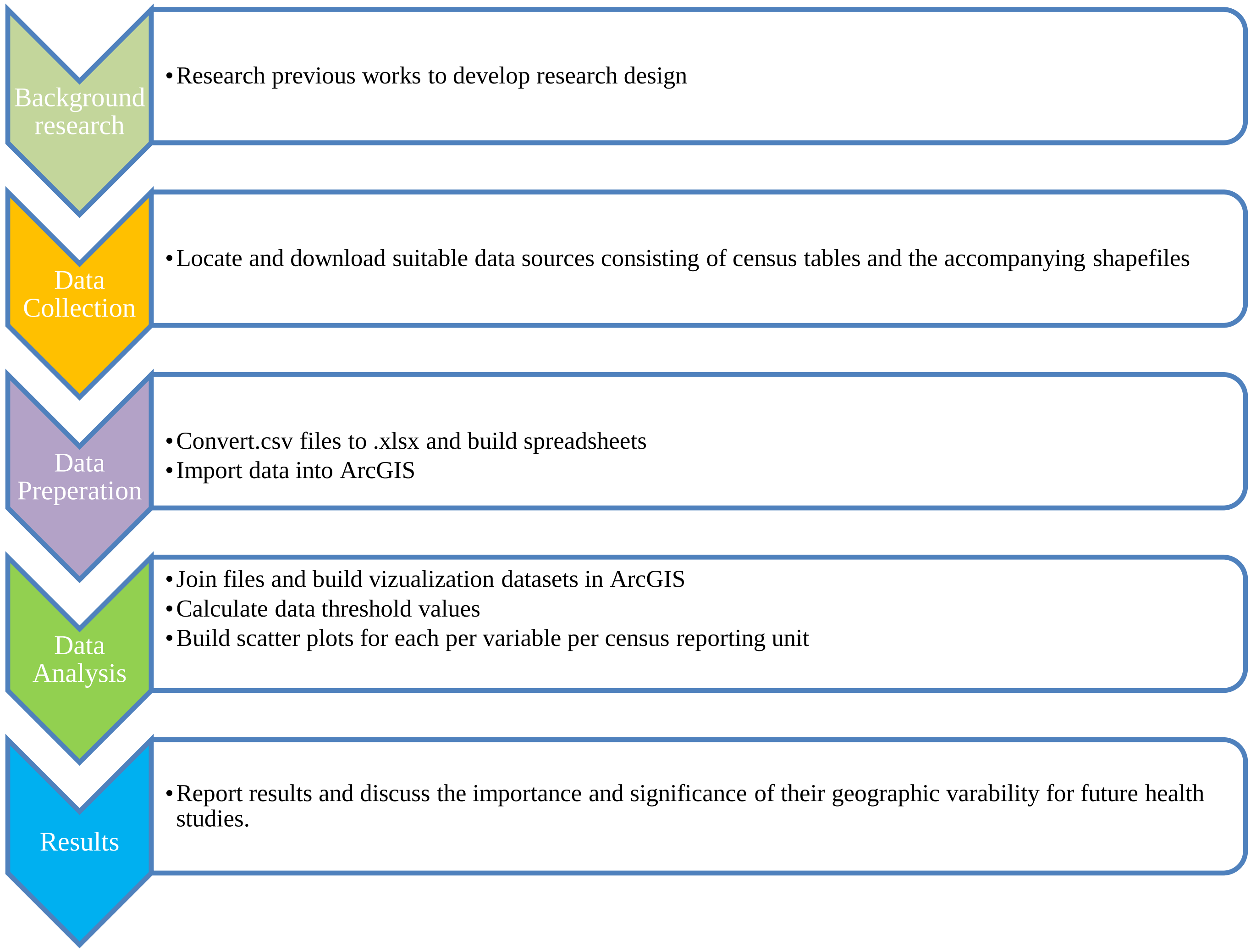

Figure 7. The workflow

This study examines the geographic variability of three variables that are commonly used in health studies: (1) Wealth by median household income; (2) race and ethnicity, the percent of the population that was Native American in this instance; and (3) vulnerable population (the percent of population) over age 65 and under age 16 in this instance.

The data for these three social determinants of health were imported into ArcGIS and threshold values were determined for each variable. Then scatterplots were built for each variable and color coded by threshold bracket. Scatterplots were used to evaluate thresholds set at ≥ 125%, ≤ 75%, and ± 25% of the corresponding county and census tract values. The calculations were conducted using ArcGIS and selecting by attributes where the formulas for each variable threshold were as follows:

A. County median values as reported by the U.S. Census Bureau were used for counties and census tracts using the following rates.

(1) ≤ 75%: Census Tract value ≤ (.75*County value)

(2) ≥ 125%: Census Tract value ≥ (1.25*County value)

(3) ± 25%: CT value ≥ (.75*County value); and ≤ (1.25*County value)

where the three classes used to classify the census tract values express the variability relative to the county estimates for each variable and where < 75% is depicted throughout in blue, > 125% is depicted throughout in red, and ± 25% is depicted throughout in green.

B. The same approach was also used for census tracts and census block groups using the following rates:

(4) ≤ 75%: Census Block Group value ≤ (.75*Census Tracts value)

(5) ≥ 125%: Census Block Group value ≥ (1.25* Census Tracts value)

(6) ± 25%: CBG value ≥ (.75*CT value); and ≤ (1.25*CT value)

where the three classes used to classify the census block group values express the variability relative to the census tract estimates for each variable and where < 75% is depicted throughout in blue, > 125% is depicted throughout in red, and ± 25% is depicted throughout in green.

The resulting units classified as outliers were grouped by county using population as a basis for the groups. Group I contain the two most populated counties in the state, Maricopa and Pima with populations of 3,072,149 and 843,746 respectively. Group II consists of seven counties with modest populations ranging from 97,000 to 179,000: Pinal, Yavapai, Yuma, Mohave, Cochise, Coconino, and Navajo. Lastly, Group III contains the six most rural and least populous counties with totals ranging from 8,000 to 69,000: Apache, Gila, Santa Cruz, Graham, La Paz, and Greenlee. There were no counties with a population that fell between Groupings.

Graphs were made for each variable by county to describe the variability from threshold values representing units classified as outliers and introduced by the choice of census reporting unit. Maps were also prepared but are displayed for just six of Arizona's 15 counties in the thesis itself. Two counties from each of the aforementioned groups were selected for this purpose as follows: Maricopa and Pima in Group I, Coconino and Pinal in Group II, and Apache and Santa Cruz in Group III.

The following chapter reports the results and reveals information that has far reaching consequences for health researchers who want to learn how their choice of census reporting unit may potentially influence the resulting outcomes of studies that consider social determinants of health.

Chapter 4 : Results

This chapter is detailed by section and investigates how the spatial measurement of a locale can affect the geographic reporting of health determinants. Three independent variables are compared at three commonly utilized variations of neighborhood: county, census tract, and census block group. In 2000, Arizona possessed 15 counties, 1,108 census tracts, and 3,570 census block groups. Following are the trends discovered when examining each variable by neighborhood for that temporal reporting.

4.1. Median Household Income

Median Household Income is a significant indicator of human well-being and a key health determinant. Theoretically, median household income reflects a household's ability to access resources, such as sustenance and medical care, and indicates overall quality of life.

4.1.1. County Level

Median Household Income by county is displayed in Figure 8. The values demonstrated are as reported by the U.S. Census Bureau's 2000 decennial census. Arizona is depicted throughout this thesis at a scale of 1:4,275,000.

The counties demonstrating the highest median household incomes, depicted on the map in red, have multiple factors contributing to their higher incomes (Figure 8). Maricopa County contains the greater Phoenix Metro area, the largest metropolitan area in Arizona. The greater Phoenix Metro is a major service hub not just for Arizona but also the entire southwestern U.S. providing population and infrastructure to support an economy that can provide higher wages and thus higher household incomes in comparison to the State's rural areas. Coconino and Greenlee, the remaining two counties demonstrating high median household income, reflect their affluence from industry. Median household income in these counties is not necessarily associated with high population or service areas.

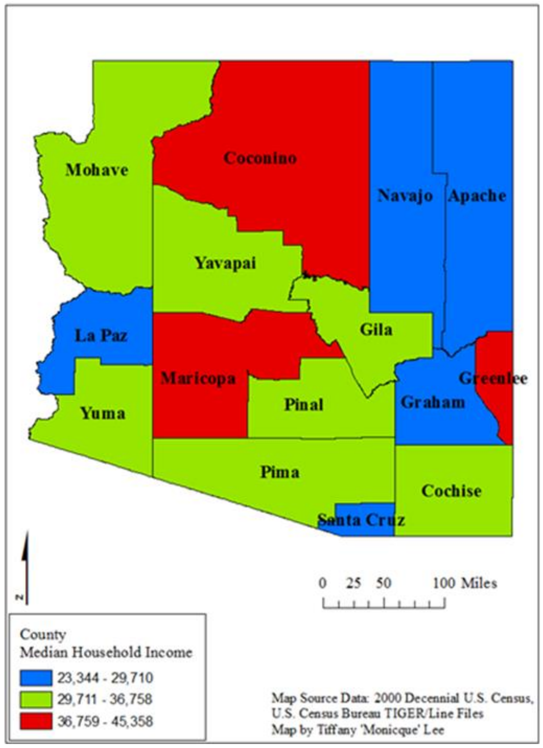

Figure 8. Median Household Income for each Arizona county

The counties in green represent the mid-level classification and for the purposes of this thesis follow the State of Arizona's Median Household Income. There are seven counties that comprise this level, making this grouping the largest of the three classes for the variable.

The counties with the lowest median household incomes, colored blue on the map, have limited service centers and county populations that did not exceed 100,000 (Figure 8). These five counties are more rural in nature and do not possess any substantial form of industry. Services are limited and few even across large geographic areas. Additionally, these counties also have substantial portions allotted as Native American Tribal Areas (Figure 6d).

County Median Household Income varies considerably from the low economic class (counties in blue) to the high economic class (counties in red). The counties in blue are primarily the most rural; however, rurality does not dictate income level as Coconino and Greenlee counties are largely rural and yet are still classified as two of the State's wealthiest counties. The most populous county, Maricopa County, is also one of the most prosperous.

4.1.2. Census Tract Level

Viewing median household income by census tract provides an entirely different perspective of the household income distribution throughout the various counties and the State of Arizona (Table 2). A comparison of Figures 8 and 10 reveals that census tracts with median household incomes classified as outliers > 125% (depicted in red) and < 75% (depicted in blue) are relatively confined to specific areas within each county and are not equally distributed throughout those counties.

Three counties; Maricopa, Pima, and Santa Cruz counties, exceed the States' average of 52.2% of census tracts classified as outliers (Table 2). Two of these counties, Maricopa and Pima contain the two largest metropolitan areas within the State, the Phoenix and Tucson Metros, respectively. Maricopa County has a population of 3,072,149 and Pima County has a population of 843,746. Santa Cruz County, however, has a relatively low county population of 38,381 and is the smallest county in terms of area within Arizona (U.S. Census Bureau 2015).

The remainder of the counties fall under the 52.2% state average; however, five counties have fewer than 20% of their census tracts classified as outliers (Table 2). Four of these counties fall between 12.5 and 20%. Graham County has a population of 33,489 with 12.5% of its census tracts classified as outliers. Yavapai has a population of 167,517 with 15.4% of its census tracts classified as outliers. Mohave contains a population of 155,032 and has 16.7% of its census tracts classified as outliers, and Gila has a population of 51,335 with 20% of its census tracts classified as outliers.

Table 2. Counts and percentages of census tracts with median household incomes ≤ 75% and ≥ 125% of county median values.

County	Population	No. of census tracts	No. of census tracts with MHI	Nos. and % of outlier census tracts *
Apache	69,423	14	14	5 + 1 = 42.9
Cochise	117,755	21	21	4 + 5 = 42.9
Coconino	116,320	28	27	8 + 5 = 46.4
Gila	51,335	15	15	2 + 1 = 20
Graham	33,489	8	8	1 + 0 = 12.5
Greenlee	8,547	3	3	0, 0
La Paz	19,715	6	6	0 + 2 = 33.3
Maricopa	3,072,149	663	659	181 + 194 = 56.6
Mohave	155,032	30	30	3 + 2 = 16.7
Navajo	97,470	23	23	6 + 3 = 39.1
Pima	843,746	198	198	52 + 70 = 61.6
Pinal	179,727	33	32	9 + 4 = 39.4
Santa Cruz	38,381	7	7	2 + 2 = 57.1
Yavapai	167,517	26	25	3 + 1 = 15.4
Yuma	160,026	33	32	6 + 6 = 36.4
Totals	5,130,632	1,108	1100	296 + 282 = 52.2

* No. of census tracts with MHI <75%, >125%, and the sum of the two classes of outliers as a percentage of total.

Greenlee County, with the smallest population in the State of 8,547, was the only county in Arizona that did not have any census tracts classified as outliers (Table 2).

The remaining seven counties fall into the mid-level classification with the percentages of census tracts exceeding the county thresholds ranging from 33 to 46% (Table 2). These counties are as follows: La Paz County, the least populous within this category with a population 19,715 and 33.3% of the census tracts classified as outliers; Yuma County, the second most populous county with a population of 160,026 and 36.4% of the census tracts classified as outliers; Navajo County with a population of 97,470 and 39.1% of the census tracts classified as outliers; Pinal County, the most populous county in this grouping with a population of 179,727 and 39.4% of its census tracts classified as outliers; Apache County, the poorest county in the State with a census tract median household income average of $21,497.29 and a population of 69,423 and 42.9% of its census tracts classified as outliers; Cochise County with a population of 117,755 and 42.9% of its census tracts classified as outliers; and lastly, Coconino County, the wealthiest in this grouping, with a median household income of $39,066, a population of 116,320, and 46.4% of its census tracts classified as outliers.

Two counties within Arizona contain 70% of the state's census tracts. These counties are Maricopa and Pima with 663 and 198 census tracts, respectively. Government apportioning of spaces by population dictates that the metro areas have smaller reporting units to account for their higher populations.

Comparison between county and census tract is implemented by evaluating how median household income is reflected within specified thresholds of the county median income, classifying values outside of these threshold values as outliers (Table 2 and Figure 9). Red dots in Figure 9 reflect census tracts that have values ≥ 125% of their county's median value, indicating areas that exceed county median incomes (similar to Figure 8). Blue dots indicate those census tracts that have a value ≤ 75% of their county's median value, denoting less wealthy areas that may be susceptible to poverty. Green dots show which census tracts are within ± 25% of their county's median value. Figures 9 and 10 reveal that median household income is not spread equally across a county and is specific to census tracts within those counties.

Figure 9 reports the variability by census tract and county graphically and reveals that Arizona possesses some very wealthy census tracts in some counties as well as some very poor census tracts. Furthermore, by examining the graph it is evident that some of these extremes are found within the same county, e.g. Maricopa County (Figure 10b).

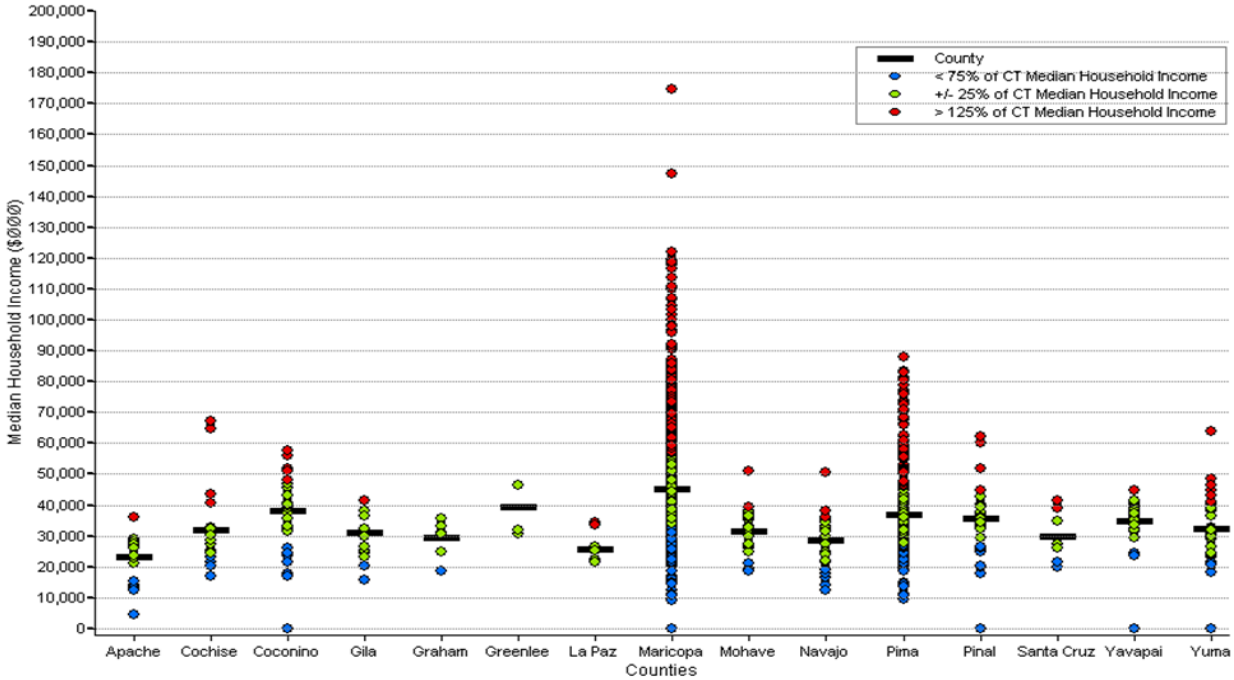

Figure 9. Median household income by census tract within each county.

Figure 10a visually demonstrates the geographic structure of the variability found using the state's census tracts as the reporting units. The greater Metro areas are also shown at a larger scale to reveal the census tract variability within these zones (Figures 10b and 10c).

Figure 10b shows part of Maricopa County's variability in and near the greater Phoenix metro region. The large zones that are the poorer areas of the county are the furthest from the greater metropolitan area. Pima County demonstrates the highest variability within the state (Table 2 shows 61.6% of the census tracts classified as outliers). This county demonstrates a high amount of variability and a segregation of economic classes at the census tract level where the greater Tucson metro area shows distinct groupings of low, middle and high economic classes (Figure 10c). Again however, as with Maricopa County, the census tracts that demonstrate the highest incomes are in or near to the greater Metro area with distant census tracts almost all colored blue

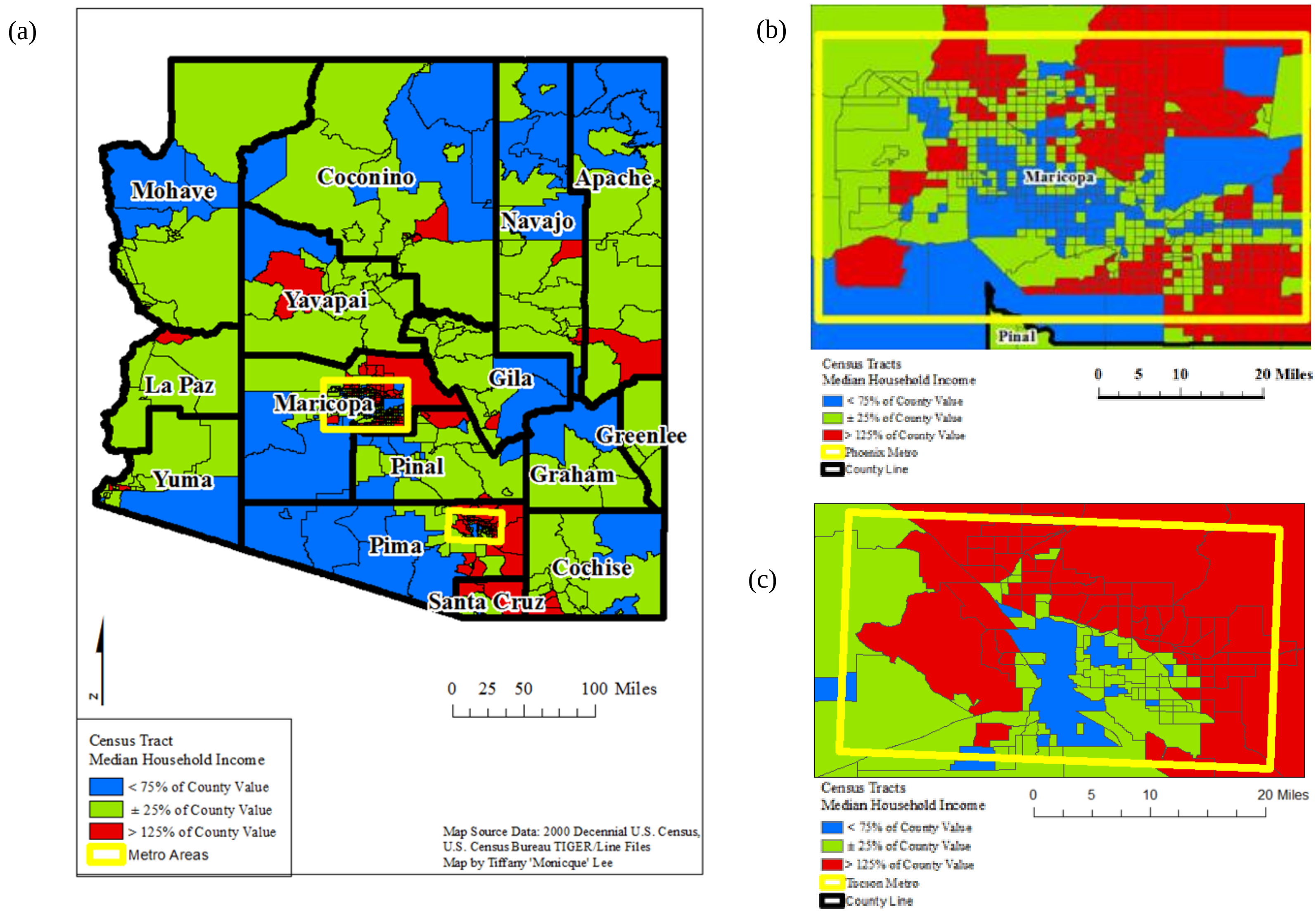

Figure 10. Median Household Income for each Arizona Census Tract (a) and the Phoenix Metro (b) and Tucson Metro areas (c).

Santa Cruz County, a relatively sparsely populated county (population of 38,381), shows people of wealth near to neighboring Pima County's Tucson Metro area. The census tracts that show the highest median household incomes also fall around the major transportation corridor (not shown) from Tucson to Mexico. The higher incomes are also likely representative of the regions other historic industry, mining. Cochise County, (population 117,755), bordering Santa Cruz County also shows considerable variability (42.9% of the census tracts classified as outliers, Table 2). The wealthiest areas, however, are located near the transportation corridor, which provides economic opportunity. The remaining parts of the county reflect largely average median household incomes, with remote zones being the least wealthy.

Apache, Navajo, and Coconino Counties (with populations of 69,432, 97,470, and 116,320, respectively) have the largest geographic areas dedicated as Native American Reservation Areas (AIRs) and percentages of census tracts classified as outliers ranging from 39.1% (Navajo) to 46.4% (Coconino). These counties are diverse, largely rural with few resource centers, and economically affected by unique challenges not found in other parts of the state.

Yavapai County has a relatively high population (167,517) with few census tracts classified as outliers (15.4%). The wealthiest census tracts are located near the primary resource community, Prescott, providing infrastructure and services. The least wealthy fall between Prescott and AIRs.

4.1.3. Census Block Group Level

Median household income was also examined at the census block group level. Census block groups classified as outliers are significant when comparing census tract to census block group because of the increase in spatial units at census block group level going from 1,108 census tracts and partitioning those locales into 3,570 "neighborhoods" as represented by census block groups.

Group I consists of Maricopa and Pima Counties, respectively (Figures 11-12). Maricopa County contains most of the Phoenix Metro area and is the highest populated region in the state. Pima County encompasses the Tucson Metro area, and although considerably smaller than Phoenix Metro, is the second largest metropolitan area in Arizona.

