INTEGRATING SPATIAL VISUALIZATION TO IMPROVE PUBLIC HEALTH UNDERSTANDING AND COMMUNICATION

by

Corina Chung

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)

December 2014

DEDICATION

For Baba, Mama, Jiejie, and Victor – thank you for all that you do, I am forever grateful.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my thesis committee chair, Dr. Katsuhiko Oda of the Spatial Sciences Institute at the University of Southern California. Without your continuous guidance, encouragement, and kindness, I could not have done this; it has been such an honor and a privilege to work with you. Thank you also to the other members of my thesis committee, Dr. Karen Kemp and Dr. Yao-Yi Chiang, for your expertise and support.

Last, and certainly not least, I would like to thank the people I love the most and owe so much to. Thank you to Baba for the loud, cheerful telephone greetings and car washings; to Mama for the most delicious food and amusing texts; to Jiejie for the ceaseless phone calls and snack deliveries; and to Victor for doing 90 percent of the chores and keeping me on task. Most of all, thank you for the never-ending encouragement, support, and love. What I have done to deserve such a wonderful family, I do not know, but I am grateful to be so blessed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACA	Patient Protection and Affordable Care Act
ACS	American Community Survey
ATSDR	Agency for Toxic Substances and Disease Registry
BRFSS	Behavioral Risk Factor Surveillance System
CDC	Centers for Disease Control and Prevention
CHIP	Children’s Health Insurance Program
CPS	Current Population Survey
CMS	Centers for Medicare & Medicaid Services
DDT	Division of Diabetes Translation
FPL	Federal poverty level
GIS	Geographic information systems
HIV	Human immunodeficiency virus
IOM	Institute of Medicine
IT	Information technology
MAUP	Modifiable areal unit problem
OSHPD	Office of Statewide Health Planning & Development
PCA	Principal component analysis
TIGER	Topologically integrated geographic encoding and referencing
TB	Tuberculosis
WHO	World Health Organization

ABSTRACT

Communication and information technology are critical in facilitating the processes in which public health stakeholders understand and utilize health information. Spatial visualization enables public health practitioners to effectively present geographic phenomena and detect spatial patterns in maps that may remain otherwise undiscovered in tabular form. Although there are many public health practitioners integrating spatial visualization into their work, there are few resources dedicated to instructing how to best visualize health data. Mapmakers will find that, among the wealth of resources on cartography and visualization best practices, few are specific to how health data can be best spatially visualized. Communication of such data is critical in understanding public health issues and developing prevention and intervention programs. This study aimed to 1) document best practices for visualizing public health data using thematic mapping techniques and 2) demonstrate how spatial visualization can be integrated into public health studies to facilitate understanding and communication of findings. A process for identifying suitable thematic mapping techniques for public health studies is discussed, in addition to best practices for employing such techniques, which includes choropleth, proportional symbol, dot density, and nominal point mapping techniques. A case study is presented to demonstrate how spatial visualization can be successfully integrated into public health studies; sociodemographic risk factors of uninsurance were identified using principal component analysis and later mapped using choropleth mapping best practices. Best practices for visualizing health outcomes, social determinants of health, and health care access, key areas of concern in improving public health, are also provided. This study addresses the gap in cartographic resources for the public health industry and aids public health practitioners in their ability to spatially visualize their data and improve communication of their findings.

CHAPTER 1: INTRODUCTION

In 1979 Surgeon General Julius B. Richmond released *Healthy People: Surgeon General's Report on Health Promotion and Disease Prevention*, beginning a decades-long national initiative to establish science-based objectives to improve health and eliminate health disparities (U.S. Department of Health, Education and Welfare 1979). Since 1990 and every ten years following, benchmarks have been established to evaluate progress in promoting community collaborations, improving health literacy, and measuring impact of prevention. Healthy People 2020, the latest set of benchmarks established, includes the goal of "us[ing] communication strategies and health information technology (IT) to improve population health outcomes and health care quality, and to achieve health equity" (Healthy People 2013, 1). Communication, information, and technology are critical in facilitating the processes in which public health stakeholders understand and utilize health information, which directly affects health decisions and outcomes (Healthy People 2013).

Public health practitioners can use geospatial information technology to respond to Healthy People 2020's aspiration to harness the potential of information technology for positively impacting health. Spatial representation of information has major benefits; in particular, the use of geographic information systems (GIS) yields three distinct advantages: 1) they can record and store data, 2) they can be used to identify and explore spatial patterns, and 3) they can effectively present and communicate information (World Health Organization 2014).

These benefits can be applied to understanding health care barriers that prevent individuals from accessing quality health care. A 2003 Institute of Medicine (IOM) report documented that one such barrier includes attaining health insurance coverage, which impacts health care access and quality. Several initiatives recognize that reducing health disparities requires addressing the unequal access to health care (Healthy People 2013; Kibon and Ahmed 2013; Paek and Lim 2013). On March 23, 2010 President Barack Obama signed into law the Patient Protection and Affordable Care Act (ACA), intending to increase the number of insured individuals by providing more affordable insurance options and expanding eligibility for public health insurance. Although the Congressional Budget Office estimated that ACA will help thirty-two million uninsured people gain health insurance coverage by 2019, many millions will still remain uninsured (Clemans-Cope et al. 2012).

There is widely documented research asserting that the likelihood of having health insurance coverage is related to one's sociodemographic characteristics (Dubay and Kenney 2004; Nelson et al. 2004; Angel, Frias, and Hill 2005; Pylypchuk 2009; Clemans-Cope et al. 2012; John et al. 2013). Deeper understanding of these sociodemographics, particularly which characteristics are more influential and where are areas with high prevalence of these characteristics, would be helpful in designing and implementing public health programs aimed at increasing health care access.

1.1 Using Geographic Information Systems to Spatially Visualize Data

The most useful functions of GIS in public health settings are spatial database management, analysis, and visualization. Spatial database management builds its foundation on a data model: the structure of data and the ability to store, retrieve, and manipulate data are at the heart of the power of GIS databases (Tomlinson and Boyle 1987; Bolstad 2012; Cromley and McLafferty 2012). Spatial analysis involves varying levels of analysis of geographic data, ranging from exploratory and descriptive analysis to statistical analysis and modeling (O'Sullivan and Unwin 2010). Spatial visualization of data is potentially the most valuable function of GIS in the eyes of the public health industry because it enables public health practitioners to effectively present geographic phenomenon and detect patterns in maps that may remain otherwise undiscovered in tabular form (Cromley and McLafferty 2012).

Public health studies have used GIS to study where people live and how this affects their health (McLafferty 2003; Cromley and McLafferty 2012). Studies have covered a wide range of applications such as health outcomes, social determinants of health, access and utilization of health care services, health behaviors, and environmental hazards (McLafferty 2003; Higgs 2005; Higgs 2009; Phillips et al. 2009; Cromley and McLafferty 2012).

In studies of health outcomes, GIS has played a vital role in analyzing clusters of disease incidence and their relationship to the geographic environment (Cromley 2003; Cromley and McLafferty 2012). One study mapped nativity of individuals with dengue in Germany, revealing that most infections were acquired in South and Southeast Asia (Jansen et al. 2008). Another study that used GIS found that children with asthma tended to live near high-traffic roads (English et al. 1999). The findings of some studies using GIS led to direct interventions to address health issues; one study mapped blood lead concentrations of children and revealed that more than one-third of the children who tested positive for lead poisoning lived in homes with high concentrations of lead, which were then prioritized for immediate lead hazard remediation (Reissman et al. 2001).

Population distribution and characteristics are also frequent areas of interest in public health studies integrating geospatial information because of well-documented relationships between social characteristics and health (Cromley and McLafferty 2012). More specifically, social determinants of health, which are the conditions in which individuals are born, live, grow, study, and work, impact health outcomes; the lower the socioeconomic status, the worse off one's health (Marmot 2005; Marmot et al. 2008; Noone 2009). Maps displaying distribution of different socioeconomic groups can indicate the likelihood of positive or negative health outcomes. For example, a study of race and ethnicity revealed that areas with a higher percentage of blacks reported poorer health outcomes – much more than areas with a higher percentage of whites (Do et al. 2008). Another study demonstrated that areas with lower median income reported higher incidence of diabetes (Cox et al. 2007).

Other public health studies applying GIS methods have focused on health care access and utilization. Many studies have mapped health services to determine spatial distribution of community resources (Pearce et al. 2007; Macintyre, Macdonald, and Ellaway 2008), and some have used distance and travel time measures to determine level of accessibility to health services (Cromley and McLafferty 2012). One study used GIS to measure distance between the centroid of ZIP codes to the nearest physician and compared variations in health care access across the study area (Mayer 2006). A health care utilization study used GIS to map and analyze hospital choices of patients by socioeconomic characteristic. The results suggested that patients relying on public health insurance coverage more frequently chose the nearest hospital to their home, while patients with higher income chose hospitals that were farther away and served more patients with higher socioeconomic status (Tai, Porell, and Adams 2004).

Although there are many public health practitioners using GIS in their work, and the number is still growing, there are few resources dedicated to instructing how to best visualize health data. There is a wealth of resources on cartography and visualization best practices, but mapmakers will find few that are specific to how public health data can be spatially visualized. Furthermore, public health data, including health insurance coverage variables, are often available only at national and state levels. Geographic trends seen at national or state geographic levels may be very different on a county level as data are disaggregated into smaller units.

1.2 Study Overview

This study sought to demonstrate how GIS can be used to better understand and communicate health data through spatial visualization. This study aimed to 1) document best practices for visualizing public health data using thematic mapping techniques and 2) demonstrate how spatial visualization can be integrated into public health studies to facilitate understanding and communication of findings.

To achieve the first aim, this study provided a process for identifying suitable thematic mapping techniques for public health studies and thoroughly discussed best practices for employing such techniques. This study's methodologies involved documenting best practices for visualizing health outcomes, social determinants of health, and health care access. Communication of such data is critical in understanding public health issues and developing prevention and intervention programs; therefore, it is important for public health practitioners to visualize the data in ways that facilitate understanding and use. In cartography, best practices for visualizing spatial phenomena have been extensively examined; however, there a few studies that explicitly examined cartographic best practices in public health despite the growing field of geographic visualization and use of GIS in public health.

For the second aim, this study reviewed existing mapping practices by public health practitioners and documented a case study to demonstrate how health care studies can expand understanding and communication of health issues through spatial visualization. The review of existing mapping practices describes how public health practitioners used various mapping methods and focuses on the applications of health outcomes, social determinants of health, and health care access. The case study identified sociodemographic risk factors for uninsurance in Alameda County using principal component analysis (PCA). The PCA revealed that adults without a high school degree, individuals identifying as Hispanic or Latino and Other Race (per the U.S. Census Bureau's race categories), immigrants without citizenship, individuals living under 200 percent of the federal poverty level (FPL), and unemployed adults were more likely to be uninsured. These risk factors were mapped using cartographic best practices; the spatial visualization of these socioeconomic characteristics clearly revealed that western Alameda County, especially Oakland, is likely to have prominent numbers of individuals at risk for being uninsured. These methodologies can be applied to a dataset for any particular geographic area to identify a customized set of risk factors and the prevalent areas with such risk factors. Public health practitioners can model their own studies after this particular case study to integrate spatial visualization to further explore and communicate their findings.

CHAPTER 2: THE HEALTH INSURANCE LANDSCAPE

Efforts to address health and well-being, such as the implementation of the Patient Protection and Affordable Care Act (ACA), will benefit from improved understanding through spatial visualization. ACA and other initiatives seeking to increase health insurance coverage will improve health care access and eventually reduce health disparities, but an important step in reaching these milestones to improve population health is understanding the characteristics of the most vulnerable communities, where they live, and how they are impacted (Healthy People 2013; Kibon and Ahmed 2013; Paek and Lim 2013).

Health insurance coverage has long been a major concern of the public health field because of its widely documented effect on health (Ayanian et al. 1993; Ayanian et al. 2000; Institute of Medicine 2002; Nelson et al. 2004). As a key component of health care access, health insurance coverage is important for individuals to maintain overall good health. Without financial means, many individuals lose the ability to use health care services, and uninsured individuals are more likely to have poor health outcomes (Ayanian et al. 2000; Baker et al. 2001; Dor, Sudano, and Baker 2006; Chatterjee and Nielson 2011; Collins et al. 2011). In 2012, 48 million Americans in the U.S. had no health insurance coverage; more than one in five non-elderly adult Americans were uninsured (The Henry J. Kaiser Family Foundation 2013; U.S Census Bureau 2013). All people should be able to have access to quality care without having to face major financial burdens (Chatterjee and Nielson 2011).

Although there is abundant literature on health insurance coverage, much of the data focus on national-level analysis. This may mask key differences in insurance status rates seen on lower geographic levels such as at the state and county levels which may have different trends (Nelson et al. 2004). Each state may have unique populations of varied sociodemographic characteristics and different health care systems, all of which impact the number of uninsured individuals (Mills 2002). While there is ample evidence of certain sociodemographic characteristics having strong relationships to insurance status on the national level, there is much less known on the state and smaller geographic levels (Berki et al. 1985; CDC 1995; Nelson et al. 2004).

2.1 Characteristics of the Uninsured

Most of the uninsured population in the U.S. comes from working families, but a higher percentage of unemployed individuals are uninsured compared to employed individuals. 63 percent of uninsured Americans have at least one full-time worker in the family and 16 percent have a part-time worker (The Henry J. Kaiser Family Foundation 2013). Using logistic regression analyses, including multinomial logistic regression, Legerski (2012), Ahluwalia and Bloen (2008), and Nelson et al. (2004) documented the increasing risk of being uninsured for employed individuals. Most of these individuals came from blue collar jobs in service-oriented industries (John et al. 2013). Although employment may meet the needs to secure health insurance for some individuals, many found that gaining full-time employment resulted in loss of public insurance benefits, which was not necessarily replaced by employer-sponsored coverage (Legerski 2012). While most of the uninsured are employed, unemployment is still a risk factor for lacking health insurance coverage. There is a higher proportion of uninsured among the unemployed compared to the employed (Collins et al. 2011), and according the U.S. Census Bureau (2012a) unemployed individuals (38.9 percent) are three times more likely to be uninsured compared to their employed counterparts (13.6 percent).

Educational attainment is also related to health insurance status. Having limited educational attainment was a risk factor for being uninsured, according to numerous studies

(Schwartz, Marcotte, and McBride 1993; Ahluwahlia and Bloen 2011; Chatterjee and Nielson 2011; John et al. 2013; Angel, Frias, and Hill 2005). In 2012, 27.2 percent of individuals with less than a high school diploma and 20.3 percent of high school graduates were uninsured, much higher than 6.8 percent of individuals with bachelor's degree or higher (U.S. Census Bureau 2012).

Most of the uninsured population comes from low-income families. Individuals living below the federal poverty level (FPL) make up 38 percent of all uninsured, and individuals living below 400% FPL (i.e., low- and middle-income families) make up 90 percent (The Henry J. Kaiser Family Foundation 2013). A family of four with a family income no higher than $23,850 is considered living below FPL (U.S. Department of Health & Human Services 2014). Both Callahan and Cooper (2004) and Nelson et al. (2004) reported an increasing likelihood of being uninsured for both low- and middle-income families in studies using logistic regression.

Adulthood is another risk factor for being uninsured (Nelson et al. 2004; The Henry J. Kaiser Family Foundation 2013). Prior to the implementation of ACA, low-income adults between eighteen and sixty-four years had relatively few public health insurance options. Many did not meet eligibility requirements for Medicaid, a federal health care program that provides free or lost-cost health care options. Medicare, another federal health program, covers only adults sixty-five years or older, adults under sixty-five years that have certain disabilities, and individuals with End-Stage Renal Disease (CMS 2013). These limited programs left many adults without health insurance coverage. Among adults, young adults ages eighteen to thirty-four were most likely to be uninsured: more than one in four young adults were uninsured in 2012 (U.S. Census Bureau 2013). Ahluwalia and Bloen (2008) and Callahan and Cooper (2004) used logistic regression to demonstrate that being a young adult, especially under twenty-five years, increases the likelihood of being uninsured. Many young adults were unable to afford health insurance premiums or assumed that their youthfulness translates into good health and therefore did not need health insurance coverage (Young Invincibles 2012; The Henry J. Kaiser Family Foundation 2013).

Although children have more health insurance coverage options than adults, there are still many children at risk for being uninsured. Children from low-income families qualify for more options than adults through Medicaid and the Children's Health Insurance Program (CHIP), which provides health coverage to children from families whose incomes are too high to qualify for Medicaid (Centers for Medicare & Medicaid Services, n.d.). Nonetheless, 53 percent of uninsured children eligible for Medicaid or CHIP were not enrolled (The Henry J. Kaiser Family Foundation 2013). Children with Hispanic or Latino immigrant parents who have low educational attainment are especially likely to be uninsured (Angel, Frias, and Hill 2005). Enrollment barriers for children may include families being unaware of the program or their eligibility, as well as complex and burdensome enrollment and renewal processes (The Henry J. Kaiser Family Foundation 2013).

Nativity is also related to insurance status: while 13.0 percent of native-born individuals in the U.S. were uninsured in 2012, 32.0 percent of foreign-born were uninsured. Among the foreign-born, 43.4 percent of non-citizens were uninsured compared to 18.3 percent of naturalized citizens (U.S. Census Bureau 2013). The John et al. (2013) and Angel, Frias, and Hill (2005) logistic regression studies revealed being an immigrant was a risk factor for being uninsured; immigrants that have not been naturalized are even stronger risk factors (Pylpchuk 2009). DeNavas-Walt et al. (2008) documented that immigrants were more than twice as likely to lack health insurance than native-born Americans, and Chatterjee and Nielson's (2011) crosssectional study indicated that even when controlling for sociodemographic and occupational characteristics immigrants were 13.6 percent less likely to be covered under employer-based health insurance plans than native-born Americans.

Stark differences in insurance coverage across different race and ethnic groups have also observed. People of color were more likely to be uninsured than non-Hispanic Whites (Table 1). According to the U.S. Census Bureau, in 2012 33.5 percent of individuals who identified as Other Race (per the U.S Census Bureau's race categories) were uninsured, more than any other group. Hispanic or Latinos were the next highest group to report being uninsured at 29.9 percent. Hispanic or Latino identity was the strongest race- or ethnicity-based risk factor for being uninsured, according to Ahluwalia and Bloen's (2008) and Angel, Frias, and Hill's (2005) logistic regression studies. In addition to being Hispanic or Latino, being Black has also been documented as contributing to the likelihood of being uninsured (Ahluwalia and Bloen 2008; DeNavas-Walt, Proctor, and Smith 2011; Clemans-Cope et al. 2012).

Table 1 Percentage of Uninsured by Race in U.S., 2012

Race/ethnicity	Estimated	Percent
Other Race	4,871,562	33.5%
Hispanic or Latino	15,304,828	29.9%
American Indian and Alaska Native	700,936	28.1%
Native Hawaiian and Other Pacific Islander	92,968	18.1%
Black or African American	6,721,402	17.7%
Asian	2,315,360	15.4%
Two or More Races	1,240,776	14.5%
White	20,651,147	10.6 %

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

2.2 Contributing Factors to Uninsurance

The largest barrier to obtaining insurance for uninsured individuals was often affordability (Baicker et al. 2012; John et al. 2013; The Henry J. Kaiser Family Foundation 2013). The Kaiser

Commission on Medicaid and the Uninsured reported that nearly two-thirds of uninsured adults age eighteen to sixty-four years were uninsured because of high costs or loss of employment; only 1.5 percent of uninsured said they do not need coverage (Figure 1).

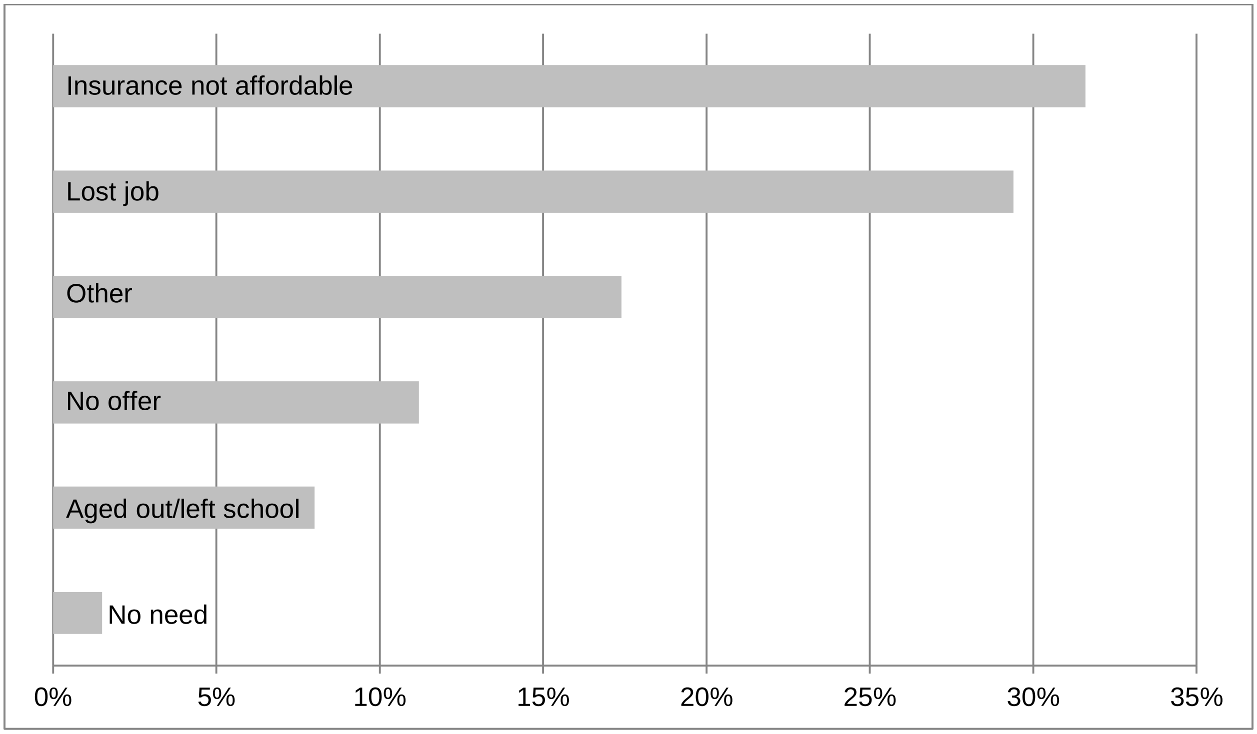


Figure 1 Reasons for Being Uninsured among Uninsured Nonelderly in U.S., 2013
Source: Kaiser Commission on Medicaid and the Uninsured analysis of 2013 National Health Interview Survey data

Uninsured workers more often are self-employed or work at small businesses where health insurance is less likely to be provided. Low-income workers that are offered insurance and want to enroll may not be able to afford the high premiums for themselves and their families (Cunningham et al. 2008; State Health Access Data Assistance Center 2013). The Great Recession from 2007 to 2009 contributed to a surge in uninsured rates when employers changed the cost-sharing of health insurance with employees, making premium costs unaffordable. Millions of Americans lost the option to enroll in employer-sponsored coverage when they lost their jobs; the unemployment rate peaked at 10.0 percent (Robert Wood Johnson Foundation 2009; Collins et al. 2011; Bureau of Labor Statistics 2012; Bureau of Labor Statistics 2014).

2.3 The Impact of Uninsurance on Health Care Access and Health Outcomes

Research consistently documented that uninsured individuals have less health care access and poorer health outcomes than their insured counterparts (Ayanian et al. 2000; Baker et al. 2001; Dor, Sudano, and Baker 2006; Chatterjee and Nielson 2011; Collins et al. 2011). Without health insurance, they are less likely to overcome financial barriers to utilizing health care services that could help maintain their overall health.

Uninsurance has negative consequences related to diminished health care access. Uninsured individuals were less likely to use both preventive and treatment care, including sufficient care for chronic health conditions (Hadley 2007; Chatterjee and Nielsen 2011; CDC 2012a; John et al. 2013). They were also more likely to delay visiting a provider and seeking treatment for chronic illnesses. Regular primary care utilization increases the number opportunities to screen for chronic diseases and increases health education in patients to follow healthier lifestyles (John et al. 2013).

Obtaining health care is especially critical for the uninsured population because they were less likely to be healthy than their insured counterparts (IOM 2002; John et al. 2013). O’Hara (2004) noted that individuals with better health status were more likely to be insured; uninsured individuals were more likely to suffer from acute illnesses and premature diseases and be diagnosed with later stage cancer.

Lack of insurance coverage does not only affect individuals without insurance. A 2003 Institute of Medicine report suggested that communities with high rates of uninsurance were more likely to face health care quality and access barriers. Even individuals with health insurance coverage may experience a spillover effect from their fellow community members being uninsured (Pauly and Pagan 2007; Pauly and Pagan 2009; Sabik 2012). The financial burden of providing care to uninsured individuals affects health care services to the whole community because of the lower use of services by the uninsured and the uncompensated costs of providing care to the uninsured who do use services. The cost and quality of health care services were more likely to be unfavorable as a result (Pauly and Pagan 2009).

CHAPTER 3: THEMATIC MAPPING METHODS

Spatial visualization of data enables public health practitioners to effectively present geographic phenomenon and detect patterns in maps (Cromley and McLafferty 2012). However, cartographers must thoroughly consider the nature of the data or risk miscommunicating information to their audience (Dent, Torguson, and Hodler 2008). This chapter discusses how to visualize public health data through various thematic mapping methods and includes a process for selecting an appropriate method.

3.1 Data Sources

To create maps about health outcomes, social determinants of health, and access to health care services, this study used datasets delivered by the Centers for Disease Control and Prevention (CDC), U.S. Census Bureau, and the California Office of Statewide Health Planning & Development (OSHPD), respectively.

The CDC is the nation's health protection agency and one of the major offices under the Department of Health and Human Services; the CDC also provides the Behavioral Risk Factor Surveillance System (BRFSS), which is the nation's largest system of health telephone survey to collect state-level data on health risk behaviors, chronic health, and use of preventive health care services (BRFSS 2014). As the premier survey system in behavioral and chronic disease surveillance, BRFSS receives sponsorship from many agencies including most divisions under the CDC National Center for Chronic Disease Prevention and Health Promotion, many other CDC centers, Health Resources and Services Administration, Administration of Aging, Department of Veterans Affairs, Substance Abuse and Mental Health Services Administration, and many other federal agencies. State health departments collect data using a standardized core questionnaire with technical and methodological assistance from CDC (BRFSS 2013a). This study used variables on obesity prevalence and self-reported health status from the 2010 BRFSS geographic information systems (GIS) dataset. Other health outcomes data used in this study were prepared by the CDC's Division of Diabetes Translation (DDT), which is also part of the National Center for Chronic Disease Prevention and Health Promotion (DDT 2014). This study used the Diagnosed Diabetes Prevalence Datasets, whose data also comes from the BRFSS (DDT 2013).

Social determinants of health were mapped using the 2010-2012 American Community Survey (ACS) 3-year Estimates data, 2000 Census Summary File 1, and 2010 Census Summary File 1. The ACS dataset pooled data from surveys conducted by the U.S. Census Bureau in 2010 to 2012. Approximately, 2.1 million households responded at a 97.5 percent response rate each year and answered the same demographic questions seen on decennial census, in addition to a multitude of the ACS's socioeconomic and housing questions. These questions included the following topics: age, sex, education, health insurance, income, employment, industry and occupation, language, citizenship, rent, housing value, and home tenure (i.e., owners and renters). The data were aggregated to single- and multi-year levels, which included three or five years of data to achieve reliable estimates. The estimates were provided on a variety of geographic levels: nation, states, congressional districts, counties and county equivalents, metropolitan and micropolitan statistical areas, urban areas, school districts, indigenous American areas, ZIP code tabulation areas, census tracts, census block groups, and other places such as cities, towns, and census-designated places (U.S. Census Bureau 2008). This study used the Selected Population Profile dataset for U.S. states, California counties, and Alameda County census tracts. The 2010 and 2000 Census Summary File 1 comes from the decennial census which collected data from households using mailed surveys or in-person surveying. Unlike ACS, the survey asks only ten questions on age, sex, race/ethnicity, and household information, although the data is available at many different geographic levels (U.S. Census Bureau 2014a). This study used state-level data on race categories.

Maps of both health outcomes and social determinants of health also included Topologically integrated geographic encoding and referencing (TIGER)/Line Shapefiles from the U.S. Census Bureau. These datasets contain geographic boundaries, roads, and water features, as well as geographic entity codes (or GEOIDs) which can be used to join to other datasets without geographic attributes, such as the ACS data (U.S. Census Bureau 2014b). This study used 2012 TIGER/Line Shapefiles for the following geographies: U.S. states, California counties, and Alameda County census tracts and places (i.e., cities).

OSHPD is a state agency that provides other agencies with information on California health care delivery systems. OSHPD collects data on the state and local health care systems, monitors hospitals and skilled nursing facilities, and provides loan insurance to nonprofit health care facilities (OSHPD 2013). Data obtained from the data portal were used for creating maps of health care access. The data included point features representing the locations of licensed health care facilities in California. Even though the spatial accuracy of the recorded locations was not sufficient for emergency services and driving navigation, they are adequate for administrative and reporting purposes. In addition, the attribute table of this data included facility names, addresses, license numbers and types, and the number of licensed beds.

3.2 Thematic Mapping Methods

In order to select the appropriate mapping method to visualize data, the characteristics of the data should be carefully considered. The subject, purpose, and audience of the map should dictate the creation of the map (Dent, Torguson, and Hodler 2008). This study provides information about how to select a suitable thematic mapping method for different studies and data considerations for these methods. The focus of this section includes choropleth, proportional symbol, dot density, and nominal point maps because of their prevalence and usefulness in public health studies. Surface, flow, and cartogram maps are not discussed because of their limited use in public health studies. A flow chart is provided to assist users in identifying the most appropriate mapping method based on the characteristics of their data (Figure 2).