Scatterplots vary from one another as the number of census tracts and corresponding census block groups differ based on county population (Figures 11-14). The scatterplots show census block groups classified as outliers around the median census tract values. Maricopa County (Figure 11) contains 663 census tracts and 2,113 census block groups, whereas Pima County (Figure 12) contains 198 census tracts and 617 census block groups.

Census block groups classified as outliers exceed 24% in both Maricopa and Pima Counties. Figure 11 shows numerous outliers ≥ 125% of the corresponding census tract value, indicating specific census block groups of wealth, as well as some census block groups with income ≤ 75 % of the corresponding census tract values, indicating relatively impoverished areas. The two Metro areas differ as Pima County and the greater Tucson Metro area do not demonstrate the same extremes of Median Household Income evident in Maricopa County and the greater Phoenix Metro area, by both census tract and census block group (Figures 11 and 12).

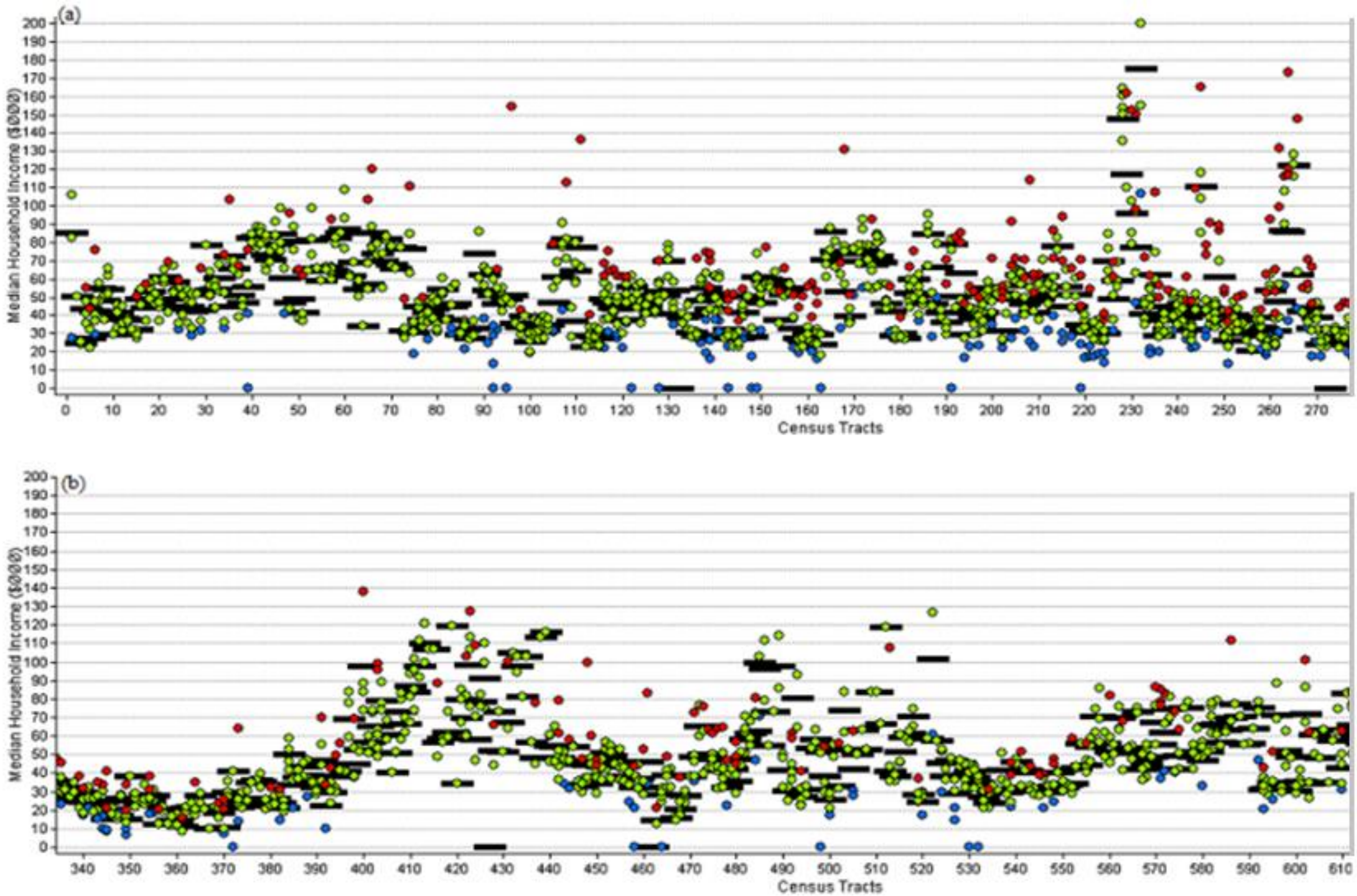

Figure 11. Group I, census block group Median Household Income by census tracts 1-333 (a) and census tracts 334-663 (b) in Maricopa County.

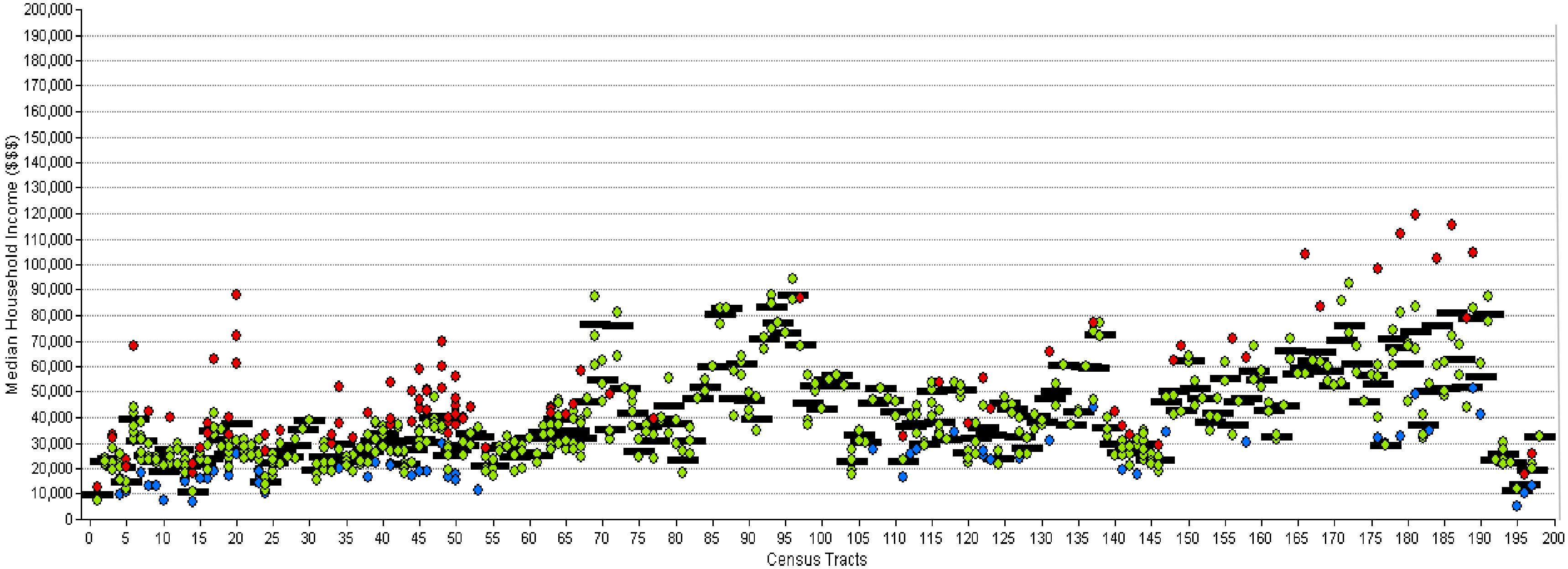

Figure 12. Group I (Continued), census block group median household income by census tract in Pima County.

Group II includes the seven Arizona counties with medium sized populations and between 21 and 33 census tracts. These seven counties, in order of highest to lowest populated county, are Pinal, Yavapai, Yuma, Mohave, Cochise, Coconino, and Navajo Counties respectively (Figures 13a-d and 14e-g). Each of the counties in this grouping have several census block groups with incomes exceeding ≥ 125% of the corresponding census tract. Pinal, Yavapai, Yuma, and Coconino Counties include census block groups with zero values which indicate census block groups which lacked sufficient resident reporting income data (Figures 13a-c and 14f).

Four of the seven counties reported outlier census block groups over the state average of 24.8%. Those counties were Coconino (34%), Pinal (31%), Navajo (29.7%), and Cochise Counties (27.8%). The remaining three counties that fell under the state average were Yuma (24.5%), Yavapai (18.6%), and Mohave (16.8%) Counties (Table 3).

Coconino County, the county that had the largest number of outliers, also had the highest outlier census block group median household income, with two outlier census block groups reporting > $80,000 (Figure 14f). Mohave County, the county with the smallest number of outlier census block groups, had the lowest census block group median household income within Group II with no outlier census block groups reporting > $50,000 (Figure 13d).

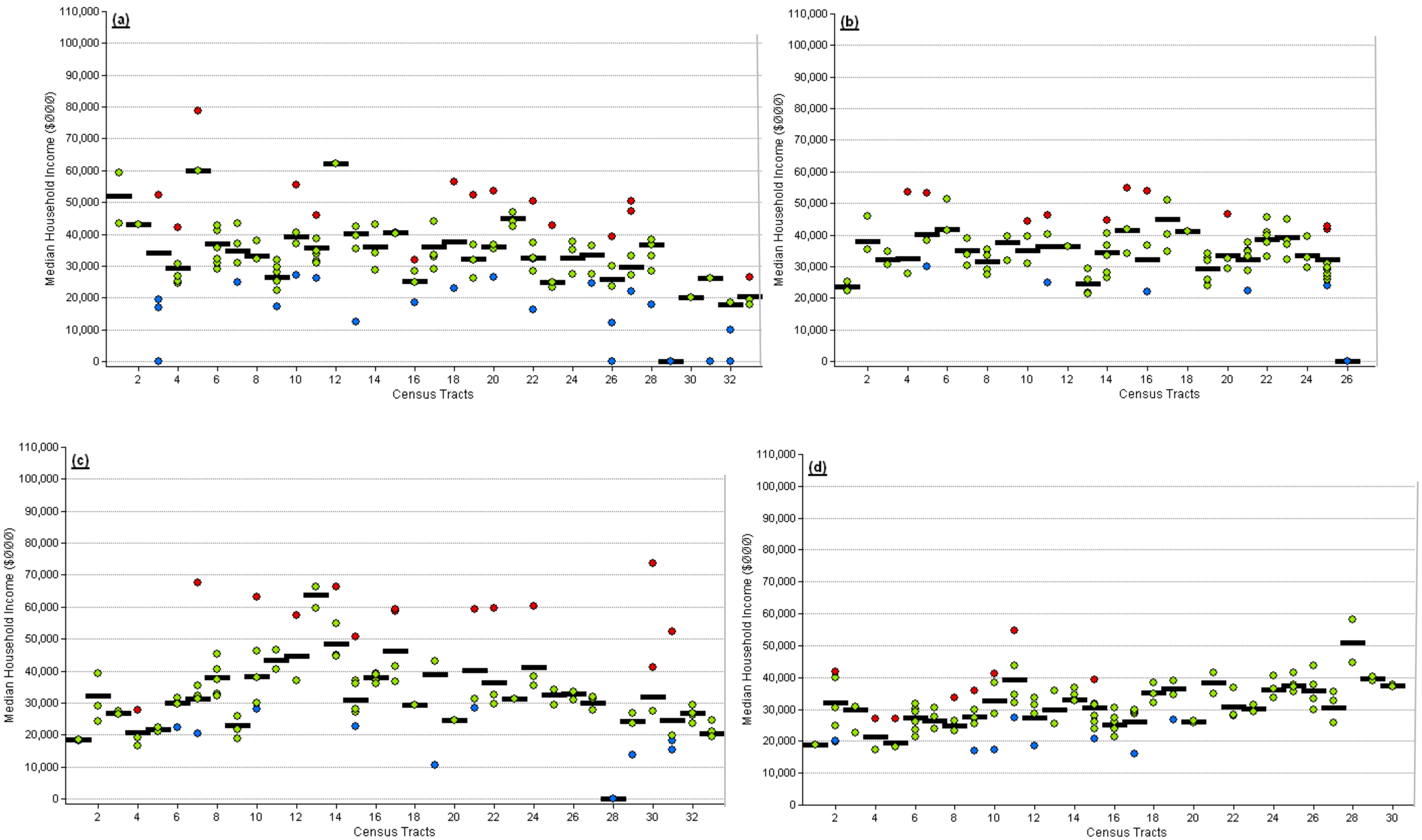

Figure 13. Group II, census block group Median Household Income by census tract for: (a) Pinal; (b) Yavapai; (c) Yuma; (d) Mohave Counties.

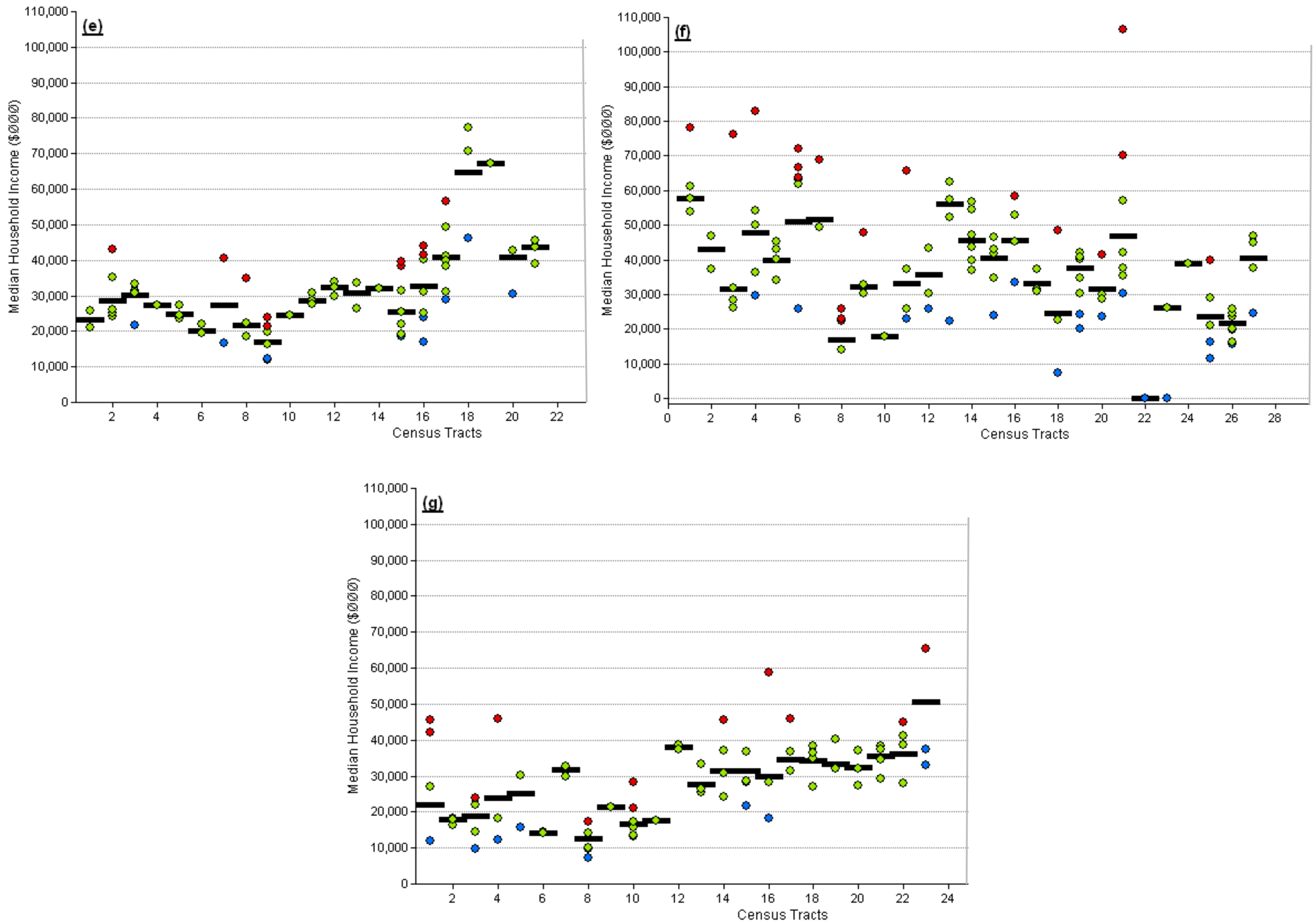

Figure 14. Group II (Continued), census block group Median Household Income by census tract for: (e) Cochise; and (f) Coconino; and (g) Navajo Counties.

Group III includes the counties with the smallest populations, (< 70,000) which contain between three and 15 census tracts. All counties in this grouping have at minimum one census block group that exceeds the ≥ 125% threshold; however, Greenlee County (Figure 14f) does not have any census blocks groups ≤ 75% of the corresponding census tract median value. None of the Group III census block groups exceed $60,000 annual median household income.

There is lesser amount of variability when census block groups are used in place of census tracts (Table 3) than exists between census tracts and county median values (Table 2). The largest numbers of census block groups classified as outliers occurred in Apache (46.3%), Coconino (34%), and Santa Cruz (65%) counties. Greenlee (12.5%), La Paz (17.4%), and Mohave (16.8%) counties had the least number of outliers. The two counties containing Arizona's metro areas, Maricopa and Pima, are the closest to the State's average.

Table 3. Counts and percentages of census block groups with median household incomes ≤ 75% and ≥ 125% of census tract values.

County	Population	No. of census block groups	No. of census block groups with MHI	Nos. and % of outlier census block groups *
Apache	69,423	54	54	11 + 14 = 46.3
Cochise	117,755	72	72	10 + 10 = 27.8
Coconino	116,320	106	104	18 + 18 = 34
Gila	51,335	55	55	7 + 4 = 20
Graham	33,489	27	27	2 + 3 = 18.5
Greenlee	8,547	8	8	1 + 0 = 12.5
La Paz	19,715	23	22	2 + 2 = 17.4
Maricopa	3,072,149	2,113	2,088	301 + 206 = 24
Mohave	155,032	101	101	8 + 9 = 16.8
Navajo	97,470	74	74	13 + 9 = 29.7
Pima	843,746	617	617	93 + 56 = 24.1
Pinal	179,727	116	111	15 + 21 = 31
Santa Cruz	38,381	20	20	7 + 6 = 65
Yavapai	167,517	86	85	10 + 6 = 18.6
Yuma	160,026	98	97	14 + 10 = 24.5
Totals	5,130,632	3,570	3,535	512 + 374 = 24.8

* No. of census block groups with MHI <75%, >125%, and the sum of the two classes of outliers as a percentage of total.

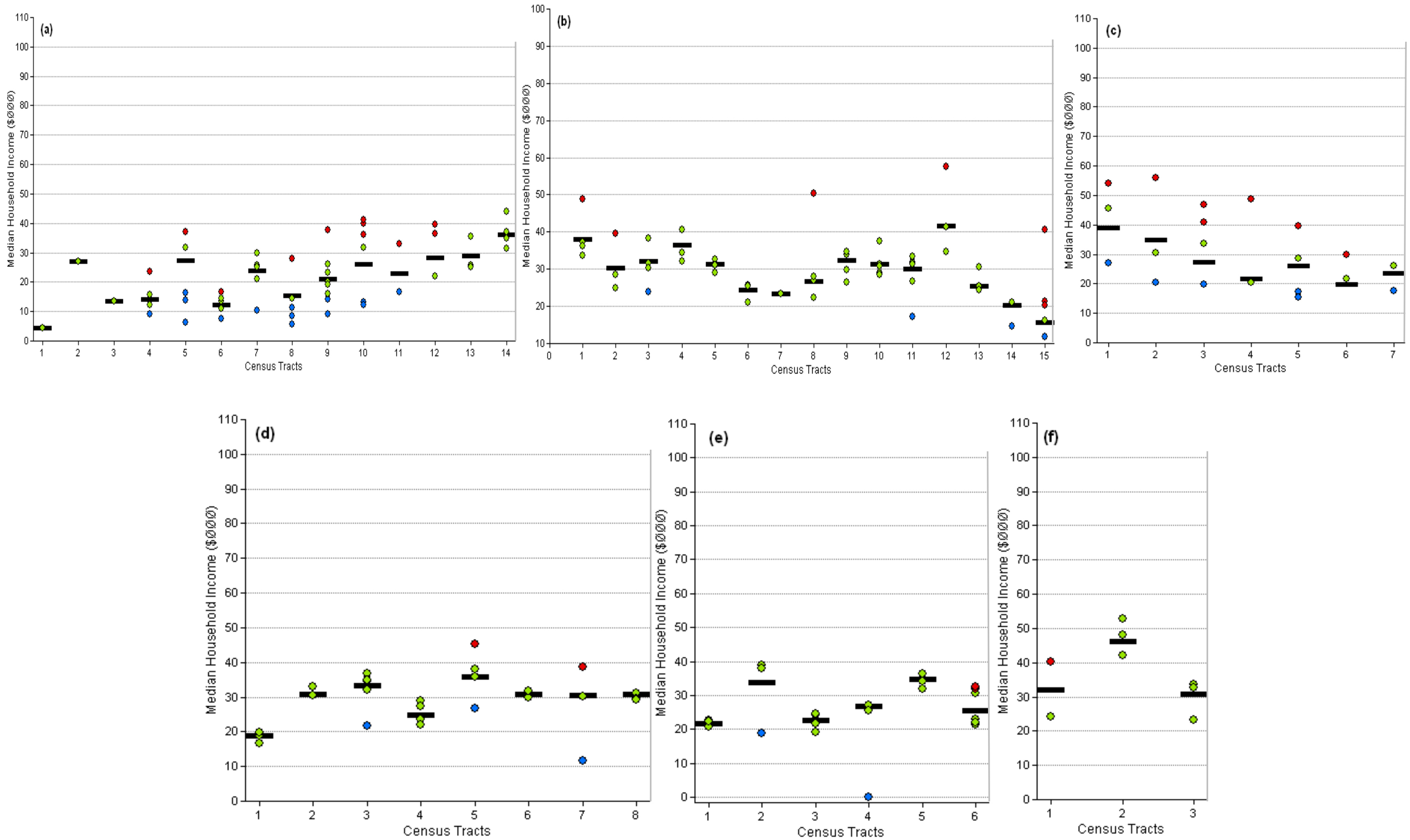

Figure 15. Group III, census block group median household income by census tracts for (a) Apache, (b) Gila, (c) Santa Cruz, (d) Graham, (e) La Paz, and (f) Greenlee Counties in Arizona.

The scatterplots show how median household income was distributed throughout counties in 2000 using census tracts and census block groups as the geographic reporting unit. When census block groups are used in place of census tracts, the trends discovered often change, as seen in Figures 16-19. The color scheme remains the same throughout where red depicts census block groups with ≥ 125% above the Median Household Income in the corresponding census tracts, green indicates census block groups with Median Household Income values within ± 25% of the corresponding value and blue represents census block groups with Median Household Income values that are ≤ 75% of census tract value.

The two counties in Group I with the highest populations in Arizona are displayed in Figures 16 and 17. The counties are depicted at a scale of 1: 2,000,000 for the county maps and the Metro areas are depicted at a scale of 1:700,000 in the Metro area inset maps. As neighborhoods become smaller using census block groups as the reporting unit, the Median Household Income estimates display patterns that were indiscernible at the county and census tract levels.

The Maricopa County maps reproduced in Figure 16 show how the Median Household Income would be over-and/or under-estimated depending on how a neighborhood is defined. When median household income is examined at the census block group level and compared to the corresponding census tract estimate, the census block group estimates tend to display patterns of wealth and poverty dispersed throughout the greater Metro area and surrounding rural areas.

The Pima County maps reproduced in Figure 16 also show differences resulting from how Median Household Income is reported. Figure 16a depicts low income census tracts in the south-central part of the Tucson Metro area and high-income areas to the west, north, and east, which cover more than 50% of the Metro area. Figure 16b shows how those distributions are representative of specific census block groups and not the standard across large portions of the greater Tucson Metro area.