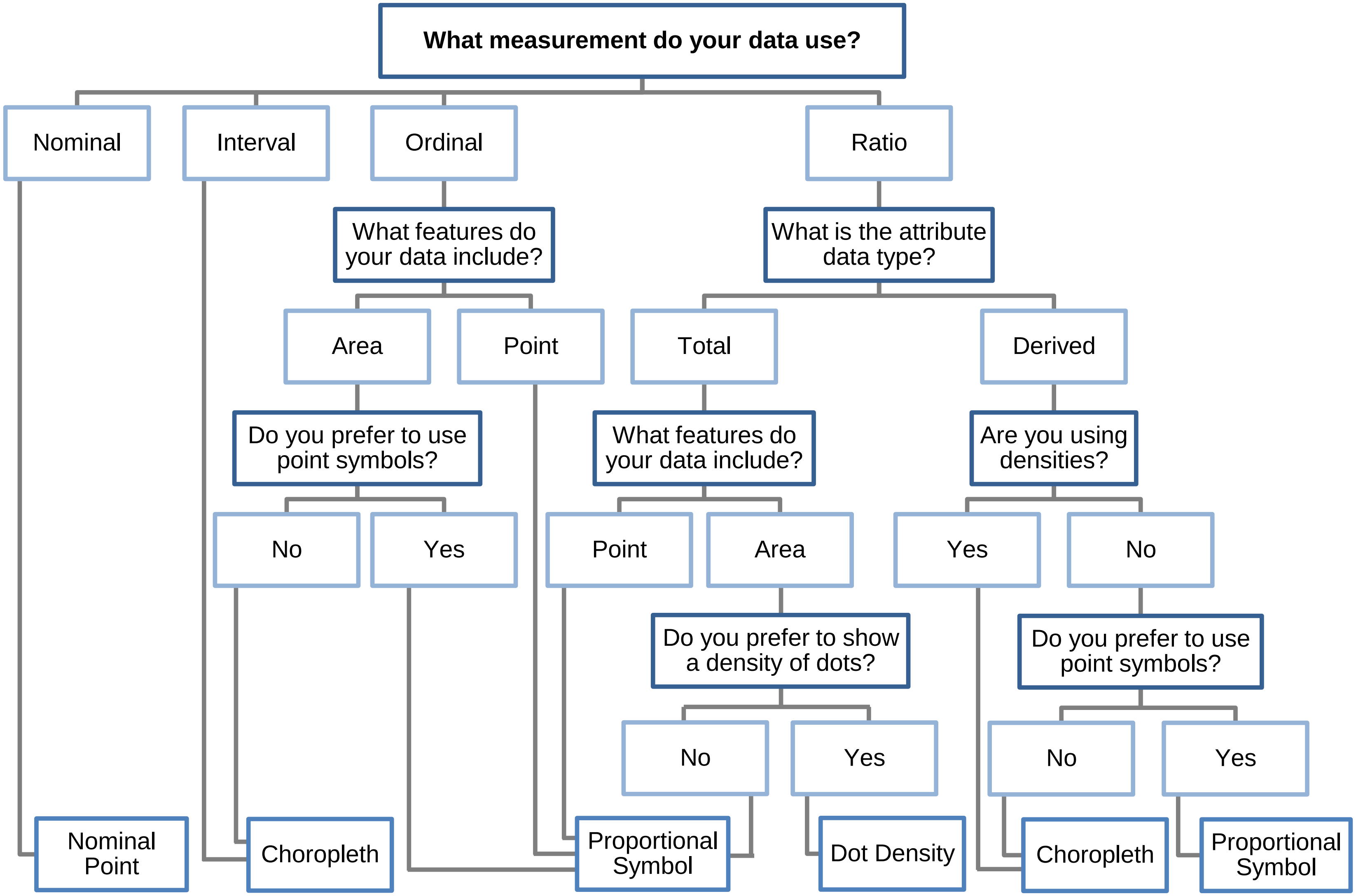


Figure 2 Selecting a Thematic Mapping Method

In order to effectively use the Selecting a Thematic Mapping Method flow chart (Figure 2), certain data characteristics and their definitions must be understood. The following section describes these data characteristics, organized by the type of measurement, which categorizes data based on the mathematical attribute.

Nominal data are at times thought to be a qualitative measurement because it describes features of a certain type. Mathematical operations cannot be done between the different types, or classes, of a nominal dataset. An example of nominal data would be a dataset of the different types of health care services, such as a hospital, free clinic, or long-term care facility. In public health studies, most nominal data are attributes of point features. Point features are specific locations, where each feature has unique coordinates. These phenomena are considered to be zero-dimensional because of the lack of spatial extent.

Area features have also been used to present nominal data, which is often done erroneously. Area phenomena have length and width and are therefore two-dimensional. While some maps present nominal data using enumeration area features, this is an improper practice. Enumeration units that are larger in size may not have large populations; nonetheless, they have stronger visual impact. This is commonly seen in U.S. maps of presidential election results where states have one of two colors: blue to represent Democrat-won states or red to represent Republican-won states. In the 2012 election, many states with a large area but small population were won by the Republican Party, such as Montana. Many states with a small area but large population were won by the Democratic Party, such as Maryland. Large states are more visually dominant than small states despite that they may have a small number of electoral votes (Montana has only three electoral votes compared to Maryland's ten). Although there appears to be much more red color than blue in these maps, suggesting a win for the Republican Party, this was not the case. When population distribution is not taken into account, the audience may mistakenly assign the most importance to the most visually dominant areas, whether or not they have large or small populations (Field 2013). It would be more appropriate to present nominal attributes of area features in tabular form or use area features for representing quantitative data (Dent, Torguson, and Hodler 2008).

Ordinal data group values into a hierarchy or ranking system where it is not known what the difference is between each class. For example, in an ordinal dataset that groups counties based on disease burden, a county with a score of five does not necessarily have five times the disease burden as a county with a score of one (Mitchell 2005; Dent, Torguson, and Hodler 2008; Slocum et al. 2009). Data from a patient indicating his or her health status using a Likert scale would also use an ordinal measurement. Unlike with nominal data, it is acceptable for public health studies to use either area or point features when showing ordinal data. While point data should only employ the proportional symbol, area data can be visualized through either the proportional symbol or choropleth technique.

The interval measurement type ranks values so that the distance between each class is known and equal throughout, but there is no natural origin or zero. For example, percent change in population by county, which may include negative and positive values, uses interval measurement. In public health data, interval data would likely be more available for area features, specifically enumeration units; therefore, the choropleth mapping technique is the only appropriate method to use for interval data among the four methods discussed in Figure 2 (Dent, Torguson, and Hodler 2008). The purpose of the nominal point and dot density techniques are to show precise point data and imply visual density, respectively, both of which are inappropriate for presenting interval data. The proportional symbol technique is also inappropriate because of the lack of a zero in interval data (Dent, Torguson, and Hodler 2008).

Ratio data order values with known differences in between values and have a zero as the natural origin. For example, the number of individuals per census tract uses ratio measurement (Mitchell 2005; Dent, Torguson, and Hodler 2008; Slocum et al. 2009). When using ratio data, it is important to distinguish the attribute context between total and derived data because certain thematic mapping techniques are not appropriate for both types. Total data are the true counts or amounts of features, such as the number of uninsured individuals by county (Mitchell 2005; Dent, Torguson, and Hodler 2008; Slocum et al. 2009). Total data can be represented using proportional symbol maps for point features or by dot density maps for area features. Areal attributes can be measured by one value that applies to the entire area, such as the minimum wage across a state. Alternatively, attributes can be calculated based on a series of individual points within the area, such as calculating the median household income for a state based on incomes for each household within the state (Dent, Torguson, and Hodler 2008; Slocum et al. 2009). Derived data have been mathematically calculated, such as the percent of uninsured individuals by county. Derived data are often standardized or normalized to adjust for the different sizes of enumeration units (Dent, Torguson, and Hodler 2008; Slocum et al. 2009). Common derived indices are ratios, proportions, percentages, and rates. Ratios, not to be confused with ratio measurement, compare the relationship between two data entities, such as the number of children to adults. Proportions are ratios of the count of features to the total, whereas percentages are calculated multiplying the proportion by one hundred (e.g., 0.55 versus 55 percent). Rates are similar to percentages, with the difference being that the proportion is multiplied by a much larger value (e.g., twenty five cases of heart disease per 1,000). These differences are important because the choropleth technique is suitable for mapping any type of derived data, but the proportional symbol technique should not be used for mapping densities, which is a type of ratio that compares a quantity to the unit area. Densities are traditionally visualized best by choropleth maps (Dent, Torguson, and Hodler 2008).

Public health practitioners must keep in mind two challenges of using spatial data when mapping area data. The modifiable areal unit problem (MAUP) occurs as a result of the common practice of aggregating spatial data. Spatial data collected at a detailed geographic level are aggregated during analysis and reporting, often at a new arbitrary geographic level, which often impacts how data are interpreted. Related to the MAUP is the ecological fallacy, which occurs when trends observed at one level of geography are incorrectly assumed to be observed at another level (O'Sullivan and Unwin 2010). Areal aggregations of data will impact how a map is interpreted by its audience, and therefore it is the responsibility of the mapmaker to make note of the areal unit in the interpretation or description (Monmonier 1996).

When understanding one's data through the previously mentioned contexts, a public health practitioner will be able to successfully select an appropriate thematic mapping method to visualize his or her data using the Selecting a Thematic Mapping Method flow chart (Figure 2). This flow chart focuses on data most prevalent in public health studies and therefore excludes line features. These features have a linear path and considered to have length but no width, thus considered to be one-dimensional. An example of linear phenomena would be a travel route from a patient's home to his or her primary provider. Line features are used less frequently compared to point and area features. The flow chart (Figure 2) and forthcoming sections focus on some of the most popular and useful thematic mapping methods and data characteristics associated with these methods. Best practices of choropleth, proportional symbol, dot density, and nominal point techniques are presented in this chapter, and examples are provided to demonstrate appropriate (and for comparison, inappropriate) cartographic practices.

3.2.1 Choropleth Mapping

Choropleth maps use color gradation to indicate varying values of what are usually administrative areas (enumeration units); these maps are one of the most popular – and incorrectly used – of the thematic maps. Darker colors are typically used for higher values whereas lighter colors represent lower values; and data should always be in the derived form, unless enumeration units are similar in size and shape (Dent, Torguson, and Hodler 2008; Slocum et al. 2009). Once data have been deemed appropriate for using the choropleth mapping method, additional considerations must be made in designing the map. Cartographers must pay attention to the attribute context, data classification, and color (Monmonier 1996; Dent, Torguson, and Hodler 2008; Slocum et al. 2009).

The attribute context of the data is an extremely important consideration. A common and very erroneous practice in using the choropleth mapping technique is to map total values, such as in Figure 3A where the number of foreign-born are mapped in California by county. Compared to Figure 3B, where the percentage of foreign-born are mapped, the spatial patterns appear very different.

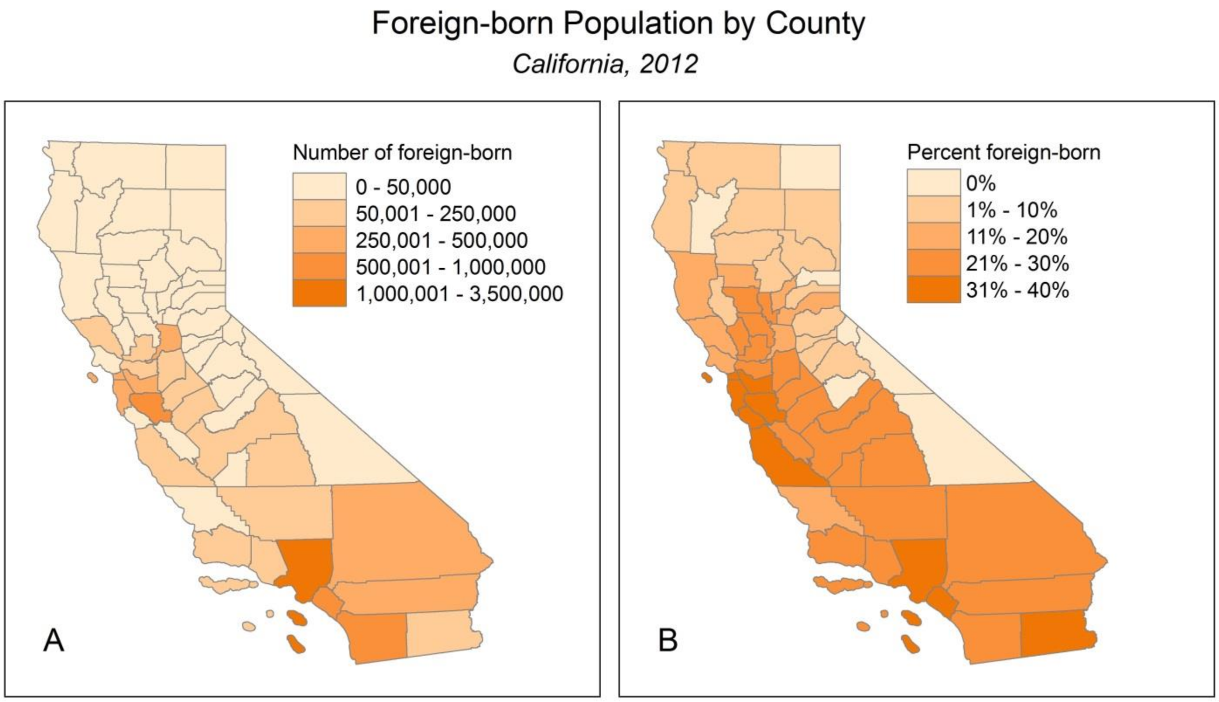


Figure 3 Total (A) and Derived (B) Values
Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Selected Social Characteristics, Table DP02, 2012 Census TIGER/Line Shapefiles

While Figure 3A appears to show that only a small number of counties in northern-central and southern California have large numbers of foreign-born, Figure 3B shows that in many counties throughout California about 21 to 40 percent of the population is foreign-born. Figure 3A masks the fact that the percentage of foreign-born is high in many counties, and this can mislead its audience into underestimating the spread of foreign-born populations. Mapping total values using the choropleth technique should be avoided because 1) this technique does not show the variation within each enumeration unit, 2) most boundaries of enumeration units are drawn arbitrarily and do not reflect the patterns of the phenomena being visualized, and 3) most enumeration units are different in size (Dent, Torguson, and Hodler 2008; Slocum et al. 2009).

Using derived data, such as in Figure 3B, adjusts for the issues faced when working with enumeration units.

One of the most important choropleth mapping technique issues that must be carefully considered and planned is the classification of data (Dent, Torguson, and Hodler 2008). Data classification is the grouping of data values into classes, and the key objective is to create classes in which observations within a class are similar and different classes are unalike (Monmonier 1996; Dent, Torguson, and Hodler 2008; Slocum et al. 2009). There is a multitude of ways to classify data, and common data classification methods include the equal interval, quantile, natural breaks (Jenks), and standard deviation methods. Cartographers should carefully examine the distribution of data and consider the advantages and disadvantages of the classification methods before selecting one (Table 2) because each way may present a unique spatial pattern, some of which may mislead audiences.

Table 2 Comparison of Data Classification Methods

Data classification method	**Advantage**	**Disadvantage**
Equal interval	Groups values in equally sized ranges, easy to interpret	May hide distribution of values and show little variation
Quantile	Evenly distributes values into classes, convenient for mapping ordinal data	Similar values may be placed in separate classes, may create illusion of diversity of data
Natural breaks (Jenks)	Groups similar values into the same class, separates dissimilar values	Large number of values may cluster into one or two classes, ranges may seem arbitrary to audience
Standard deviation	Shows how values deviate from mean	May confuse audiences that are unfamiliar with statistics

Source: Dent, Torguson, and Hodler 2008; Slocum et al. 2009

To illustrate the importance of selecting an appropriate data classification method, a histogram and maps of the percentage of individuals with a disability in California by county are provided (Figures 4 and 5). Figure 4 is a histogram of the distribution of percentages, which shows that all counties have between a minimum of 8 percent and a maximum of 23 percent of individuals with disabilities. The distribution has a positive skew, where the highest numbers of counties have approximately 11 percent and the mean percentage is 13 percent. There is a gap in the distribution, where no counties have 15 percent of individuals with disabilities.

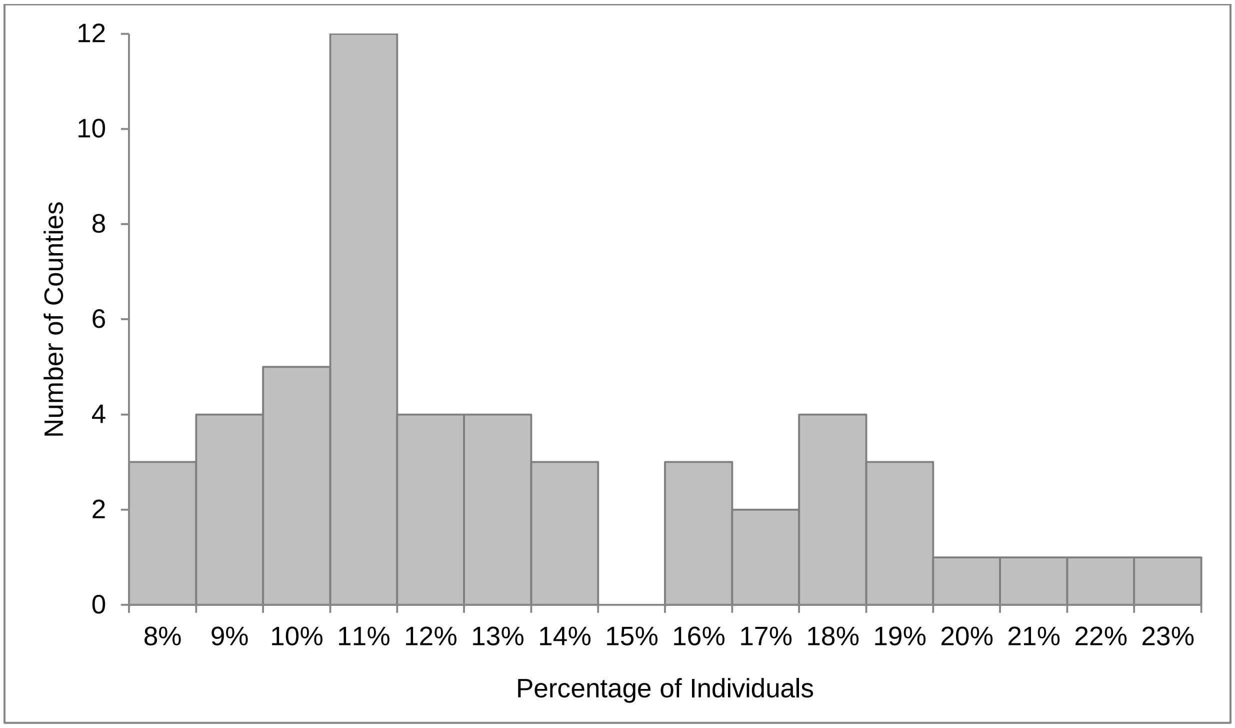


Figure 4 Percentage with Disability in California, 2012
Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Selected Social Characteristics, Table S2701

Figure 5 presents four maps of the percentage of individuals with disabilities. Each map is visualized using one data classification method: the equal interval, quantile, natural breaks, or standard deviation data classification method. Despite using the same dataset each map has a unique color pattern.

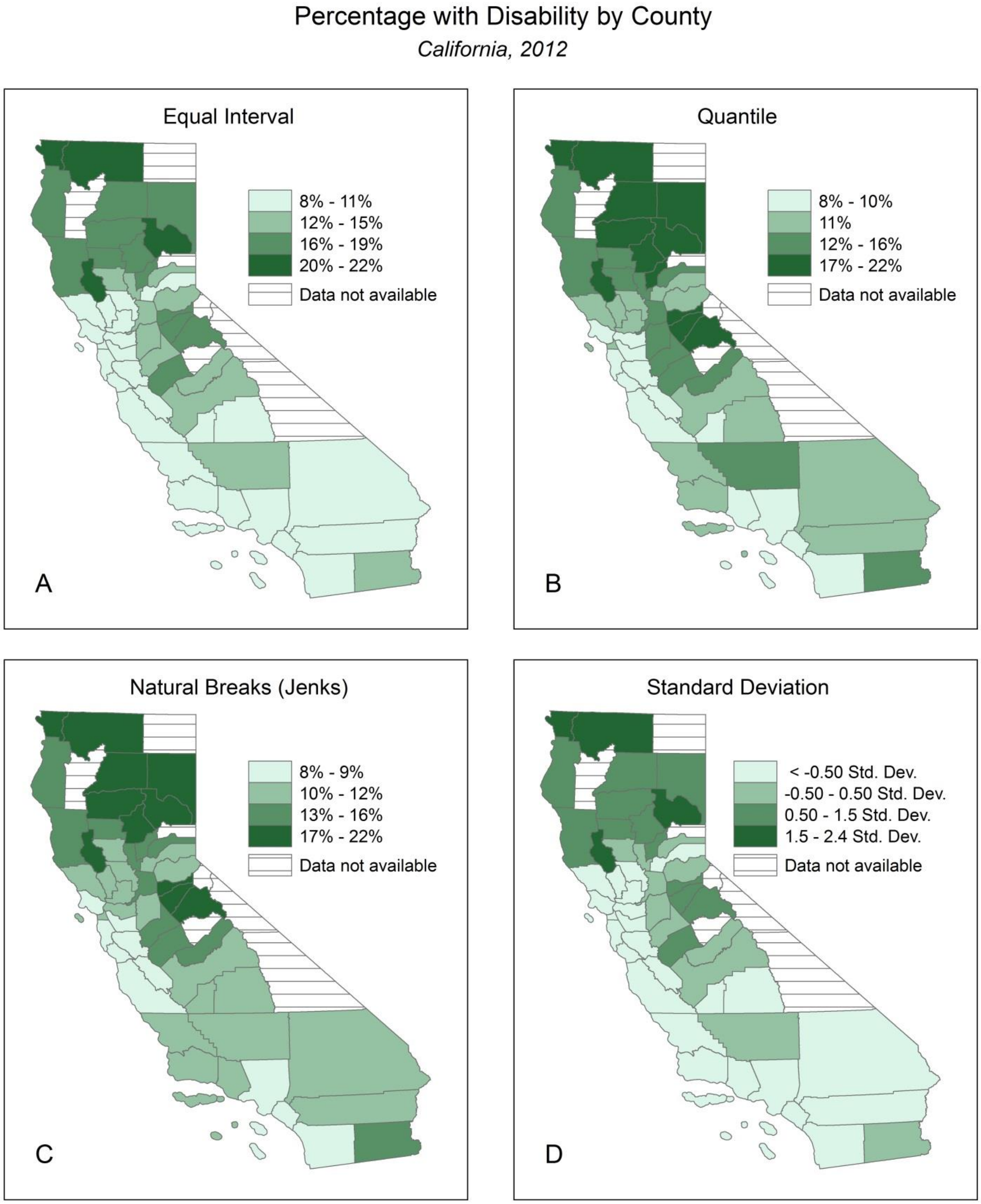


Figure 5 Comparison of Data Classification Methods
Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Selected Social Characteristics, Table S2701, 2012 Census TIGER/Line Shapefiles

The map using equal intervals, which groups values into equally sized ranges, shows the least variability among the maps that visualize percent values (Figure 5A). Nearly half of the counties fall into the lowest class (8 to 11 percent), mostly in the south whereas counties in the north appear to fall into the higher classes.

The quantile map, which groups approximately the same number of counties for each class, shows the most variability, with thirteen counties throughout northern and northern-central California having percentages that fall in the highest range (Figure 5B). Counties in central and southern California have low to moderate percentages. The most common percent in the distribution, 11 percent, has a class of its own; and the largest data values (17 to 22 percent) are grouped into one class whereas the equal interval uses two classes for the largest data values.

The variability in spatial pattern of the map using natural breaks more closely resembles the quantile map than the equal interval map (Figure 5C). Like the equal interval map, most southern counties have low percentages of individuals with disabilities, whereas the color pattern of the natural breaks map in the northern counties more resemble the quantile map. The distribution of data values into classes is also more similar between the natural breaks and quantiles maps.

The spatial pattern appears different across the various data classification methods despite the fact that the dataset is the same, demonstrating the impact of classification on audience interpretation. The map using the final data classification method, standard deviation, actually has a similar color pattern to the equal interval map (Figure 5D), but it should be noted that diverging color schemes are recommended for maps showing standard deviation. Standard deviation should also be mapped only when the data is normally distributed and when the audience understands basic statistical concepts (Dent, Torguson, and Hodler 2008; Slocum et al. 2009).

Just as there is no one right data classification method, there is no one way of selecting the number of classes and colors. Too few classes risk hiding patterns in the data, whereas too many confuses the reader, as demonstrated in Figure 6. As the number of classes increase, the spatial pattern of the map becomes more complex and requires the audience to explore the data in more depth. However, when the same color is used with varying lightness, the number of classes should never exceed seven because it is challenging for the human to perceive differences in any more than seven classes with different lightness. The default for number of classes is typically four or five (Dent, Torguson, and Hodler 2008; Slocum et al. 2009). Although using varying hues and lightness of color can create complex spatial patterns, cartographers should always keep in mind that facilitating understanding of the data should take priority over aesthetics of the map (Peterson 2009).

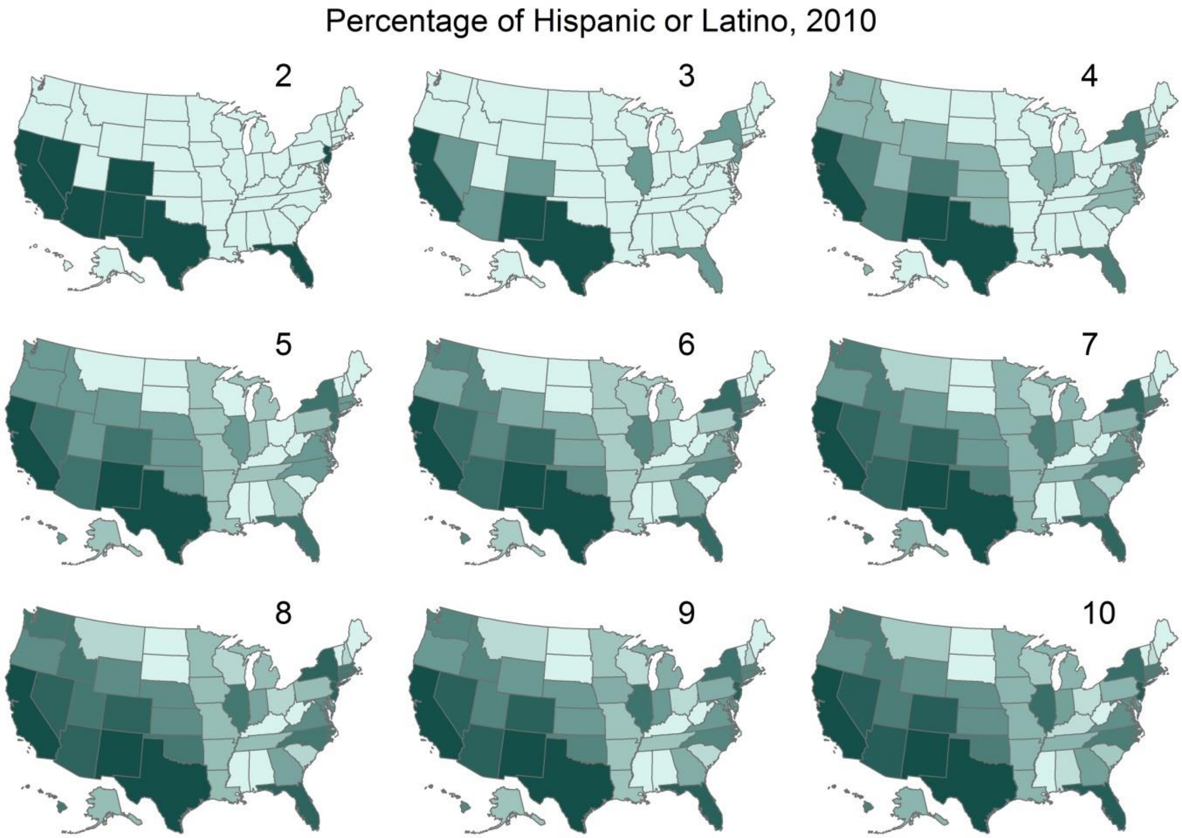


Figure 6 Comparison of Numbers of Data Classes
(Number of data classes indicated by number above each map)
Source: CDC, 2010 Behavioral Risk Factor Surveillance System GIS Data

The color scheme should be carefully chosen to appropriately match the data. Two common types of color schemes used in choropleth maps are the sequential and diverging schemes. When color schemes do not match the data, there is high potential for audience misunderstanding (Dent, Torguson, and Hodler 2008; Slocum et al. 2009). For example, using a diverging color scheme to present unipolar data like percentage of individuals who report low health status can be confusing (Figure 7A). The audience may not carefully read the legend and could easily assume that the yellow hue represents a mean or threshold value or feel confused regarding which hue (red or blue) represents high percentages of unhealthy people. Instead, such data should be mapped using a sequential color scheme, where the color value increases in one direction, just as the data do (Dent, Torguson, and Hodler 2008; Slocum et al. 2009). Mapping the percentage of individuals who report fair or poor health using a red color scheme (Figure 7B) leaves little room for misunderstanding what the darker color value represents. Unipolar data should always be visualized with sequential schemes.