Group II is represented in Figure 18 by Coconino and Pinal counties. At the census tract level, Coconino County shows a small central area of higher Median Household Income, a pronounced area of low income to the east, and large areas falling near the county mean (Figure 18a). The map which shows Median Household Income by census block group on the other hand reveals a dissimilar, varied and dispersed pattern throughout the county (Figure 18b).

The Pinal County maps reproduced in Figures 18c and 18d show different locations of high and low incomes using census tracts and census block groups as reporting units notwithstanding the presence of mid-range values covering much of this county.

Group III (Figure 19) is represented by Apache and Santa Cruz counties. Apache County contains census tracts classified as outliers at the north and southern ends of the county (Figure 19a) but the substitution of census block groups for census tracts reveals a diverse pattern of Median Household Income values dispersed throughout the county (Figure 19b).

Santa Cruz County reveals similar trends (Figures 19c and d). At the census tract level Median Household Income is split regionally (Figure 19c). High income census tracts are found in the northern and eastern portions of the county, with a large mid-range area in the southwest and one small area of low income values within the midrange in the southcentral part of the county. Figure 19d once again shows that using Census Block Groups as the reporting unit produces a more varied and dispersed pattern of Median Household Income across the county and/or individual census tracts.

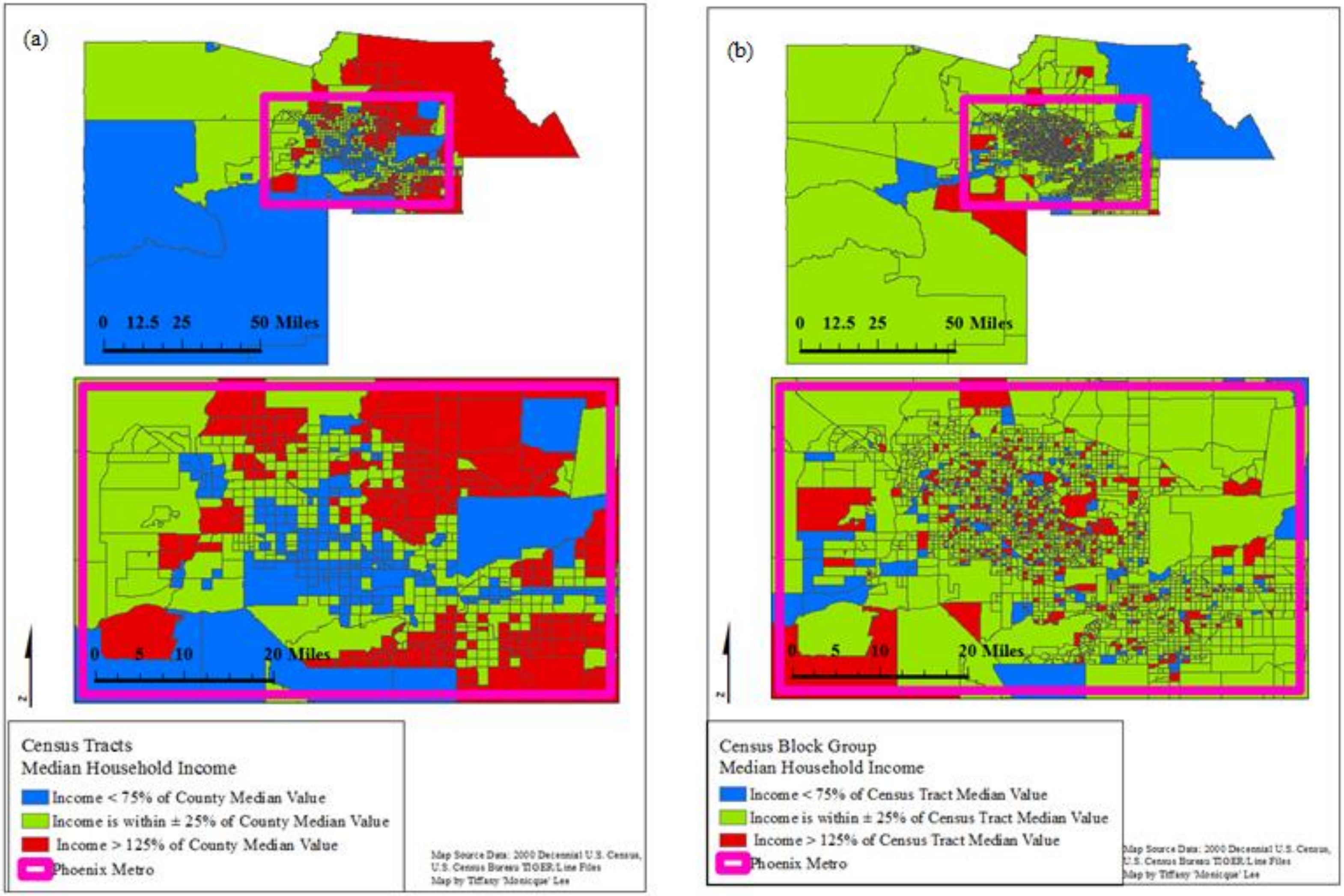

Figure 16. Group I, Median Household Income in Maricopa County by: (a) census tract; and (b) census block group. Both pairs of maps show the entire county on the top with the greater Phoenix Metro highlighted and shown at a larger scale on the bottom.

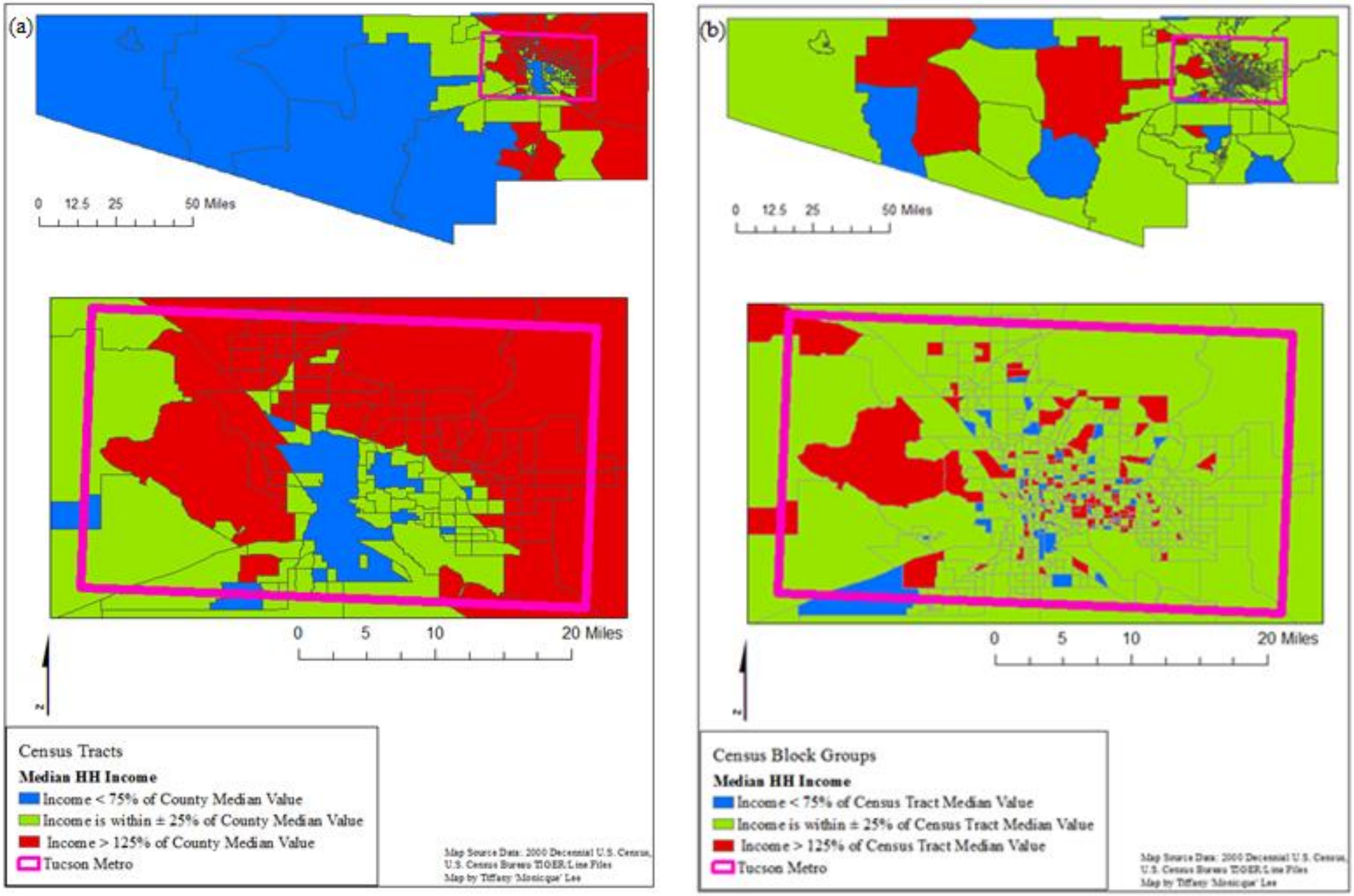

Figure 17. Group I (Continued), Median Household Income in Pima County by: (a) census tract; and (b) census block group. Both pairs of maps show the entire county on the top with the greater Tucson Metro highlighted and shown at a larger scale on the bottom.

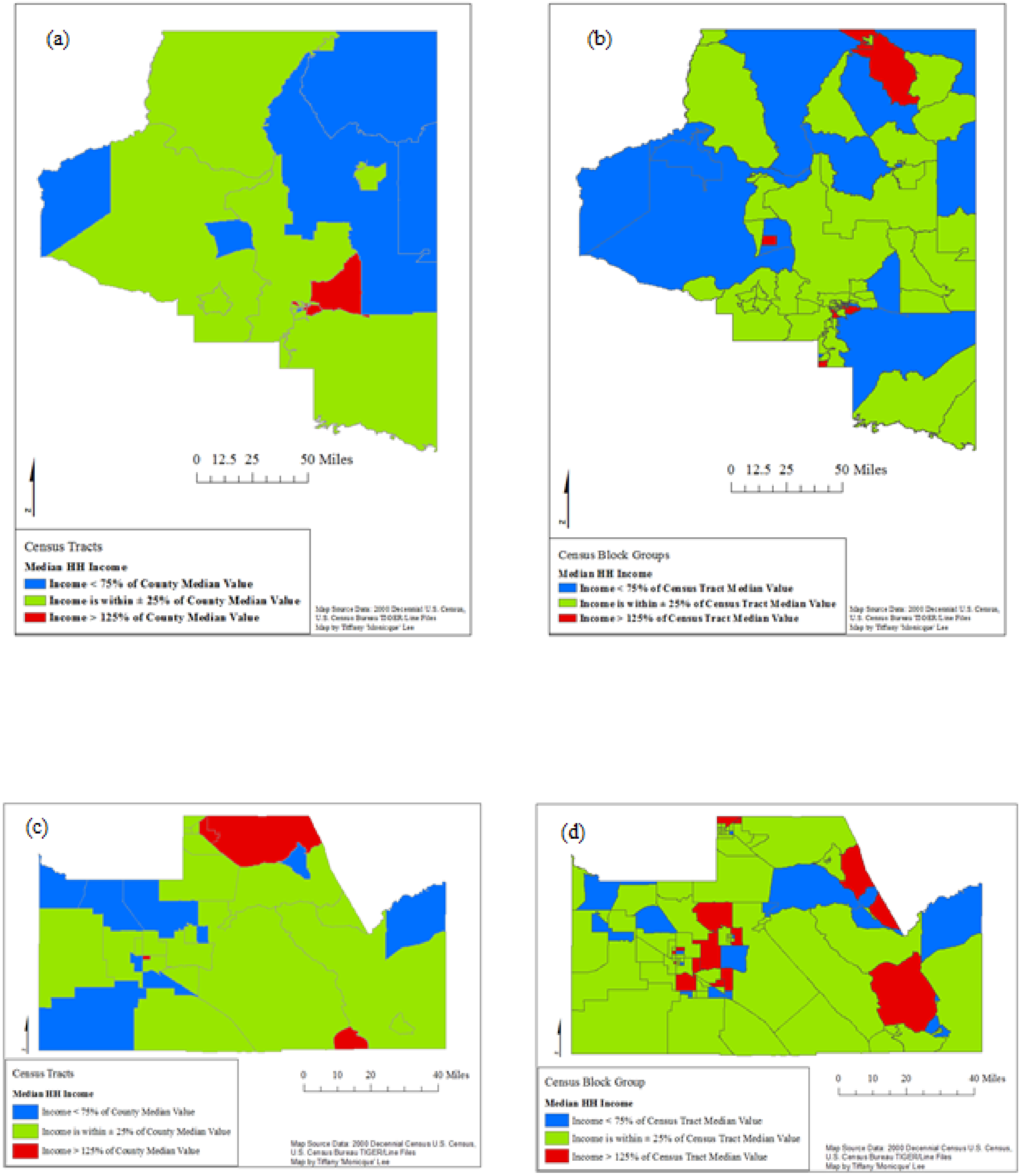

Figure 18. Group II, Median Household Income in: (a) Coconino County by census tract; (b) Coconino County by census block group; (c) Pinal County by census tract; and (d) Pinal County by census block group.

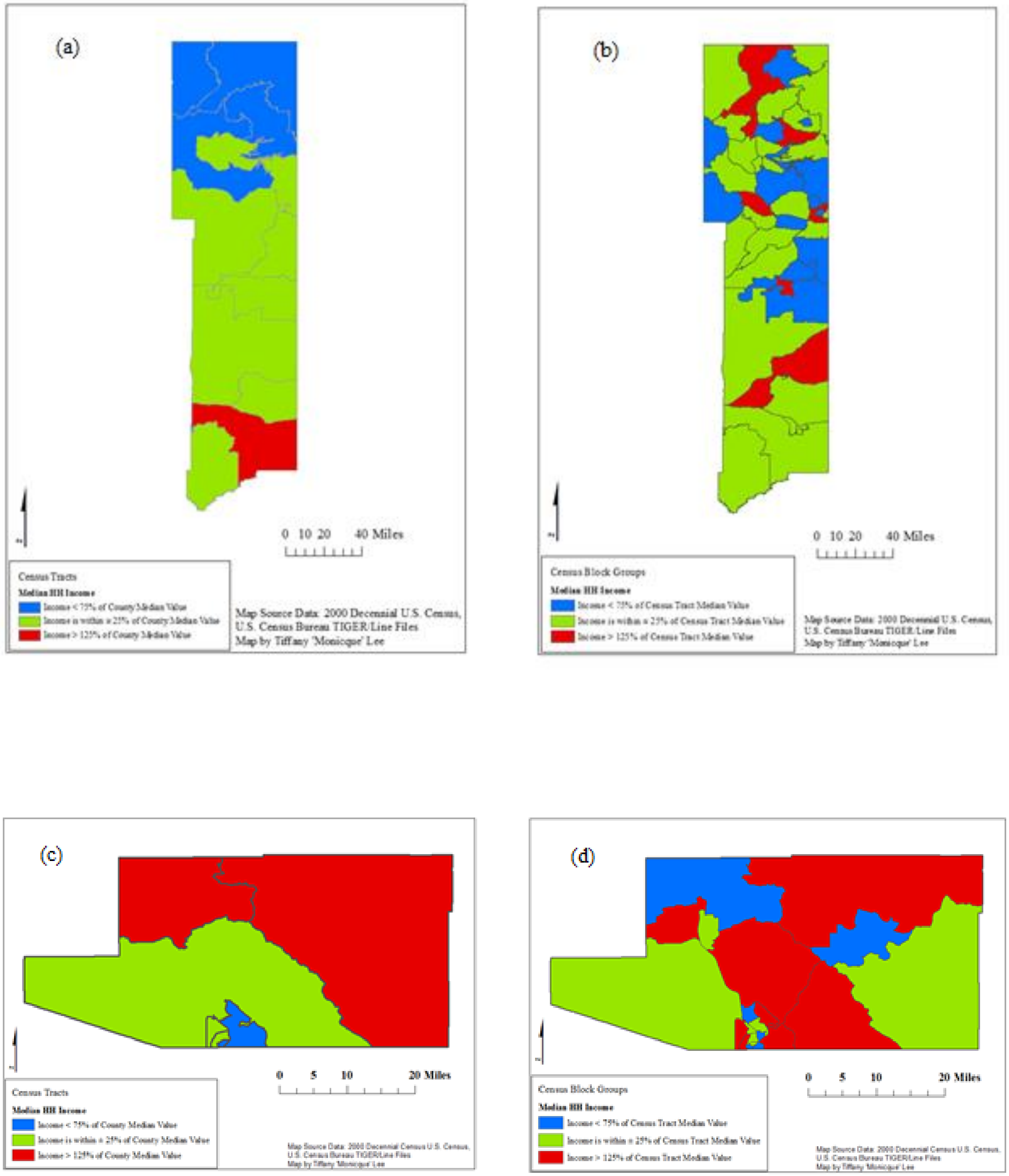

Figure 19. Group III, Median Household Income in: (a) Apache County by census tract; (b) Apache County by census block group; (c) Santa Cruz County by census tract; and (d) Santa Cruz County by census block group.

4.2. **Children and the Elderly**

The children (< 16 years old) and elderly (≥ 65 years old) represent the more vulnerable population given that this group often requires more public services than the non-vulnerable population, such as school, medical, and transportation. For the purposes of this study, vulnerability is conveyed by percent.

4.2.1. County Level

The size of the vulnerable populations by county is shown in Figure 20. The counties with the highest percentages of children and the elderly are predominantly rural and with one exception, cover the northern two-thirds of Arizona. The three counties that display the lowest vulnerability are dispersed and six of the eight counties that comprise the southern third of the state have similar percentages of children and elderly to the State of Arizona as a whole (Table 4).

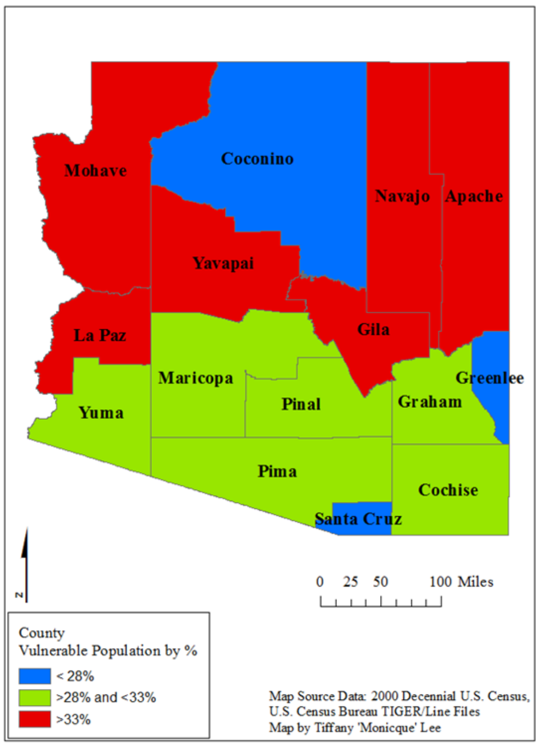

Figure 20. Vulnerability represented by the percentages of children and the elderly by county in the State of Arizona.

4.2.2. Census Tract Level

The percentages of children and the elderly by census tract are displayed in Figure 21 for the whole state as well as the Phoenix and Tucson Metro areas, respectively. The maps provide a stark contrast to the regional values displayed at the county level in Figure 20. As Figure 22 demonstrates, there are many census tracts with estimates < 75% (depicted in blue) or ≥ 125% (depicted in red) of the values at the county level. The most conspicuous outlier was a Census Tract in which all of the residents were 65 years or older, notwithstanding the county (Coconino) was one of the least vulnerable counties in Arizona.

Table 4 shows there were fewer census tracts classified as outliers when evaluating Vulnerability in place of Median Household Income in most of Arizona's counties as well as the state as a whole.

Table 4. Counts and percentages of census tract Vulnerability ≤ 75% and ≥ 125% of county median values.

County	Population	No. of census tracts	No. of census tracts with Vulnerability	Nos. and % of outlier census tracts *
Apache	69,423	14	14	0 + 2 = 14.3
Cochise	117,755	21	21	1 + 1 = 9.5
Coconino	116,320	28	27	6 + 7 = 46.4
Gila	51,335	15	15	0 + 2 = 13.3
Graham	33,489	8	8	1 + 2 = 37.5
Greenlee	8,547	3	3	0,0
La Paz	19,715	6	6	1 + 3 = 66.7
Maricopa	3,072,149	663	658	57 + 138 = 29.4
Mohave	155,032	30	30	0 + 1 = 3.3
Navajo	97,470	23	23	0 +1 = 4.3
Pima	843,746	198	198	18 + 52 = 35.4
Pinal	179,727	33	31	2 + 7 = 27.3
Santa Cruz	38,381	7	7	0,0
Yavapai	167,517	26	24	0 + 3 = 11.5
Yuma	160,026	33	33	4 + 15 = 57.6
Totals	5,130,632	1,108	1,098	90 + 234 = 29.2

* No. of census tracts with Vulnerability <75%, >125%, and the sum of the two classes of outliers as a percentage of total.

(a)

(b)

(c)

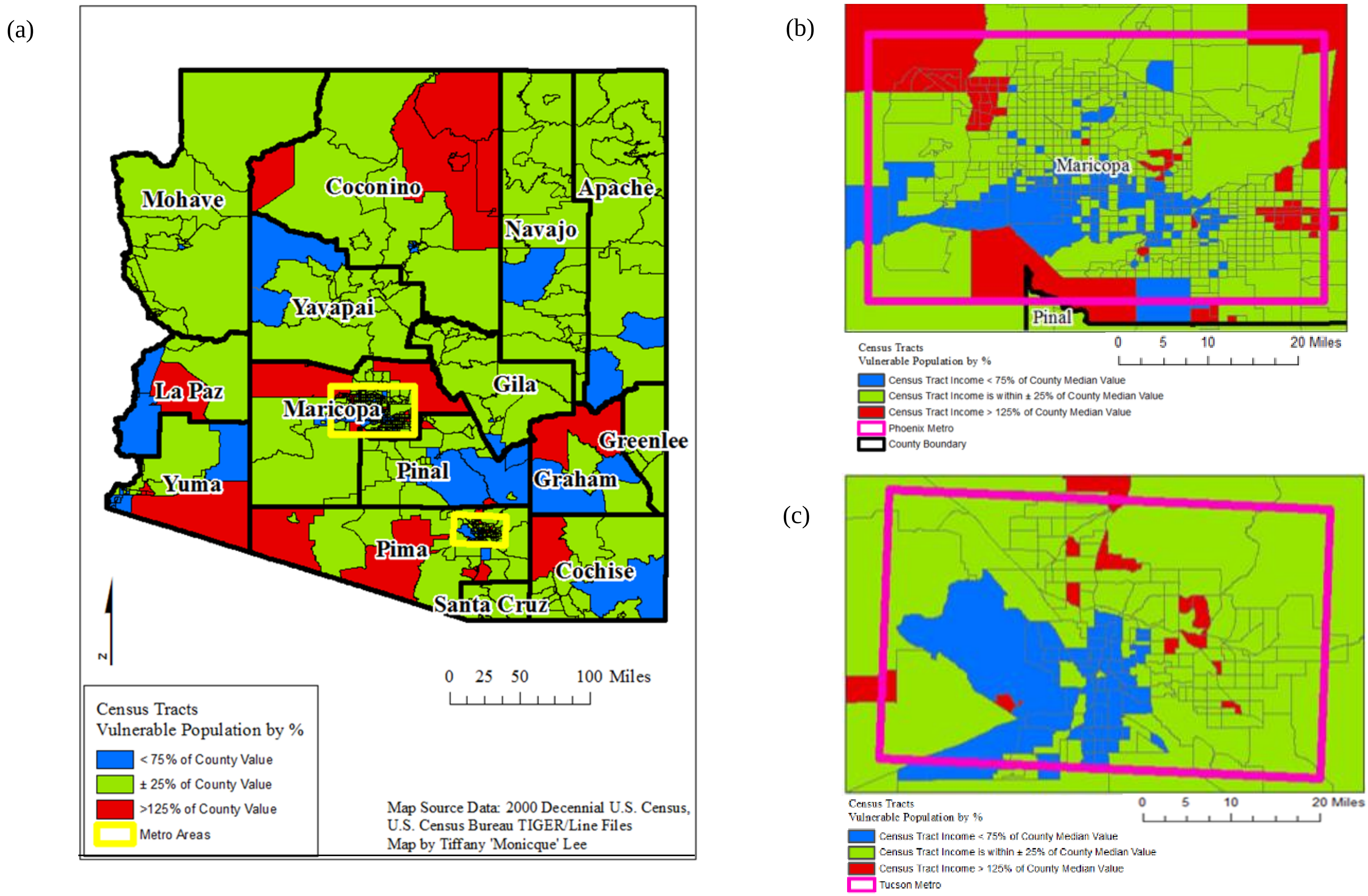

Figure 21. Maps showing the percentages of children and the elderly for each Arizona census tract (a) and the Phoenix Metro (b) and Tucson Metro (c) areas.