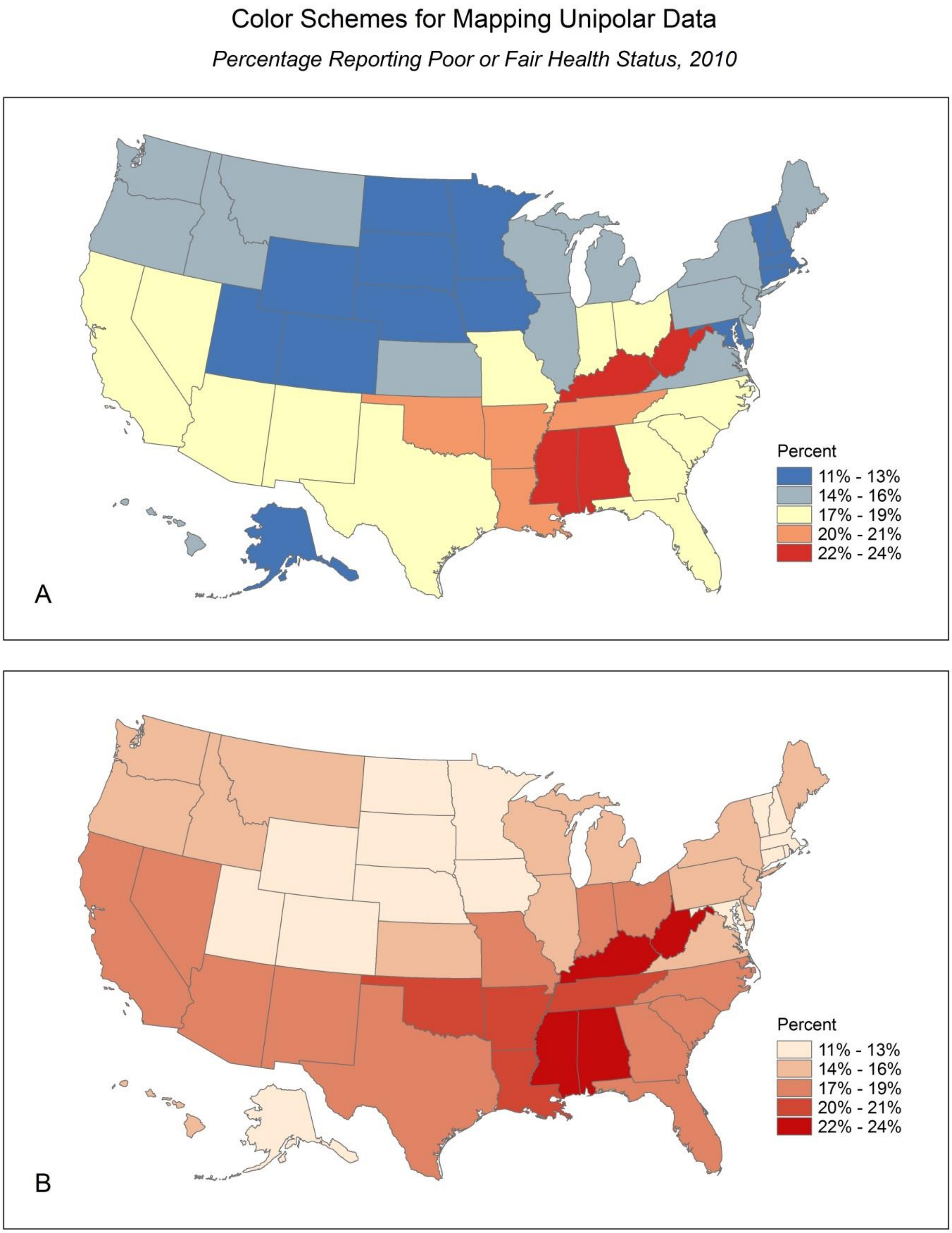


Figure 7 Diverging (A) and Sequential (B) Color Schemes for Mapping Unipolar Data
Source: CDC, 2010 Behavioral Risk Factor Surveillance System GIS Data

Diverging schemes are intended for emphasizing both ends of a data range in a dataset that has values increasing in both directions from a middle point (Dent, Torguson, and Hodler 2008; Slocum et al. 2009; Brewer 2005). Because two different hues increase in value from a light hue, diverging schemes are excellent for visualizing percent change and standard deviation (Slocum et al. 2009). Negative values can be symbolized using one hue and positive values with the other, such as in Figure 8.

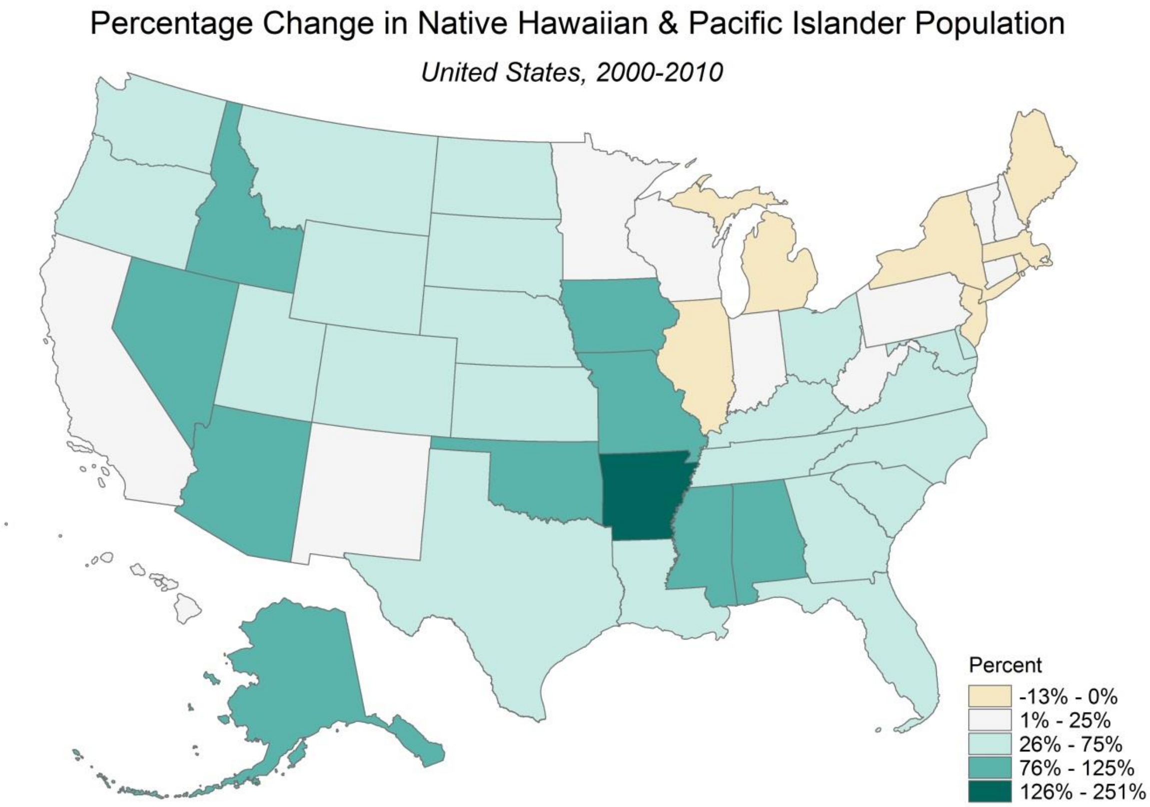


Figure 8 Appropriate Visualization of Bipolar Data Using Diverging Color Scheme
Source: U.S. Census Bureau, 2000 Census Summary File 1, 2010 Census Summary File 1, 2010 Census TIGER/Line Shapefiles

As the most popular thematic mapping technique, choropleth maps are recognizable and easy to read by most audiences. However, it is critical for cartographers creating these maps to follow mapping standards to ensure that the maps convey spatial patterns and do not mislead.

When using the choropleth mapping technique, cartographers should pay special attention to determining the data classification method and selecting the color scheme.

3.2.2 Proportional Symbol Mapping

The proportional symbol mapping technique involves varying the size of symbol forms in proportional to the quantities they represent. This technique can be used to visualize many different types of data: true and conceptual point data, data using ordinal or ratio measurements, as well as both total and derived values. When using proportional symbols, cartographers should be careful to avoid using inappropriate data and pay special attention to data classification and symbol design (Dent, Torguson, and Hodler 2008; Slocum et al. 2009).

The proportional symbol map can be adapted for both true point and conceptual point data (Slocum et al. 2009). In true point data, point symbols are located at the precise location of data collection. What can also appear as a true point is actually a conceptual point, where point symbols represent the area in which the data are collected and aggregated (Slocum et al. 2009). Selecting between these two data types will depend on what data is available (although point features can be spatially joined to area features to calculate the sum, mean, minimum, maximum, etc. of point features and attributes that fall within the boundaries of each area feature). In Figure 9, two proportional symbol maps are presented, the first showing the number of hospital beds where each point is placed at the actual hospital location (Figure 9A) and the second showing the number of hospital beds per city (Figure 9B). Figure 9A emphasizes the precise distribution of the hospitals and the number of beds available in each hospital. Figure 9B represents the overall distribution and range of hospital beds for each city. Each of the two types of point data provides different perspectives and information; therefore, each map has potential to provide valuable information, but cartographers should first consider the purpose of visualization and communication.

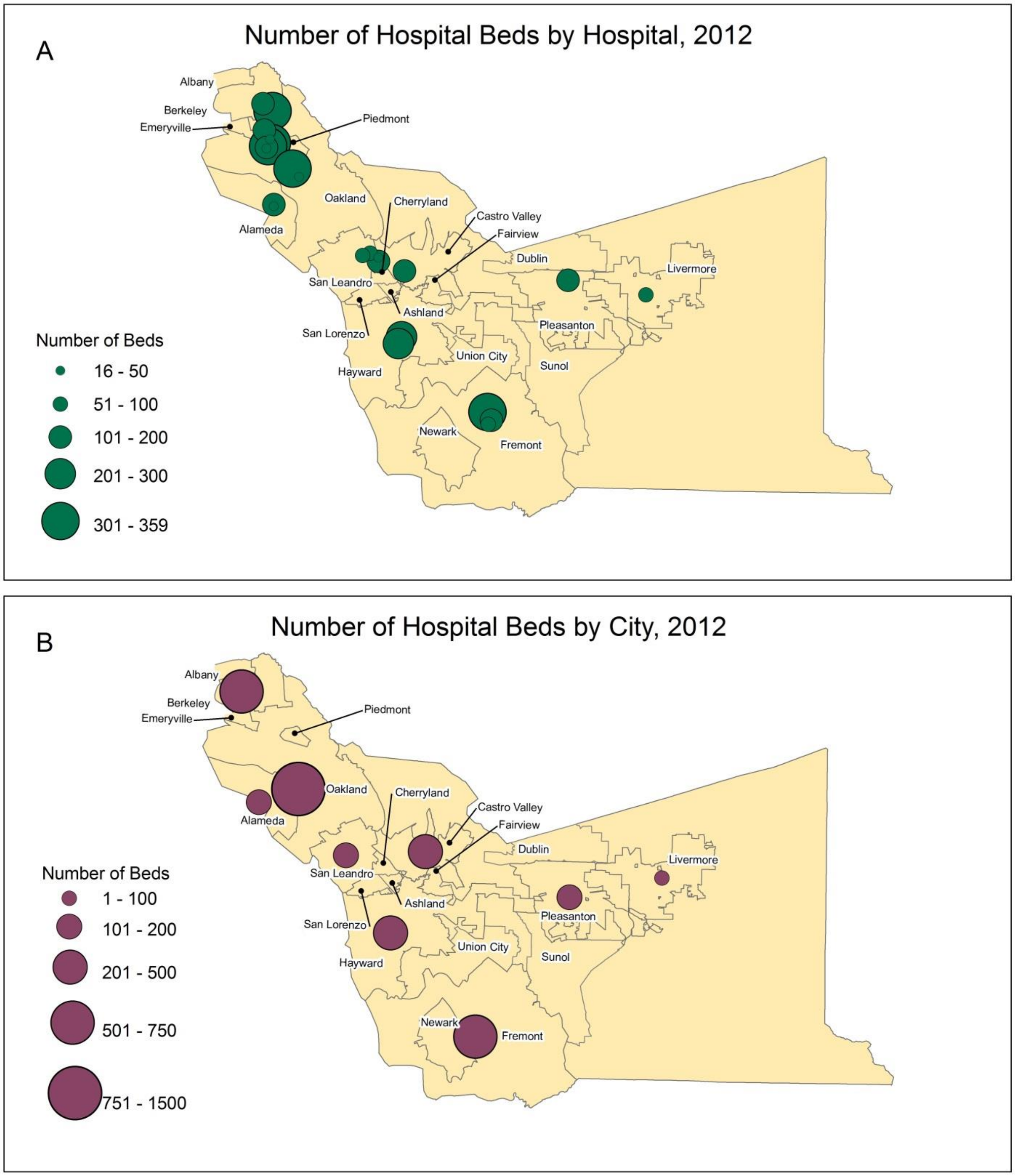


Figure 9 True Point (A) and Conceptual Point (B) Data
Source: Office of Statewide Health Planning & Division, 2012 Health Care Facilities; U.S. Census Bureau, 2012 Census TIGER/Line Shapefiles

Interval data and densities should always be mapped using a mapping technique other than proportional symbols. If proportional symbols were used to map interval data, whenever there is no symbol the audience normally assumes a zero value. However, interval data lack a zero or natural origin, and this may confuse the audience. Densities are traditionally mapped using the choropleth method.

Data with relatively small data ranges should not be mapped with proportional symbols because the resulting spatial pattern will appear monotonous (Dent, Torguson, and Hodler 2008). Figure 10A is a map that is technically an accurate representation of the geographic distribution of individuals with a history of heart disease, but this map does not appear to provide any useful findings. The percentage of people with such a health history is relatively small in range across all of the United States. There appears to be no variation in the percentages, and the map would not effectively visualize the spatial distribution of the percentages. Figure 10B is a map with a larger data range; the percentage of individuals who suffer from obesity clearly differ across the U.S. The map reveals that the Midwest and South have higher percentages of obesity, while the West Coast and the Northwest have much lower percentages, demonstrating that proportional symbol maps with larger data ranges are more visually informative.

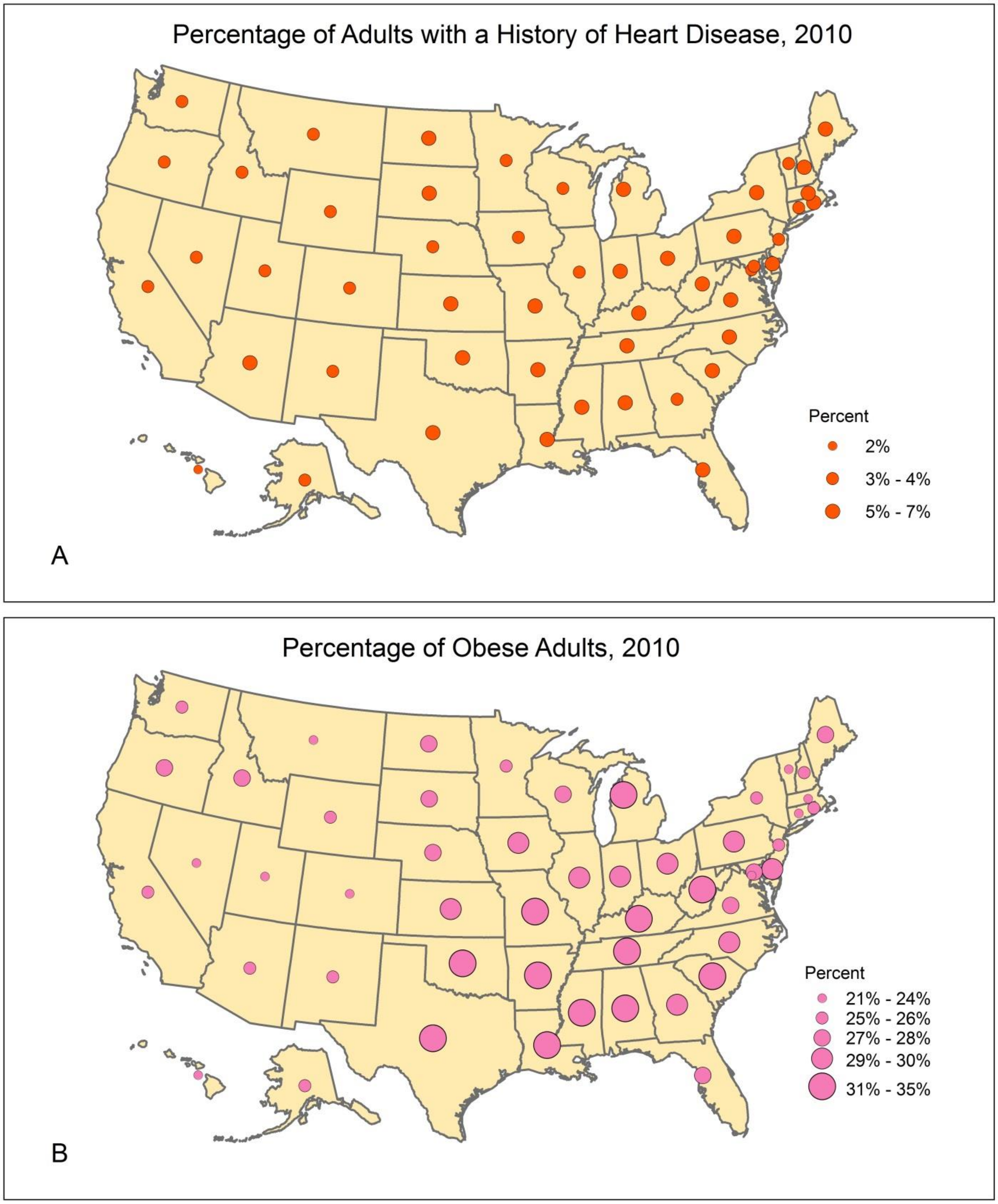


Figure 10 Inappropriate (A) and Appropriate (B) Data Ranges for the Proportional Symbol Mapping Technique

Source: CDC, 2010 Behavioral Risk Factor Surveillance System GIS Data

When using the proportional symbols mapping technique, cartographers can use unclassed or classed symbols. Unclassed symbols are sized in proportion to the data value they represent; they are also known as proportional symbols. Classed symbols are similar to choropleth classes, where the number of classes is set and data values are grouped into classes. These are also known as graduated symbols, and the symbol size is usually not proportional to the data value (Dent, Torguson, and Hodler 2008; Slocum et al. 2009). Figure 11 compares the unclassed and classed symbols using the same dataset of Asian population by state. The unclassed symbols in Figure 11A show much more variation in symbol size compared to Figure 11B's classed symbols, and Figure 11A's spatial pattern is more readily understandable. However, maps with unclassed symbols may have more overlap, such as with the states with the largest Asian populations on the west and east coast, requiring the map's audience to discern the overlapping symbols.

Although there is no one right way of classifying data, cartographers should consider the data range of their dataset. When data ranges are large and include several extremely large data values, cartographers should be wary using unclassed symbols where symbols may overly dominate the map and cover up other features. This can be avoided when using classed symbols, which offers more control in setting symbol size.

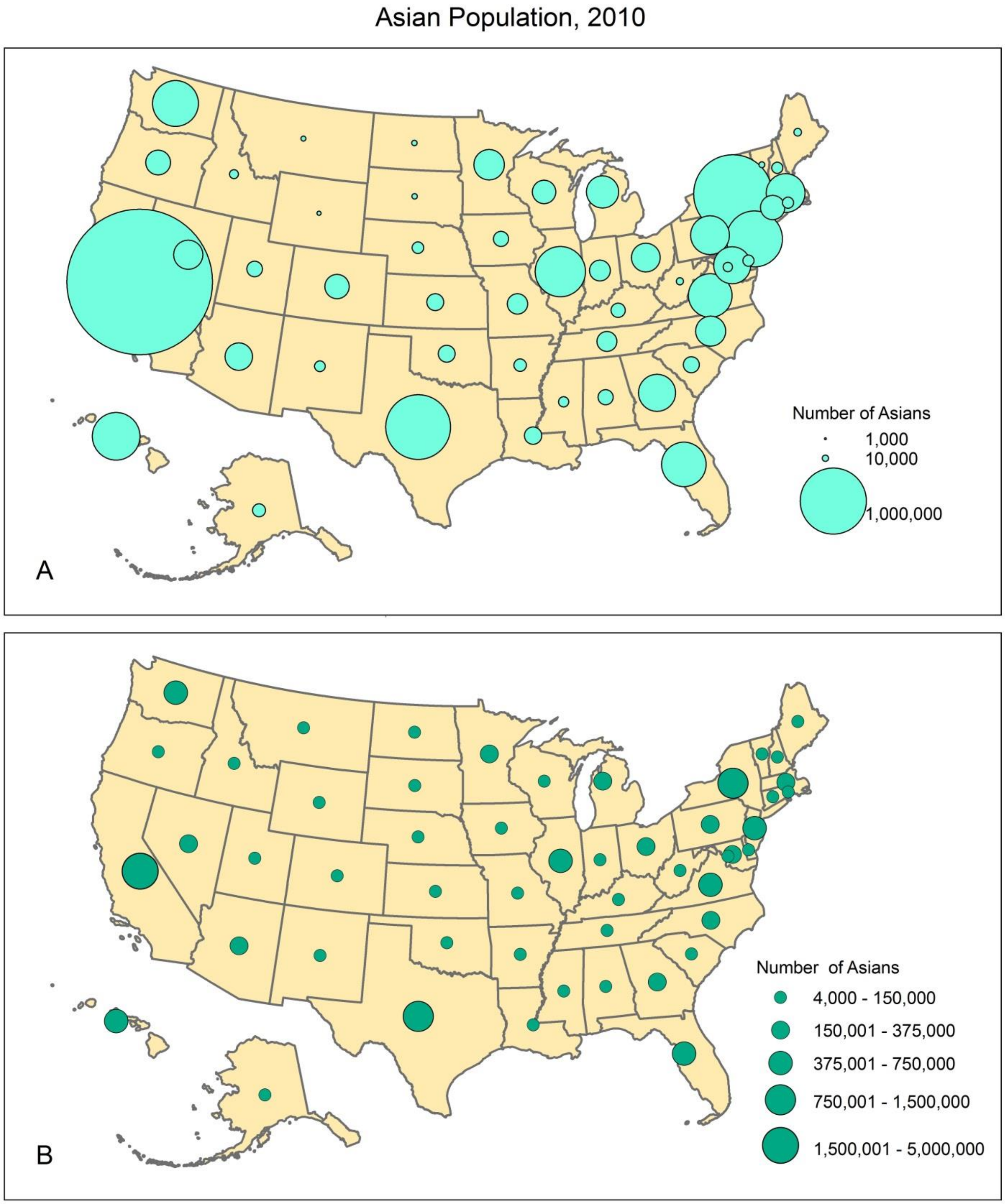


Figure 11 Unclassed (A) and Classed (B) Symbols
Source: U.S. Census Bureau, 2000 Census Summary File 1, 2010 Census Summary File 1, 2010 Census TIGER/Line Shapefiles

Setting the size of symbols can be a challenging process because it is strongly related to handling overlapped symbols. Figure 10A is dominated by many small symbols; it appears empty and shows little variation and spatial pattern. However, large symbols can create too much clutter, as seen in Figure 12A (Slocum et al. 2009). The map showing the number of kindergartners in California by county looks crowded, and many county boundaries cannot be discerned because of large symbol size and excessive symbol overlap. Figure 12B presents the same data but with smaller symbol size and less symbol overlap; it is easier to read than Figure 12A and much less likely to give the audience the wrong impression. Whereas the audience may incorrectly infer from Figure 12A that there are large numbers of kindergartners in each county, the audience will be able to correctly discern that the largest numbers of kindergartners are in the Bay Area and southern California. In addition to using opaque overlapping symbols, cartographers may also want to consider the use of transparent symbols which can achieve a clean appearance while still providing a valuable spatial pattern to the audience. While opaque symbols promote visual hierarchy because the symbols are more likely to stand out, transparent symbols allow for more background to show. Cartographers should use opaque symbols when they want to emphasize the proportional symbols, and they should use transparent symbols when they want to show background information, such as boundaries of enumeration units (Slocum et al. 2009).

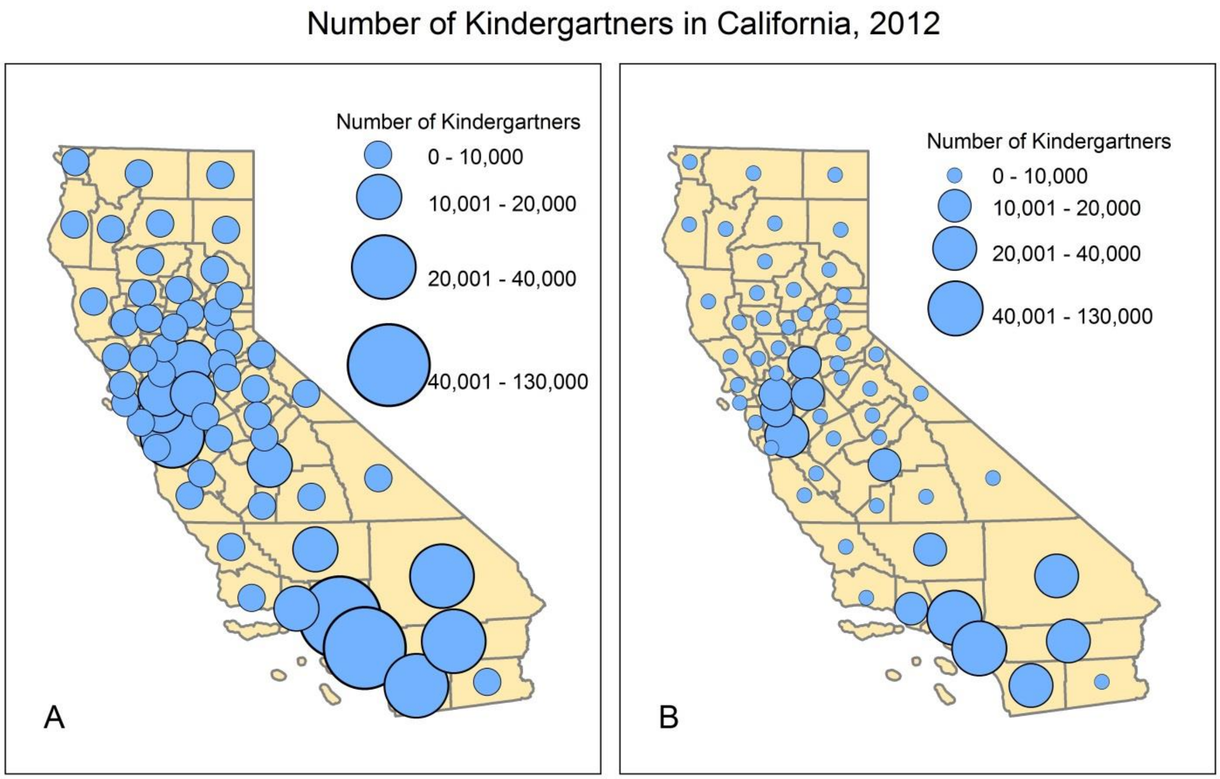


Figure 12 Inappropriate (A) and Appropriate (B) Overlapping Symbols
Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Selected Social Characteristics, Table S2701, 2012 Census TIGER/Line Shapefiles

Another consideration of symbol design is the symbol type. Geometric symbols are the most common, with circles being the most popular among squares, triangles, and other shapes, as well as pictorial symbols. Circles provide several advantages: they are compact and visually stable and they can overlap while still conveying magnitude of data (Dent, Torguson, and Hodler 2008). Other symbols can be used as well, but cartographers should take care not to select symbols that prevent audiences from easily understanding the magnitude of the data. Figure 13A's use of a pictorial symbol looks cluttered and the irregular shape does not clearly convey the differing data values. In this case, cartographers should use more simple shapes, such as squares, as seen in Figure 13B. The advantage of square symbols is that readers are more able to easily discern spatial patterns and magnitude of data, but the square shape is not as visually stable as the circle. Furthermore, the aesthetics of the angularity of the square is not popular among cartographers (Dent, Torguson, and Hodler 2008; Slocum et al. 2009).

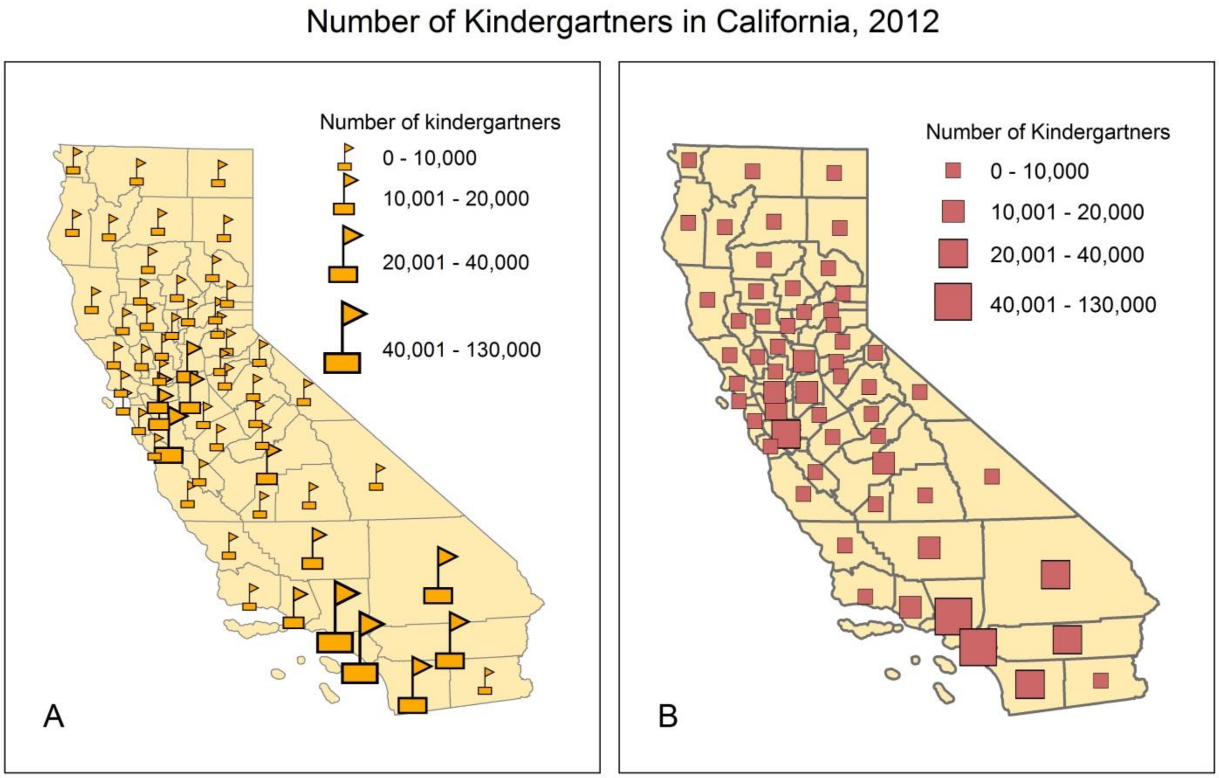


Figure 13 Pictorial (A) and Geometric (B) Symbols
Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Selected Social Characteristics, Table S2701, 2012 Census TIGER/Line Shapefiles

The proportional symbol mapping technique can visualize many different types of data. However, cartographers must remember that there are specific cases in which data is not suitable for using this particular technique (e.g., interval data, densities, and data with small data ranges). Cartographers should also carefully consider how they classify their data and set symbol size and shape in order to create a map that clearly communicates spatial patterns and is aesthetically pleasing (Dent, Torguson, and Hodler 2008; Slocum et al. 2009).

3.2.3 Dot Density Mapping

The dot density mapping technique is a simple method where dots are used to represent total values of enumeration units, displaying hypothetical patterns to show variation. Dot density mapping, also known as dot mapping, involves the selection of dot size and value to represent raw totals by dots (e.g., one dot equals 1,000 events). Dots are placed within the enumeration unit in one of three ways: uniform, geographically weighted, or geographically based. The uniform approach places dots within the enumeration unit with uniformity. The geographically weighted approach considers the intensity of the neighboring enumeration areas and place more dots near the boundaries of the areas that have higher intensity. The geographically based approach places dots only within areas where phenomena are possibly observed by using ancillary information, which is information used to minimize error in mapping enumeration data.

These maps are excellent for visualizing the intensity of quantity because they are easy to understand and effectively communicate spatial density (Dent, Torguson, and Hodler 2008; Slocum et al. 2009). For a cartographer to successfully use the dot density mapping technique, special attention must be paid to the type of data. It is important to not map derived data including densities because the density of dots implies the magnitude of raw total values. The data range must not be too large or too small, and the dot symbols should be carefully designed. Also, ancillary information should be included to increase accuracy in dot placement (Dent, Torguson, and Hodler 2008; Slocum et al. 2009).

Determining appropriate dot value is vital in creating a dot density map that the audience can correctly interpret. Figure 14 demonstrates the impact of the dot value on a map’s readability. A high value may result in a map with little spatial pattern, such as Figure 14A’s map of the percentage of women who gave birth in California. The distribution of dots in more populous counties looks acceptable but less populous enumeration units have no dots. This falsely implies that these enumeration units had no women who gave birth. A low value may create an excessive number of dots that renders the map unreadable, as seen in Figure 14B. The dots in the less populous appears can be seen, but the more populous areas become unreadable. Furthermore, when enumeration units contain too many dots, an outline of the enumeration unit may be perceptible by an audience, which is inappropriate. Enumeration unit boundaries of the data used to create a dot density map should never be discernible by the audience (Dent, Torguson, and Hodler 2008). Assigning a dot value to data with a large range is especially challenging, and such data would be better mapped using the choropleth or proportional symbol mapping techniques (Dent, Torguson, and Hodler 2008).

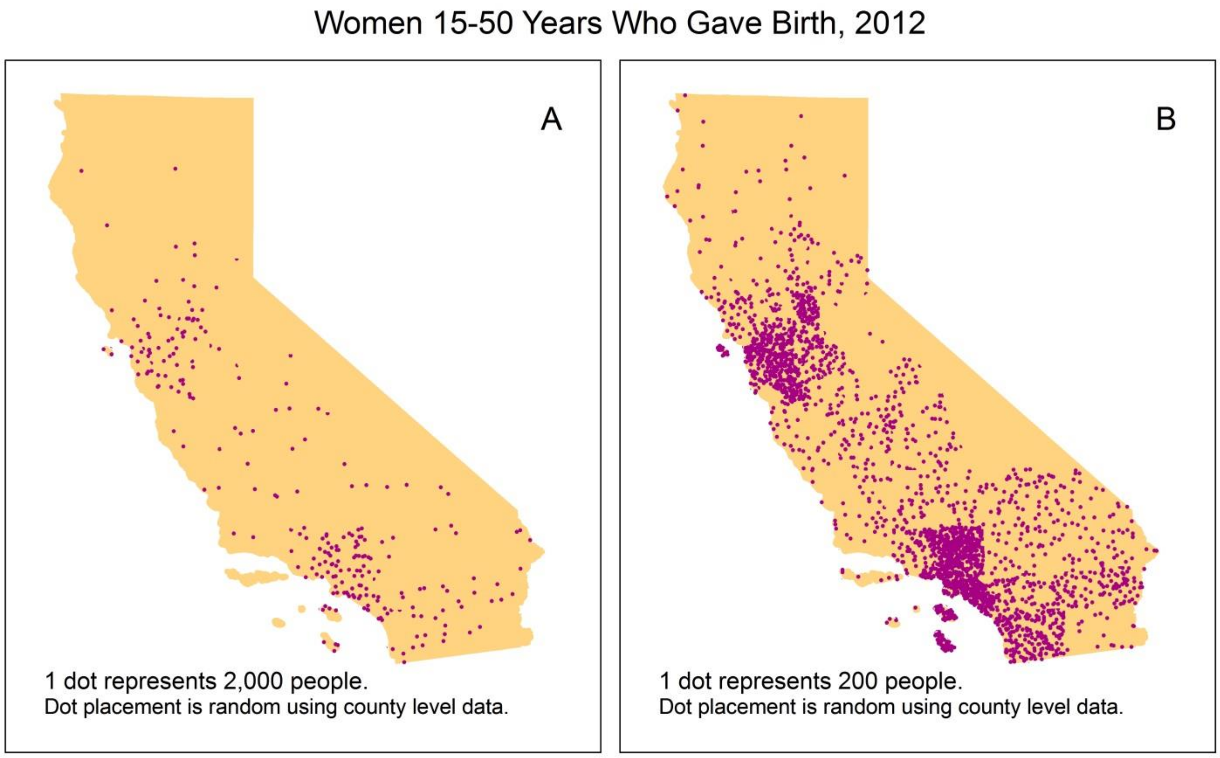


Figure 14 Inappropriate Data Range for Dot Density Maps
Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Selected Social Characteristics, Table S2701, 2012 Census TIGER/Line Shapefiles

Dot size is also important because of its effect on the readers' ability to read the map and understanding of the spatial patterns. The size of the dot should not be so large that there is too much overlap, which impedes discerning the pattern, such as in Figure 15A. The spatial pattern in Figure 15B is much more understandable because of the adequate dot size. The value and size of the dot should properly be set so that the enumeration unit with the smallest quantity should have two or three dots and the enumeration unit the largest quantity should have moderate overlapping of the dots (Dent, Torguson, and Hodler 2008).

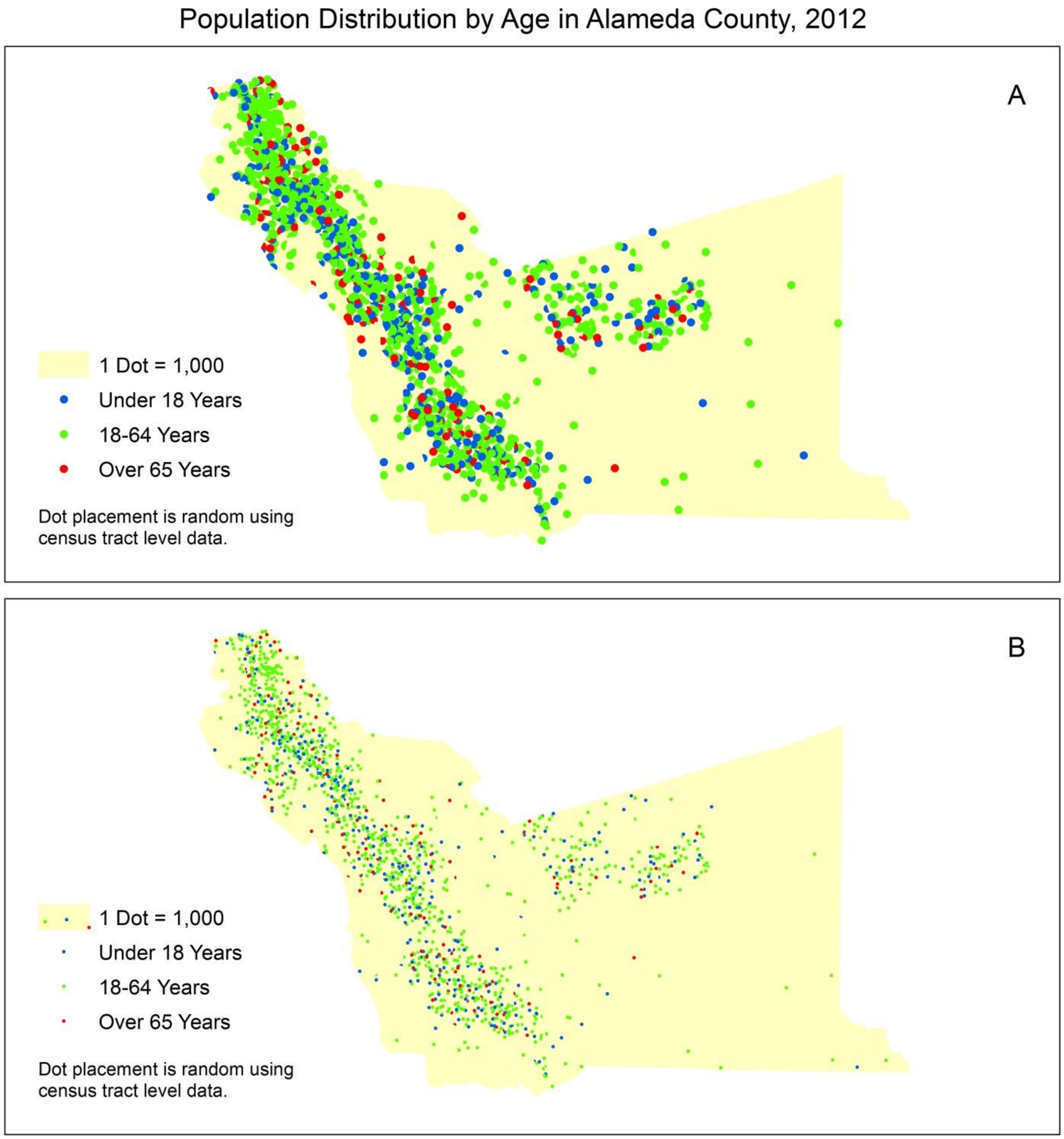


Figure 15 Inappropriate (A) and Appropriate (B) Dot Size
Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Selected Social Characteristics, Table S2701, 2012 Census TIGER/Line Shapefiles

Since dots are randomly placed within each enumeration unit, there is reasonable room for error in dot placement. Ancillary information can be extremely helpful in improving dot placement (Slocum et al. 2009). For example, for the most part, people do not live in the Great Lakes of North America; however, in a map of individuals with diabetes in the U.S. in Figure 16A, there appear to be dots placed all throughout Michigan, including where Lake Superior, Lake Michigan, and Lake Huron are located. Figure 16B takes into account that individuals are unlikely to live on the Great Lakes and only includes dots on land. Other ancillary information may also include land use and zoning data, which would indicate areas where people are unlikely to reside, such as parks or industrial areas. With the ancillary information, the dot placement can be adjusted so that dots are more accurately placed in areas where individuals are likely to be.

Dot density maps are easy to understand for most audiences; therefore, they are very useful for visualizing total values by enumeration unit. When care is taken to ensure that cartographic standards are practiced for dot value, size, and placement, cartographers can present a map that is both informative and interesting.

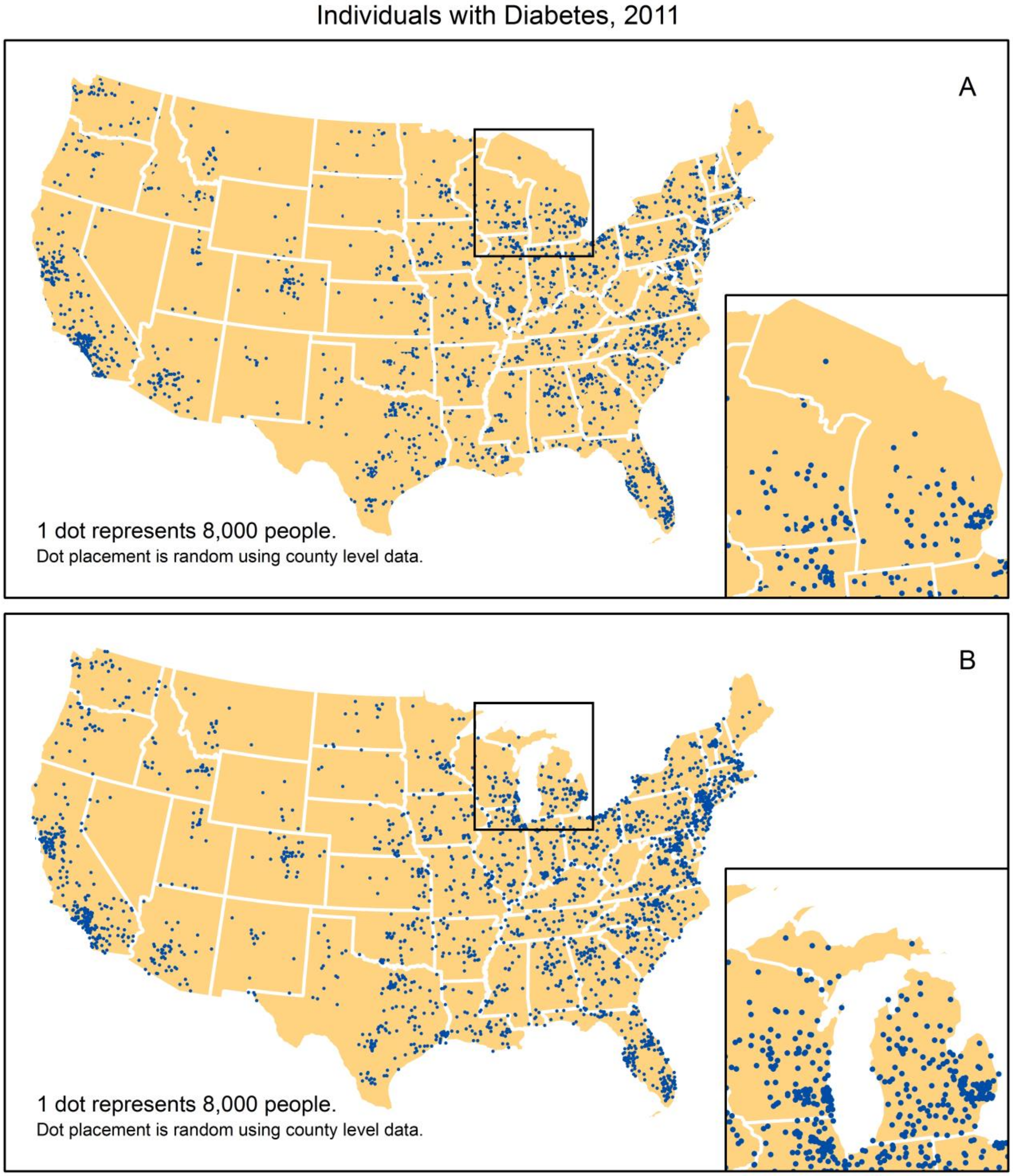


Figure 16 Lack (A) and Use (B) of Ancillary Information
Source: DDT, Diagnosed Diabetes Prevalence

3.2.4 Nominal Point Mapping

The nominal point mapping technique is also known as one-to-one mapping, where the objective is to show precise locations of events or objects. Each point symbol represents the precise location in which the data were collected. Nominal point maps use qualitative data and emphasize the specific location of the point features. When using the nominal point technique, cartographers must consider the symbol type and scale in order to create a successful map (Dent, Torguson, and Hodler 2008). Cartographers should also take care to ensure their nominal point map is not mistaken for a dot density map, which sometimes share similarities.

Compared to the other thematic mapping techniques discussed previously in this study, nominal point maps are unique. They are the only technique that uses qualitative data, which are collected at specific points. This requires the point symbols to be placed precisely in the location representing where the data was collected. One way to ensure that precise point locations are discernible to the audience is to select the appropriate scale. A map that is scaled too small may result in too many overlapping point symbols, which becomes challenging for the audience to understand the data. Figure 17A shows that a map scaled to the county level is too small to see individual locations of health care facilities, whereas Figure 17B shows a map using a much larger scale. Focusing on a smaller area solves the issue of overlap and allows each point to be clearly seen.

Health Care Facilities in Alameda County, 2012

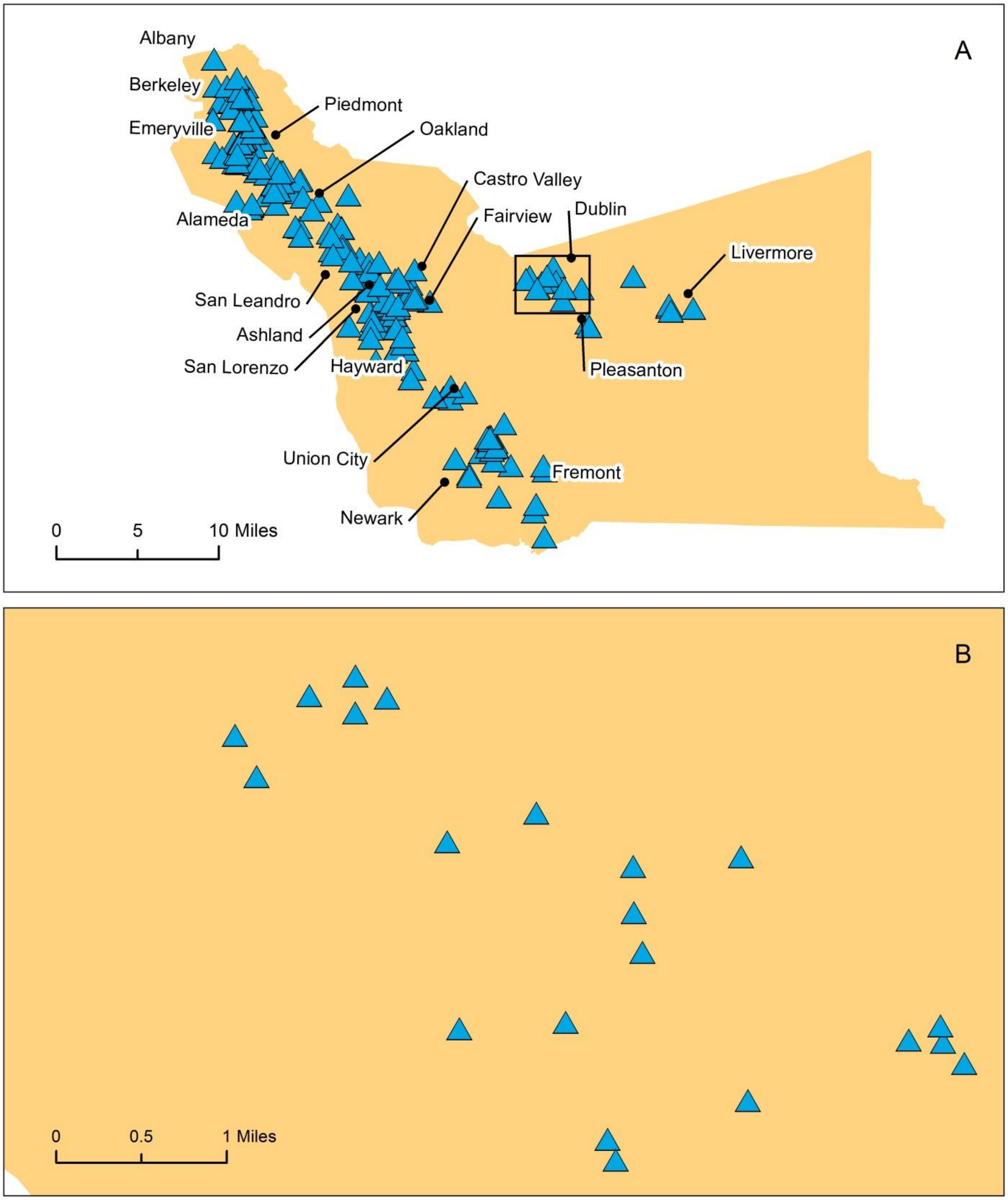


Figure 17 Small Scale (A) and Large Scale (B)
Source: Office of Statewide Health Planning & Division, 2012 Health Care Facilities; U.S. Census Bureau, 2012 Census TIGER/Line Shapefiles

The wide variety in selecting a symbol shape and color creates the opportunity to map multiple nominal categories, with one unique symbol for each category. There are many different point symbols that can be used, similar to proportional symbol maps. Geometric shapes, such as circles and triangles, and pictorial shapes, such as buildings and hospital signs, are common symbol types (Figure 18). Figure 19 is a map of health care facilities, using a unique color to represent each type of facility. Using different colors can be very useful for public health practitioners interested in mapping several different qualitative attributes.

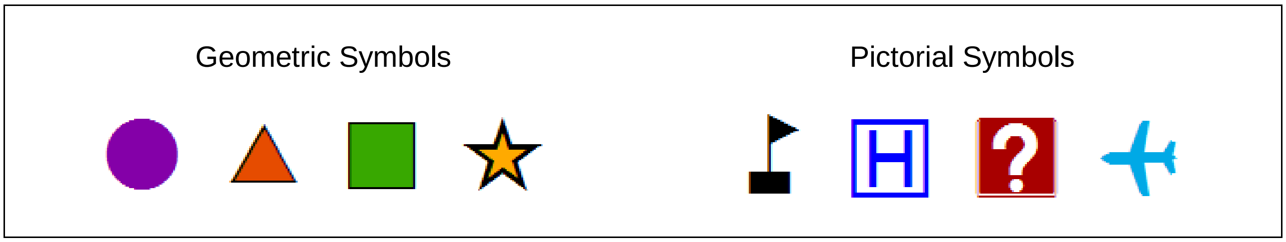


Figure 18 Point Symbol Types

53

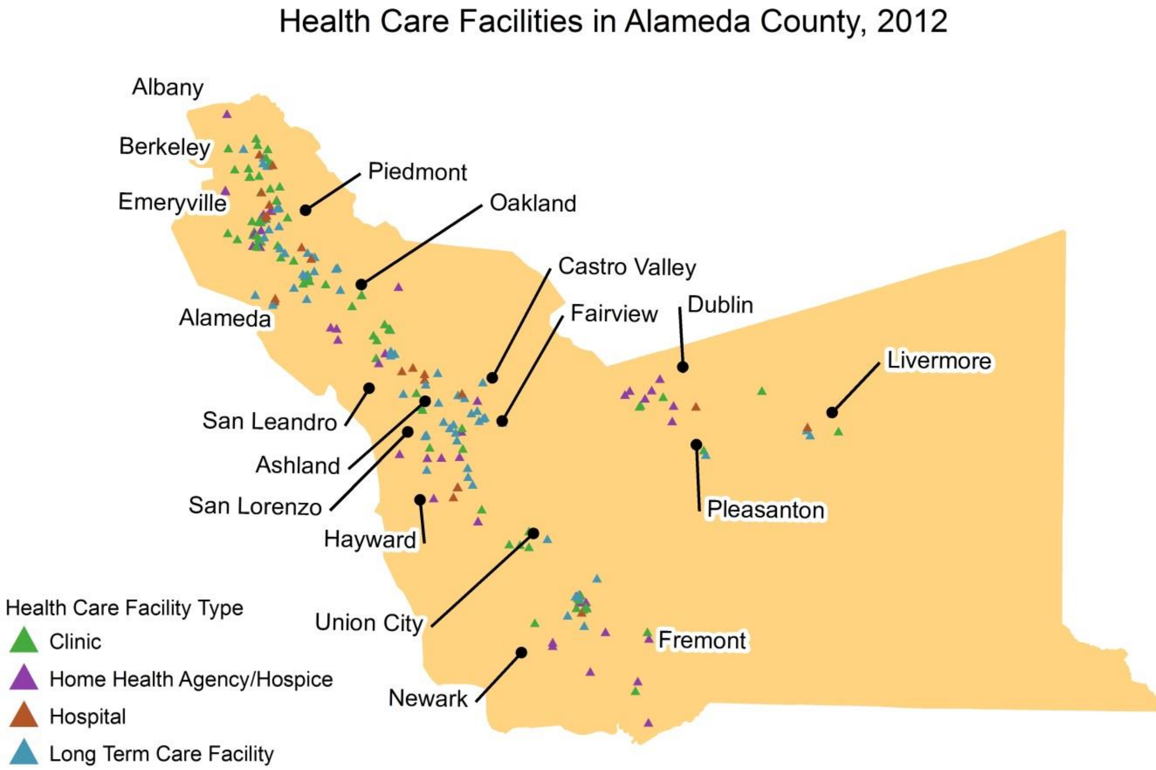


Figure 19 Multiple Nominal Categories
Source: Office of Statewide Health Planning & Division, 2012 Health Care Facilities; U.S. Census Bureau, 2012 Census TIGER/Line Shapefiles

The use of point symbols in nominal point maps may give the impression that they are similar to dot density maps. Although the data used for each thematic map are very different (quantitative data of enumeration units for dot density maps and qualitative data collected at specific points for nominal point maps), the symbols can be very similar. Both mapping techniques use dot symbols, which creates a potential risk for a nominal point map to be mistaken as a dot density map. To avoid any confusion, cartographers may consider avoiding the use of dot symbols in their nominal point maps (Dent, Torguson, and Hodler 2008).

The flexibility in selecting the point symbol shape and color provides many visualization options for cartographers whose objective is to map precise locations of buildings, events, and other places. The nominal point thematic technique is frequently used when mapping access to health care services for this reason; many nominal point maps are used to show locations of hospitals, clinics, and other health care providers.

3.2.5 Summary of Best Practices for Using Thematic Mapping Techniques

Spatial visualization of data through maps is one of the most important functions of GIS in public health (Cromley and McLafferty 2012). Data appearing complicated to lay audiences can become visually striking maps that communicate easily understandable information for any stakeholder (Phillips et al. 2000). However, maps that do not follow cartographic best practices risk communicating false information to their audience (Koch 2005). Public health practitioners will find this chapter helpful in avoiding erroneous map-making. They will be able to use the flow chart, Selecting a Thematic Mapping Method (Figure 2), to safely select an appropriate method based on the nature of their data and read the relevant discussions of thematic mapping methods to guide their map-making process.

CHAPTER 4: SPATIAL VISUALIZATION PRACTICES IN PUBLIC HEALTH STUDIES

The limited number of cartographic best practices dedicated to health hinders Healthy People 2020's advancement in achieving the goal of employing communication, information, and technology strategies to improve health outcomes and health care quality. Many studies have documented how geographic information systems (GIS) have been used to spatially visualize health data; however, few studies have described how visualizations of health-related patterns can be optimized to facilitate understanding (McLafferty 2003; Cromley and McLafferty 2012).

This chapter discusses the use of spatial visualization in public health studies. The first section of this chapter discusses the existing practices of visualization in the public health field. Factors related to lack of insurance will be discussed, in addition to how data on health outcomes, social determinants of health, and access to health care services have been visualized in the public health industry. The second section of this chapter demonstrates how public health practitioners can integrate geographic information systems (GIS), specifically the spatial visualization function, into their studies on health insurance coverage. This section presents a case study on exploring sociodemographic risk factors for being uninsured in Alameda County, California. The case study visualizes the risk factors using best practices for thematic mapping techniques. This case study also provides a potential method for analyzing the characteristics of uninsured individuals to better understand a health care access issue.

4.1 Current Spatial Visualization Practices in Public Health

More and more public health studies are integrating maps to communicate their findings. Many of the studies focus on health outcomes, social determinants of health, access and utilization of health care services, health behaviors, and environmental hazards (McLafferty 2003; Higgs

2005; Higgs 2009; Phillips et al. 2009; Cromley and McLafferty 2012). Unfortunately, there are few cartographic models for mapping health-related data; there is little literature on best practices for the spatial visualization of public health data. This section discusses existing practices of visualization of various themes related to health insurance coverage and other health care access issues: health outcomes, social determinants of health, and access to health care services. While reviewing this section, public health practitioners should note how their colleagues followed best practices in selecting the appropriate thematic mapping technique, as documented in the previous chapter. The techniques chosen effectively present the data given the characteristics of the data; these methods can be compared to the Selecting a Thematic Mapping Method flow chart (Figure 2).

4.1.1 Mapping Health Outcomes

In recent decades, maps of health outcomes have made major contributions to the field of public health with the aid of GIS (Koch 2005; Cromley and McLafferty 2012). The relationship between health and location raises important questions about the impact of spatial differences in health, and visualizing such patterns using GIS can open new opportunities for exploration. Visualizing health outcomes typically involves the distribution of disease incidence and rates, such as the spread of polio cases or the human immunodeficiency virus (HIV).