Table 2 shows how 52.2% of the census tracts were classified as outliers for Median Household Income, whereas just 29.2% of the census tracts were classified as outliers for percentages of children and the elderly (Table 4).

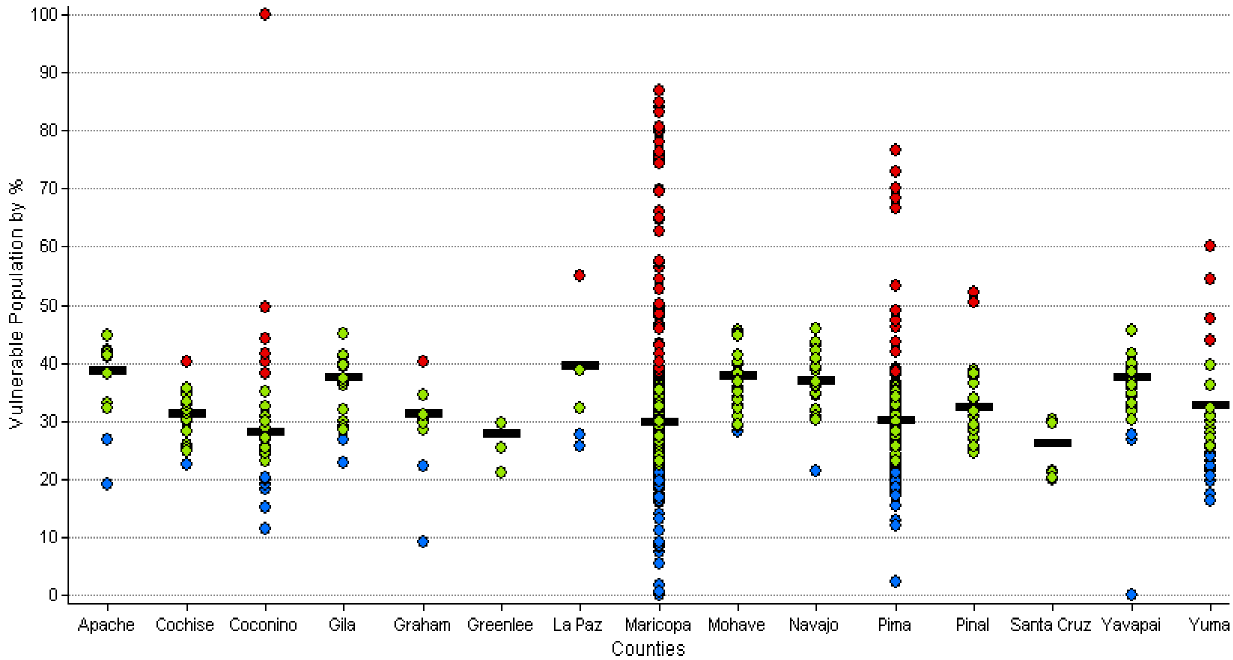

Figure 22. Arizona census tract estimates within each county for the percentages of children (≤ 16 years) and the elderly (≥ 65 years).

4.2.3. Census Block Group Level

When Census Block Group was used as the geographic reporting unit, in place of Census Tract, the pattern continued to change. Figures 23-27 show the census block groups classified as outliers around the census tract values. The distributions of high and low estimates around census tract values display which counties contained varying numbers of children and the elderly over short distances. Coconino County included several census block groups in a single census tract with all residents > 65 years as expected (Figure 26f).

The metrics summarized in Table 5 show that the outliers increased when using census block groups in place of census tracts, as compared to the pattern with county and census tract values. Table 5 reports that 45.1% of the census block groups were classified as outliers, compared to the 29.2% of the census tracts reported in Table 4.

Seven of the counties exceeded the State average of 45.1%: Maricopa (45.9%) and Pima (52.7%) in Group I, and La Paz (56.6%), Pinal (68.1%), Graham (70.4%), Yuma (78.6%), and Santa Cruz (85%) in Groups II and III. Table 5 also denotes the sparse population of Greenlee County in Group III and how it had no census block groups classified as outliers.

Figures 28-31 demonstrate how census tract and census block group outliers for Vulnerability are dispersed across the 15 counties and the State of Arizona as a whole. Depending on the geographic reporting unit examined, differing patterns of vulnerability emerge.

When the counties in Group I are visualized at the census tract level, large census tracts classified as outliers (≥125%) are found at some distance away from the Metro areas (Figures 28a and 29a). When the reporting unit becomes smaller, at the census block group level, patterns showing areas with high percentages of children and elderly are seen within the Metro areas (Figures 28b and 29b).

Group II and Group III shows counties that are largely rural with distinct patterns of vulnerability that reflect proximity to economic centers in both census tracts and census block groups (Figures 30a-d and 31a-d).

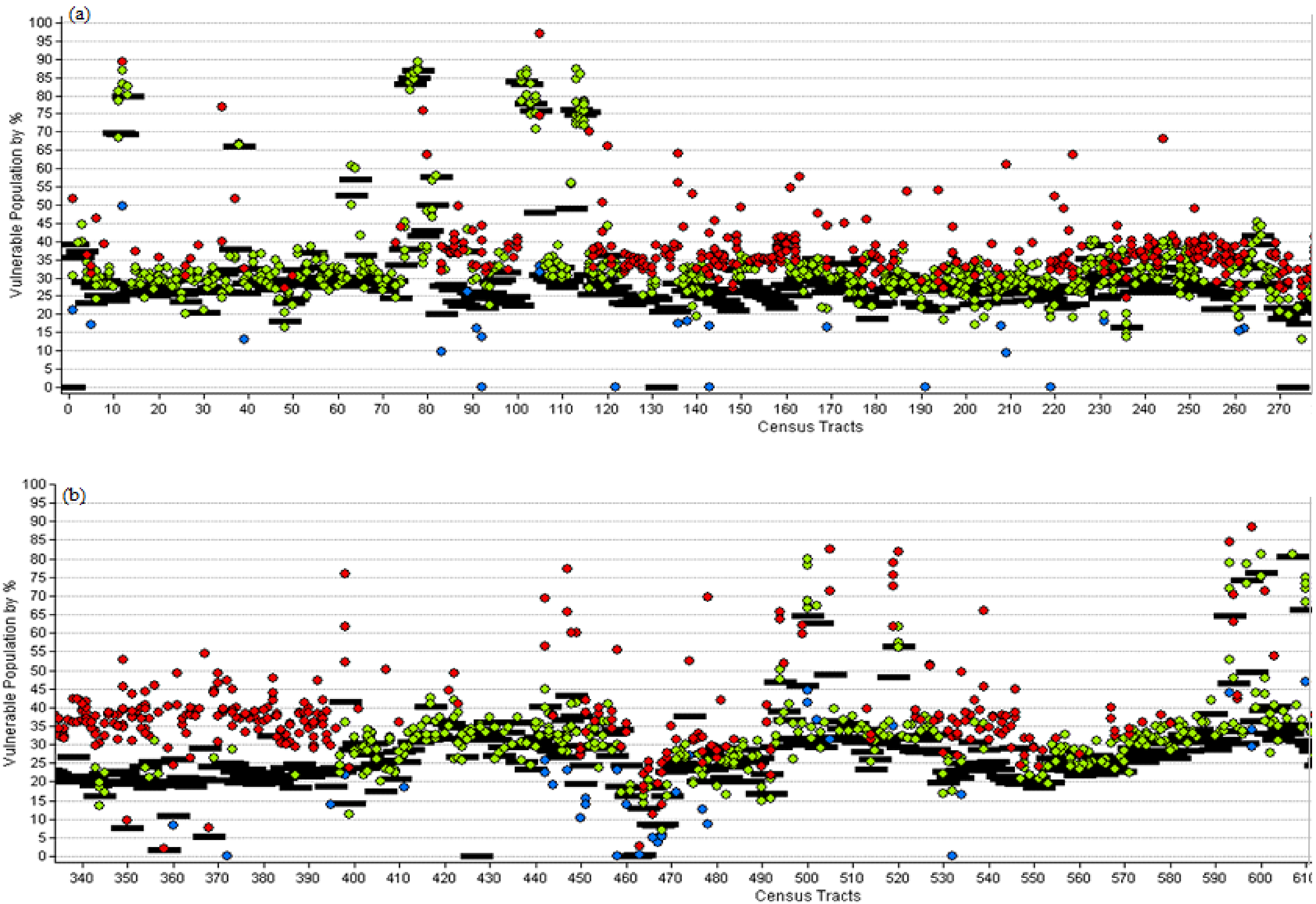

Figure 23. Group I: Census block group Vulnerable Population by census tracts 1-333 (a) and census tracts 334-663 (b) in Maricopa County.

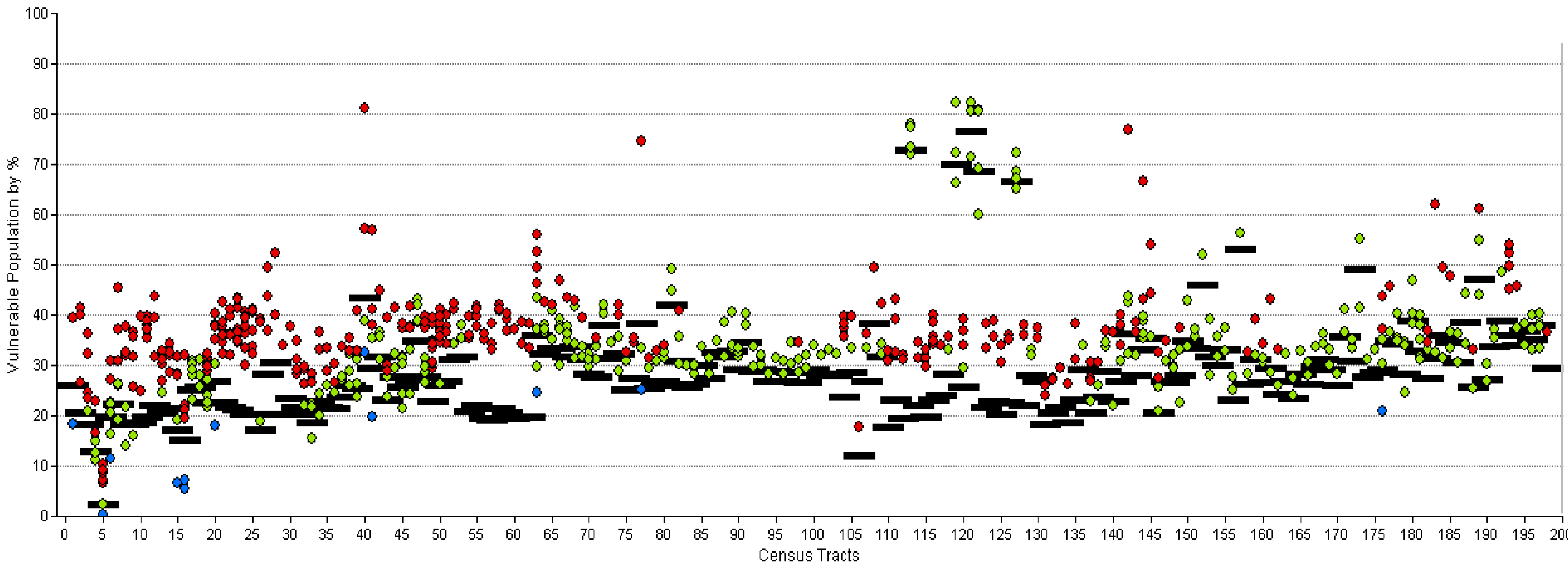

Figure 24. Group I (Continued): Census block group vulnerable population by census tract in Pima County.

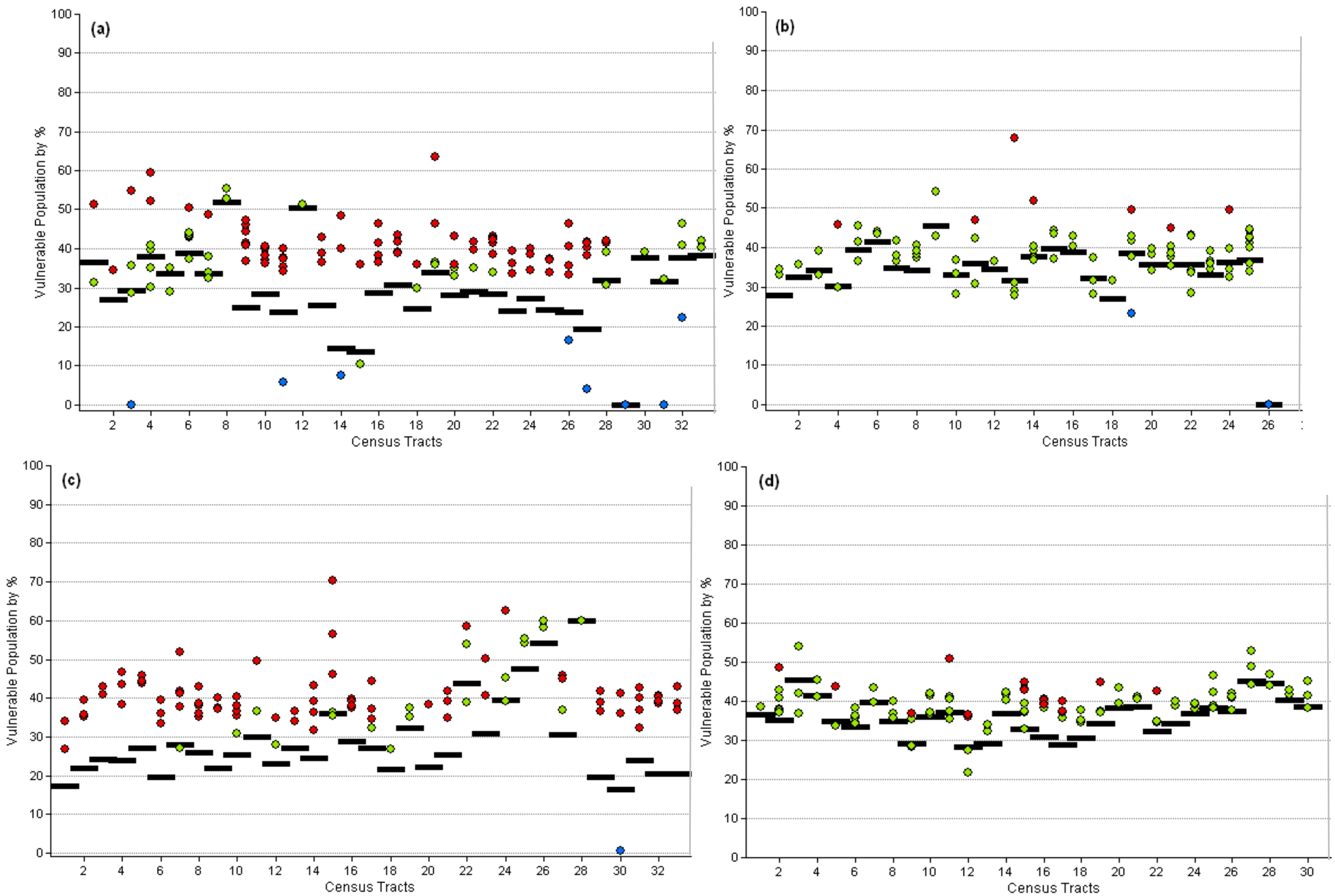

Figure 25. Group II: Census block group vulnerable population by census tract for: (a) Pinal; (b) Yavapai; (c) Yuma; (d) Mohave.

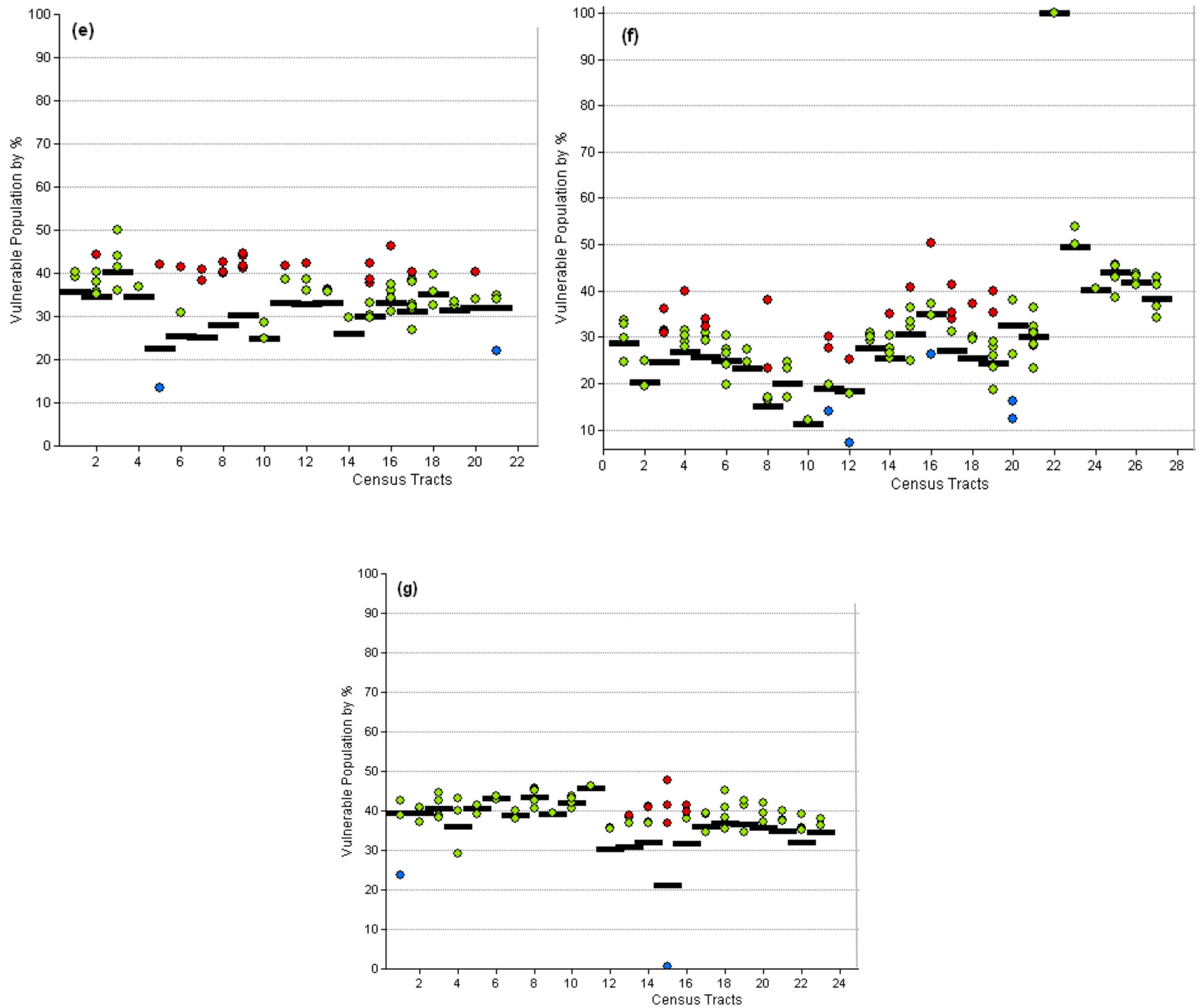

Figure 26. Group II (Continued): Census block group vulnerable population by census tract for: (e) Cochise; (f) Coconino; and (g) Navajo counties.

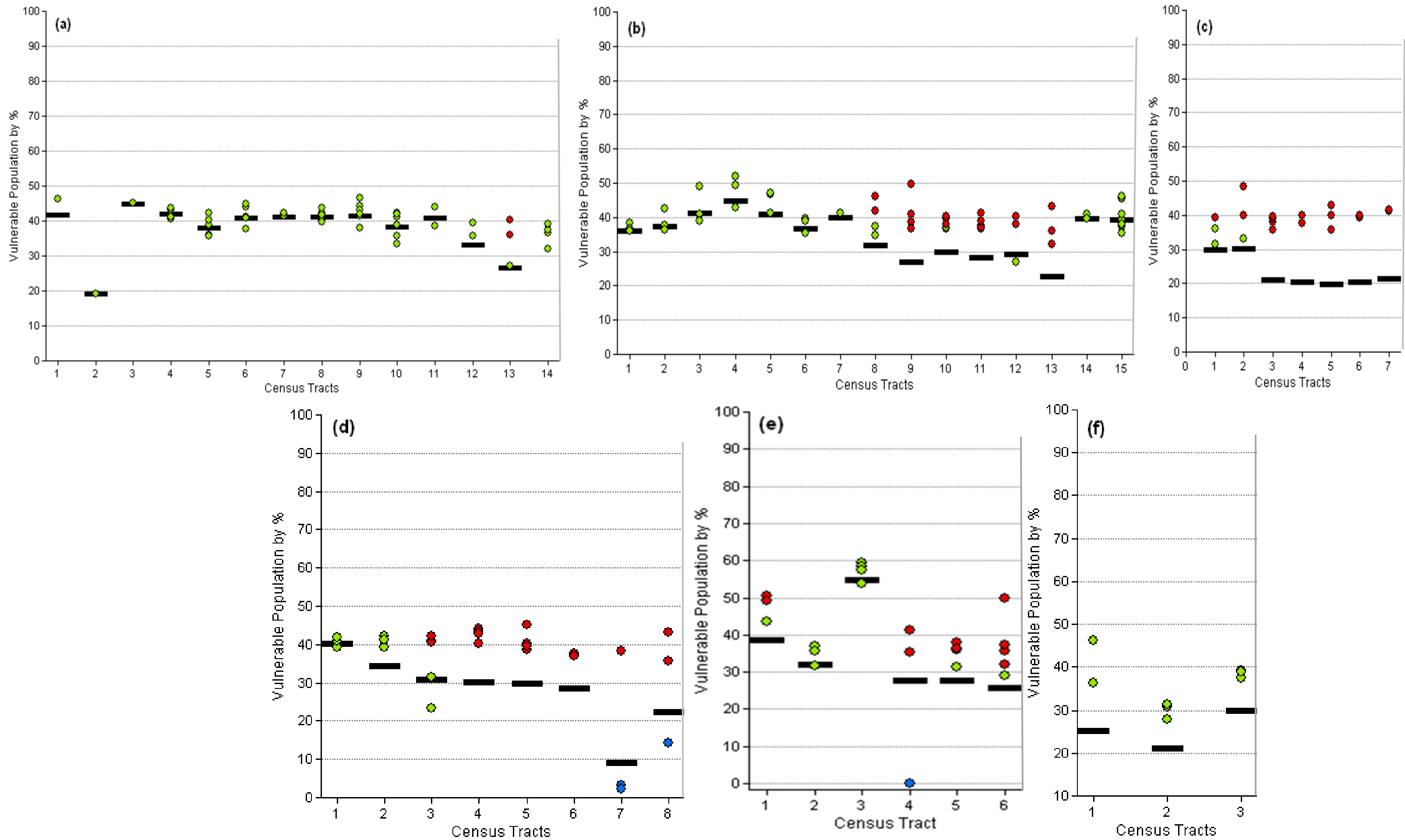

Figure 27. Group III: Census block group Vulnerable Population by census tracts for: (a) Apache; (b) Gila; (c) Santa Cruz; (d) Graham; (e) La Paz; and (f) Greenlee Counties.

Table 5. Counts and percentages of census block group Vulnerability that are ≤ 75% and ≥ 125% of the corresponding census tract values.

County	Population	No. of census block groups	No. of census block groups with Vulnerability	Nos. and % of outlier census block groups *
Apache	69,423	54	54	2 + 0 = 3.7
Cochise	117,755	72	72	23 + 2 = 34.7
Coconino	116,320	106	106	21 + 5 = 24.5
Gila	51,335	55	55	20 + 0 = 36.4
Graham	33,489	27	27	16 + 3 = 70.4
Greenlee	8,547	8	8	0, 0
La Paz	19,715	23	22	12 + 1 = 56.5
Maricopa	3,072,149	2,113	1,102	901 + 69 = 45.9
Mohave	155,032	101	101	18 + 0 = 17.8
Navajo	97,470	74	74	8 + 2 = 13.5
Pima	843,746	617	617	311 + 14 = 52.7
Pinal	179,727	116	113	71 + 8 = 68.1
Santa Cruz	38,381	20	20	17 + 0 = 85.0
Yavapai	167,517	86	85	7 + 2 = 3.7
Yuma	160,026	98	98	76 + 1 = 34.7
Totals	5,130,632	3,570	2,554	1,503 + 107 = 45.1

* No. of census block groups with Vulnerability <75%, >125%, and the sum of the two classes of outliers as a percentage of total.

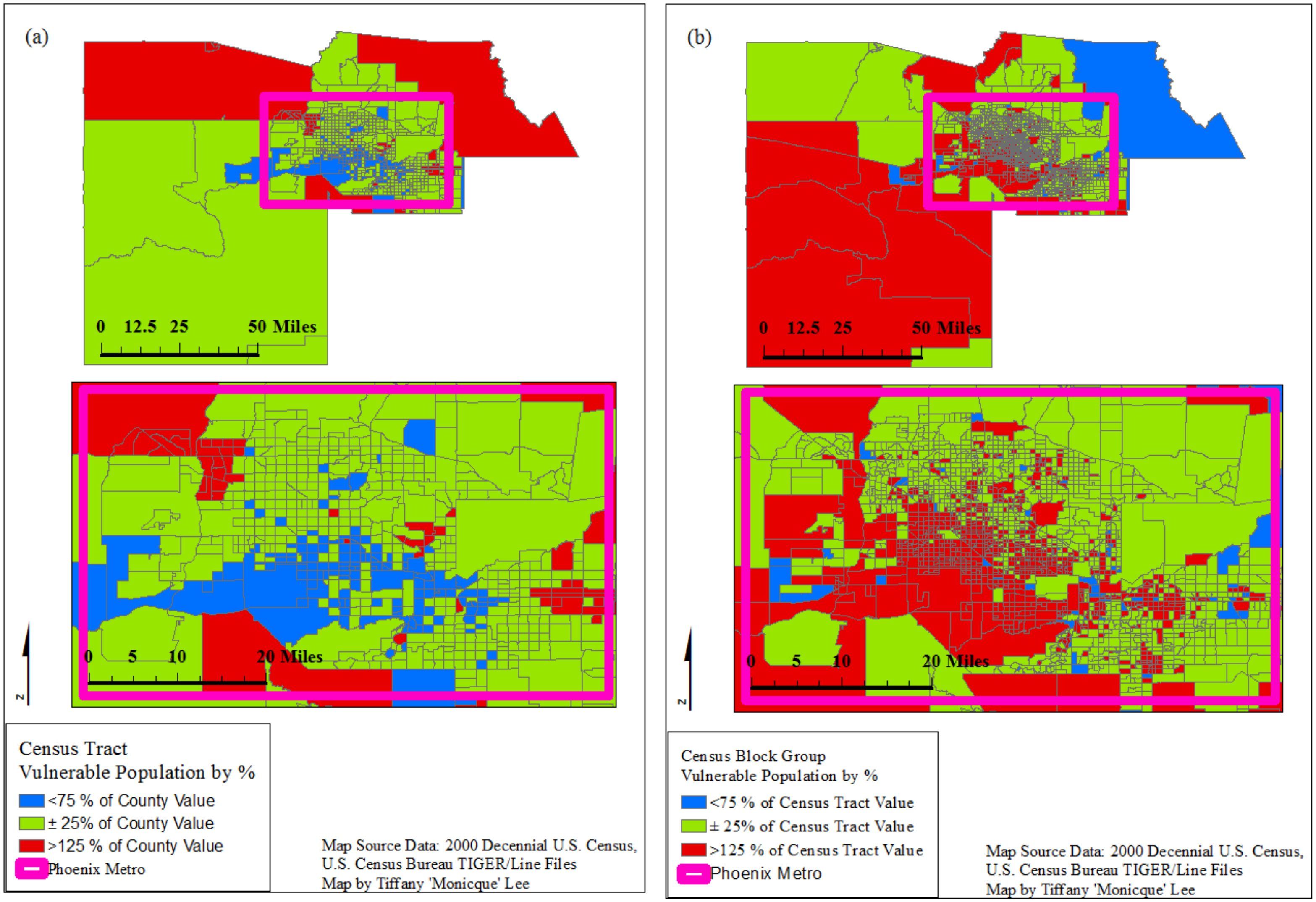

Figure 28. Group I, Percent Vulnerable Population in Maricopa County by: (a) census tract; and (b) census block group. Both pairs of maps show the entire county on the top with the greater Phoenix Metro area highlighted and shown at a larger scale on the bottom.

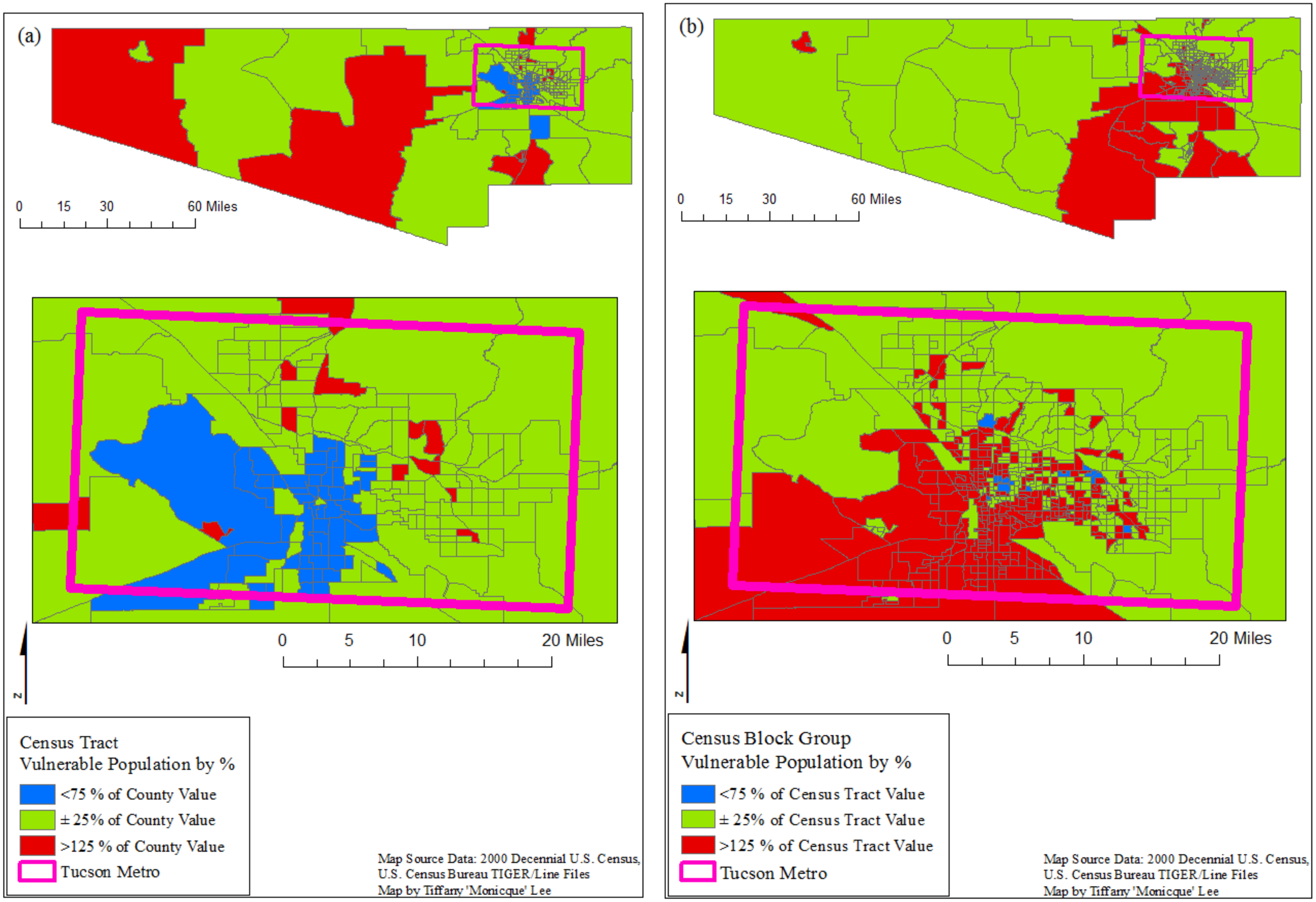

Figure 29. Group I (continued), Percent Vulnerable Population in Pima County by: (a) census tract; and (b) census block group. Both pairs of maps show the entire county on the top with the greater Tucson Metro area highlighted and shown at a larger scale on the bottom.

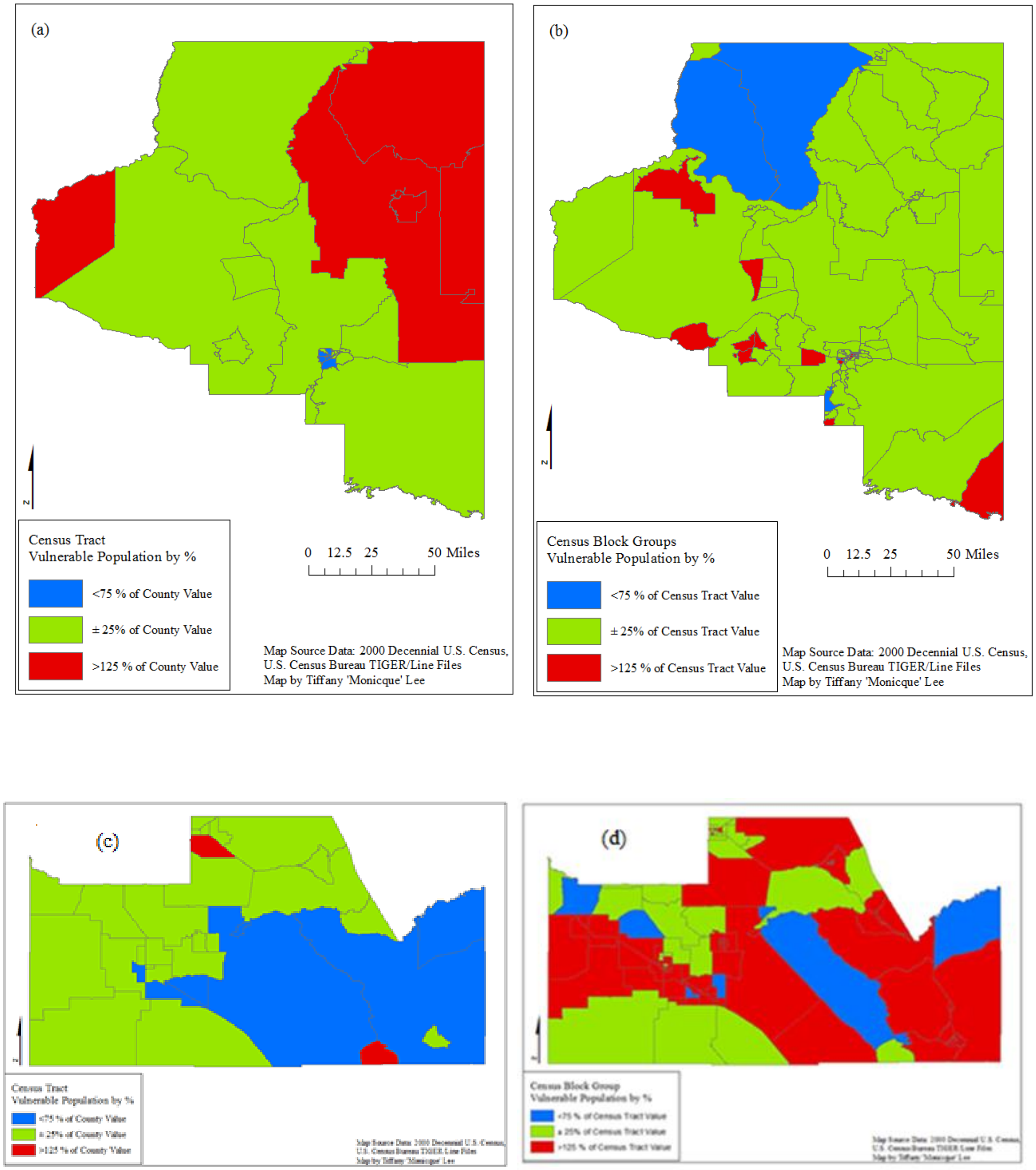

Figure 30. Group II, Percent Vulnerable Population in (a) Coconino County by census tract; (b) Coconino county by census block group; (c) Pinal County by census tract; (d) Pinal County by census block group.

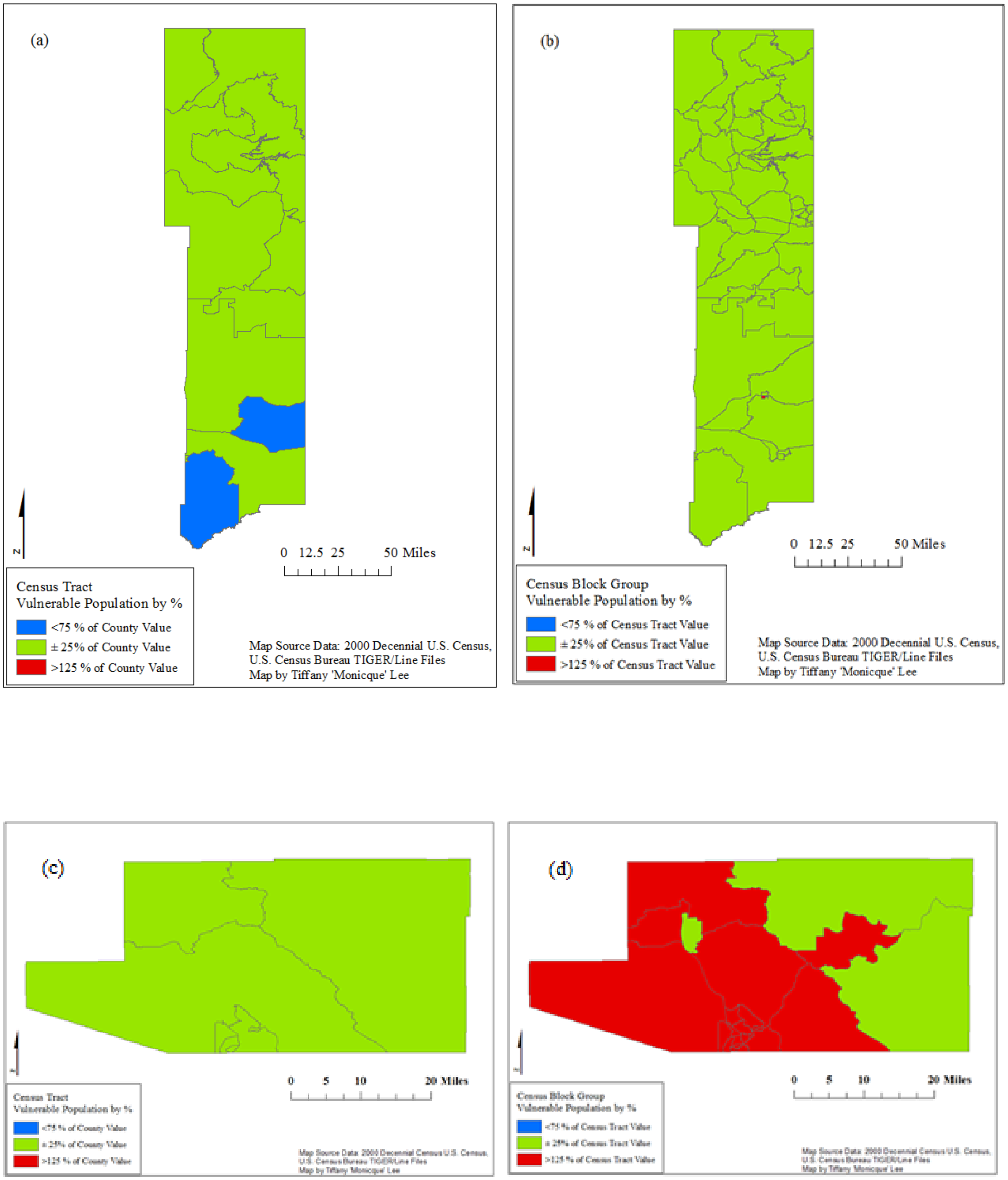

Figure 31. Group III, Percent Vulnerable Population in: (a) Apache County by census tract; (b) Apache County by census block group; (c) Santa Cruz County by census tract; and (b) Santa Cruz County by census block group.

4.3. **Native American Population**

Ethnicity is also a commonly studied health determinant as it is indicative of specific needs within certain populations and/or spatial zones. The overall Native American population in Arizona is relatively small yet significant, evident by the presence of 22 Federally recognized AIRs that cover a large proportion of the State, as well as the strong cultural identity that their presence represents in terms of Arizona's history and heritage.

4.3.1. County Level

Figure 32 displays the county distribution of the State's Native American population by percentage. The two counties in red, denoting the counties with the highest Native American populations, are counties where approximately half of their total area is AIRs. The counties in blue and green also contain AIRs, however, apart from Pima County, the reservations tend to be smaller in physical area and are more dispersed throughout these counties.

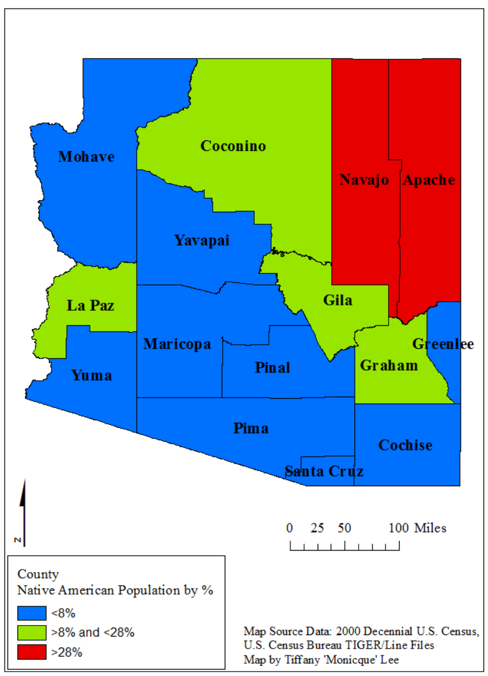

Figure 32. Native American Population by percent for each Arizona county.

4.3.2. Census Tract Level

The scatterplot in Figure 33 shows Native American census tract populations compared to county population values. Census tracts are colored to highlight census tracts classified as outliers (where < 75% of census tracts classified as outliers are depicted in blue, > 125% of census tracts classified as outliers are depicted in red, and ± 25% of census tracts classified as outliers are depicted in green) and show that numerous census tracts in multiple counties across the state have Native American population of 50% or more. Eight of the 15 counties have one or more census tracts with Native American census tract populations at or near 100%.

Figure 34 shows the census tract Native American population distribution relative to county values across Arizona. When the spatial reporting unit is decreased from county to census tract, large numbers of census tracts and therefore areas are classified as outliers. The numbers in Table 6 confirm this result, showing that 50% or more of the census tracts in every county were classified as outliers and 75% or more of the census tracts in 11 of 15 counties were classified as outliers (i.e. with substantially smaller or larger Native American populations, than the appropriate county as a whole).

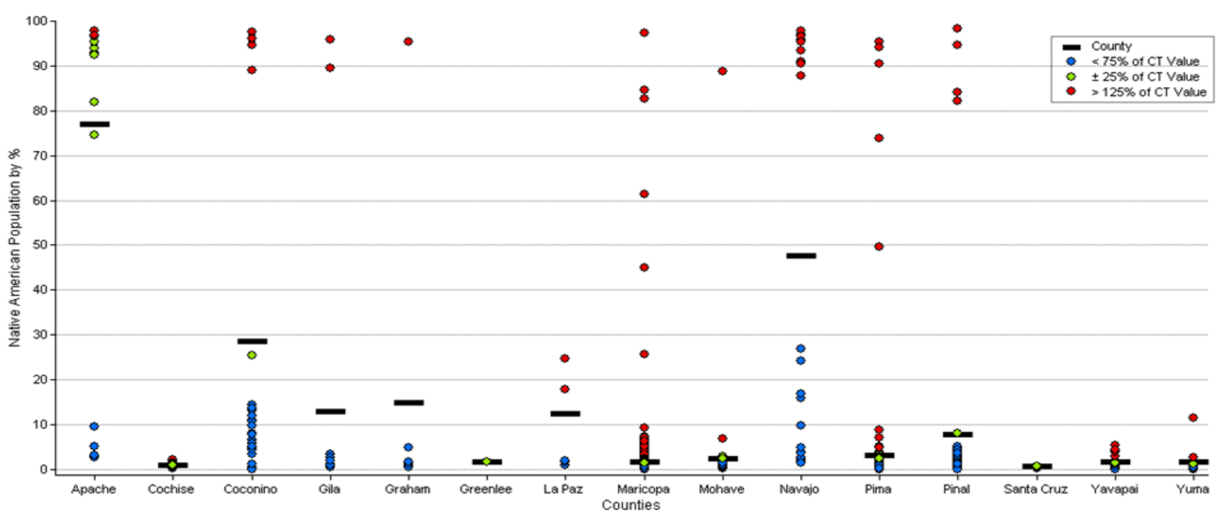

Figure 33. Arizona Census Tract estimates within each county for Native American population by percentage.