Health outcomes have been represented using the nominal point mapping technique. One of the most famous epidemiology maps that used point symbols is John Snow's map of cholera outbreak in London. This presents a cluster of cholera cases around a pump and reveals the pump to be the source of the vector-borne disease (Koch 2005; Cromley and McLafferty 2012). In this map, point symbols are used to represent locations where the cases were actually observed. In another nominal point map, the California Department of Public Health (2012) compared infection rates of each hospital to the state average using different colors and shapes for symbols. Each hospital is represented by a point symbol indicating whether its infection rate is lower than, comparable to, or higher than the California average.

Proportional symbol mapping can present ratio data, both total and derived values (Slocum et al. 2009). The Council on Foreign Relations' (2014) interactive map of vaccine-preventable outbreaks demonstrates how proportional symbols can be used to visualize the magnitude of disease frequency. This interactive map also shows base information to provide geographic context of the study area. The Illinois Department of Public Health (2013) also created a proportional symbol map to present colorectal cancer mortality rates by county. This map also used a choropleth mapping technique to present colorectal cancer incidence rates in the same map, which results in a multivariate map.

Like proportional symbol maps, choropleth maps are also commonly used to map ratio data. For example, the Nebraska Department of Health & Human Services (2014) used a choropleth mapping method and a sequential color scheme to present heart disease mortality rates by county in Nebraska. Huston (2011) also used choropleth mapping to visualize five chronic disease rates, which were cancer, asthma, coronary heart disease, diabetes, and stroke, for the state of Maine by public health district. Each chronic disease rate was visualized with a choropleth map, showing which quartile did each public health district fall into; and a final choropleth showed the overall score, taking into consideration how many times each public health district fell into the highest quartiles for chronic disease rates.

There are additional methods to visualize health outcomes that are seen less frequently in public health studies, such as spatiotemporal maps. Spatiotemporal maps show a time series, which can be viewed either through an animated map or collection of static maps (such as small multiples). The Centers for Disease Control and Prevention (CDC) (2012b) presented a series of maps on youth pregnancy rates across fifteen years, allowing users to compare five-year increments. The Council on Foreign Relations' (2014) map is an example of an interactive spatiotemporal map, where users can slide a scroll bar to adjust the time view outbreaks.

4.1.2 Mapping Social Determinants of Health

Data related to social determinants of health are regularly collected by the U.S. Census Bureau through survey programs such as the decennial census, American Community Survey (ACS), and Current Population Survey (CPS). To protect the confidentiality of survey participants, the U.S. Census Bureau releases aggregated data for enumeration units such as states, counties, census tracts, block groups, and blocks (U.S. Census Bureau 2012c). That being said, the data are still useful for visualizing area differences in demographic characteristics and socioeconomic status, both of which influence people's health status (Cromley and McLafferty 2012). Therefore, maps showing distribution of different socioeconomic groups can reveal unique health trends across neighborhoods, cities, counties, or larger geographic areas. This enables public health practitioners to not only see under what conditions people are born, live, grow, study, and work, but also where. For example, income is a variable commonly used in health disparities research because of its widely understood impact on health. Income provides the resources to obtain food, housing, education, and health care; therefore, income has a positive correlation with health status (Jones, Duncan, and Twigg 2004; Pickett and Pearl 2001). Health insurance status, poverty, occupation, educational attainment, nativity, immigration status, and race and ethnicity are other social determinants of health commonly studies (Do et al. 2008; Duncan et al. 2002; Krieger et al. 2003).

Sociodemographic variables, which are strongly related to social determinants of health, are typically presented as derived data through choropleth maps. Examples of sociodemographic choropleth maps include CDC's map on percent of population under sixty-five years without health insurance (2013) and the U.S. Census Bureau's maps on percent population distribution and change (2011a), maps on percent vacancy rates (2011b), and maps on percent Asian population (2012d). The U.S. Census Bureau (2011b) map on change in percentage of vacancy rates is especially engaging because of the appropriate selection of a diverging color scheme. The increase and decrease in vacancy rates are illustrated well with differing hues, each hue representing a positive or negative change.

Dot density and proportional symbol maps have been used to effectively visualize total data using a ratio measurement. For example, the Weldon Cooper Center for Public Service, University of Virginia (2013) used raw totals to create a dot density map on population by race. Each of the racial categories is represented by a different color in the map. As for proportional symbol maps, the U.S. Census Bureau presents raw population totals by county (2011a) and raw totals of the largest Asian group by state (2012d). One noticeable point found in these map is that the proportional symbols are conceptual point data rather than true point data.

4.1.3 Mapping Access to Health Care Services

Health care access has a large impact on health outcomes because the utilization of health care services enables individuals to maintain their health. When there is not access to health care services individuals are more likely to have poorer health (Carrillo et al. 2011). There are some factors that positively or negatively impact on individuals' abilities to seek health care services. The factors are the proximity of patients to providers, patients' transportation and socioeconomic characteristics, population in need of health services, as well as location, type, and quality of services. Visualizing these factors allow patients and public health practitioners to understand where to access services, the characteristics of those services, and where people and communities in need of health services are.

In maps of access to health care services, locations of health services are frequently represented by nominal point symbols. For example, The Alameda Health Consortium (2014) created an interactive web map representing locations of health centers under the Alameda Health Consortium. This map allows users to click an individual point symbol to view detailed information about the health center while referring to base information such as streets and major points of interests.

Some nominal point maps of health care service locations integrate other thematic mapping methods to provide additional information. Some of the health access maps that use point symbols to present health service locations also represent the socioeconomic characteristics of people in need of health care services by choropleth mapping. For example, Gerahty et al. (2010) calculated distances of diabetic patients to their primary care providers and showed the distances in their choropleth maps to represent areas in need for diabetic treatment services. Similarly, Dunlin et al. (2010) created a series of choropleth maps of sociodemographic and health care utilization characteristics with point symbols representing hospitals and clinics to visualize areas in need of primary care services.

4.1.4 Limited Resources for Mapping Public Health Data

Although there are many resources on cartography and visualization, few are targeted specifically at the public health community. CDC and the Agency for Toxic Substances and Disease Registry (ATSDR) formed the Geographic and Geospatial Science Working Group, which collaboratively organized the Public Health and the Cartography Ad Hoc Committee when they recognized the necessity for cartography best practices. The committee's subsequent report, *Cartographic Guidelines for Public Health* (CDC 2012b), is one of few cartographic resources designed to provide best practices for public health purposes.

The report discussed recommendations for symbolizing data using visual variables while paying attention to how the audience comprehends the presented information. For example, the report explained that the selection of color hues and color values influence the audience's interpretation of the map (e.g., red hues are often associated with negativity or urgency and light color values usually represent smaller numbers). Another major topic the report discussed was the selection of thematic mapping methods including choropleth, dasymetric, dot density, graduated symbol, isopleth, cartogram, and spatiotemporal mapping. The report provided suggestions for when each type would be appropriate to use. The report also provided an overview of map elements, scale and generalization, projection, visual hierarchy, and confidentiality issues.

While this report is a useful document for reviewing key concepts in public health mapping, in-depth discussion of existing practices and potential real-world scenarios are not provided. The public health community and practitioners will benefit from an expansion of the suggested best practices for better mapping. This study sought to present a thorough discussion of cartographic best practices tailored to public health practitioners. This study provides a process to select a suitable thematic mapping method for different studies (Figure 2), as well as in-depth discussions on how the thematic mapping methods should be implemented in accordance with conventional cartographic standards. Also included is a case study demonstrating how spatial visualization can be integrated into a study to facilitate better understanding and communication of findings and best practices for mapping data on health outcomes, social determinants of health, and access to health care services.

4.2 Case Study: Integrating Spatial Visualization

In this case study, principal component analysis (PCA) was conducted to identify sociodemographic risk factors for being uninsured in Alameda County, California. The results of the PCA were used to create choropleth maps showing the spatial patterns of these risk factors, following best practices of the choropleth mapping technique.

4.2.1 Study Area

This study used Alameda County, California as the study area because of its diverse population. Alameda County, located in northern-central California and comprising a large area of East Bay in San Francisco Bay Area, is home to more than 1.5 million individuals sprawled across 743 square miles in fourteen incorporated cities and six unincorporated communities (Figure 20). The most urbanized areas in Alameda County are the cities of Oakland, the county seat, and Berkeley. Areas surrounding Oakland and Berkeley, extending eastward and southward are suburban, and the Livermore-Amadore Valley’s heavily agricultural area in the eastern-most side of Alameda County have become more suburban in the past several years (Superior Court of Alameda n.d.). The Port of Oakland, located in west Oakland, is the fourth busiest container port in the U.S.

Figure 20 Study Area: Alameda County, California
Source: U.S. Census Bureau, 2012 Census TIGER/Line Shapefiles

According to the U.S. Census Bureau, as of 2012 there were approximately 540,000 households in Alameda County, 354,000 of which were family households. 34.4 percent of households included at least one person who is under eighteen years of age, and 22.9 percent had at least one person who is sixty-five years and over. The average household had 2.76 individuals, and the average family had 3.39 individuals. 51.0 percent of individuals living in Alameda County were female, and 18.6 percent were under eighteen years. Young adults from nineteen to twenty-five years made up 10.0 percent of the population; remaining adults under sixty-five years made up 56.3 percent while individuals 65 years or over made up 11.5 percent (U.S. Census Bureau 2012b).

64

As one of the most racially and ethnically diverse counties in the nation, Alameda County has over fifty-three languages spoken and no majority racial or ethnic group (Alameda County 2012). Most of the population identifies as White, Asian, or Hispanic or Latino (Table 3).

Table 3 Population by Race/Ethnicity in Alameda County, 2012

Race/Ethnicity	Estimate	Percent
White	515,525	33.6%
Asian	405,569	26.5%
Hispanic or Latino	346,799	22.6%
Black or African American	182,335	11.9%
Two or More Races	61,278	4.0%
Native Hawaiian and Other Pacific Islander	12,924	0.8%
Other Race	4,578	0.3%
American Indian and Alaska Native	4,303	0.3%
Total	1,533,311	100%

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Selected Social Characteristics, Table DP02

In 2012, 30.7 percent of the population were foreign-born and 48.1 percent of the foreign-born population were not naturalized U.S. citizens. More than half of the population (56.8 percent) spoke only English at home, and 8.1 percent of individuals spoke English less than "very well," a common indicator of limited English proficiency (U.S. Census Bureau 2012b).

In 2012 a majority of adults in Alameda County had attended some college or completed college programs. Two-thirds of the population had at least attended some college: 41.2 percent of the population 25 years or over in Alameda County had a bachelor's degree or higher and 25.5 percent attended some college or had an associate's degree. The remaining third of the population had only a high school diploma or equivalent (19.5 percent) or did not have a high school diploma (13.6 percent).

Most adults in Alameda County were part of the labor force; 68.0 percent of the population eighteen years or older were part of the labor force, 10.6 percent of which were unemployed. Of the population who was employed, 41.4 percent worked full-time while 26.3 percent worked part-time. The employed population mostly comprised individuals working in educational services, health care, and social assistance, followed by professional, scientific, management, and administration and waste management services, manufacturing, and retail trade (Table 4).

Table 4 Population by Industry in Alameda County, 2012

Industry	Estimate	Percent
Educational services, health care, and social assistance	163,259	22.6 %
Professional, scientific, management, and administrative and waste management services	117,496	16.2%
Manufacturing	79,648	11.0%
Retail trade	72,774	10.1%
Arts, entertainment, recreation, and accommodation and food services	64,988	9.0%
Finance, insurance, and real estate and rental/leasing	43,646	6.0%
Other services, except public administration	37,265	5.1%
Construction	32,324	5.0%
Transportation, warehousing, and utilities	36,262	5.0%
Public administration	26,573	3.7%
Information	21,901	3.0%
Wholesale trade	21,110	2.9%
Agriculture, forestry, fishing and hunting, and mining	2,454	0.3%
Total employed population 16 years and over	723,700	100%

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Selected Economic Characteristics, Table DP03

In Alameda County, 12.9 percent of the civilian noninstitutionalized population did not have health insurance coverage (noninstitutionalized individuals do not reside in penal or mental facilities or in homes for the elderly or infirmed). Of those who had health insurance coverage, 69.5 percent had private insurance while 25.9 percent had public insurance (U.S. Census Bureau 2012a).

4.2.2 Method of Principal Component Analysis

The PCA used the same American Community Survey (ACS) and Topologically integrated geographic encoding and referencing (TIGER)/Line Shapefile datasets from the U.S. Census Bureau discussed in Chapter 3. The 2010-2012 ACS 3-year Estimates datasets included census-tract level Health Insurance Coverage Status data, which was used to identify risk factors for uninsurance in the PCA. This allowed the author to examine risk factors for being uninsured at a census-tract level in Alameda County. The PCA analyzed variables on age, sex, race and ethnicity, nativity and citizenship status, educational attainment, employment status, work experience, and poverty level. The TIGER/Line Shapefile included census-tract level data for Alameda County.

PCA is a variable reduction procedure used to address the issue of redundancy in a dataset with many variables (Suhr 2005; O'Rourke and Hatcher 2013; SAS Institute Inc. 2013). Redundancy occurs when there are multiple correlated variables that measure the same construct (O'Rourke and Hatcher 2013). Such variable correlation eliminates the use of techniques such as regression to identify relationships between a dependent and many independent variables. PCA allows users to organize the observed variables into smaller sets of redundant variables known as principal components. In this case, PCA was used to identify the set of variables that together are correlated to uninsurance. PCA is conducted through the following processes: 1) conduct initial PCA; 2) determine the number of principal components to keep; 3) rotate to a final solution; 4) interpret rotated solution; and 5) report findings (Suhr 2005; O'Rourke and Hatcher 2013).

This study ran the PCA in JMP Pro 11.0.0 software to identify which sociodemographic variables are associated with the uninsurance variable. The initial PCA was first run to extract several components, which is always equal to the number of variables being analyzed. Not all components in PCA are considered meaningful enough to analyze and interpret in the penultimate step. The number of meaningful components, which account for the most amount of variance or trends seen in the data, can be identified by the scree plot from the outputs of the initial PCA (Figure 21). The scree test method graphs the accumulated amount of variance, or eigenvalues, associated with each principal component ordered from highest value to lowest. Using this method, the number of components to carry forward into the analysis was determined by identifying when the amount of variance accounted for drops to a low level.

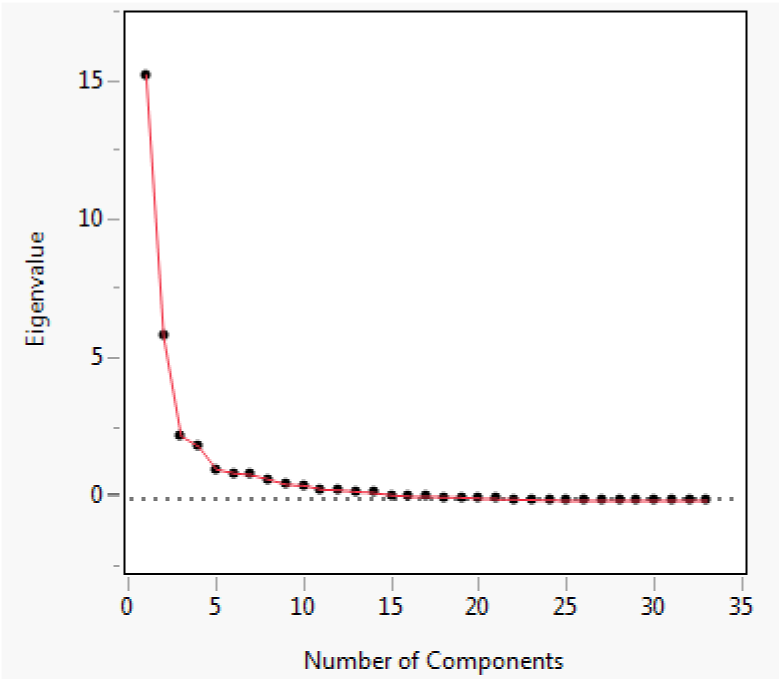


Figure 21 Scree Plot from Principal Component Analysis

Based on the scree plot, three meaningful components were identified to include in the next step, rotating to a final solution. In this step, the loading matrix is rotated to one where each variable is heavily weighted on one component, also known as a factor (Lorenzo-Seva 2003). Coefficient values indicate the weight each variable "loads" on a component, or contributes to the variance of the component (O'Rourke and Hatcher 2013; SAS Institute Inc. 2013). Once the rotation was complete, the PCA results were analyzed, interpreted, and then visualized using the choropleth mapping technique.

4.2.3 Results and Discussion of Principal Component Analysis

The PCA generated a table of rotated factor loadings which was interpreted to identify which sociodemographic variables correlated with the uninsurance variable (Table 5). The rotated factor loading showed coefficient values for each variable of each component. The uninsurance variable had significant loading on the second component, indicated by its high coefficient value of 0.887631. Other high loadings on the second component indicated seven variables that strongly correlate with the uninsurance variable. These variables can therefore be considered risk factors for uninsurance. These variables included adults without a high school degree (0.897138), individuals identifying as Hispanic or Latino (0.877299), individuals living under 138 percent of the FPL (0.827015), individuals identifying as Other Race (0.814032), individuals living between 138 and 199 percent of the FPL (0.782964), foreign-born individuals who have not been naturalized (0.754004), and individuals who are part of the labor force but not employed (0.708006). These variables and their respective loadings are bolded in Table 5. The correlation between uninsurance and each of the risk factor variables can be seen in Figure 22. The scatterplots show a positive correlation between uninsurance and each of the other variables, further demonstrating that these sociodemographic characteristics are indeed risk factors for being uninsured.

The findings of this case study indicated that adults without a high school degree, individuals identifying as Hispanic or Latino and Other Race (per the U.S. Census Bureau's race categories), immigrants without citizenship, individuals living under 200 percent of the FPL, and unemployed adults are at risk for being uninsured. All the risk factors identified in this case study have been documented as risk factors for being uninsured in previous studies. However, some of the characteristics that have been established as correlates of uninsurance were not identified in the PCA, such as adulthood (especially young adulthood). The lack of adulthood being identified as a correlate of uninsurance may be a result of Alameda County's adult population having protective factors against uninsurance, such as higher educational attainment; in Alameda County nearly half of adults have a bachelor's degree or higher and are therefore more likely to be employed with a job that provides employer-sponsored health insurance plans. Individuals who come from working families or are employed in blue-collar jobs are also likely to be uninsured according to previous research, but the PCA did not include these variables. In further studies, a PCA study can include these variables to determine if these populations would also be identified as having potential risk factors of uninsurance.

Table 5 Rotated Factor Loading from Principal Component Analysis

Variable	**Factor 1**	**Factor 2**	**Factor 3**
Insurance Status			
Uninsured	0.194748	**0.887631**	-0.026517
Age			
Under 18 years	0.586252	0.472525	0.358719
18 to 64 years	0.869661	0.345964	0.246135
65 years and older	0.673160	-0.121346	0.233113
19 to 25 years	0.377808	0.397459	-0.145720
Sex			
Male	0.841272	0.362242	0.343566
Female	0.882306	0.349506	0.274900
Race/Ethnicity			
American Indian and Alaska Native	0.038241	0.321647	0.135579
Asian	0.313643	0.002508	0.851998
Black	0.038577	0.503235	-0.382903
Native Hawaiian or Pacific Islander	0.039527	0.402630	0.139320
White	0.843758	-0.150422	-0.195841
Other Race	0.074587	**0.814032**	0.109784
Two or More Races	0.534505	0.244407	0.193327
Hispanic or Latino	0.205356	**0.877299**	0.025298
Nativity and Citizenship Status			
Native born	0.968737	0.132160	-0.081407
Foreign born	0.321779	0.531850	0.750870
Naturalized	0.380030	0.118407	0.851550
Not naturalized	0.146744	**0.754004**	0.377796
Educational Attainment			
Less than high school diploma	0.057286	**0.897138**	0.110311
High school graduate	0.461508	0.629955	0.275145
Some college	0.753498	0.244083	0.090916
Bachelor’s degree or higher	0.586312	-0.589285	0.288580
Employment Status			
Labor force	0.883653	0.221274	0.321692
Employed	0.883653	0.121897	0.343173
Unemployed	0.408363	**0.708006**	0.027948
Not in labor force	0.716185	0.314348	0.098796
Work experience			
Worked full-time in the past 12 months	0.813377	0.029473	0.447408
Did not work full-time in the past 12 months	0.770989	0.326515	-0.050089
Did not work	0.716061	0.452341	0.157589
Poverty level in the past 12 months			
Under 138% FPL	0.094464	**0.827015**	-0.226348
138% to 199% FPL	0.236623	**0.782964**	-0.006883
200% FPL or higher	0.824836	-0.132550	0.477513

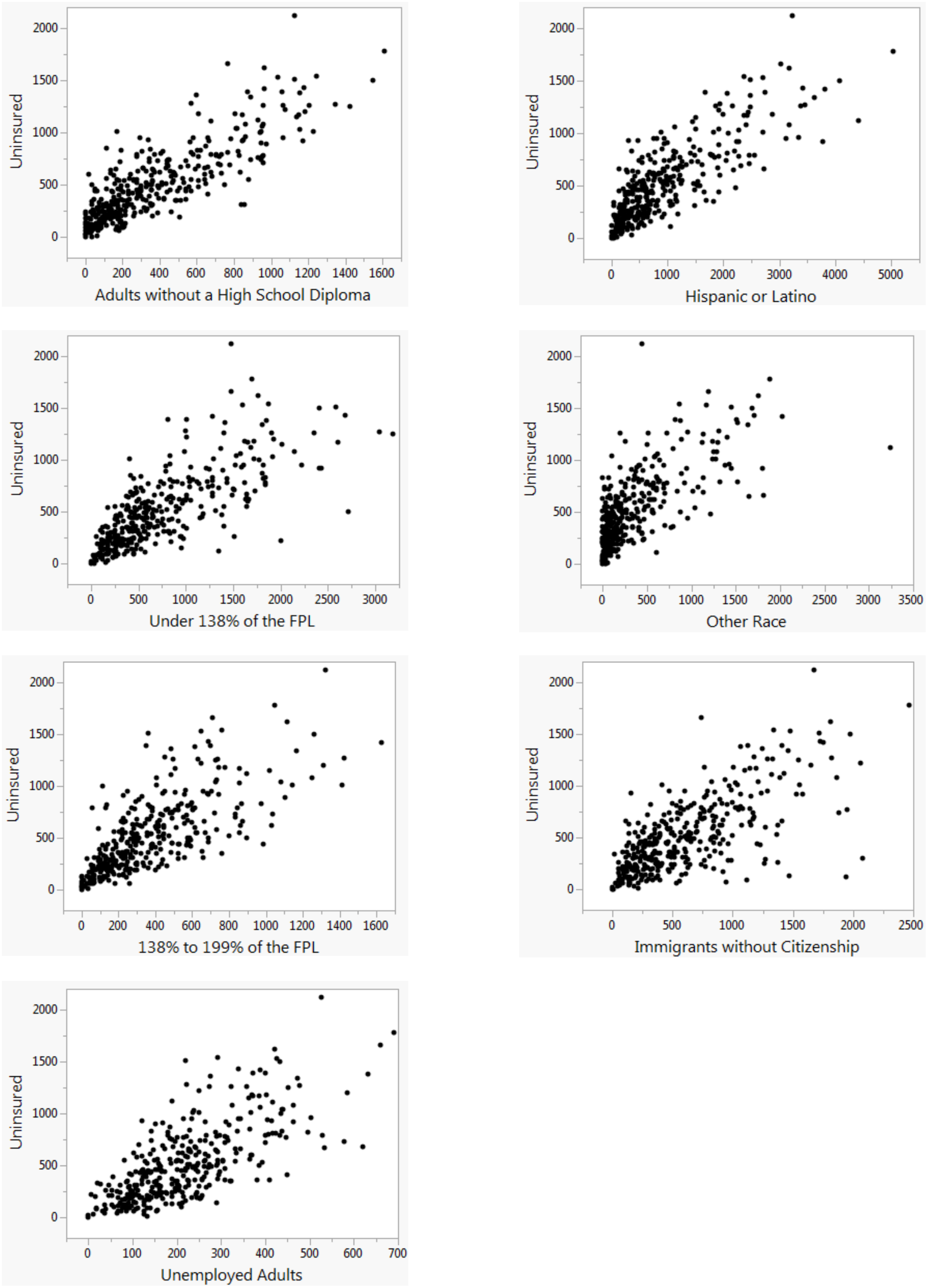


Figure 22 Correlation between Uninsurance and Risk Factors in Alameda County, 2012

4.2.4 Visualizing Results of Principal Component Analysis

The PCA results can be used to facilitate deeper understanding of the populations at risk for uninsurance through spatial visualization. The percentage of individuals with each risk factor was mapped on the census tract level for Alameda County using the choropleth technique. Each variable was visualized using two maps to show where there are high and low concentrations of these individuals, likely indicating where there are pockets of communities with high and low rates of uninsurance. The first map of each variable uses the same data classification system; the classes were manually set using equal intervals, where each map has the same class ranges (regardless of the variable's data range) so that variables can be compared to one another. The second map uses the quantile method for data classification, where each of the six classes has approximately one-sixth of the total number of counties. Although the class ranges are different across all the risk factor maps and can therefore not be directly compared, using the quantile method allows the audience to detect differences in finer detail for each variable's spatial pattern.

Uninsured individuals live in the highest concentration in west Oakland, followed by Emeryville, and east and central Oakland (Figure 23A). However, almost all throughout the western corridor of Alameda County there were high concentrations of uninsured individuals, with the exception of Union City and more southern areas of Newark and Fremont. Cities with moderate percentages of uninsured included parts of Ashland, Hayward, and San Lorenzo; lower percentages can be seen in the city of Alameda, Castro Valley, Fairview, and San Leandro. Dublin, Pleasanton, and Livermore had mostly low percentages of uninsured. Figure 23B, which uses the quantile classification scheme, shows similar spatial patterns and highlights the lowest percentage of uninsured throughout Piedmont and Fremont, where most census tracts have no more than 5 percent who are uninsured.

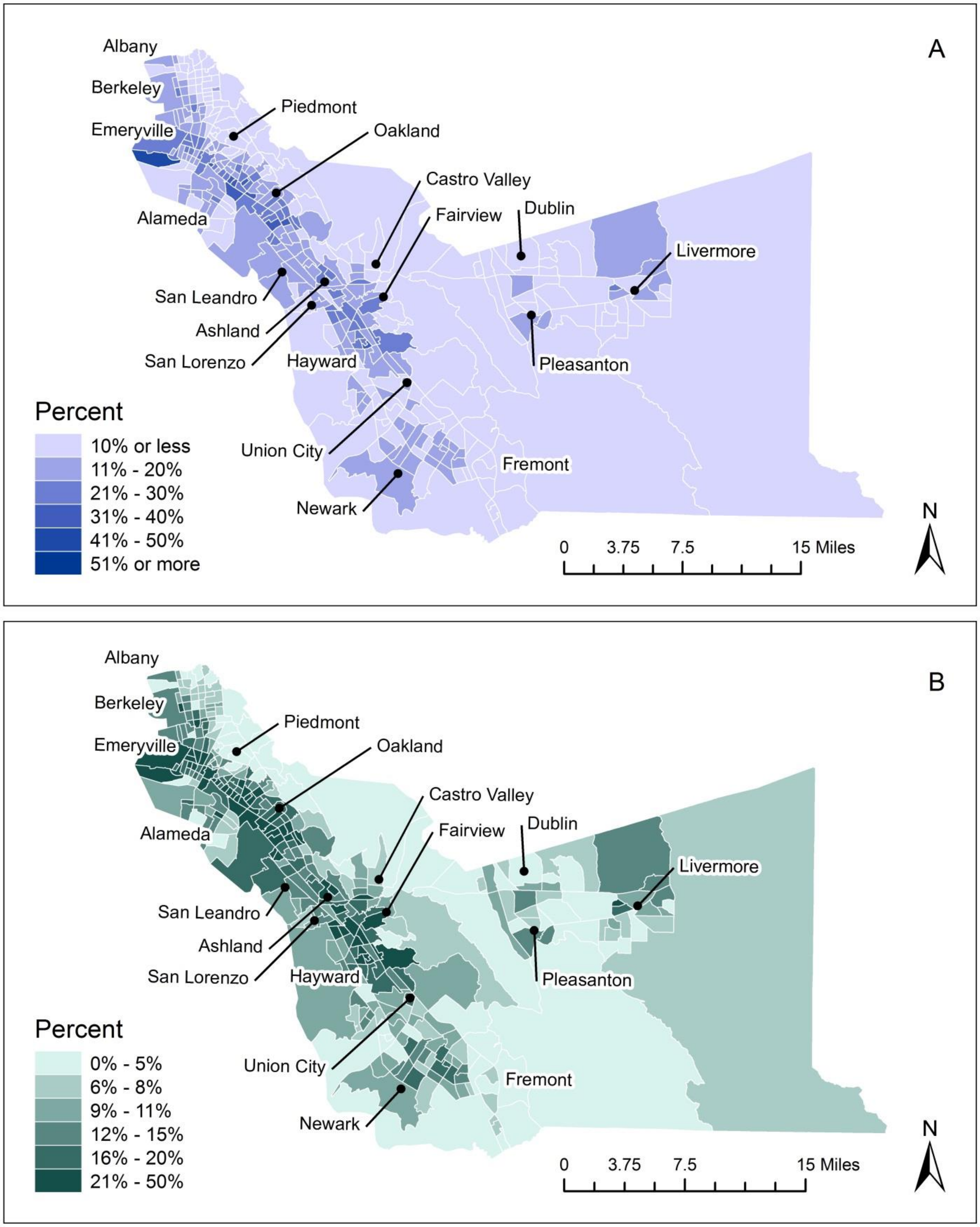


Figure 23 Percentage of Uninsured Using Manual (A) and Quantile (B) Data Classification Schemes in Alameda County, 2012

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

The sociodemographic risk factor that loaded most strongly on the second component along with uninsurance was adults (25 years or older) without a high school diploma (Figure 24A). Individuals with this characteristic were distributed similarly to uninsured individuals but with less prevalence. Again, the western corridor of Alameda County had consistently the highest percentage of adults who did not complete a high school education. These adults mostly resided in Emeryville, Ashland, and east, west, and central Oakland; according to Figure 24B, between 18 and 41 percent of these areas did not have a high school diploma. The areas with the next highest concentration of adults without a high school degree were San Leandro, San Lorenzo, Hayward, and Union City. There were very low percentages living in Albany, Berkeley, Fremont, and especially in Dublin, Piedmont, and Pleasanton. In Figure 24A, census tracts in these cities had no more than 10 percent of adults without a high school diploma. Figure 24B shows that Piedmont and parts of Dublin and Pleasanton had percentages as low as 0 to 2 percent.