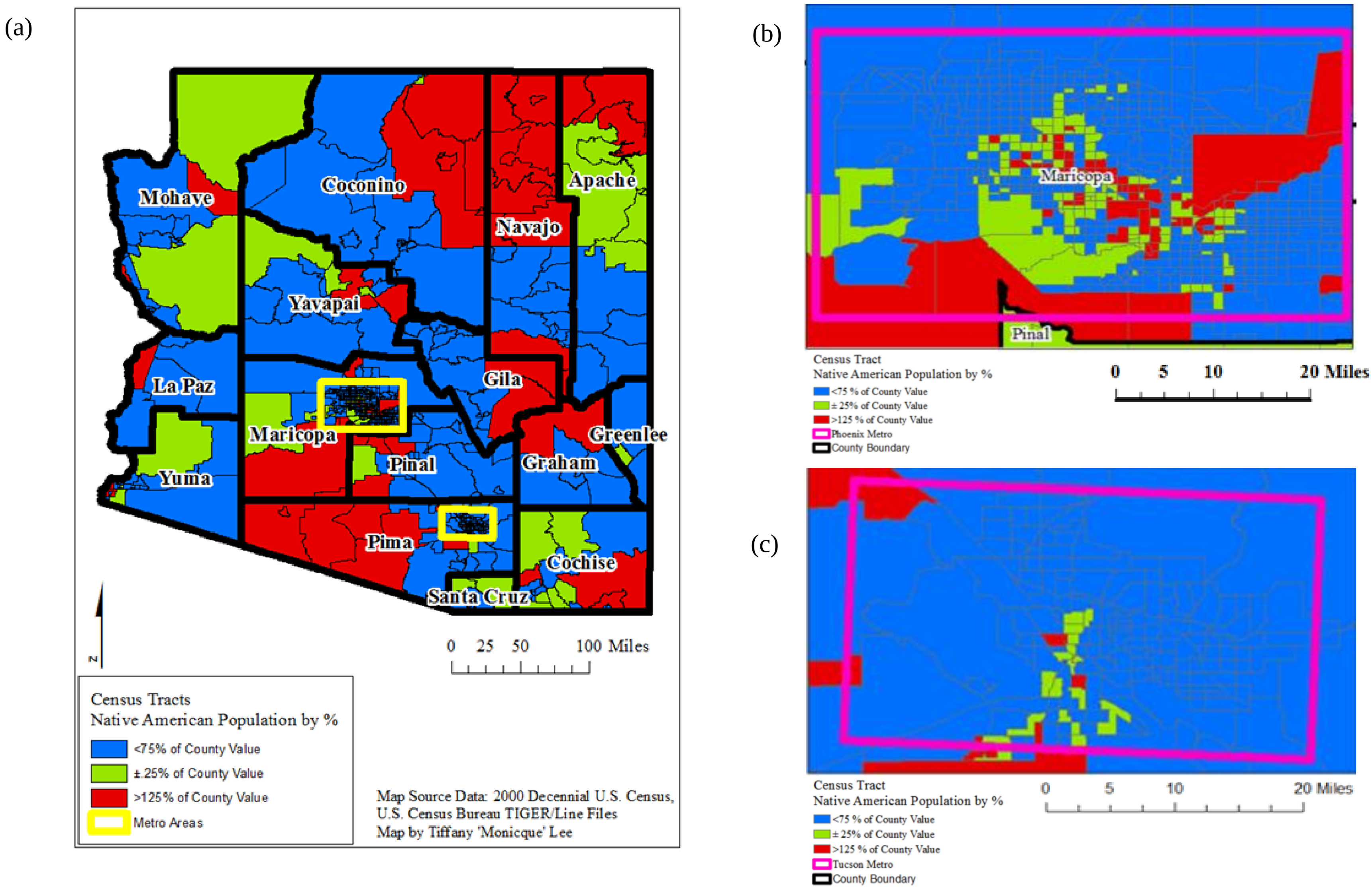

Figure 34. Native American Population by Percentage for each Arizona census tract (a) with separate maps showing the Phoenix Metro (b), and Tucson Metro (c) areas.

Table 6. Counts and percentages of census tract Native American Population by percentage examining outliers ≤ 75% and ≥ 125% of county median values.

County	Population	No. of census tracts	No. of census tracts with Native American population	Nos. and % of outlier census tracts *
Apache	69,423	14	14	3 + 4 = 50.0
Cochise	117,755	21	21	3 + 11 = 66.7
Coconino	116,320	28	27	5 + 22 = 96.4
Gila	51,335	15	15	2 + 13 = 100.0
Graham	33,489	8	8	1 + 7 = 100.0
Greenlee	8,547	3	3	0 + 2 = 66.7
La Paz	19,715	6	6	2 + 4 = 100.0
Maricopa	3,072,149	663	656	71 + 476 = 82.5
Mohave	155,032	30	30	2 + 25 = 90.0
Navajo	97,470	23	23	11 + 12 = 100.0
Pima	843,746	198	198	10 + 170 = 90.9
Pinal	179,727	33	31	4 + 28 = 97.0
Santa Cruz	38,381	7	7	0 + 5 = 71.4
Yavapai	167,517	26	25	4 + 18 = 84.6
Yuma	160,026	33	32	2 + 23 = 75.8
Totals	5,130,632	1,108	1,096	120 + 820 = 84.8

* No. of census tracts with Native American Population by % < 75%, > 125%, and the sum of the two classes of outliers as a percentage of total.

4.3.3. Census Block Group Level

The scatterplots in Figures 35-39 show the Native American census block group population percentages relative to the corresponding census tract values, with colors showing the same outlier groupings as earlier. The scatterplots show how the distribution of the Native American population numbers vary across the state when different spatial reporting units are chosen.

The scatterplots for Group I (Figures 35 and 36) show that although these counties are the two highest populated counties in the State, only three census tracts in Maricopa County had Native American populations of > 25%. Table 6, however, indicates that that the number of census block groups classified as outliers were relatively high in both Maricopa (69.1%) and Pima (63%) counties.

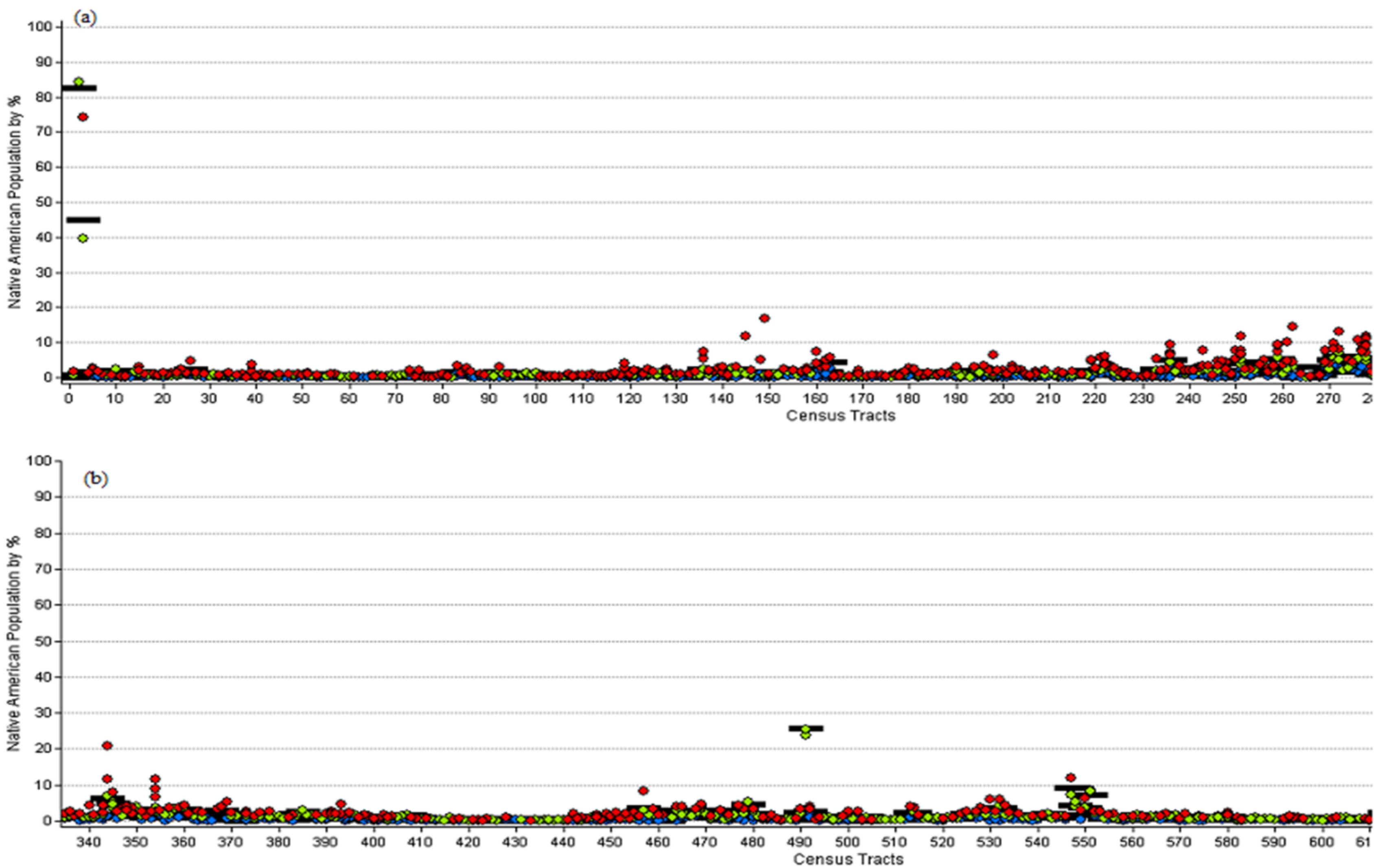

Figure 35. Group I: Census block group percentage Native American population by census tract for Maricopa County: (a) census tracts 1-333; and (b) census tracts 334-663.

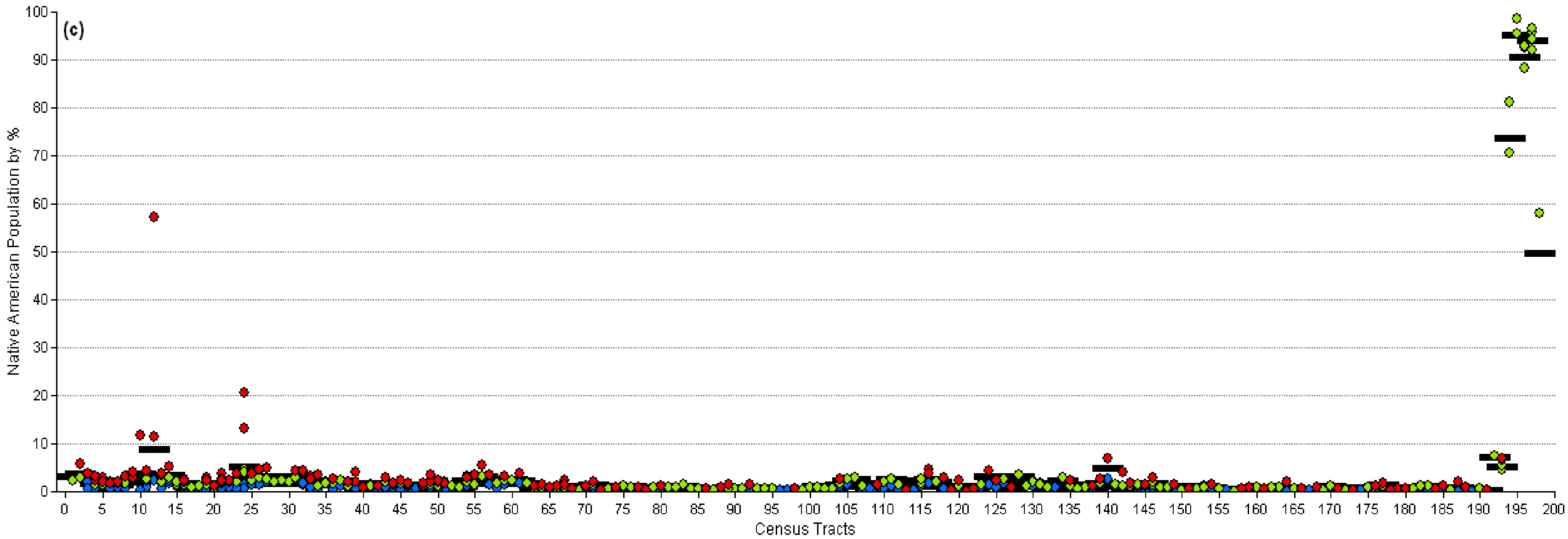

Figure 36. Group I (Continued): Census block group percentage Native American population by census tract in Pima County.

The scatterplots in Figures 37, 38 and 39 show several of the counties with medium-sized and small populations have one or more census block groups with large Native American populations (see Figures 37a, 38f and g, and Figures 39a, b, and c for examples) and that the switch from census tracts to census block groups highlights numerous outliers, or census block groups with relatively high or low Native American populations relative to the corresponding census tract estimates.

The three counties with the most census block groups classified as outliers were La Paz (73.9%), Gila (74.5%), and Graham (81.5%) and those with the lowest number of census block groups classified as outliers were Apache (14.8%), Navajo (36.5%), and Greenlee (50.0%). All but Navajo County are in Group III, the group with the smallest populations across the State. Navajo County is the least populated county in Group II.

Table 7 indicates that there were fewer outliers going from census tract to census block group (65.9%) than there was going from county to census tract (84.8%) (Table 6) for this variable. The total number of census block groups classified as outliers is still substantial compared to the first two variables examined when switching the geographic unit from census tract to census block group: Median Household Income, Table 3 (24.8%) and Vulnerability, Table 5 (45.1%).

The maps in Figures 40-43 visually demonstrate the outliers when switching the geographic reporting units from census tract to census block group and how the patterns vary by county, population, and AIRs. The Group I counties, Figures 40 and 41, exhibit an increase in variability at the census block group level and the inset maps show that the outliers are especially evident in the metropolitan areas in both counties. The Group II counties, Figure 42, demonstrate similar patterns switching from census tract to census block group.

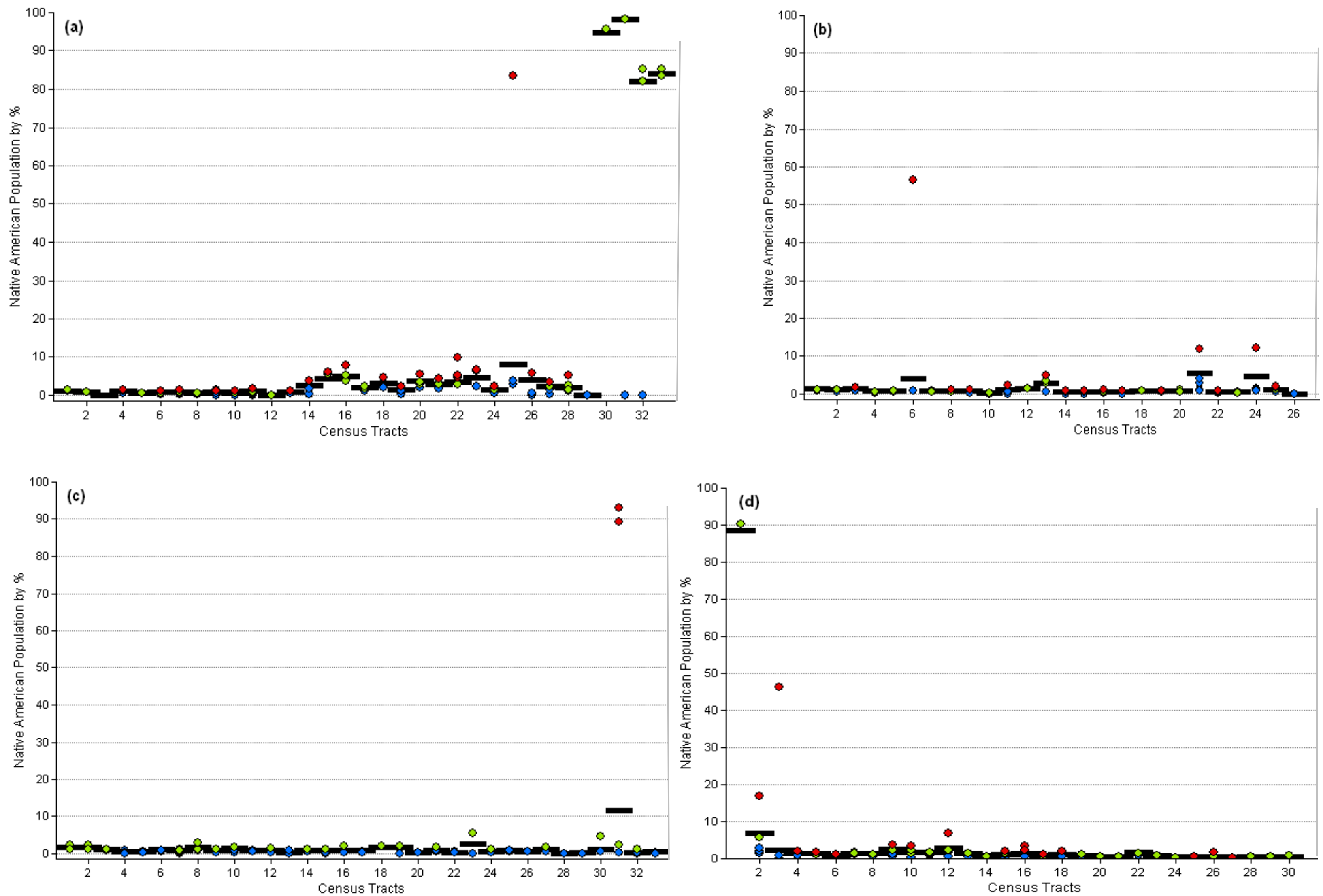

Figure 37. Group II: Census block group percentage Native American Population compared to census tract percentage Native American Population for: (a) Pinal; (b) Yavapai; (c) Yuma; (d) Mohave.

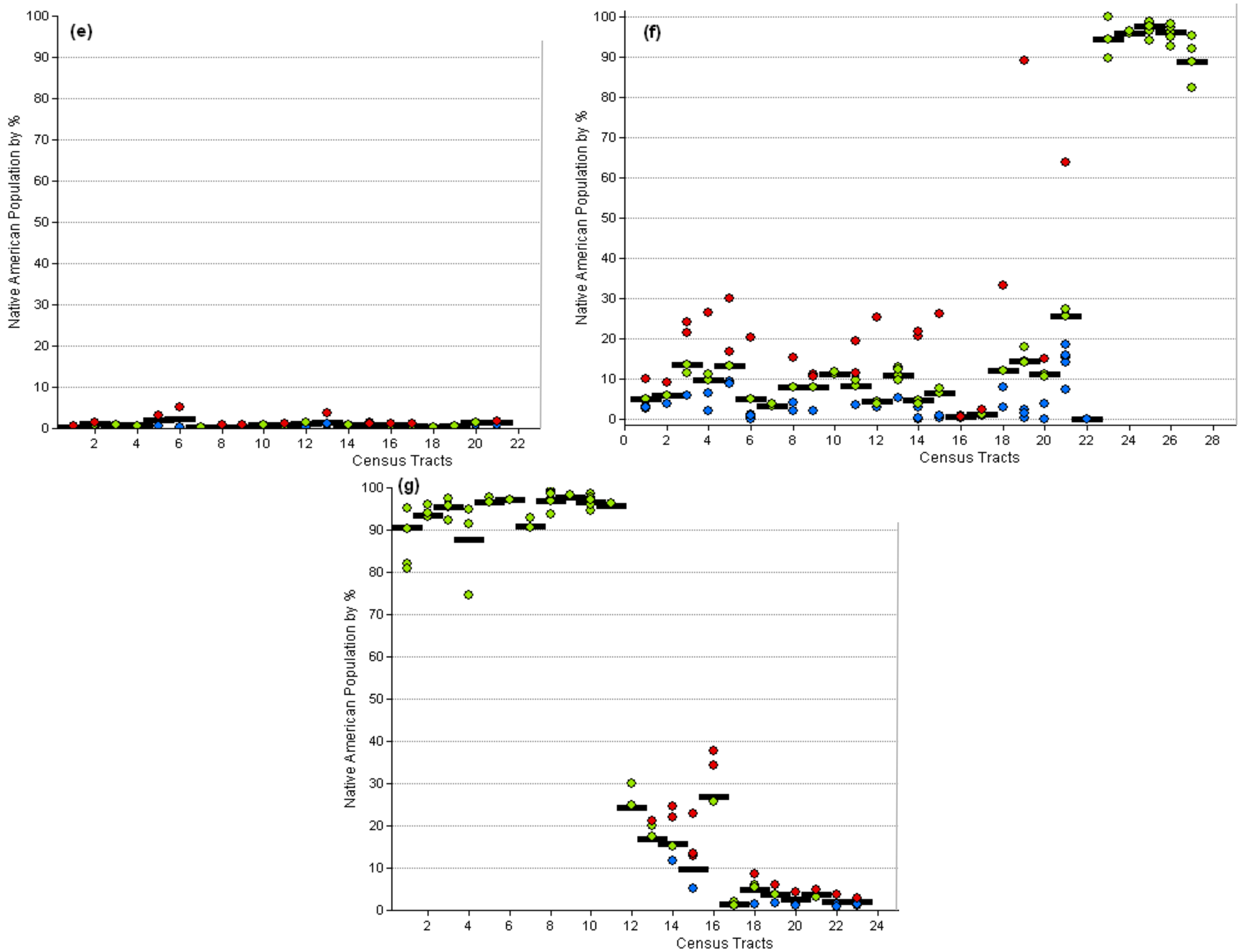

Figure 38. Group II (Continued): Census block group percentage Native American Population compared to census tract percentage Native American Population for: (e) Cochise; (f) Coconino; and (g) Navajo counties.

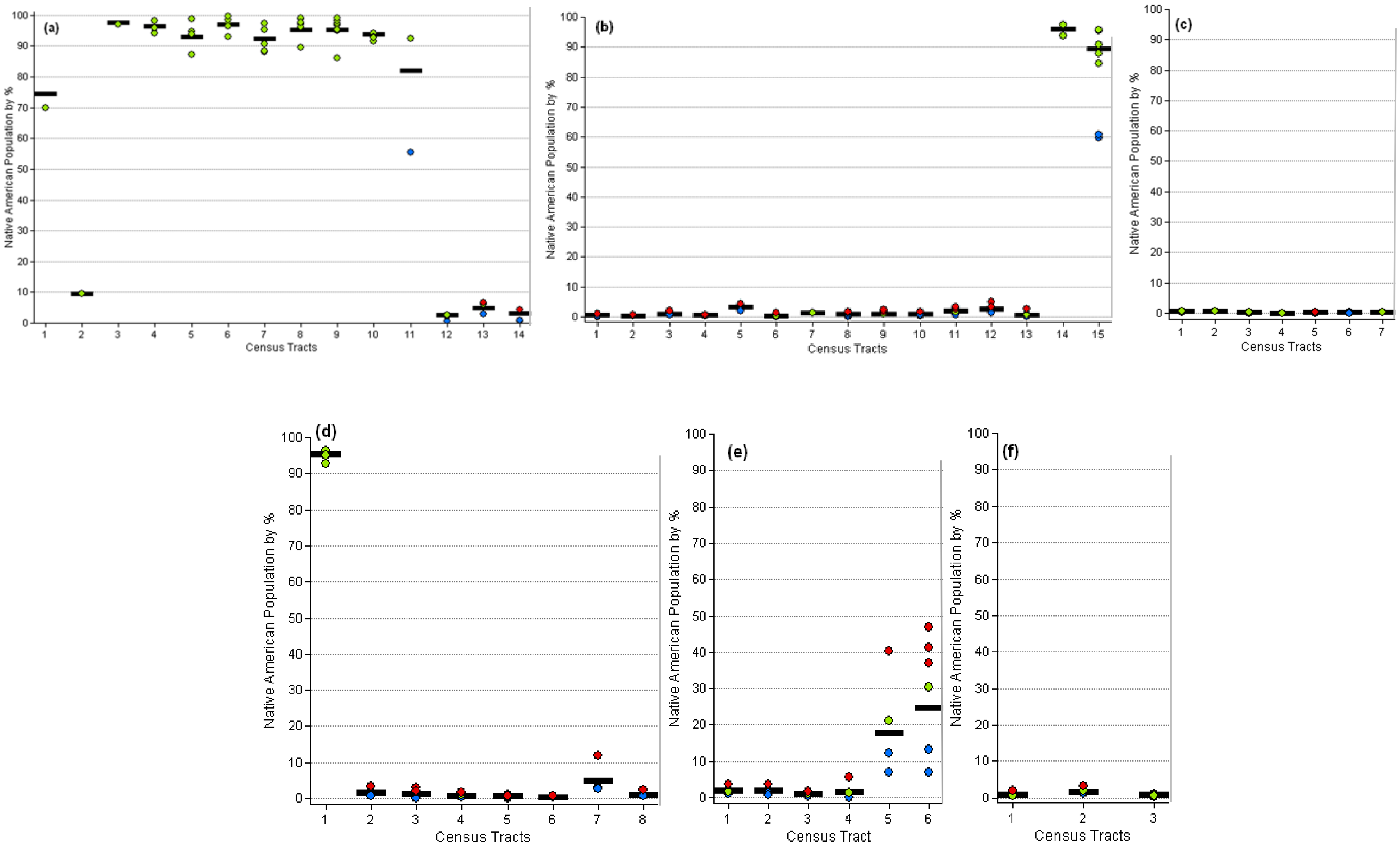

Figure 39. Group III: Census block group Native American population by percentage compared to census tract Native American population by percentage for: (a) Apache; (b) Gila; (c) Santa Cruz; (d) Graham; (e) La Paz; and (f) Greenlee Counties.

The Group III counties in Figures 43, visually demonstrate the modest numbers of census block groups classified as outliers compared to the more populated counties.

At the census tract level, there is little diversity within reporting zones, as seen in Figures 40-43, and there are sizable percentages of census tracts classified as outliers (Table 6). As the reporting unit becomes smaller, variability becomes evident, especially in and around AIRS, throughout Metro areas and economic centers.

Table 7. Counts and percentages of census block group Native American population by % ≤ 75% and ≥ 125% of census tract values.

County	**Population**	**No. of census block groups**	**No. of census block groups with Native American population**	**Nos. and % of outlier census block groups ***
Apache	69,423	54	54	3 + 5 = 14.8
Cochise	117,755	72	71	17 + 33 = 69.4
Coconino	116,320	106	102	25 + 41 = 62.3
Gila	51,335	55	51	19 + 22 = 74.5
Graham	33,489	27	26	9 + 13 = 81.5
Greenlee	8,547	8	8	2 + 2 = 50.0
La Paz	19,715	23	22	8 + 9 = 73.9
Maricopa	3,072,149	2,113	1,915	594 + 867 = 69.1
Mohave	155,032	101	100	16 + 43 = 58.4
Navajo	97,470	74	74	15 + 12 = 36.5
Pima	843,746	617	578	140 + 249 = 63.0
Pinal	179,727	116	104	27 + 49 = 65.5
Santa Cruz	38,381	20	15	1 + 10 = 55.0
Yavapai	167,517	86	83	19 + 34 = 61.6
Yuma	160,026	98	92	23 + 45 = 69.4
Totals	5,130,632	3,570	3,295	918 + 1,434 = 65.9

* No. of census block groups with Native American population <75%, >125% of the corresponding census tract value, and the sum of the two classes of outliers as a percentage of total.

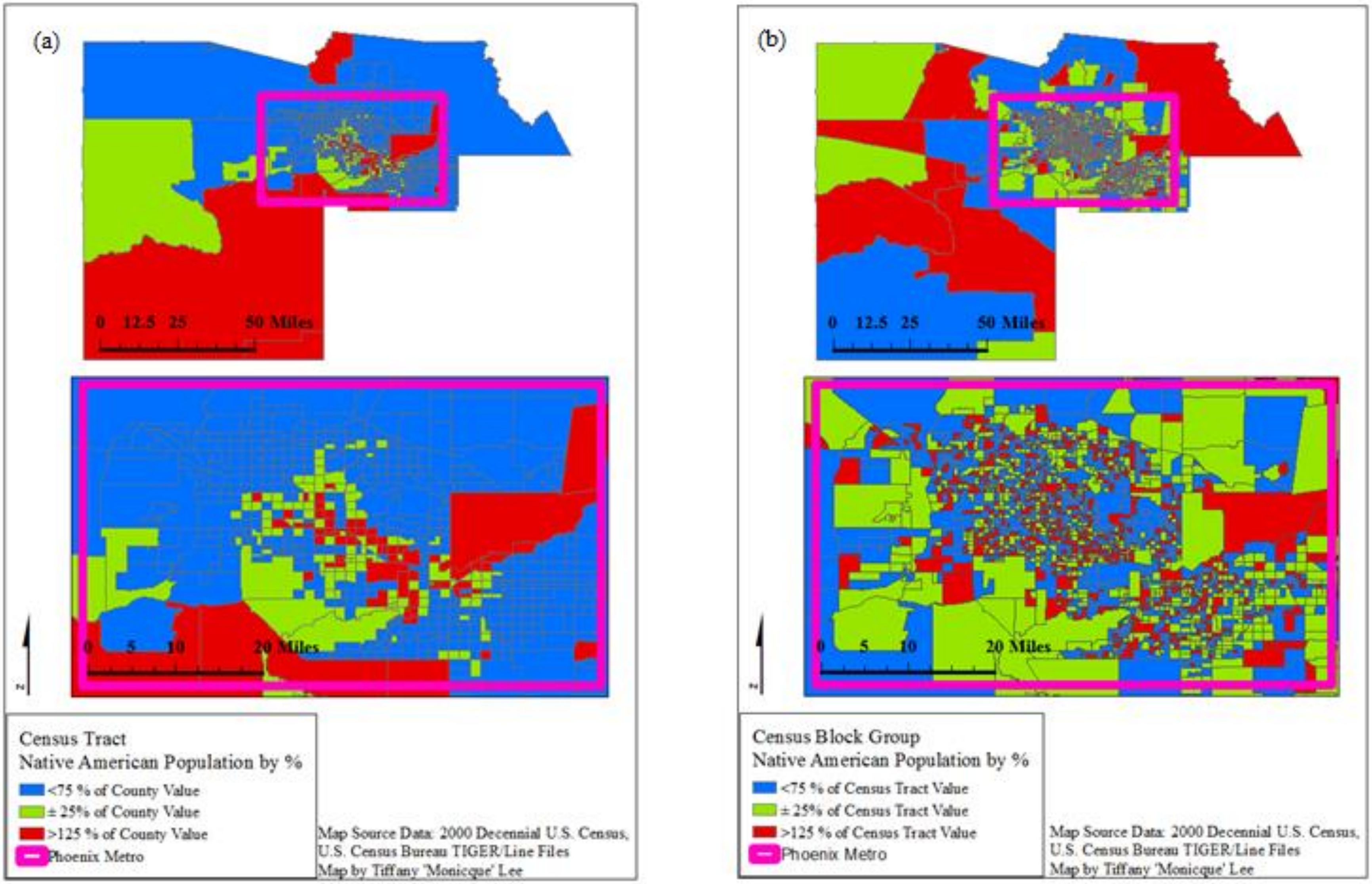

Figure 40. Group I, Percent Native American in Maricopa County by: (a) census tract; and (b) census block group. Both pairs show the entire county on the top with the greater Phoenix Metro highlighted and shown at a larger scale on the bottom.

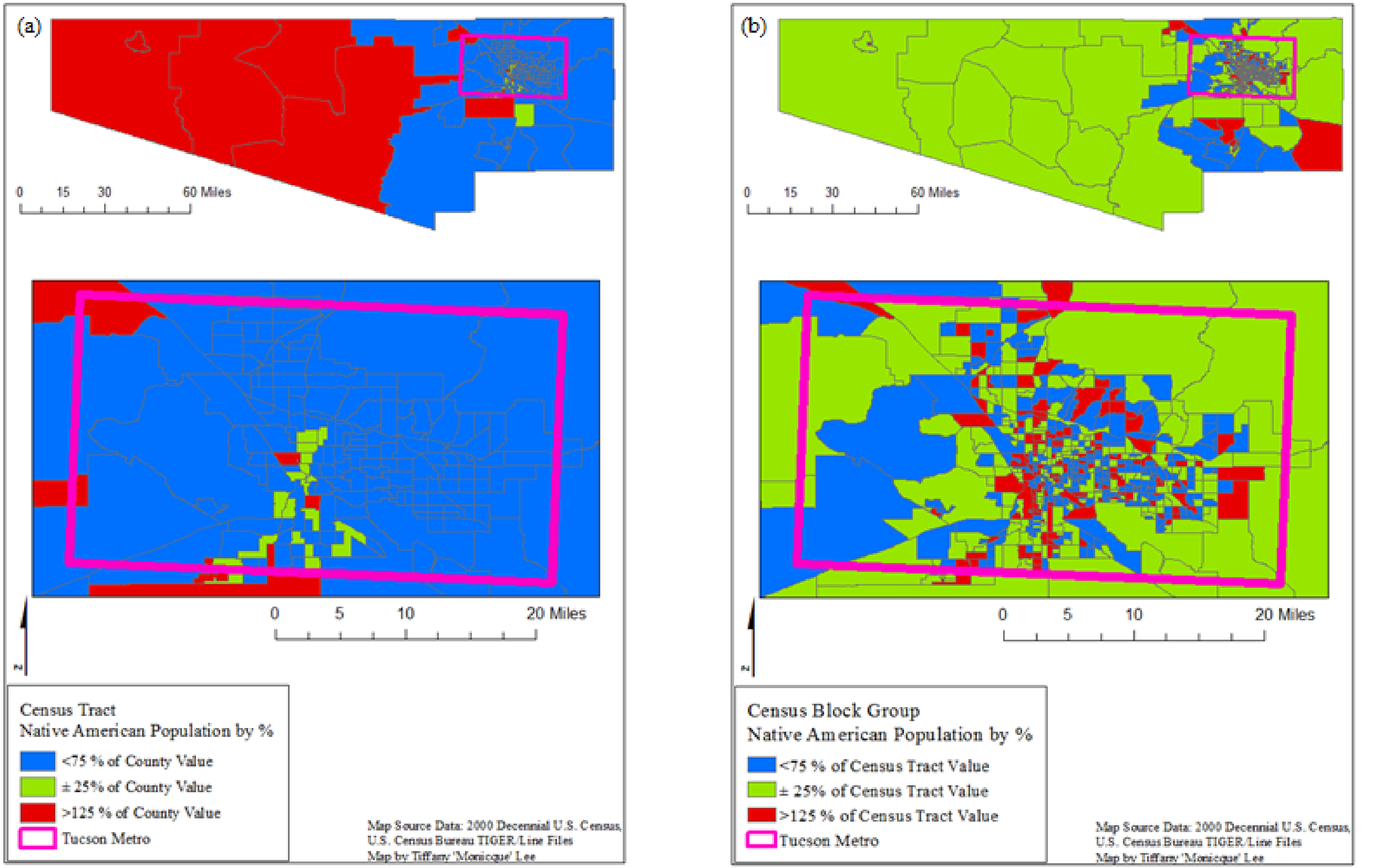

Figure 41. Group I (Continued), percent Native American in Pima County by: (a) census tract; and (b) census block group. Both pairs of maps show the entire county on the top with the greater Tucson Metro highlighted and shown at a larger scale on the bottom.

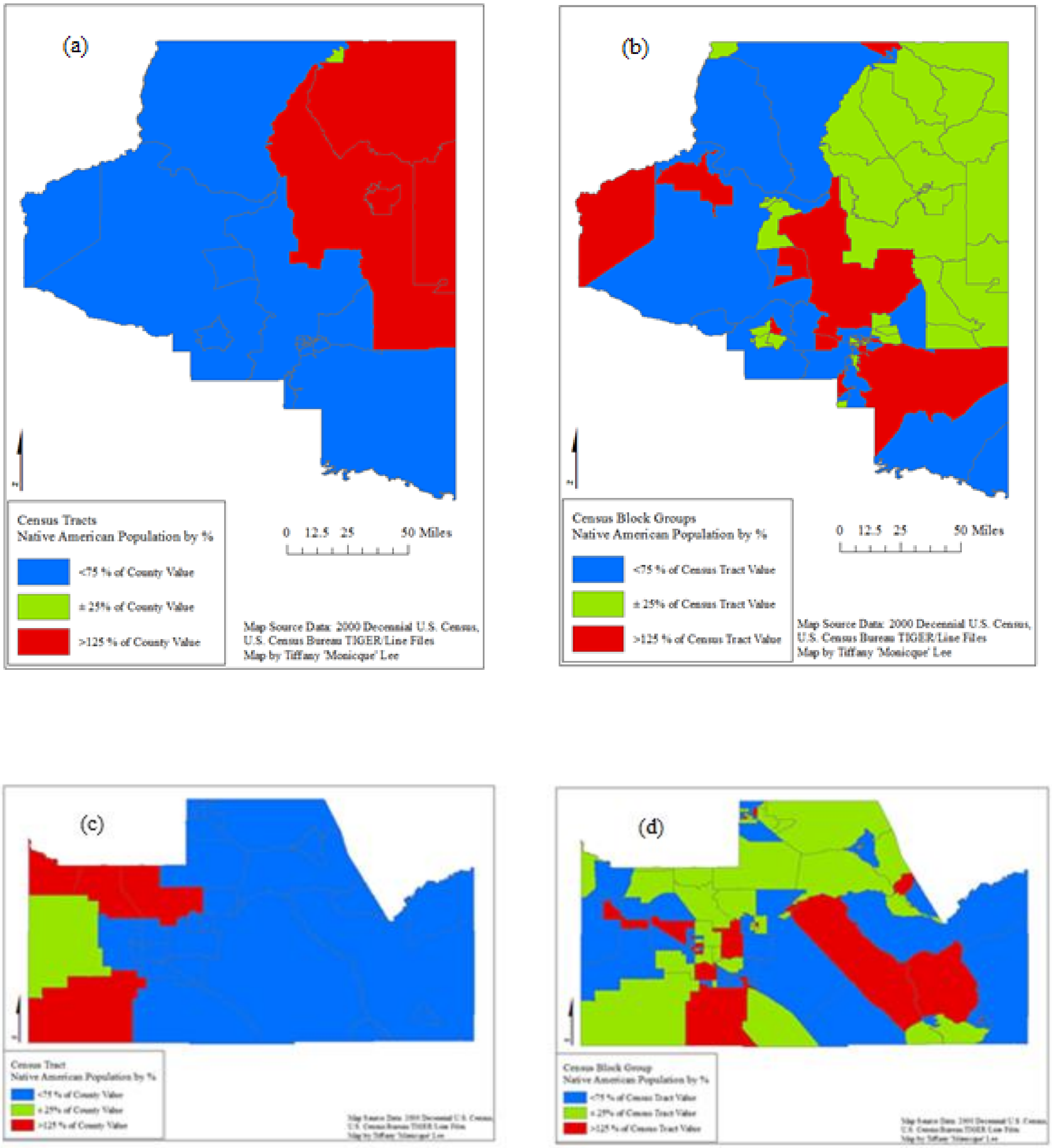

Figure 42. Group II, Percent Native American population in: (a) Coconino County by census tract; (b) Coconino County by census block group; (c) Pinal County by census tract; and (d) Pinal County by census block group.

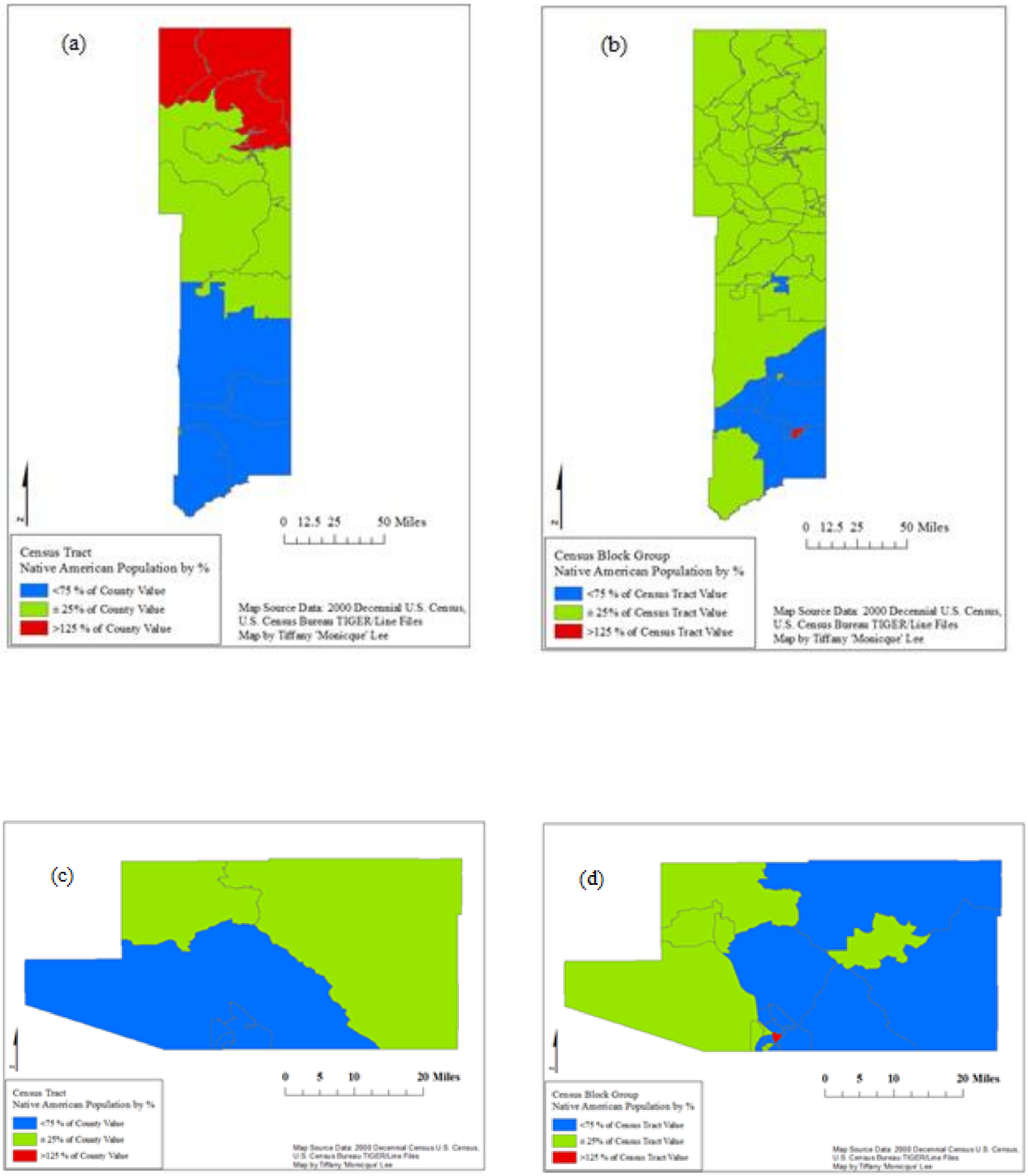

Figure 43. Group III: Percent Native American population in: (a) Apache County by census tract; (b) Apache County by census block group; (c) Santa Cruz County by census tract; and (d) Santa Cruz County by census block group.

Chapter 5 : Discussion and Conclusions

County, Census Tract, and Census Block Group estimates were compared for three commonly used social determinants of health. County estimates are often used by researchers and officials to broadly describe population health. This study tested the efficacy of this supposition.

The findings show that county areas, whether small or large in geographic extent and with rural and/or metropolitan populations, provide very generalized population descriptions and as such subject the resulting assumptions and/or the statistical interpretations to both MAUP affects and ecological fallacy. Rurality, often identified by geographic population density, can also be denoted by perception-based attributes such as regional lifestyle choices, and also cultural factors. Geographic areas considered rural within Maricopa and Pima Counties may not experience the geographic isolation found in some of the low population density counties such as those in Groups II and III; however, the residents may still experience limited infrastructure and are equally subject to ecological bias introduced by aggregation issues. Rurality is therefore found to affect resulting research outcomes in the gamut of rural settings as inconsistencies in neighborhood definition subject data to potential misreporting error.

When social determinants of health are examined at the census tract and census block group levels, nuances within larger geographic areas of the county begin to appear. Depending on the social determinant of health under examination and the geographic reporting unit used, this variability can sometimes be considerable. This variability occurs regardless of what data compilation method is utilized and proved present in the 2000 Decennial Census, as demonstrated in Chapter 4, and the 2010-2015 ACS data, as demonstrated in Table 8 below.

The total number of geographic reporting units changed between the 2000 Decennial Census and the 2010-2015 ACS creating not only potential MAUP effects but also congruency issues. Table 8 compares data between the two reporting's for three counties, one from each of the aforementioned groupings: Maricopa (Group I); Coconino (Group II); and Apache (Group III) Counties. Census block groups classified as outliers increased in two-thirds of the comparisons with the most significant change being in Maricopa County where the Median Household Income estimates increased from 24% in the 2000 Decennial Census to 63.6% in the 2010-2015 ACS.

The 2000 Decennial Census was used for this project; however, temporal issues have consequences as this static portrait of population means that the estimates that were derived are somewhat dated. Nonetheless, the ability to use this Census reporting was extremely beneficial in being able to conduct the spatial unit investigations and comparisons to evaluate if population reporting techniques might affect research outcomes.

Table 8. Specific county variability comparisons at Census Block Group Level between decennial census data and ACS data.

(a) 2000 Decennial Census

County	No. of census block groups	Median Household Income - Estimates <75% or ≥125% of census tract values	Vulnerability comparison - Estimates <75% or ≥125% of census tract values	Native American Population - Estimates <75% or ≥125% of census tract values
Apache	54	46.3	3.7	14.8
Coconino	106	34	24.5	62.3
Maricopa	2,113	24	45.9	69.1

(b) 2010-2015 ACS

County	No. of census block groups	Median Household Income - Estimates <75% or ≥125% of census tract values	Vulnerability comparison - Estimates <75% or ≥125% of census tract values	Native American Population - Estimates <75% or ≥125% of census tract values
Apache	55	27.3	7.3	18.2
Coconino	98	46.9	32.7	62.2
Maricopa	2,505	63.6	19.5	85.3

The thesis project used three separate means of visualizing variability across the spatial units studied. Each method built on the other; however, all contributed to the variability assessment in the following ways.

Threshold values were used to identify the places where the social determinant of health estimates varied by ≥ 25% from the previous estimate at the next highest level of aggregation (i.e. county for census tract and census tract for census block group). This means of visualizing areas with little or no change (green ± 25% of median values) and ≥ 25% change (blue represented as median values classified as outliers < 75% and red represented as median values classified as outliers > 125%) across different geographic reporting units were used on all maps and scatterplots to unlock spatial patterns that might otherwise not be apparent. The choice of this approach was used to highlight the sensitivity of the social determinants of health estimates to the choice of geographic reporting units across the State of Arizona.

The scatterplots graphs were essential for analyzing the actual values behind the map visualizations. The scatterplots provided intricate insights as to how many spatial units fell inside or outside the previously noted thresholds. The color coding of spatial units per threshold bracket helped as a visual aid to assess the range of variability documented within this thesis. The scatterplots along with the maps revealed the true variability present within Arizona and provided story lines that might otherwise not have been evident when looking at a table of values or a map by itself.

The maps, on the other hand, provide intricate insights into the locations across the state where the choice of geographic reporting unit fell inside or outside the previously noted thresholds. This information is helpful in showing these parts of the state where social determinants of health vary over short distances as well as places where similar characteristics cover large areas.

This study has demonstrated that in general, the variability in social determinants of health increased as the number of spatial units increased. As spatial units were examined at differing sizes, and units over a given area were increased by spatial partitioning, the percentages of census units classified as outliers outside of a median value range also changed as indicated in the summary tables (Tables 2 - 8) throughout Chapters 4 and 5. This finding suggests that the null hypothesis should not be accepted since the results showed that social determinants of health estimates changed and affected research results as smaller (census tracts) and smaller (census block groups) geographic reporting units were used in place of counties.

This thesis therefore sets a basis for future work to continue in the exploration of how choice of neighborhood effects research outcomes in health studies. Further investigation into neighborhood effects might include study into what social determinants of health are the most likely to show the greatest variability across geographic reporting units such as those across Native American Indian Areas. Additionally, of interest is if and how the independent social determinant of health variables as selected for this thesis, might be connected and if there is any overlap of spatial patterning within areas classified as outliers.

Future work might also use this project as a basis to explore how proximity buffers are affected by the choice of geographic reporting units for acquisition of census data. Specifically, how those buffers might report what population has potential exposure to a pollution source, e.g. a target population within a certain distance to a major roadway or gas and oil well. Environmental exposure assessments are frequently dependent on proximity factors which are directly influenced by choice of neighborhood making further investigation on how neighborhood is defined for health research of vital importance for health researchers and professionals.

References

Adams, Paul C., Steven Hoelscher, and Karen E. Till. 2001. *Textures of Place : Exploring Humanist Geographies*. NED - New edition ed. Minneapolis: University of Minnesota Press. doi:10.5749/j.cttttg77. http://lib.myilibrary.com?ID=522749.

Apparicio, Philippe, Jeremy Gelb, Anne-Sophie Dube, Simon Kingham, Lise Gauvin, and Eric Robitaille. 2017. "The Approaches to Measuring the Potential Spatial Access to Urban Health Services Revisited: Distance Types and Aggregation-Error Issues." *International Journal of Health Geographics* 16. doi:10.1186/s12942-017-0105-9.