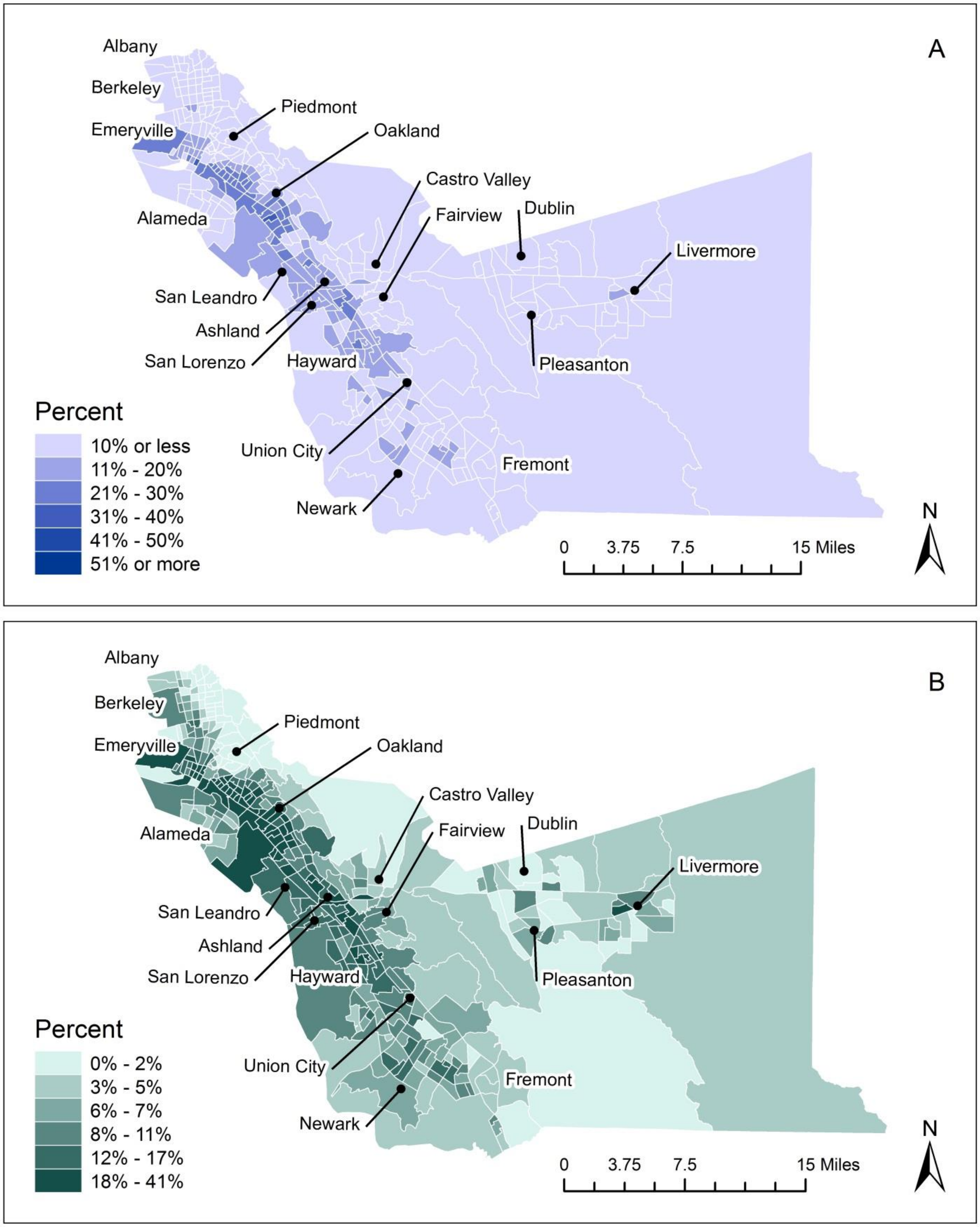
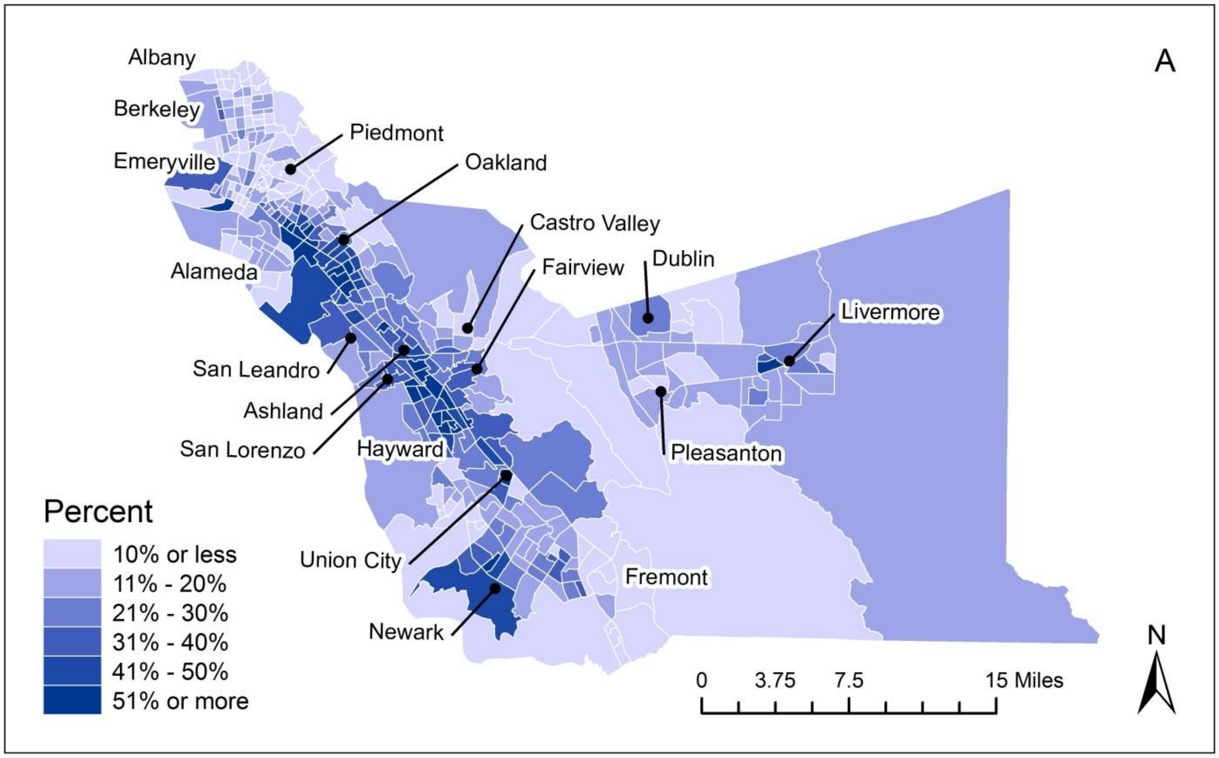


Figure 24 Percentage of Adults without a High School Diploma Using Manual (A) and Quantile (B) Data Classification Schemes in Alameda County, 2012

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

Being Hispanic or Latino was another sociodemographic characteristic highly associated with being uninsured. Individuals who identify as such lived in high concentrations all throughout the western corridor of Alameda County and to a lesser degree eastern Alameda County (Figure 25A). East and south Oakland, San Leandro, San Lorenzo, Hayward, Newark, and Livermore had census tracts with the highest percentage of Hispanics or Latinos. Emeryville, Ashland, Fairview, and Union City contained areas of moderately high concentrations, while Piedmont, Pleasanton, Berkeley, and Albany had the lowest concentrations. Figure 25B shows a very similar spatial pattern, with some small differences in the color pattern. While Figure 25A's lowest class groups values of 0 to 10 percent together, Figure 25B's lowest class only includes values up to 7 percent. The smaller class range in Figure 25B's lowest class allows the audience to better identify which census tracts have the lowest concentration of Hispanics or Latinos.



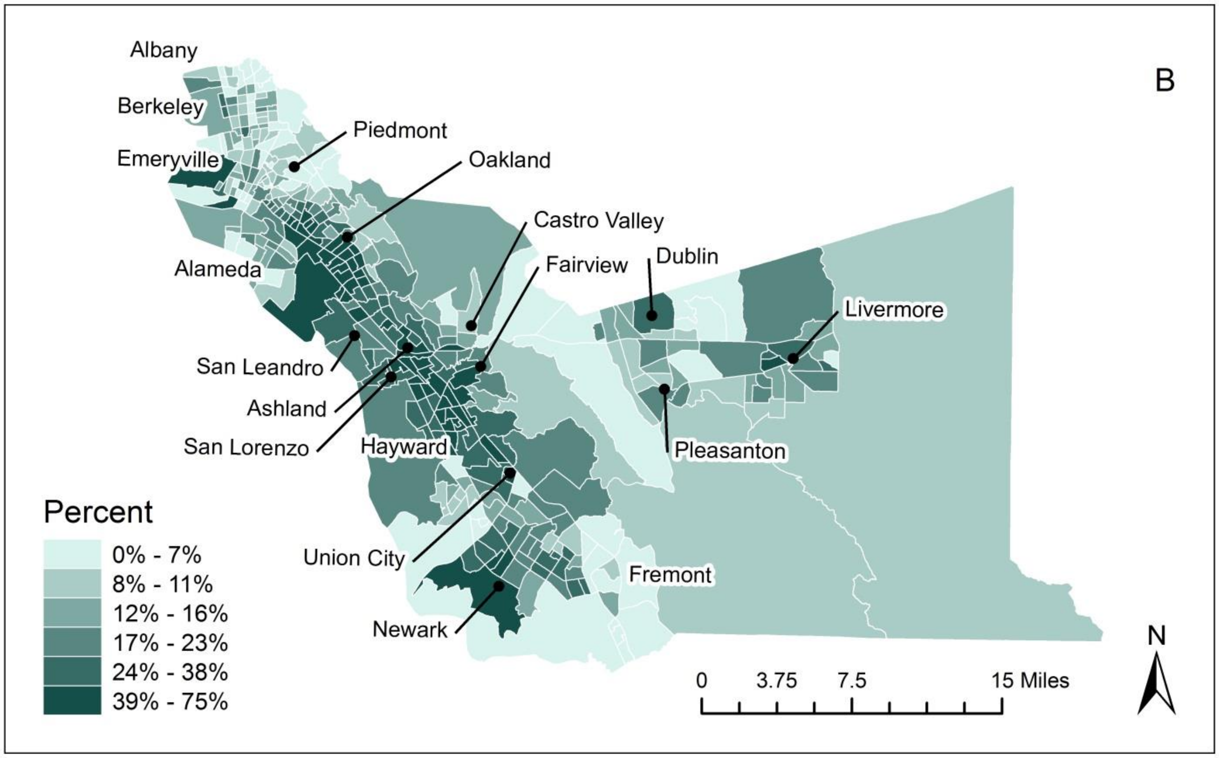


Figure 25 Percentage of Hispanic or Latino Using Manual (A) and Quantile (B) Data Classification Schemes in Alameda County, 2012

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

The risk factor with the next highest coefficient was living under 138 percent of the FPL, and individuals with this risk factor were very similarly distributed to uninsured individuals (Figure 26A). Individuals in high poverty lived in the highest concentrations in the northwest side of Alameda County, stretching from Berkeley to Oakland, especially in Emeryville and east and south Oakland. South of Oakland, from San Leandro to Newark, contained mostly moderate percentages of these individuals, although Union City had more areas of low poverty rates. Dublin and Livermore, like Union City, contained few census tracts with moderate concentrations of individuals living under 138 percent of the FPL. Figure 26B shows a slightly different color pattern because of the different class ranges. While values up to 10 percent make up the first class in Figure 26A, they make up the first two classes in Figure 26B. However, Figure 26A uses four classes for values above 20 percent, while Figure 26B approximately uses two classes. The impact of the different class ranges results in the implication that there is a higher concentration of individuals living under 138 percent of the FPL throughout western and eastern Alameda County in Figure 26B, compared to Figure 26A.

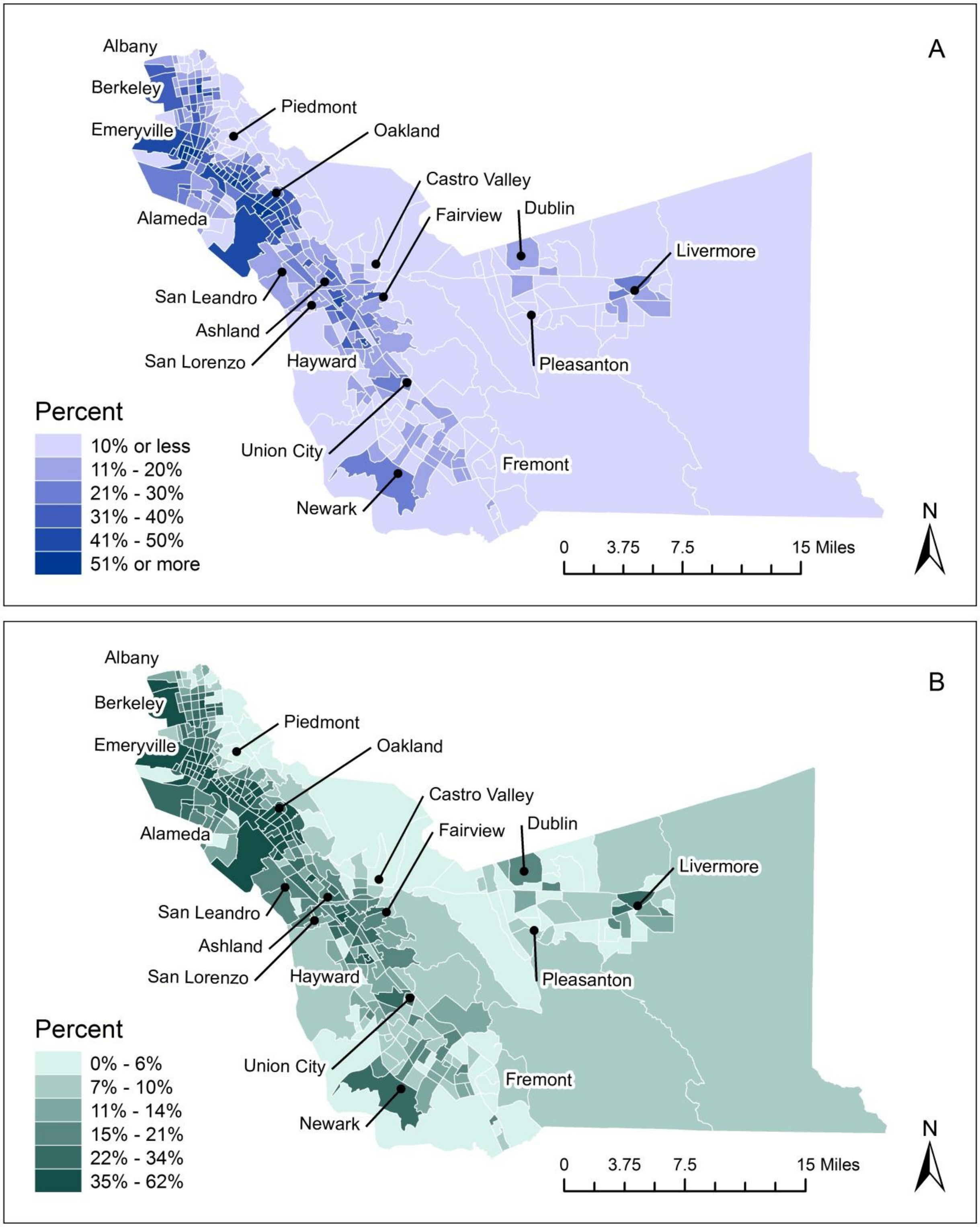


Figure 26 Percentage of Individuals Living Under 138% of the Federal Poverty Level Using Manual (A) and Quantile (B) Data Classification Schemes in Alameda County, 2012
Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

Individuals identifying as Other Race, another risk factor associated with uninsurance, lived in much lower concentrations than uninsured individuals (Figure 27A). They mostly lived in Emeryville, west and east Oakland, San Lorenzo, Hayward, and Newark. Most cities had census tracts with low percentages of these individuals. Like individuals who were uninsured or possessed risk factors of uninsurance, the areas with the highest concentration of individuals who identified as Other Race also resided on the west side of Alameda County. Figure 27B presents a very different color pattern from Figure 27A because of the difference in class ranges. In Figure 27B values up to 15 percent are grouped into five classes, whereas Figure 27A groups values up to 20 percent into only two classes. It should be noted that Figure 27B's highest class includes values from 16 to 55 percent, and the resulting color pattern may mislead the audience. Without a careful examination of Figure 27B's legend, the audience may assume the color pattern implies a high concentration of Other Race individuals.

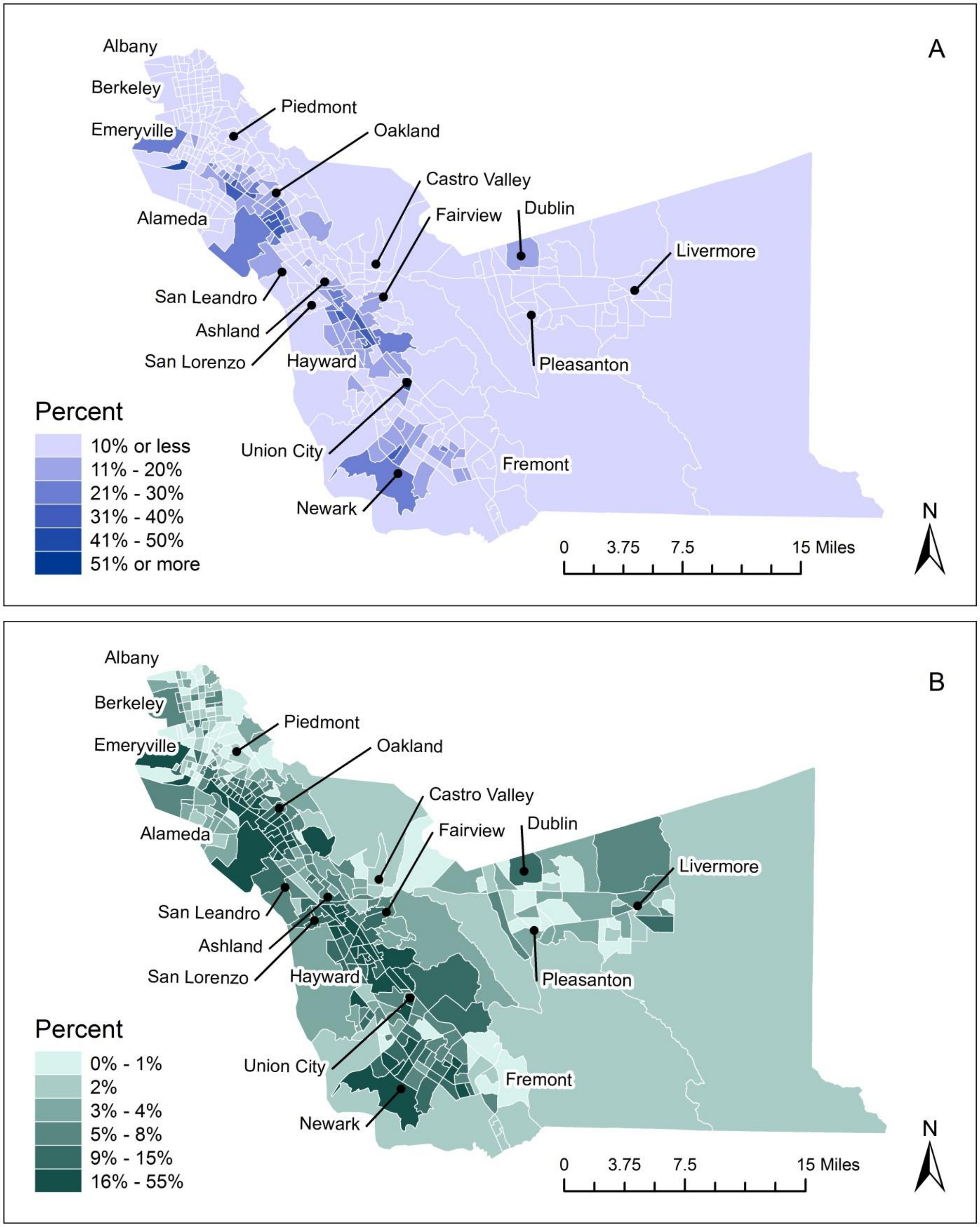


Figure 27 Percentage of Other Race Using Manual (A) and Quantile (B) Data Classification Schemes in Alameda County, 2012

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

There were low concentrations of individuals that lived in between 138 and 199 percent of the FPL (Figure 28A), despite the variable being strongly loaded in Component 2 in the PCA. The areas that had the highest percentages (but still relatively low) stretched throughout western Alameda County but were mostly in the eastern parts from Oakland to Hayward (Figure 28A). Newark, Fremont, Pleasanton, and Livermore had some areas with moderately low percentages of individuals living in moderate poverty. Similar to Figures 26 and 27, Figure 28 shows a very different color pattern between the manually set equal intervals and quantile maps because of the different class ranges. Figure 28A groups values up to 10 percent into the lowest class, whereas Figure 28B groups the same values into the lowest four classes. The data classification results in Figure 28A shows minimal variety in color pattern and Figure 28B shows much variety. Without reading the legend, Figure 28B's color pattern implies that individuals living between 138 and 199 percent of the FPL are distributed well throughout Alameda County.

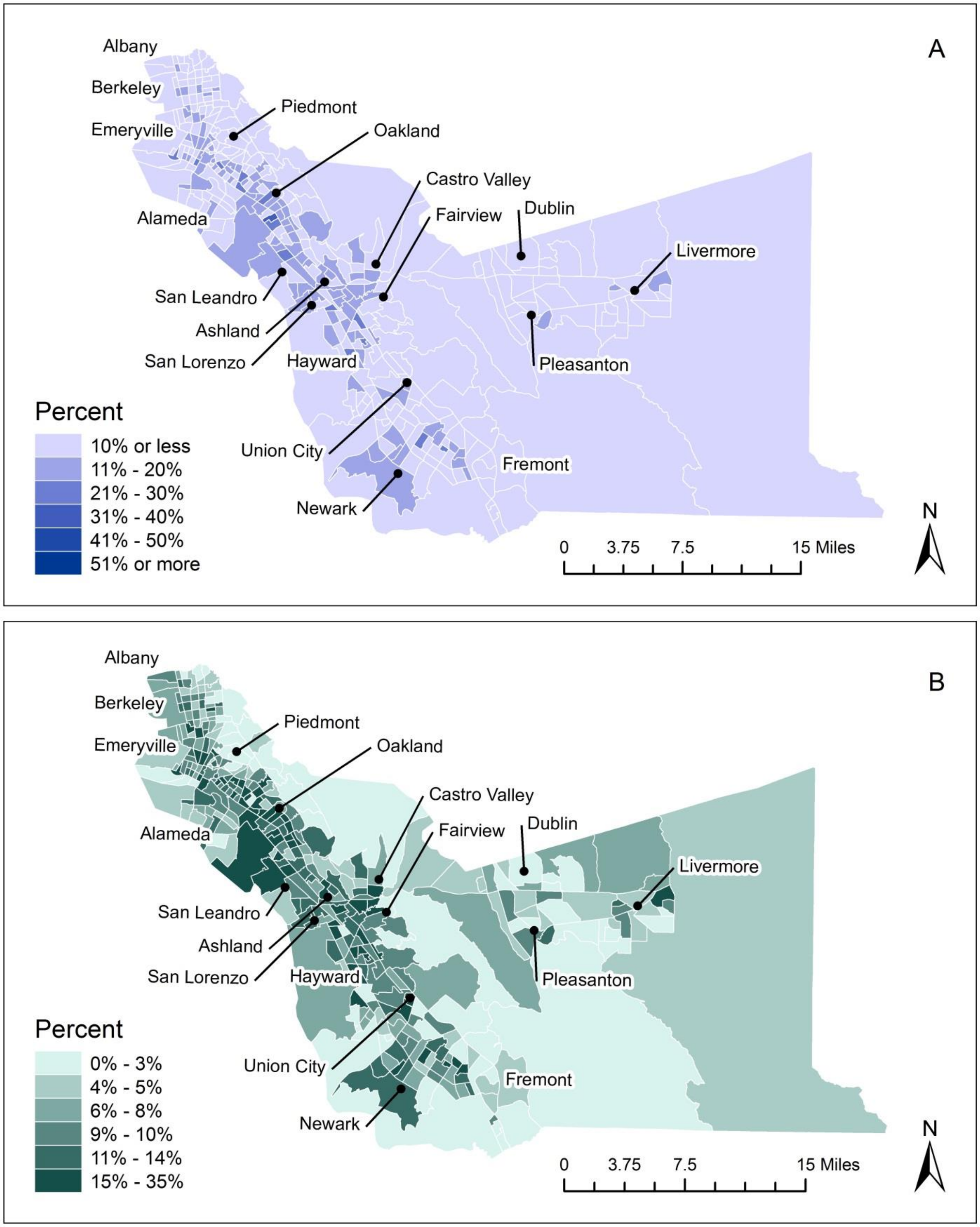


Figure 28 Percentage of Individuals Living between 138% and 199% of the Federal Poverty Level Using Manual (A) and Quantile (B) Data Classification Schemes in Alameda County, 2012
Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

Immigrants who have not been naturalized did not have as high of a coefficient in the PCA as the other risk factors but were still similarly distributed to individuals who were uninsured. All throughout the western corridor of Alameda County were moderate and high concentrations of immigrants without citizenship, with the highest concentrations in Albany and Berkeley, followed by Emeryville, central and east Oakland, San Lorenzo, Hayward, Fremont, and Dublin (Figure 29A). Cities with moderate percentages of immigrants without citizenship included San Leandro, Ashland, Union City, Newark, Pleasanton, and Livermore. Figure 29B presents a similar map but shows a higher number of census tracts falling into the highest class (24 to 56 percent) in Berkeley, Oakland, and Fremont. The color patterns between the two maps are mostly different for the lightest color values because Figure 29A uses one class for values up to 10 percent whereas Figure 29B uses approximately two classes. This results in Figure 29B appearing to show a higher distribution of immigrants who have not been naturalized. The distribution of individuals with this particular risk factor was more wide-spread than individuals who were uninsured, especially throughout southwest Alameda County.