Aylin, Paul, Ravi Maheswaran, Jon Wakefield, Samantha Cockings, Lars Jarup, Richard Arnold, Gideon Wheeler, and Paul Elliott. 1999. "A National Facility for Small Area Disease Mapping and Rapid Initial Assessment of Apparent Disease Clusters Around a Point Source: The UK Small Area Health Statistics Unit." *Journal of Public Health Medicine* 21 (3): 289-298. doi:10.1093/pubmed/21.3.289. http://www.ncbi.nlm.nih.gov/pubmed/10528956.

Barrozo, Ligia Vizeu, Reinaldo Paul Pérez-Machado, Christopher Small, and William Cabral-Miranda. 2016. "Changing Spatial Perception: Dasymetric Mapping to Improve Analysis of Health Outcomes in a Megacity." *Journal of Maps* 12 (5): 1242-1247. doi:10.1080/17445647.2015.1101403.

Bell, B. Sue., Richard Hoskins, Linda Pickle, and Daniel Wartenberg. 2006. "Current Practices in Spatial Analysis of Cancer Data: Mapping Health Statistics to Inform Policymakers and the Public." *International Journal of Health Geographics* 5 (1): 49. doi:10.1186/1476-072X-5-49.

Beyer, Kirsten M. M., Chetan Tiwari, and Gerard Rushton. 2012. "Five Essential Properties of Disease Maps." *Annals of the Association of American Geographers* 102 (5): 1067-1075. doi:10.1080/00045608.2012.659940. http://www.tandfonline.com/doi/abs/10.1080/00045608.2012.659940.

Bithell, J. F. 2000. "A classification of disease mapping methods." *Statistics in Medicine* 19, no. 17-18 (2000): 2203-2215.

Bolstad, Paul. 2012. *GIS Fundamentals: A First Text on Geographic Information Systems*. White Bear Lake, Minn.: White Bear Lake, Minn.: Eider Press.

Boruff, Bryan, Andrea Nathan, and Sandra Nijënstein. 2012. "Using GPS Technology to (Re)-Examine Operational Definitions of 'Neighbourhood' in Place-Based Health Research." *International Journal of Health Geographics* 11: 22. doi:10.1186/1476-072X-11-22.

Boscoe, Francis P. and Linda W. Pickle. 2003. "Choosing Geographic Units for Choropleth Rate Maps, with an Emphasis on Public Health Applications." *Cartography and Geographic Information Science* 30 (3): 237-248. doi:10.1559/152304003100011171.

Bravo, Mercedes A., Rebecca Anthopolos, Michelle L. Bell, and Marie Lynn Miranda. 2016. "Racial Isolation and Exposure to Airborne Particulate Matter and Ozone in Understudied US Populations: Environmental Justice Applications of Downscaled Numerical Model Output." *Environment International* 92-93: 247-255. doi:10.1016/j.envint.2016.04.008.

Chaix, Basile, Juan Merlo, and Pierre Chauvin. 2005. "Comparison of a Spatial Approach with the Multilevel Approach for Investigating Place Effects on Health: The Example of Healthcare Utilisation in France." *Journal of Epidemiology and Community Health* 59 (6): 517. doi:10.1136/jech.2004.025478.

Chaix, Basile, Juan Merlo, David Evans, Cinira Leal, and Sabrina Havard. 2009. "Neighbourhoods in Eco-Epidemiologic Research: Delimiting Personal Exposure Areas. A Response to Riva, Gauvin, Apparicio and Brodeur." *Social Science & Medicine* 69 (9): 1306-1310. doi:10.1016/j.socscimed.2009.07.018.

Chang, Bowon. 2010. *The power of geographical boundaries: Cultural, political, and economic border effects in a unitary nation*. Master's thesis. Iowa State University. http://lib.dr.iastate.edu/etd/11344

Charreire, Hélène, Romain Casey, Paul Salze, Chantal Simon, Basile Chaix, Arnaud Banos, Dominique Badariotti, Christiane Weber, and Jean-Michel Oppert. 2010. "Measuring the Food Environment using Geographical Information Systems: A Methodological Review." *Public Health Nutrition* 13 (11): 1773-1785. doi:10.1017/S1368980010000753. http://journals.cambridge.org/abstract_S1368980010000753.

Chen, Xiang and Mei-Po Kwan. 2015. "Contextual Uncertainties, Human Mobility, and Perceived Food Environment: The Uncertain Geographic Context Problem in Food Access Research." *American Journal of Public Health* 105 (9): 1734. doi:10.2105/AJPH.2015.302792.

Cruz, Hildemar. 2016. *A Geospatial Analysis of Income Level, Food Deserts and Urban Agriculture Hot Spots.* Master's thesis. University of Southern California

Cutchin, Malcolm P., Karl Eschbach, Christine A. Mair, Hyunsu Ju, and James S. Goodwin. 2011. "The Socio-Spatial Neighborhood Estimation Method: An Approach to Operationalizing the Neighborhood Concept." *Health and Place* 17 (5): 1113-1121. doi:10.1016/j.healthplace.2011.05.011.

Delmelle, Elizabeth C. 2016. "Mapping the DNA of Urban Neighborhoods: Clustering Longitudinal Sequences of Neighborhood Socioeconomic Change." *Annals of the American Association of Geographers* 106, no. 1 (n.d.): 36–56.

Diez Roux, Ana V. 2001. "Investigating Neighborhood and Area Effects on Health." *American Journal of Public Health* 91 (11): 1783-1789. doi:10.2105/AJPH.91.11.1783. http://ajph.aphapublications.org/cgi/content/abstract/91/11/1783.

Diez Roux, Ana. V. 2007. "Neighborhoods and health: where are we and were do we go from here?" *Reviews in Epidemiology Sante Publique*, *55*(1), 13–21.

Duncan, Dustin T., Ichiro Kawachi, S. V. Subramanian, Jared Aldstadt, Steven J. Melly, and David R. Williams. 2014. "Examination of how Neighborhood Definition Influences Measurements of Youths' Access to Tobacco Retailers: A Methodological Note on Spatial Misclassification." *American Journal of Epidemiology* 179 (3): 373-381. doi:10.1093/aje/kwt251.

Eagleson, Serryn, 2002. "Hierarchical Spatial Reasoning Theory and GIS Technology Applied to the Automated Delineation of Administrative Boundaries." *Computers, Environment and Urban Systems* 26 (2): 185-200. doi:10.1016/S0198-9715(01)00040-0.

Eicher, Cory L. and Cynthia A. Brewer. 2001. "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation." *Cartography and Geographic Information Science* 28 (2): 125-138. doi:10.1559/152304001782173727.

Elliott, Paul and Daniel Wartenberg. 2004. "Spatial Epidemiology: Current Approaches and Future Challenges.(Information Systems: Mini-Monograph)." *Environmental Health Perspectives* 112 (9): 998. doi:10.1289/ehp.6735.

Faber, Jacob W., and Sharkey, Patrick. 2015. "Neighborhood Effects." *International Encyclopedia of the Social & Behavioral Sciences*, 443–449. https://doi.org/10.1016/B978-0-08-097086-8.32189-4

Flint, Ellen, Steven Cummins, and Stephen Matthews. 2013. "Do Perceptions of the Neighbourhood Food Environment Predict Fruit and Vegetable Intake in Low-Income Neighbourhoods?" *Health & Place* 24: 11-15. doi:10.1016/j.healthplace.2013.07.005. http://www.ncbi.nlm.nih.gov/pubmed/23999578.

Flowerdew, Robin, David J. Manley, and Clive E. Sabel. 2008. "Neighbourhood Effects on Health: Does it Matter Where You Draw the Boundaries?" *Social Science & Medicine* 66 (6): 1241-1255. doi:10.1016/j.socscimed.2007.11.042 [doi].

Foster, Kirk A. and J. Aaron Hipp. 2011. "Defining Neighborhood Boundaries for Social Measurement: Advancing Social Work Research." *Social Work Research* 35 (1): 25-35.

Goodchild, Michael F. 2018. "Reimagining the history of GIS", *Annals of GIS*, 24:1, 1-8, DOI: 10.1080/19475683.2018.1424737

Hall Alexander, W., Schuurman Nadine, and N. Oliver Lisa. 2007. "Comparing Circular and Network Buffers to Examine the Influence of Land use on Walking for Leisure and Errands." *International Journal of Health Geographics* 6 (1): 41. doi:10.1186/1476-072X-6-41.

Hipp, J. Aaron and N. Chalise. 2015. "Spatial Analysis and Correlates of County-Level Diabetes Prevalence, 2009-2010." *Preventing Chronic Disease* 12: E08.

Holt, James B., C. P. Lo, and Thomas W. Hodler. 2004. "Dasymetric Estimation of Population Density and Areal Interpolation of Census Data." *Cartography and Geographic Information Science* 31 (2): 103.

Institute of Education Sciences, National Center for Education Statistics. 2015. “Education Demographic and Geographic Estimates (EDGE) Program; Locale Boundaries” Retrieved from https://nces.ed.gov/programs/edge/docs/EDGE_NCES_LOCALE_2015.pdf

Institute of Education Sciences, National Center for Education Statistics. 2018a. *Geographic Data*. Retrieved from https://nces.ed.gov/programs/edge/geographic.aspx

Institute of Education Sciences, National Center for Education Statistics. 2018b. *School Attendance Boundary Survey; Frequently Asked Questions*. Retrieved from https://nces.ed.gov/programs/sabs/sabsFaq.aspx

Institute of Education Sciences, National Center for Education Statistics. 2018c. *School District Geographic Relationship Files*. Retrieved from https://nces.ed.gov/programs/edge/geographicRelationshipFiles.aspx

Jacquez, Geoffrey. 2014. "Spatial Clustering and Autocorrelation in Health Events." In *Handbook of Regional Science,* 1311-334. Springer Berlin Heidelberg. doi**:** 10.1007/978-3-642-23430-9_80

Jankowska, Marta M., Jasper Schipperijn, and Jacqueline Kerr. 2014. "A Framework for using GPS Data in Physical Activity and Sedentary Behavior Studies." *Exercise and Sport Sciences Reviews* 43 (1): 48.

Jankowska, Marta M, Natarajan, Loki, Godbole, Suneeta, Meseck, Kristin, Sears, Dorothy D, Patterson, Ruth E, and Kerr, Jacqueline. 2017. “Kernel Density Estimation as a Measure of Environmental Exposure Related to Insulin Resistance in Breast Cancer Survivors.” *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 26, no. 7 (2017): 1078–1084.

Kim, Dohyeong, M. A. Galeano, Andrew Hull, and Marie Lynn Miranda. 2008. "A Framework for Widespread Replication of a Highly Spatially Resolved Childhood Lead Exposure Risk Model.(Children's Health)." *Environmental Health Perspectives* 116 (12): 1735. doi:10.1289/ehp.11540.

Kirby, Russell S., Eric Delmelle, and Jan M. Eberth. 2017. "Advances in Spatial Epidemiology and Geographic Information Systems." *Annals of Epidemiology* 27 (1): 1-9. doi:10.1016/j.annepidem.2016.12.001.

Kwan, Mei-Po. 2012a. "Geographies of Health." *Annals of the Association of American Geographers* 102 (5): 891-892. doi:10.1080/00045608.2012.687348.

Kwan, Mei-Po. 2012b. "The Uncertain Geographic Context Problem." *Annals of the Association of American Geographers* 102 (5): 958-968. doi:10.1080/00045608.2012.687349.

Lam, Nina Siu-Ngan. 2012. "Geospatial Methods for Reducing Uncertainties in Environmental Health Risk Assessment: Challenges and Opportunities." *Annals of the Association of American Geographers* 102 (5): 942-950. doi:10.1080/00045608.2012.674900. http://www.tandfonline.com/doi/abs/10.1080/00045608.2012.674900.

Leite, Maria Jalila Vieira de Figueirêdo, Aldo Dantas, and Angelo Giuseppe Roncalli. 2015. "Epidemiology and Place: A Spatial Area for Health Analysis." *Holos* 8: 357-367. doi:10.15628/holos.2015.2854. https://doaj.org/article/576df02d761c4fe390c136ffc31e1b21

Macintyre, Sally, Ellaway, Anne, and Cummins, Steven. 2002. "Place Effects on Health: How Can We Conceptualise, Operationalise and Measure Them?" *Social Science & Medicine* 55, no. 1: 125–139. doi://doi-org.libproxy1.usc.edu/10.1016/S0277-9536(01)00214-3. http://www.sciencedirect.com.libproxy1.usc.edu/science/article/pii/S0277953601002143.

Martin, David, Samantha Cockings, and Samuel Leung. 2015. "Developing a Flexible Framework for Spatiotemporal Population Modeling." *Annals of the Association of American Geographers*, 2015, 1-1.

Matthews Stephen A. 2011. "Spatial Polygamy and the Heterogeneity of Place: Studying People and Place via Egocentric Methods." In Burton, L., Matthews, S., Leung, M., Kemp, S., & Takeuchi, D. *Communities, Neighborhoods, and Health Expanding the Boundaries of Place*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-7482-2

Matthews, Stephen A., 2018. "Place and the Goldilocks Principle: Theory, Measurement and the Relevance of Places in Population and Health." Lecture. Spatial Sciences Institute of University of Southern California, February 22

Matthews, Stephen A. and Daniel M. Parker. 2013. "Progress in Spatial Demography." *Demographic Research* 28: 10.

Matthews, Stephen A, and Tse-Chuan, Yang. 2013. "Spatial Polygamy and Contextual Exposures (SPACEs)." *American Behavioral Scientist* 57, no. 8: 1057-081.

Mennis, Jeremy. 2003. "Generating Surface Models of Population using Dasymetric Mapping." *The Professional Geographer* 55 (1): 31-42. doi:10.1111/0033-0124.10042.

Mennis, Jeremy and Torrin Hultgren. 2006. "Intelligent Dasymetric Mapping and its Application to Areal Interpolation." *Cartography and Geographic Information Science* 33 (3): 179-194. doi:10.1559/152304006779077309.

Miranda, Marie, Rebecca Anthopolos, and Douglas Hastings. 2011. "A Geospatial Analysis of the Effects of Aviation Gasoline on Childhood Blood Lead Levels." *Environmental Health Perspectives* 119 (10): 1513-6. doi:10.1289/ehp.1003231.

Mu, Lan, Fahui Wang, Vivien W. Chen, and Xiao-Cheng Wu. 2015. "A Place-Oriented, Mixed-Level Regionalization Method for Constructing Geographic Areas in Health Data Dissemination and Analysis." *Annals of the Association of American Geographers* 105

(1): 48-66. doi:10.1080/00045608.2014.968910. http://www.tandfonline.com/doi/abs/10.1080/00045608.2014.968910.

Mujahid, Mahasin S., Ana V. Diez Roux, Jeffrey D. Morenoff, and Trivellore Raghunathan. 2007. "Assessing the Measurement Properties of Neighborhood Scales: From Psychometrics to Ecometrics." *American Journal of Epidemiology* 165 (8): 858-867. doi:10.1093/aje/kwm040. http://www.ncbi.nlm.nih.gov/pubmed/17329713.

Nagle, Nicholas N., Barbara P. Buttenfield, Stefan Leyk, and Seth Spielman. 2013. "Dasymetric Modeling and Uncertainty." *Annals of the Association of American Geographers*. doi:10.1080/00045608.2013.843439.

Nelson, Elisabeth S. 2010. "Dasymetric Maps." *Encyclopedia of Geography*, edited by Barney Warf, 670-671. Thousand Oaks, CA: SAGE Publications, Inc., 2010. doi: 10.4135/9781412939591.n252.

Onubogu, Carl. 2013. *Size and Shape Matter: The Modifiable Areal Unit Problem in Neighborhood Effects Research*. ProQuest Dissertations and Theses.

Openshaw, Stan and Seraphim Alvanides. 2001. "Designing Zoning Systems for Representation of Socio-Economic Data." *Time and Motion of Socio-Economic Units.* Taylor and Francis, London.

Perchoux, Camille, Basile Chaix, Ruben Brondeel, and Yan Kestens. 2016. "Residential Buffer, Perceived Neighborhood, and Individual Activity Space: New Refinements in the Definition of Exposure Areas - the RECORD Cohort Study." *Health and Place* 40: 116-122. doi:10.1016/j.healthplace.2016.05.004. https://www.sciencedirect.com/science/article/pii/S1353829216300533.

Perchoux, Camille, Basile Chaix, Steven Cummins, and Yan Kestens. 2013. "Conceptualization and Measurement of Environmental Exposure in Epidemiology: Accounting for Activity Space Related to Daily Mobility." *Health & Place* 21: 86. doi:10.1016/j.healthplace.2013.01.005.

Robertson, Colin, and Feick, Rob. 2018."Inference and Analysis Across Spatial Supports in the Big Data Era: Uncertain Point Observations and Geographic Contexts." *Transactions in GIS* 22, no. 2 (n.d.): 455–476. https://doi.org/10.1111/tgis.12321

Root, Elisabeth Dowling. 2012. "Moving Neighborhoods and Health Research Forward: Using Geographic Methods to Examine the Role of Spatial Scale in Neighborhood Effects on Health." *Annals of the Association of American Geographers* 102 (5): 986-995. doi:10.1080/00045608.2012.659621.

Rousseau, Nicolette. 1995. "What is rurality?" *Occasional Paper (Royal College of General Practitioners)*, 71, 1–4.

Rushton, Gerard. 2003. "Public Health, GIS, and Spatial Analytic Tools." *Annual Review of Public Health* 24: 43.

Sabel, Clive E., W Kihal, D Bard, and Christiane Weber. 2013. "Creation of Synthetic Homogeneous Neighbourhoods using Zone Design Algorithms to Explore Relationships between Asthma and Deprivation in Strasbourg, France." *Social Science & Medicine* 91: 110-121. doi:10.1016/j.socscimed.2012.11.018. http://www.ncbi.nlm.nih.gov/pubmed/23332654.

Santos, Simone M., Dora Chor, and Guilherme Loureiro Werneck. 2010. "Demarcation of Local Neighborhoods to Study Relations between Contextual Factors and Health." *International Journal of Health Geographics* 9 (1): 34. doi:10.1186/1476-072X-9-34.

Schüle, Steffen Andreas and Gabriele Bolte. 2015. "Interactive and Independent Associations between the Socioeconomic and Objective Built Environment on the Neighbourhood Level and Individual Health: A Systematic Review of Multilevel Studies." *PloS One* 10 (4): e0123456. doi:10.1371/journal.pone.0123456.

Sharkey, Patrick and Jacob W. Faber. 2014. "Where, when, Why, and for Whom do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects." *Annual Review of Sociology* 40: 559-579. doi:10.1146/annurev-soc-071913-043350.

Smith, George Davey. 2002. "Commentary: Behind the broad street pump: Aetiology, epidemiology and prevention of cholera in mid-19th century Britain." *International Journal of Epidemiology* 31 (5): 920-32.

Spielman, Seth E., and Logan, John R. 2012. "Using High-Resolution Population Data to Identify Neighborhoods and Establish Their Boundaries." *Annals of the Association of American Geographers* 103, no. 1 (n.d.): 67–84. doi:10.1080/00045608.2012.685049.

Stafford, Mai, Oliver Duke-Williams, and Nicola Shelton. 2008. "Small Area Inequalities in Health: Are we Underestimating them?" *Social Science & Medicine* 67 (6): 891-899. doi:10.1016/j.socscimed.2008.05.028.

Strominger, Julie, Rebecca Anthopolos, and Marie Lynn Miranda. 2016. "Implications of Construction Method and Spatial Scale on Measures of the Built Environment." *International Journal of Health Geographics* 15: 15. doi:10.1186/s12942-016-0044-x.

Swift, Andrew, Lin Liu, and James Uber. 2008. "Reducing MAUP Bias of Correlation Statistics between Water Quality and GI Illness." *Computers, Environment and Urban Systems* 32 (2): 134-148. doi:10.1016/j.compenvurbsys.2008.01.002. https://www.sciencedirect.com/science/article/pii/D0198971508000033.

Swift, Andrew, Lin Liu, and James Uber. 2014. "MAUP Sensitivity Analysis of Ecological Bias in Health Studies." *GeoJournal* 79 (2): 137-153. doi:10.1007/s10708-013-9504-z.

Tatalovich, Zaria, John P. Wilson, Joel E. Milam, Michael L. B. Jerrett, and Rob McConnell. 2006. "Competing Definitions of Contextual Environments." *International Journal of Health Geographics* 5 (1): 55. doi:10.1186/1476-072X-5-55. https://www.ncbi.nlm.nih.gov/pubmed/17156433.

Tobler, Waldo. R. 1970. "A computer movie simulating urban growth in the Detroit region." *Economic Geography* 46 : 234-40.

U. S. Census Bureau. 1990. *Geographic Areas Reference Manual*. Retrieved from https://www.census.gov/geo/reference/garm.html

U.S. Census Bureau. 2010a. *Geographic Terms and Concepts - American Indian, Alaska Native, and Native Hawaiian Areas.* Retrieved from https://www.census.gov/geo/reference/gtc/gtc_aiannha.html

U.S. Census Bureau. 2010b. *Overview of the 2010 Census Tribal Statistical Areas Program (TSAP)*. Retrieved from https://www.census.gov/geo/partnerships/tsap2010_overview.html

U.S. Census Bureau. 2010c. *What is the Census*. Retrieved from https://www.census.gov/2010census/about/

U.S. Census Bureau. 2014. *American Community Survey Design and Methodology (January 2014)*. Retrieved from https://www.census.gov/history/pdf/acsdesign-methodology2014.pdf

U.S. Census Bureau. 2015. *2010 Census Tallies of Census Tracts, Block Groups & Blocks.* Retrieved from https://www.census.gov/geo/maps-data/data/tallies/tractblock.html

U.S. Census Bureau. 2016. *Substantial Changes to Counties and County Equivalent Entities: 1970-Present.* Retrieved at https://www.census.gov/geo/reference/county-changes.html

U.S. Census Bureau. 2017. *Boundary and Annexation Survey (BAS)*. Retrieved at https://www.census.gov/programs -surveys/bas.html

U.S. Census Bureau. 2018a. *American Community Survey*. Retrieved from https://www.census.gov/history/www/programs/demographic/american_community_survey.html

U. S. Census Bureau. 2018b. *School District Boundaries*. Retrieved from https://www.census.gov/did/www/schooldistricts/data/boundaries.html

U. S. Census Bureau. 2018c. *Zip Code Tabulation Areas*. Retrieved from https://www.census.gov/geo/reference/zctas.html

U.S. Census Bureau. 2018d. *Population Estimate Table.* Retrieved from https://www.census.gov/quickfacts/AZ

U.S. Department of Homeland Security. 2017. *Indian Reservations*. Retrieved at https://catalog.data.gov/dataset/indian-reservations

U.S. Environmental Protection Agency. 2016. *2016 American Indian/Alaska Native/Native Hawaiian Areas (AIANNH) Michigan, Minnesota, and Wisconsin.* Retrieved from https://catalog.data.gov/dataset/2016-american-indian-alaska-native-native-hawaiian-areas-aiannh-michigan-minnesota-and-wisconsin

Waller, Lance A., and Gotway, Carol A. 2004. Applied Spatial Statistics for Public Health Data. Hoboken: John Wiley & Sons, Incorporated. Accessed January 21, 2018. ProQuest Ebook Central.

Xia, Gangqiang, Marie Lynn Miranda, and Alan E. Gelfand. 2006. "Approximately Optimal Spatial Design Approaches for Environmental Health Data." *Environmetrics* 17 (4): 363-385. doi:10.1002/env.775.

Yiannakoulias, Nikolaos. 2011. "Spatial Aberration vs. Geographical Substance: Representing Place in Public Health Surveillance." *Health and Place* 17 (6): 1242-1248. doi:10.1016/j.healthplace.2011.07.009.

Zandbergen, Paul A. and Drew A. Ignizio. 2010. "Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates." *Cartography and Geographic Information Science* 37 (3): 199-214. doi:10.1559/152304010792194985.