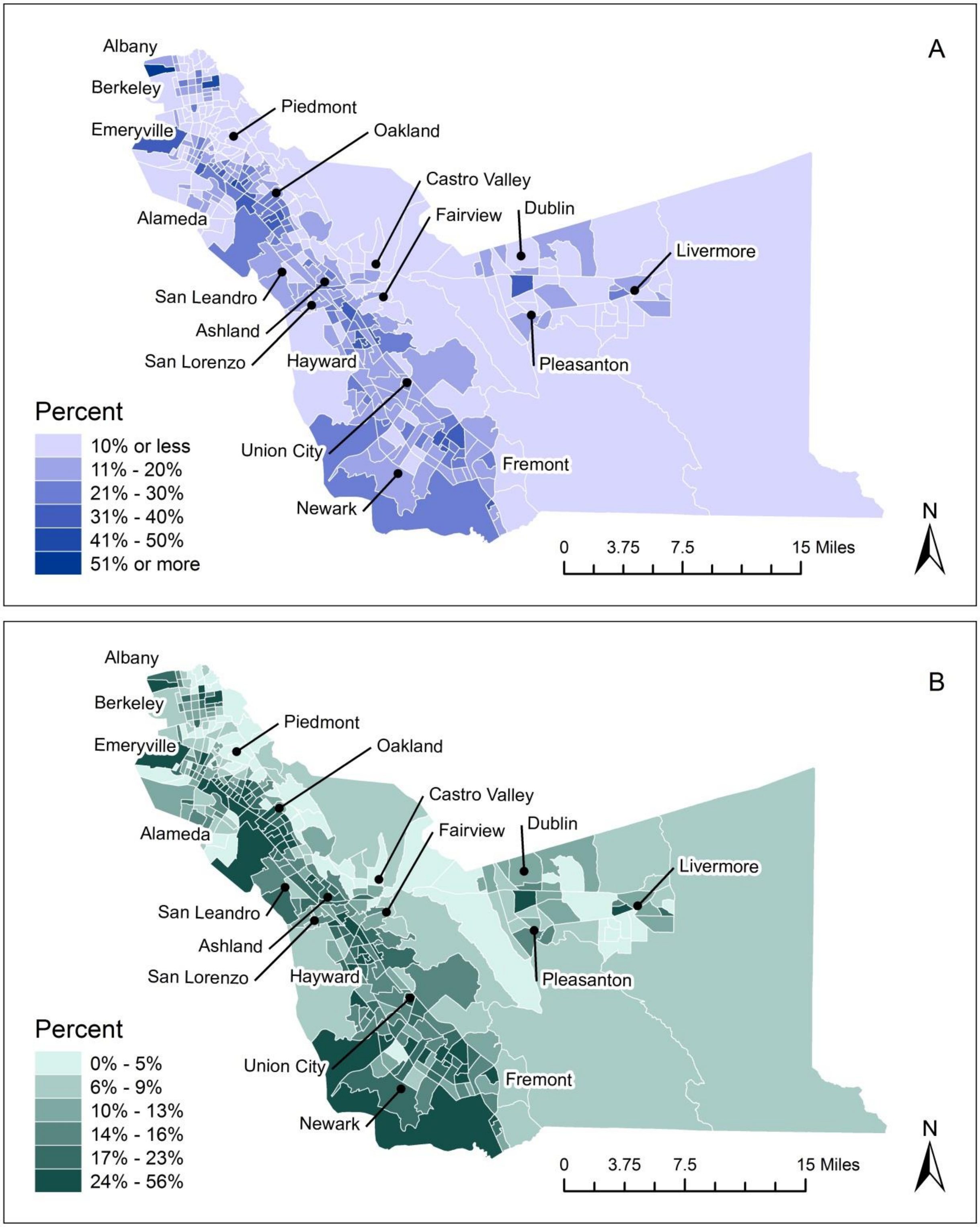


Figure 29 Percentage of Immigrants without Citizenship Using Manual (A) and Quantile (B) Data Classification Schemes in Alameda County, 2012

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

The final risk factor of uninsurance, unemployment among adults, correlated the least with uninsurance, compared to the other risk factors. Very few areas in Alameda County had moderate or high rates of unemployment, and nearly all census tracts throughout the county had unemployment percentages of 10 percent or less (Figure 30A). The only census tracts that had higher concentrations were found in central and east Oakland, Fairview, and Hayward. All of these census tracts had unemployment percentages between 11 and 20 percent, still very low in concentration compared to the other risk factors. Figure 30B shows an extremely different color pattern, showing a data range of 0 to 15 percent from the lowest to highest class (whereas Figure 30A uses only two classes to group the same values). While Figure 30A shows a dull and mostly monotonous color pattern, Figure 30B shows much more variety and allows the audience to see more detail in how different cities in Alameda County have different percentages of unemployed. Half of the census tracts in Alameda County had between 0 and 5 percent of unemployed adults, and most of these census tracts are in Piedmont, Fremont, Dublin, Pleasanton, and Livermore. Approximately one-sixth of census tracts fell into the highest class, with 8 to 15 percent of unemployed adults; most of these census tracts are in Oakland, Fairview, and Hayward. Compared to individuals who were uninsured or had the other risk factors, unemployed adults were distributed much less throughout Alameda County.

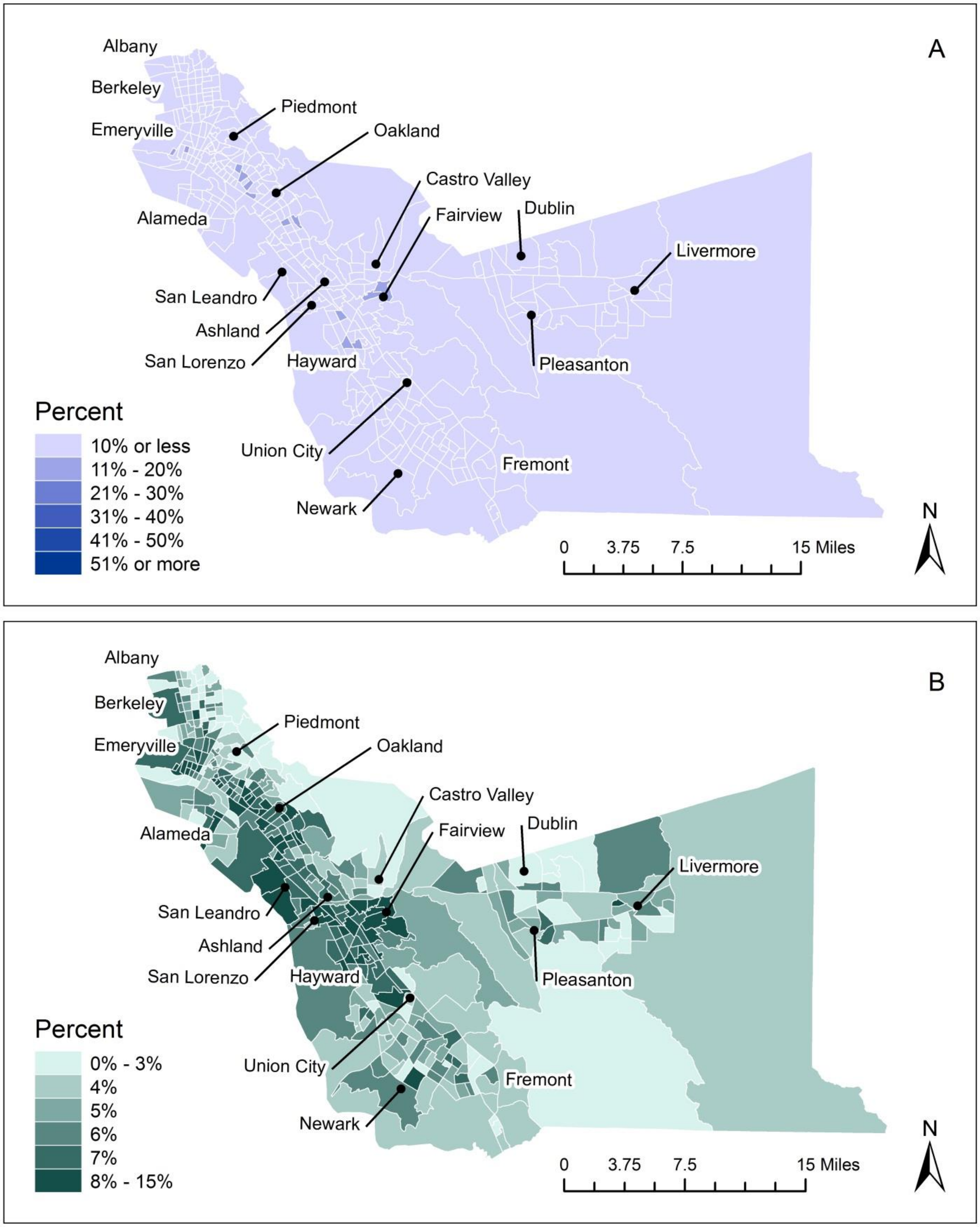


Figure 30 Percentage of Unemployed Adults Using Manual (A) and Quantile (B) Data Classification Schemes in Alameda County, 2012

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

Individuals with risk factors for uninsurance were mostly distributed in similar spatial patterns as individuals without insurance. While individuals who were uninsured or had risk factors of uninsurance were almost always mostly concentrated in western Alameda County, there were some similarities and differences in distribution across the different risk factors when looking at a composite, or small multiple, of all uninsurance and risk factor maps that used the same manual data classification scheme (Figure 31). Adults without a high school diploma and individuals living under 138 percent of the FPL were spatially distributed in the most similar patterns to uninsured individuals. Individuals who identified as Other Race or lived between 138 and 199 percent of the FPL were less distributed than uninsured individuals, while Hispanics or Latinos and immigrants without citizenship were more distributed. Unemployed individuals had the most dissimilar pattern, where very few areas in Alameda County didn’t have low concentrations, likely because of the low rate of unemployment.

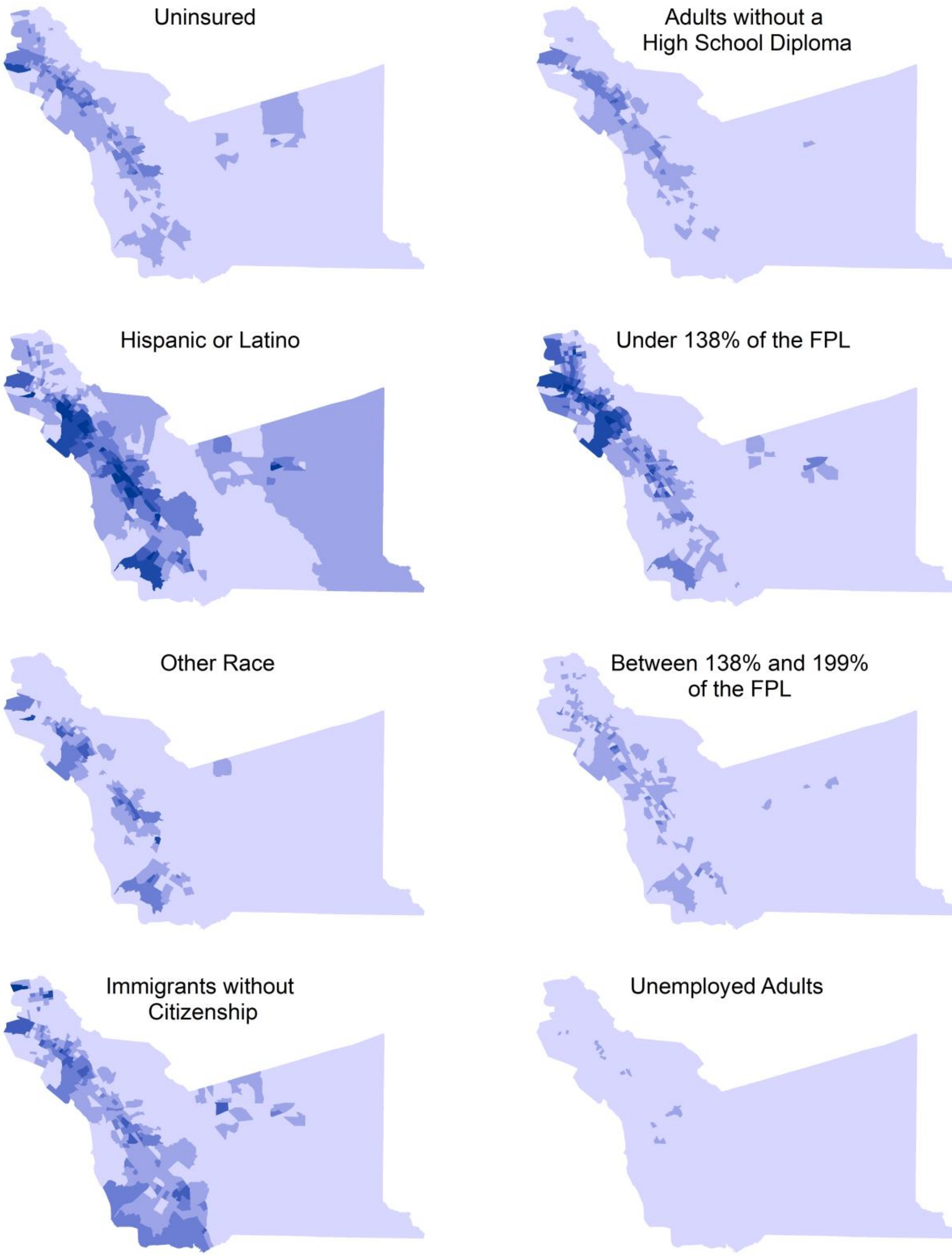


Figure 31 Spatial Pattern of Uninsurance and Risk Factors Using Manual Data Classification Scheme in Alameda County, 2012

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

When examining the small multiple using the quantile classification scheme, the spatial patterns of percentage of individuals who are uninsured and have risk factors are very similar (Figure 32). Although each map has its own unique set of classes, the small multiple shows that the variables are geographically associated and census tracts with the highest percentage of uninsurance and risk factors are all concentrated on the western side of Alameda County, mostly in the Oakland area.

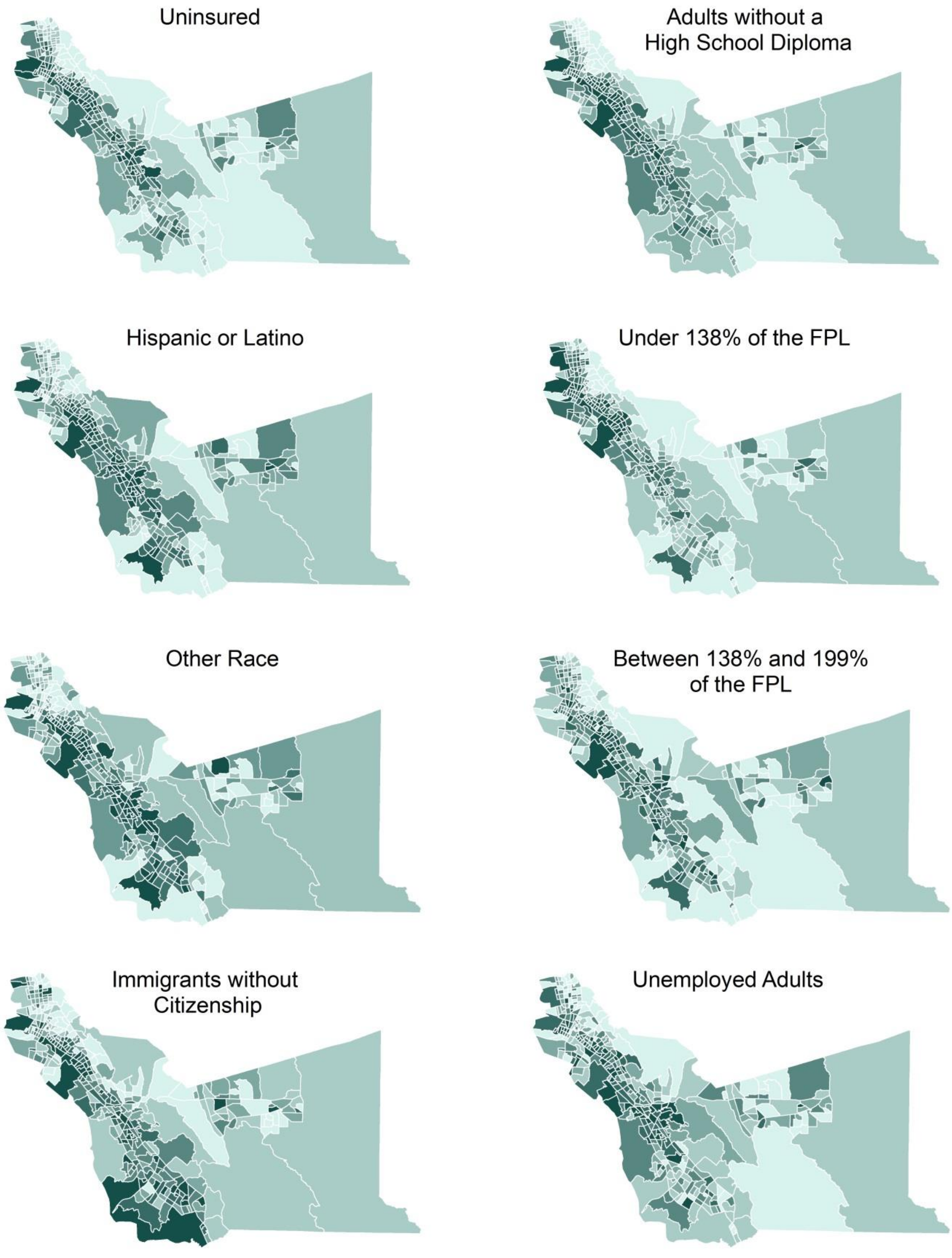


Figure 32 Spatial Pattern of Uninsurance and Risk Factors Using Quantile Data Classification Scheme in Alameda County, 2012

Source: U.S. Census Bureau, 2010-2012 American Community Survey 3-year Estimates, Health Insurance Coverage Status, Table S2701

This case study demonstrated how spatial visualization can be integrated into a study to facilitate understanding and communication of findings. The data characteristics, potential thematic mapping methods, data classification methods, and color schemes were all carefully considered in order to create maps that effectively communicated the results of the PCA.

The choropleth mapping technique was chosen to visualize the concentration of risk factors throughout Alameda County by census tract. The author considered the nature of the data and chose the appropriate mapping technique using the flow chart, Selecting a Thematic Mapping Method (Figure 2). The dataset used ratio measurement and included total value attributes of each risk factor for enumeration; since the attributes were in total value form, which is inappropriate for using the choropleth method, the variables were standardized using the total population attribute as the denominator to calculate percentages.

Two sets of maps were created using manually set equal interval and quantile data classification schemes in order to adequately compare the spatial patterns of the variables, as well as explore the spatial patterns in fine detail. Each method offered advantages and disadvantages for conveying spatial patterns. In the manually set equal interval maps, each map had identical class ranges, despite the difference in data ranges. By using the same classes, the spatial pattern of the percentages can be compared across the uninsurance and risk factor variables. The audience can easily compare and identify census tracts that have the lowest or highest percentages. However, risk factors that had small data ranges, such as the percentage of unemployed adults, show little variation in the spatial pattern (Figure 30A).

The quantile maps address this issue, showing more variability within each map by grouping the same number of counties into each class. Variability can be seen even in the map of unemployed adults (Figure 30B). Here, the audience is able to identify nuances in the color pattern and easily discern areas in the map that have the lowest or highest percentage values. However, the quantile maps can mislead the audience into believing that there are similar concentrations of risk factors if the audience does not carefully read the legend. For example, the maps of percentages of Hispanic or Latino (Figure 25B) and individuals living between 138 and 199 percent of the FPL (Figure 28B) have similar color patterns, where census tracts in the highest class (i.e., having the darkest color value) inhabit the same areas in both maps; a closer inspection of the legends reveals that the highest class in the Hispanic or Latino map contains values of 39 to 75 percent whereas it contains values of 15 to 35 percent in the map of individuals living between 138 and 199 percent of the FPL.

To balance the advantages and disadvantages of these two data classification methods and to further facilitate the audience's ability to compare the maps, small multiple maps were created. Small multiples display many variables at once and allow the audience to easily scan each map and compare the spatial patterns (Gemignani 2010). The small multiple of maps using the manually set equal intervals data classification scheme shows areas that overlap in high percentages of the uninsurance and risk factor variables, while the small multiple of maps using the quantile scheme shows areas where there are the lowest and highest concentrations of each variables, essentially how they are geographically associated. Each data classification method can be useful, but the cartographer must consider the purpose of visualizing data and who the audience is in order to select the best method.

Cartographers should also consider how different data classification methods impact the use of small multiples. In the maps using manual data classification, there are no borders to distinguish census tracts from one another (Figure 31). Unlike the maps in Figures 23-30, the small multiple maps in Figure 31 are small enough that using borders would prevent the audience from clearly seeing the smallest census tracts. However, the small multiple maps using the quantile data classification method include a white border because of the purpose of this data classification method (Figure 32). Using the quantile method allows the audience to approximate the number of census tracts being grouped into each class. Without the white border to distinguish each census tract, the audience may mistakenly believe several census tracts of the same class that are next to one another to form one census tract.

The final design of the maps involved setting the color scheme and adding additional map elements. Because the data are unipolar, sequential color schemes were used, with lower values of color representing lower percentages and higher values of color representing higher percentages. The manual data classification maps all used the same blue color ramp to allow for easy comparison; if each map used a unique color, the audience may find it challenging to compare percentages. Similarly, the quantile maps all used the same color of green. The manual and quantile maps used different colors so that the audience would avoid confusing which map used which data classification scheme. A legend was added to the maps for the audience to understand what the color patterns represent, although the small multiples do not use a legend because of the intent to communicate broad spatial patterns. Lastly, city labels, a scale bar, and north arrow were added to provide geographic context to Alameda County, which allows the audience to better understand the maps.

These maps demonstrated the impact of spatial visualization on understanding and communicating findings. Whereas the tabular results of the PCA only showed the strength of correlation between uninsurance and the risk factors for all of Alameda County, mapping the results showed spatial patterns on the census tract level. The census tract level data of percentage of uninsurance and risk factors allow public health practitioners to explore what neighborhoods contain higher concentrations of communities with risk factors, while statistical outcomes only provide a broad overview of what the risk factors of uninsurance are for Alameda County. Public health practitioners can also see what variables are associated geographically. The choropleth mapping technique applied in this study clearly conveys the patterns of high uninsurance and risk factor rates throughout western Alameda County, and the visualization methods of this study could be replicated and adapted to future studies of health insurance coverage and populations at risk for being uninsured.

Public health practitioners can use this case study as a model for understanding how to integrate spatial visualization into studies, especially those addressing health insurance coverage barriers and other health care access issues. Based on the findings of this case study, health programs or initiatives can strategically prioritize geographic areas with communities that should receive more outreach and support to increase health insurance coverage.

CHAPTER 5: BEST PRACTICES AND DISCUSSION ON MAPPING PUBLIC HEALTH DATA

As part of Healthy People 2020's goals to improve health and eliminate health disparities, integrating communication and health information technology was prioritized to elevate community health (U.S. Department of Health, Education and Welfare 1979; Healthy People 2013). Communication and information technology are critical in facilitating the processes in which public health stakeholders understand and utilize health information, which directly affects health decisions and outcomes (Healthy People 2013). Public health practitioners can use geographic information systems (GIS) to respond to Healthy People 2020's aspiration to harness the potential of information technology for impacting health. Spatial visualization, a distinct advantage that GIS offers, enables public health practitioners to effectively present geographic phenomenon and detect patterns in maps that may remain otherwise undiscovered in tabular form (Cromley and McLafferty 2012).

This study sought to demonstrate how GIS can be used to better understand and communicate health data through spatial visualization. This study aimed to 1) document best practices for visualizing public health data using thematic mapping techniques and 2) demonstrate how spatial visualization can be integrated into public health studies to facilitate understanding and communication of findings.

Public health studies have used GIS to study where people live and how this affects their health (McLafferty 2003; Cromley and McLafferty 2012). Studies have visualized a wide range of applications such as health outcomes, social determinants of health, and access to health care services (McLafferty 2003; Higgs 2005; Higgs 2009; Phillips et al. 2009; Cromley and McLafferty 2012). The most common thematic mapping methods used for visualization are choropleth mapping, followed by proportional symbol and dot density mapping. Although there are many public health practitioners using GIS in their work, and the number is still growing, there are few resources dedicated to instructing how to best visualize health data. There is a wealth of resources on cartography and visualization best practices, but mapmakers will find few that are specific to how public health data can be spatially visualized.

This study addressed the gap in resources by first providing a process for identifying suitable thematic mapping techniques for public health studies, in addition to best practices for employing such techniques. A flow chart, Selecting a Thematic Mapping Method (Figure 2), is provided to help public health practitioners identify potential methods appropriate for visualizing their data based on their data's characteristics. Once identifying a thematic mapping method, they can use the accompanying discussions of each method to guide their map-making process. The choropleth, proportional symbol, dot density, and nominal point mapping techniques were discussed in detail to specify what data are appropriate for each technique and how to best visualize the data. The choropleth map is ideal for mapping derived data of enumeration units, and as one of the most popular and commonly seen of the different thematic maps, it is easy to understand for most audiences. The proportional symbol technique is flexible and can be used to map both point and area data, as well as either total or derived values. The dot density map uses enumeration data like the choropleth map but it is used to visualize total values so that the cluster of dots implies density. Lastly, the nominal point map is used to visualize precise locations of point features.

To demonstrate how spatial visualization can be integrated into a public health study, this study reviewed existing mapping practices by public health practitioners and presented a case study that examined and mapped risk factors of uninsurance. The review of existing mapping practices describes how public health practitioners used various mapping methods and focuses on the applications of health outcomes, social determinants of health, and health care access. The case study used principal component analysis (PCA) and identified the following sociodemographic risk factors: adults without a high school degree, individuals identifying as Hispanic or Latino and Other Race (per the U.S. Census Bureau's race categories), immigrants without citizenship, individuals living under 200 percent of the federal poverty level (FPL), and unemployed adults. These risk factors were visualized using the choropleth mapping best practices, showing how study findings can be better communicated in cartographic form than tabular.

Lastly, this study documented best practices for visualizing health outcomes, social determinants of health, and health care access. This final chapter describes the most suitable mapping techniques for visualizing commonly used data in these three application areas. The following section first discusses common geographic phenomenon for each public health application and the data characteristics public health practitioners are likely to find. Based on this information, best practices are provided, suggesting which thematic mapping methods would be useful and describing particularly important issues that public health practitioners should pay attention to when using such methods.

This chapter concludes with a discussion of the study's implications, significance, limitations, and future directions. The purpose of this study was to address the gap in cartographic resources for the public health sector, and although the study focuses on a limited number of thematic mapping methods and public health applications, the study provides in-depth guidelines that will certainly help public practitioners working in the field of population health improve their ability to create maps of their findings. This study will also have a larger impact on the broader public health sector by demonstrating the value of spatial visualization and encouraging public health practitioners to integrate more cartography into their studies, which is one step toward increasing the amount of cartographic resources and creating a stronger relationship between the public health and GIS communities.

5.1 Best Practices for Mapping Public Health Data

Spatial visualization of data enables public health practitioners to effectively present geographic phenomenon and detect patterns in maps (Cromley and McLafferty 2012). In order to select the appropriate mapping method to visualize data, the characteristics of the data, as well as the subject, purpose, and audience of the map, should be carefully considered (Dent, Torguson, and Hodler 2008). The following sections discuss how to best visualize and enhance understanding of health outcomes, social determinants of health, and access to health care services.

5.1.1 Mapping Health Outcomes

In studies of health outcomes, clusters of disease incidence are often analyzed and visualized to study their relationship to their geographic environment (Cromley 2003; Cromley and McLafferty 2012). However, health outcomes encompass much more than disease; they include length and quality of life (County Health Rankings 2014). Variables on length of life, or mortality, often include number of deaths, death rates, life expectancy, infant mortality, and causes of death. Quality of life focuses on health-related impacts on quality of life, such as overall health, as well as physical and mental health (County Health Rankings 2014). Overall health variables include self-reported health status, such as the number or percent of individuals who report poor or fair health status (Figure 7B). Physical and mental health variables may include chronic diseases (e.g., asthma, cancer, diabetes, obesity), disabilities (Figure 5), and depression.

Many of the datasets with these health outcome variables are available from government agencies as enumeration units. The Centers for Disease Control and Prevention (CDC) contains a vast amount of health-related data, although state and local health departments can also be excellent sources of health data as well. Data are typically quantitative in nature as total or derived values. They are often found as state-level enumeration units and sometimes county-level; they are occasionally available in a GIS data format but more often in a tabular format. When data are compiled in a tabular form such as an Excel or a comma delimited text file, public health practitioners will need to join the table to a GIS dataset such as the Topologically Integrated Geographic Encoding and Referencing (TIGER)/Line Shapefile data to visualize the tabular data in the mapping software environment.

Based on the typical nature of available health outcome data, there are several suitable thematic mapping methods that can be used. Choropleth, proportional symbol, and dot density maps are all potential options for visualizing the data. The choropleth technique is very useful for mapping derived values, such as percentages by enumeration unit (Figure 3B), but cartographers should pay special attention to certain elements of the map when using this technique. The data classification method is extremely important because of its major impact on audience interpretation. Cartographers should carefully consider which method (equal intervals, quantiles, natural breaks, and standard deviation) is best (Table 2), and this can only be done with a close examination of the data (Figures 4 and 5). Regardless of the data classification method, the number of classes should never exceed seven when one color of varying lightness is used; any more would be too challenging for the audience to perceive differences (Figure 6). To facilitate further understanding of spatial patterns, cartographers should correctly match the color scheme type to the data; sequential schemes should be used for unipolar data and diverging schemes for bipolar data (Figures 7B and 8, respectively). When the choropleth technique is correctly practiced, a cartographer can present a visually engaging map that most audiences can understand and enjoy.

The proportional symbol technique can be used to map both total and derived values (Figures 11A and 10B, respectively); many types of health outcome datasets can be successfully visualized using this technique. However, not all data ranges are appropriate; small data ranges should not be visualized using proportional symbols. Health outcomes that have small data ranges are often ones with low percentages, such as heart disease (Figure 10A). Large data ranges can also be troublesome if cartographers aim to use unclassed symbols, which can result in too much symbol overlap. Proportional symbols should be sized and placed so that there is some overlap to show variety in symbol size (Figure 10B), but not so much that the maps become unreadable. Other than these considerations, cartographers will find that proportional symbols can be adapted to most datasets to successfully visualize health outcomes.

The dot density map should only be used for mapping total values of health outcomes (Figure 16B). Similar to the proportional symbol technique, the dot density technique should only be used under the right data range circumstances. While the proportional symbol map should not show data with a small range, the dot density map should not show data with a large range. Enumeration units with smaller or larger values would show almost no pattern or one that is unreadable (Figure 14). Dot value and size are also extremely important because of their impact on the map's spatial pattern and readability, as well as the audience's interpretation of the map. The value and size of the dot should properly be set so that the enumeration unit with the smallest quantity should have two or three dots and the enumeration unit with the largest quantity should have moderate overlapping of the dots. Lastly, when using the dot density technique, cartographers should use ancillary information to improve the accuracy of the map so that dots are not placed in areas unlikely to have individuals residing there (such as the Great Lakes as seen in Figure 16B). The dot density map is relatively easy to understand for most audiences, and applying the previously mentioned best practices will result in a map that is easy to read.

5.1.2 Mapping Social Determinants of Health

The well-documented relationship between social characteristics and health has resulted in the popularity of examining population characteristics in public health studies (Cromley and McLafferty 2012). Sociodemographic characteristics include both demographics and socioeconomic statuses of individuals. These include age (Figure 15B), sex, race and ethnicity (Figures 8, 11 and 25), educational attainment (Figure 24), employment (Figure 30), language use, health insurance coverage (Figure 23), poverty (Figures 26 and 28), immigration and citizenship status (Figure 29), and many more.

Datasets with these variables are almost always available only as enumeration units to protect privacy and confidentiality (Cromley and McLafferty 2012), and the largest data source is like the U.S. Census Bureau. This federal government agency has several survey program that collects hundreds of variables, many of which are available for public use on a large range of geographic levels. Geographies include region, state, county, census tract, block group, block, congressional district, and several others. Social determinants of health datasets typically have quantitative attributes, either as total or derived values. Similar to health outcomes data, they can be found in a GIS data format but are more often found in only tabular form. In this case, they would have to be joined to GIS data, to be mapped.

The choropleth, proportional symbol, and dot density technique can all be used for mapping social determinants of health, under the right conditions. The same considerations taken when mapping health outcomes should also be applied to mapping social determinants of health.

5.1.3 Mapping Access to Health Care Services

Many studies have mapped health services to determine spatial distribution of community resources and access to those resources (Pearce et al. 2007; Macintyre, Macdonald, and Ellaway 2008). Access to health care services mostly focuses on the locations of services, such as hospitals, clinics, and other health care providers.

Locations of health care services can come from many different data sources. Government agencies such as the California Office of Statewide Health Planning & Division (OSHPD) may track data from health care facilities, especially local government agencies that oversee their own facilities. Other potential sources of health care services data include nonprofit agencies that focus on health care services, as well as private companies that track data on local businesses. All of these data sources typically provide addresses of health care facilities, while some datasets may be available as GIS data. If GIS data are not available, a spreadsheet of addresses will need to be geocoded in mapping software in order to visualize each location, which will be represented by a point. In addition to address data, health care service datasets may also contain attributes that reflect their capacity (such as the number of health care providers or number of hospital beds) or their patients (such as the number of patients).

Nominal point maps are appropriate for mapping precise locations of health care services. The attributes of these point features should be qualitative in nature to use this thematic mapping technique. When using this technique, all features in the dataset can be visualized using the same symbol; alternatively, features with different nominal attributes can be represented by different shapes or colors (Figure 19). When using the nominal point technique to visualize precise locations of health care services, cartographers must select a scale that is large enough for the symbols to be clearly read by the audience (Figure 17B). A scale that is too small may result in too many overlapping symbols, which makes the map harder to read (Figure 17A). Applying these best practices will allow the audience to successfully identify locations of health care services.

The proportional symbol technique can also be used to map the precise locations of health care services and their quantitative attributes. Also known as true point data, the point symbols would be placed so that they represent the exact location of the health care services and they would be in varying sizes to represent the values of the attributes (Figure 9A). Mapping such data provides the audience not only locations of health care services, but also additional information about the capacity or patients of the services.

Aggregating the number of health care services or their attributes by enumeration unit, such as city, county, or state, allows the audience to more easily compare the number of health care services across different enumeration areas. The proportional symbol technique can be used to map either total values, such as the number of hospital beds per city (Figure 9B), or derived values such as sums, means, medians, minimums, and maximums of the data. Dot density maps can be used to map total values, such as the number of health care providers, while choropleth maps should only be used to map derived values, such as the number of providers per patient. These three mapping techniques should be used if the purpose of the map is to show quantitative data about the health care services and not merely where they are located.

5.1.4 Summary of Best Practices for Mapping Public Health Data

Mapping health outcomes, social determinants of health, and access to health care services – and many other topics – can be effectively visualized using several thematic mapping techniques. There is no one right method to map the data, but there are potentially many wrong ways to use any given method. Before making any decisions about mapping data, cartographers should carefully consider the purpose of the map and the characteristics of the data to be visualized. These two pieces have an extremely large impact on how the data should be visualized and how the audience understands and interprets the map. Once the purpose is understood, cartographers can use the Selecting a Thematic Mapping Method flow chart (Figure 2) to help them identify an appropriate method to use based on the nature of their data. When these are carefully considered and mapping techniques implemented using best practices, cartographers will very likely be able to create a map that is engaging and successfully communicates findings.

5.2 Implications, Significance, and Future Directions

Despite the growing popularity in using spatial visualization to convey research findings in the public health sector, public health practitioners seeking guidance will find little that are specifically tailored to the field. The purpose of this study was to address that gap in public health resources. This study discussed best practices for only three applications in public health – health outcomes, social determinants of health, and access to health care services – which is a limitation of this study. However, these three applications are all listed in Healthy People 2020’s topic priorities to improve health and eliminate health disparities, and they are also common topics in studies on population health (Healthy People 2014). Since these applications are most related to population health, the author chose to discuss thematic mapping techniques that are most likely to be useful given the type of datasets commonly seen in these applications.

Choropleth, proportional point, dot density, and nominal point maps were discussed in this study, while surface, cartogram, and flow maps were not. An additional limitation of this study is the lack of discussion of layout design. Although layout design is important in cartography, this study focused on how public health data affects the use of thematic mapping techniques. Selecting how to design the layout relies much less on the nature of the data, and there are many cartographic resources that discuss layout design in depth (such as Slocum et al.'s *Thematic Cartography and Geovisualization*; Dent, Torguson, and Hodler's *Cartography: Thematic Map Design*; and Peterson's *GIS Cartography*).

Public health practitioners working in the field of population health will find this study's cartographic guidelines helpful in improving their ability to spatially visualize their data. The flow chart, Selecting a Thematic Mapping Method (Figure 2), is a step-by-step process that not only helps public health practitioners select what method they should use to visualize their data, but also instills the importance of how data characteristics should dictate the methods of spatial visualization. This key lesson is emphasized during the discussion of best practices for using the choropleth, proportional symbol, dot density, and nominal point mapping techniques. These best practices provide essential tips public health practitioners should follow when employing the mapping techniques in order to best facilitate communication and understanding of their findings. What is unique about this cartographic resource is the specific public health focus that few other cartography materials have; the best practices discussion on mapping health outcomes, social determinants of health, and access to health care services will help public health practitioners nurture their ability to visualize such data. The case study is especially useful for individuals who are new to mapping their findings and unsure how to proceed. The case study is a helpful introduction to integrating spatial visualization, demonstrating how sociodemographic data can be analyzed through PCA to identify risk factors of uninsurance and then mapped to show how much more depth successful implementation of thematic mapping techniques, specifically the choropleth technique, can introduce to interpretation of findings.

This study also has a larger impact on the broader public health sector. Although this study focused on addressing the gap of cartographic resources for the individual public health practitioner, this study also sought to demonstrate the importance of spatial visualization to the field. Maps allow the audience to examine geographic phenomenon and detect patterns that may remain otherwise undiscovered in tabular form. When maps are successfully created, their usefulness is magnified. Although the number of population health studies using GIS for spatial visualization is increasing, the number is still relatively small compared to other disciplines, such as geography. There is much potential for further growth if individual public health practitioners, and eventually the field as a whole, begin to integrate spatial visualization and implement cartographic best practices.

To achieve this growth, additional work is needed from both the public health and GIS communities. This study seeks to address the lack of cartographic resources; however, additional guidelines are needed to support and encourage the spatial visualization capacity of public health practitioners. Future guidelines should include best practices for other thematic mapping techniques that would be useful for public health application, such as the dasymetric method. The guidelines should be written in ways that resonate with public health practitioners, such as using relevant real-world examples that they can understand. Additional best practices for mapping other public health applications, such as health behaviors and environmental hazards, would encourage studies to integrate more spatial visualization. Lastly, the bridge between the GIS and public health communities needs to be better developed. While many public health practitioners are aware of what GIS is, relatively few understand the major benefits GIS offers. A stronger connection between the two communities will result in not only help the public health field integrate more tools and methods to address population health concerns, but also build a larger network of GIS professionals and cartographers that can support one another.

REFERENCES

Ahluwalia, I.B. and J. Bolen. 2008. Lack of Health Insurance Coverage Among Working-age Adults, Evidence From the Behavioral Risk Factor Surveillance System, 1993-2006. Journal of Community Health 33:293-296.4

Alameda County. 2012. About us. *Alameda County, CA*. http://acgov.org/about/ (last accessed on 8 April 2014).

Alameda Health Consortium. 2014. Map of Alameda Health Consortium Member Health Center Locations. *Alameda Health Consortium*. http://www.alamedahealthconsortium.org/community-health-centers/map-of-ahc-member-health-centers/ (last accessed on 12 May 2014).

Angel, R. J., S. M. Frias, and T. D. Hill. 2005. Determinants of Household Insurance Coverage Among Low-Income Families from Boston, Chicago, and San Antonio: Evidence from the Three-City Study. *Social Science Quarterly* 86(Supplement): 1338-1353.

Ayanian, J.Z., B.A. Kohler, A. Toshi, and A.M. Epstein. 1993. The relation between health insurance coverage and clinical outcomes among women with breast cancer. *New England Journal of Medicine* 329:326-331.

———. Weissman, E.C. Schneider, J.A. Ginsburg, and A.M. Zaslavsky. 2000. Unmet health needs of uninsured adults. *Journal of the American Medical Association* 284: 2061-2069.

Baicker, K., W. J. Congdon, and S. Mullainathan. 2012. Health Insurance Coverage and Take-Up: Lessons from Behavioral Economics. *The Milbank Quarterly* 90(1): 107-134.

Baker, D.W., J.j. Sudano, J.M. Albert, E.A. Borawski, A. Dor. 2001. Lack of Health Insurance and Decline in Overall Health in Late Middle Age. *New England Journal of Medicine* 345(15): 1106-1112.

Berki, S. L. Wyszewianski, R. Lichtenstein, P.A. Gimoty, J.E. Bowlyow, M.E. Parke, T.B. Smith, S.C. Crane, and J. Bromberg. 1985. Health Insurance Coverage of the Unemployed. *Medical Care* 23(7): 847-854.

Bolstad, P. 2012. *GIS Fundamentals: A First Text on Geographic Information Systems*. Saint Paul, MN: Eider Press.

Brewer, C. 2005. Types of Color Schemes. *Color Brewer 2.0*. http://colorbrewer2.org/learnmore/schemes_full.html#diverging (last accessed on 30 July 2014).

BRFSS (Behavioral Risk Factor Surveillance System). 2013a. About BRFSS. http://www.cdc.gov/brfss/about/index.htm (last accessed on 29 May 2014).

———. 2013b. Survey Data and Documentation. *Centers for Disease Control and Prevention*. http://www.cdc.gov/brfss/data_documentation/index.htm (last accessed on 29 May 2014).

———. 2014. BRFSS Frequently Asked Questions (FAQs). *Centers for Disease Control and Prevention*. http://www.cdc.gov/brfss/about/brfss_faq.htm (last accessed on 29 May 2014).

Bureau of Labor Statistics. 2012. The Recession of 2007-2009. *Bureau of Labor Statistics*. http://www.bls.gov/spotlight/2012/recession/pdf/recession_bls_spotlight.pdf (last accessed on 8 April 2014).

———. 2014. Databases, Tables & Calculators by Subject. *Bureau of Labor Statistics*. http://data.bls.gov/timeseries/LNS14000000 (last accessed on 8 April 2014).

California Department of Public Health. 2012. Infection Rates Map. *California Department of Public Health*. http://www.cdph.ca.gov/programs/Pages/MyHospital411Infections.aspx#/@12/37.8112/-122.2389?tab=CBGB (last accessed on 13 May 2014).

Callahan, S.T. and Cooper. 2004. Gender and Uninsurance Among Young Adults in the United States. *Pediatrics* 113:291-297.

Carrillo, J.E., V.A. Carrillo, H.R. Perez, D. Salas-Lopez, A. Natale-Pereira, and A.T. Byron. 2011. Defining and Targeting Health Care Access Barriers. *Journal of Health Care for the Poor and Underserved* 22: 562-575.

CDC (Centers for Disease Control and Prevention). 1998. State-specific prevalence estimates of uninsured and underinsured persons—Behavioral Risk Factor Surveillance System,1995. *Morbidity and Mortality Weekly Report* 47: 51-55.

———. 2012a. Cancer Screening – United States, 2010. *Morbidity and Mortality Weekly Report* 61(3): 41-45.

———. 2012b. Cartographic Guidelines for Public Health. *Centers for Disease Control and Prevention*. http://www.cdc.gov/dhdsp/maps/GISX/resources/Cartographic_Guidelines.pdf (last accessed on 13 May 2014).

———. 2013. Health Care Maps – Health Insurance. *Centers for Disease Control and Prevention*. http://www.cdc.gov/dhdsp/maps/sd_insurance_2000.htm (last accessed on 12 May 2014).

Chatterjee, S. and R.B. Nielson. 2011. Employer-provided Health Insurance Coverage: A Comparison of Employed Native-born and Immigrant Americans. *Family & Consumer Sciences Research Journal* 40(1): 15-27.

Clemans-Cope, L., G.M. Kenney, M. Buettgens, C. Carroll, and F. Blavin. 2012. The Affordable Care Act's Coverage Expansions Will Reduce Differences in Uninsurance Rates By Race and Ethnicity. *Health Affairs* 31(5): 920-930.

CMS (Centers for Medicare & Medicaid Services). 2013. Medicare Program – General Information. *Centers for Medicare & Medicaid Services*. http://www.cms.gov/Medicare/MedicareGeneralInformation/MedicareGenInfo/index.html (last accessed 29 April 2014).

———. n.d. By Population. *Centers for Medicare & Medicaid Services*. http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Population/By-Population.html (last accessed 8 April 2014).

Collins, S.R., M.M. Doty, R. Robertson, and T. Garber. 2011. Help on the Horizon: How the Recession Has Left Millions of Workers Without Health Insurance, and How Health Reform Will Bring Relief. *The Commonwealth Fund*. http://www.commonwealthfund.org/~/media/Files/Surveys/2011/1486_Collins_help_on_the_horizon_2010_biennial_survey_report_FINAL_31611.pdf (last accessed on 8 April 2014).

Council on Foreign Relations. 2014. Vaccine-Preventable Outbreaks. *Council on Foreign Relations*. http://www.cfr.org/interactives/GH_Vaccine_Map/#map (last accessed on 13 May 2014).

County Health Rankings. 2014. Health Outcomes. *County Health Rankings*. http://www.countyhealthrankings.org/our-approach/health-outcomes (last accessed on 14 August 2014).

Cox, M., P.J. Boyle, P.G. Davy, Z. Feng, and A.D. Morris. 2007. Locality deprivation and Type 2 diabetes incidence: A local test of relative inequalities. *Social Science and Medicine*, 65(9): 1953-1964.

Cromley, E. 2003. GIS and Disease. *Annual Review of Public Health* 24: 7-24.

——— and S. McLafferty. 2012. *GIS and Public Health*. New York, NY: The Guilford Press.

DeNavas-Walt, C., B.D. Proctor, and J. Smith. 2008. Income, poverty and health insurance coverage in the United States. *Current Population Reports*, *P60-233* 1-67.

———. 2011. Income, poverty, and health insurance coverage in the United States: 2010. Washington, DC: Census Bureau.

Dent, B., J. Torguson, and T. Hodler. 2008. *Cartography: Thematic Map Design*. New York, NY: McGraw-Hill.

DDT (Division of Diabetes Translation). 2013. Methodology for County-Level Estimates. *Centers for Disease Control and Prevention*. http://www.cdc.gov/diabetes/atlas/countydata/County_Methods.html#countylevelestimates (last accessed on 30 July 2014).

———. 2014. About CDC's Division of Diabetes Translation. *Centers for Disease Control and Prevention*. http://www.cdc.gov/diabetes/about/index.htm (last accessed on 30 July 2014).

Do, D.P., B.K. Finch, R. Basurto-Davila, C. Bird, J. Escarce, and N. Lurie. 2008. Does place explain racial health disparities? Quantifying the contribution of residential context to the Black/white health gap in the United States. *Social Science and Medicine* 67(8) 1258-1268.

Dor, A., J. Sudano, D.W. Baker. 2006. The Effect of Private Insurance on the Health of Older, Working Age Adults: Evidence from the Health and Retirement Study: *Health Services Research* 41(3 Pt 1): 759-787.

Dubay, L. and G. Kenney. 2004. Address Coverage Gaps For Low-income Parents. *Health Affairs* 23(2): 225-234.

Duncan, G.J., M.C. Daly, P. McDonough, and D.R. Williams. 2002. Optimal indicators of socioeconomic status for health research. *American Journal of Public Health* 92(7): 1151-1157.

Dunlin, M.F., T.M. Ludden, H. Tapp, J. Blackwell, B. Urquieta de Hernandez, H.A. Smith, and O.J. Furuseth. 2010. Using Geographic Information Systems (GIS) to Understand a Community's Primary Care Needs. *The Journal of the American Board of Family Medicine* 23:13-21.

English, P, R. Neutra, R. Scalf, M. Sullivan, L. Waller, and L. Zhu. 1999. Examining associations between childhood asthma and traffic flow using a geographic information system. *Environmental Health Perspectives* 107:761-767.

Field, K. 2013. Red, Blue and Purple: mapping the 2012 US Presidential Election. *Esri*. http://blogs.esri.com/esri/arcgis/2013/04/29/red-blue-and-purple-mapping-the-2012-us-presidential-election/ (last accessed on 31 July 2014).

Geraghty, E.M., T. Balsbaugh, J. Nuovo, and S. Tandon. 2010. Using Geographic Information Systems (GIS) to Assess Outcome Disparities in Patients with Type 2 Diabetes and Hyperlipidemia. *The Journal of the American Board of Family Medicine* 23:88-96.

Hadley, J. 2007. Insurance Coverage, Medical Care Use, and Short-term Health Changes Following an Unintentional Injury or the Onset of a Chronic Condition. *Journal of the American Medical Association* 297(10):1073-84.

Healthy People. 2013. Health Communication and Health Information Technology. *Healthy People*. http://www.healthypeople.gov/2020/topicsobjectives2020/overview.aspx?topicid=18 (last accessed on 8 April 2014).

———. 2020 Topics & Objectives – Objectives A-Z. *Healthy People*. http://www.healthypeople.gov/2020/topicsobjectives2020/ (last accessed on 15 August 2014).

The Henry J. Kaiser Family Foundation. 2013. The Uninsured Primer: Key Facts about Health Insurance on the Eve of Health Reform. Menlo Park, CA: The Henry J. Kaiser Family Foundation.

Higgs, G. 2005. A Literature Review of the Use of GIS-Based Measures of Access to Health Care Services. *Health Services & Outcomes Research Methodology* 5: 119-139.

———. 2009. The role of GIS for health utilization studies: literature review. *Health Services & Outcomes Research Methodology* 9: 84-99.

Huston, Sara L. 2011. Chronic Disease Rates by Public Health District, Maine. *Centers for Disease Control and Prevention*. http://www.cdc.gov/dhdsp/maps/gisx/mapgallery/maps/detail/index.html?39#39 (last accessed on 9 August 2014).

Illinois Department of Public Health. 2013. Colorectal Cancer Incidence and Mortality Rates and Federally Qualified Health Centers, Illinois, 2006-2010. *Centers for Disease Control and Prevention*. http://www.cdc.gov/dhdsp/maps/gisx/mapgallery/maps/detail/index.html?34#34 (last accessed on 9 August 2014).

IOM (Institute of Medicine). 2002. *Health Insurance is a Family Matter*. Washington, DC: IOM.

———. 2003. Coverage Matters: Insurance and Health Care. Washington, DC: IOM.

Jansen, A., C. Frank, J. Koch, and K. Stark. 2008. Surveillance of vector-borne diseases in Germany: Trends and challenges in the view of disease emergence and climate change. *Parasitology Research* 103(1): S11-S17.

John, D.A., A.B. de Castro, B. Duran, D.P. Martin. 2013. Nativity and Occupational Class Disparities in Uninsurance and Routine Preventive Care Use Among Asian Americans. *Journal of Immigrant and Minority Health* 15:1011-1022.

Jones, K., C. Duncan, and L. Twigg. 2004. “Evaluating the absolute and relative income hypotheses in an exploratory analysis of deaths in the Health and Lifestyle Survey.” In *The geography of health inequalities in the developed world: Views from Britain and*

North America, edited by P. Boyle, S. Curtis, E. Graham, and E. Moore. Aldershot, UK: Ashgate.

Kibon, U.A. and M. Ahmed. 2013. Distribution of Primary Health Care Facilities in Kano Metropolis Using GIS (Geographic Information System). *Research Journal of Environmental and Earth Sciences* 5(4): 167-176.

Koch, T. 2005. *Cartographies of Disease: Maps, Mapping, and Medicine*. Redlands, CA: Esri Press.

Krieger, N., J.T. Chen, P.D. Waterman, D.H. Rehkophf, and S.V. Subramanian. 2003. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: A comparison of area-based socioeconomic measures – The Public Health Disparities Geocoding Project. *American Journal of Public Health* 93(10) 1655-1671.

Legerski, E.M. 2012. The Cost of Instability: The Effects of Family, Work, and Welfare Change on Low-Income Women's Health Insurance Status. *Sociological Forum* 27(3): 641-657.

Lorenzo-Seva, U. 2003. A Factor Simplicity Index. *Psychometrika* 68(1):49-60.

Macintyre, S., L. Macdonald, and A. Ellaway. 2008. Do poorer people have poorer access to local resources and facilities? The distribution of local resources by area deprivation in Glasgow, Scotland. *Journal of Health Psychology* 2(3): 315-334.

Marmot, M. 2005. Social determinants of health inequalities. *Lancet* 365: 1099-1104.

———, S. Friel, R. Bell, T.A. Houweling, and S. Taylor. 2008. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet* 372: 1661-1669.

Mayer, M. 2006. Are we there yet? Distance to care and relative supply among pediatric subspecialties. *Pediatrics* 118(6): 2313-2321.

McLafferty, S. 2003. GIS and Health Care. *Annual Review of Public Health* 24: 25-42.

Mills, R.J. 2002. *Health Insurance Coverage: 2001*. Washington, DC: U.S. Bureau of the Census; 1-19.

Mitchell, A. 11999. *The Esri Guide to GIS Analysis, Volume 1: Geographic Patterns & Relationships*. Redlands, CA: Esri Press.

———. 2005. *The Esri Guide to GIS Analysis, Volume 2: Spatial Measurements & Statistics*. Redlands, CA: Esri Press.

Monmonier, M. 1996. How to Lie with Maps. Chicago, IL: University of Chicago Press.

Nebraska Department of Health & Human Services. 2014. Nebraska Heart Disease Mortality over Time. http://www.cdc.gov/dhdsp/maps/gisx/mapgallery/maps/detail/index.html?56#56 (last accessed on 9 August 2014).

Nelson, D.E., J. Bolen, H.E. Wells, S.M. Smith, S. Bland. 2004. State Trends in Uninsurance Among Individuals Aged 18 to 64 Years: United States, 1992-2001. *American Journal of Public Health* 94(11): 1992-1997.

Noone, P. 2009. Social determinants of health. *Occupational Medicine* 59: 209.

O'Hara, B. 2004. Do mother work to support ailing husbands? *Journal of Family and Economic Issues* 25(2):179-1982

O'Rourke, N. and L. Hatcher. 2013. A Step-by-Step Approach to Using SAS® for Factor Analysis and Structural Equation Modeling, Second Edition. SAS Institute Inc., Cary, NC.

O'Sullivan, D. and D.J. Unwin. 2010. *Geographic Information Analysis*. Hoboken, NJ: John Wiley & Sons.

OSHPD (Office of Statewide Health Planning & Division). 2013. About OSHPD. *Office of Statewide Health Planning & Division*. http://www.oshpd.ca.gov/AboutUs.html (last accessed 29 May 2014).

Paek, M. and J.L. Lim. 2013. Factors Associated with Health Care Access and Outcome. *Social Work in Health Care* 51(6): 506-530.

Pauly, M.V. and J.A. Pagan. 2007. Spillovers and Vulnerability: The Case of Community Uninsurance. *Health Affairs* 26(5): 1304-1314.

———. 2009. Spillovers of Uninsurance in Communities. *Institute of Medicine*. http://www.iom.edu/~/media/Files/Activity%20Files/HealthServices/UninsuredConsequence/Commissioned%20Paper%20%20Pauly%20and%20Pagan%202008%20-%20Spillovers%20of%20Uninsurance%20in%20Communities.pdf (last accessed 8 April 2014).

Pearce, J., K. Written, R. Hiscock, and T. Blakely. 2007. Are socially disadvantaged neighborhoods deprived of community health-related resources? *International Journal of Epidemiology* 36(2): 348-355.

Peterson, G. 2009. *GIS Cartography: A Guide to Effective Map Design*. Boca Raton, FL: Taylor & Francis Group.

Phillips, R.L., E.L. Kinman, P.G. Schnitzer, E.J. Lindbloom, and B. Ewigman. 2009. Using Geographic Information Systems to Understand Health Care Access. *Archives of Family Medicine* 9: 971-978.

Pickett, K.E. and M. Pearl. 2001. Multilevel analyses of neighborhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health* 55(2): 111-122.

Pylypchuk, Yuriy. 2009. Effects of immigration on the health insurance status of natives. *Journal of Health Economics* 28: 1028-1037.

Reissman, D.B., F. Staley, G.B. Curtis, R.B. Kaufmann. 2001. Use of geographic information system technology to aid health department decision making about childhood lead poisoning prevention activities. *Environmental Health Perspectives* 109:89-94.

Robert Wood Johnson Foundation. 2009. Impact of the Economy on Health Care. *Robert Wood Johnson Foundation*. http://www.rwjf.org/en/research-publications/find-rwjf-research/2009/08/impact-of-the-economy-on-health-care-.html (last accessed on 8 April 2014).

Sabik, L.M. 2012. The Effect of Community Insurance Rates on Access to Health Care. *Health Services Research* 47:3, Part 1.

SAS Institute Inc. 2013. *JMP 11 Multivariate Methods.* Cary, NC: SAS Institute Inc.

Schwartz, K., J. Marcotte, and T.D. McBride. 1993. Personal Characteristics and Spells Without Health Insurance. *Inquiry* 30(1): 64-76.

Slocum, T.A., R. B. McMaster, F.C. Kessler, and H.H. Howard. 2009. *Thematic Cartography and Geovisualization*. Upper Saddle River, NJ: Pearson Education, Inc.

State Health Access Data Assistance Center. 2013. State-Level Trends in Employer-Sponsored Health Insurance. *Robert Wood Johnson Foundation*. http://www.rwjf.org/content/dam/farm/reports/reports/2013/ rwjf405434 (last accessed on 8 April 2014).

Suhr, D. D. 2005. Principal Component Analysis vs. Exploratory Factor Analysis. Statistics and Data Analysis Paper 203-230. In: The Proceedings of the 31th Annual SAS Users Group International Conference. SAS Institute Inc.

Superior Court of Alameda. n.d. About Alameda County. http://www.alameda.courts.ca.gov/ pages.aspx/about-alameda-county (last accessed on 8 April 2014).

Tai, W., F Porell, and E.K. Adams. 2004. Hospital choice of rural Medicare beneficiaries: Patient, hospital attributes, and the patient-hospital relationship. *Health Services Research* 39(6) 1903-1922.

Tomlinson, R.F. and A.R. Boyle. 1987. The state of development of systems for handling national resources inventory data. *Cartographica* 18(4): 65-95.

U.S. Census Bureau. 2008. A Compass for Understanding and Using American Community Survey. *U.S. Census Bureau*. http://www.census.gov/acs/www/Downloads/handbooks/ACSGeneralHandbook.pdf (last accessed 29 May 2014).

———. 2011a. Population Distribution and Change: 2000 to 2010. *U.S. Census Bureau.* http://www.census.gov/prod/cen2010/briefs/c2010br-01.pdf (last accessed on 13 May 2014).

———. 2011b. Housing Characteristics: 2010. *U.S. Census Bureau.* http://www.census.gov/prod/cen2010/briefs/c2010br-07.pdf (last accessed on 13 May 2014).

———. 2012a. Health Insurance Coverage, 2010-2012 ACS 3-year Estimates, Table S2701. *U.S. Census Bureau.* factfinder2.census.gov (last accessed on 8 April 2014).

———. 2012b. Selected Social Characteristics 2010-2012 ACS 3-year Estimates, Table DP02. *U.S. Census Bureau.*factfinder2.census.gov (last accessed on 8 April 2014).

———. 2012c. Data Protection and Privacy Policy. *U.S. Census Bureau.* http://www.census.gov/privacy/ (last accessed on 12 May 2014).

———. 2012d. The Asian Population: 2010. *U.S. Census Bureau.* https://www.census.gov/prod/cen2010/briefs/c2010br-11.pdf (last accessed on 12 May 2014).

———. 2013. Income, Poverty, and Health Insurance Coverage in the United States: 2012. *U.S. Census Bureau.* http://www.census.gov/prod/2013pubs/p60-245.pdf (last accessed on 8 April 2014).

———. 2014a. What is the Census? *U.S. Census Bureau.* http://www.census.gov/2010census/about/ (last accessed on 30 July 2014).

———. 2014b. TIGER Products. *U.S. Census Bureau.* https://www.census.gov/geo/maps-data/data/tiger.html (last accessed on 10 July 2014).

———. n.d. American FactFinder. *U.S. Census Bureau.* http://factfinder2.census.gov (last accessed on 29 May 2014).

U.S. Department of Health, Education and Welfare. 1979. Healthy People: The Surgeon General's Report on Health Promotion and Disease Prevention. *U.S. National Library of*

Medicine. http://profiles.nlm.nih.gov/ps/access/NNBBGK.pdf (last accessed on 8 April 2014).

U.S. Department of Health & Human Services. 2014. 2014 Poverty Guidelines. *U.S. Department of Health & Human Services.* http://aspe.hhs.gov/poverty/14poverty.cfm (last accessed on 8 April 2014).

Weldon Cooper Center for Public Service, University of Virginia. 2013. The Racial Dot Map. *Weldon Cooper Center for Public Service.* http://www.coopercenter.org/demographics/Racial-Dot-Map (last accessed on 13 May 2014).

WHO (World Health Organization). 2014. Maps and spatial information technologies (Geographical Information Systems) in health and environment decision-making. *World Health Organization*. http://www.who.int/heli/tools/maps/en/ (last accessed on 8 April 2014).

Young Invincibles. 2012. Health Care. *Young Invincibles*. http://younginvincibles.org/issues/health-care/ (last accessed on 8 April 2014).