THE UTILITY OF SELF-ORGANIZING
MAPS IN THE ANALYSIS OF SOCIAL
DETERMINANTS OF HEALTH

By

MIRIAM MCGAUGH

Bachelor of Science in Biology
Oklahoma City University
Oklahoma City, OK
1998

Master of Science in Epidemiology
University of Oklahoma Health Sciences Center
Oklahoma City, OK
2001

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2009

THE UTILITY OF SELF-ORGANIZING

MAPS IN THE ANALYSIS OF SOCIAL

DETERMINANTS OF HEALTH

Dissertation Approved:

Dr. Janice Miller

Dissertation Adviser

Dr. Katye Perry

Dr. Diane Montgomery

Dr. Patricia Hughes

Dr. A. Gordon Emslie

Dean of the Graduate College

ACKNOWLEDGMENTS

As this is the last part of this dissertation, it is probably the hardest. How can one adequately say thank you to all the people that helped, prodded, encouraged, and cheered me on through this process? First, I would like to express my sincere gratitude to Dr. Janice Miller, my advisor and chair. The never-ending source of knowledge, encouragement, flexibility, and patience was always deeply appreciated. I do not think I would have survived the past five years without all those wonderful traits she possessed and she is arguably one of the best statistics teachers I have ever had. I was also fortunate to have an excellent committee of talented and diverse professors that were not afraid to tackle something new. Dr. Katye Perry and Dr. Diane Montgomery provided the encouragement and understanding that is truly needed to thrive during graduate work and they both challenged me to always go a little further and to think more critically. My outside committee member, Dr. Patricia Hughes, provided me with very enlightening suggestions, comments, and conversations regarding my research. I will think back on these individuals often and use what they have taught me, in and out of the classroom, throughout my career.

In addition to the support from the faculty in the College of Education, including many who I have not named here, I have been privileged to work with some wonderful fellow graduate students, one of which will always hold a special place in my heart. Amy Morse provided an anchor and sounding board when I needed one, was always available to talk through a sticky point in a project, and often helped me deal with a paperwork situation because I was too far away to get to campus in time. With a sad heart, we had to say goodbye to Amy this year. She lost her battle with Melanoma in September. I will forever miss her smile and laughter, but I know she is looking down, smiling of course, and saying "It is about time you got that thing done!"

Finally, I wish to acknowledge the generous support of my friends and family. My husband, John, was particularly generous with his patience and understanding, especially as this dissertation neared its end and the time consumed increased exponentially. While my son, Wesley, was not particularly into the statistics and research thing like mommy as he was only born when I started this program, he was an exceptionally good baby and even escorted me to class a few times just to make sure I made it. I cannot thank John or Wesley enough for the time that was sacrificed to reach this goal, and I eagerly anticipate much more time together in the future. My parents, Jim and Margarett, instilled in me the value of hard-work, determination, and shear stubbornness. Without their support, I would have never completed college in the first place. My in-laws, John Sr. and Audrey, provided many hours of free babysitting and daycare runs so I could attend class, study, write, or just have a night out with my husband. The sacrifices will always be appreciated and never forgotten. I would also like to thank both of my sisters for doing so many things for me during this entire experience. Nikki allowed me to drop my son off with her on more than one occasion just so I could get some work done. She will never know how much that helped! In the tradition of saving the best for last, I could have never finished this dissertation without the tremendous support of my sister, Heather. The hours that she spent reading, editing, commenting, and then doing it all over again multiple times is truly appreciated. The "Friday Cheers" that I received every week for almost two years will be cherished forever, along with the unselfish and unrivaled support. Thank you.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

According to the World Health Organization (2008), health inequalities are pervasive throughout the world. Marmot (2007) states health inequalities occur both across and within country borders. Global, national, and local level factors affect health outcomes, according to the Commission on Social Determinants of Health (Marmot et al., 2007). Access to societal resources, including nutritious foods, safe physical activity outlets, and transportation, as well as access to medical care, influence health outcomes within a community (Link & Phelan, 1995). The growing body of research indicates that society and the surrounding environment play a significant role on health outcomes (Berkman & Kawachi, 2000; Cohn, 2007; Hill & Peters, 1998; Marmot, 2007; Raphael, 2003; Watt, 2002). Promotion of healthy lifestyles and increased spending on health care will not change health outcomes; policy and social issues must be addressed (Raphael, 2003).

Social determinants of health (SDOH) refer to a broad range of social exposures that interact and cumulatively relate to a person's and a society's health (Link & Phelan, 1995; Marmot & Wilkinson, 1999). According to Raphael (2003), SDOH structure lifestyle choices and predict individual and population health better than individual health behaviors. Further, social or health policies and programs may alter these societal exposures and conditions (Anderson, Scrimshaw, Fullilove, Fielding, & the Task Force on Community Preventive Services, 2004).

Existing models related to social determinants of health are often developed to be theoretical and very broad. On one end of the continuum, models range from being so broad that they do not indicate testable links between social determinants of health and health outcomes. At the opposite end, models focus solely on how social determinants of health alter the human body at the biological or elemental level (Andersonet al., 2004; Gehlert et al., 2008; Marmot & Wilkinson, 1999). In between, there is a myriad of models that are disease-specific and not easily adaptable. This wide variety of models poses a problem for community and social epidemiologists who are often constrained by time, finances, and other limited resources. For example, because community and social epidemiologists traditionally examine issues on smaller scales, it is extremely difficult to collect real-time or immediate, current data on all the issues that affect a community's health. Working with even one community would impose such a burden on a single analyst that the task would be insurmountable in a timely manner. Leveraging existing archival data systems to examine health in a societal and environmental context is paramount to timely and efficient information dissemination. Community and social epidemiologists would benefit from a SDOH model that is easily adaptable and points to specific indicators to assist in evaluating a variety of health problems. Additionally, exploring new methods of analyzing the myriad of data that represent SDOH could be of vital importance to community and social epidemiologists in order to explore the full picture of health. The purpose of this study is to present a statistical method to be used in the analysis of data that are readily available nationwide via multiple public sources within a health context. All variables for this study were represented at the county level.

Background of Problem

Along with most social determinants of health models designed to be theoretically broad, many do not indicate testable links between social determinants of health and health outcomes. Community and social epidemiologists often work with limited state or local budgets, hindering their ability to acquire new data from expensive sources or through new data collection efforts. Data elements are collected in an increasingly standardized manner and archived for use at state, county and local levels, but remain underutilized in evaluating health issues. Most common epidemiologic techniques are inadequate for analyzing large volumes of data in one comprehensive analysis, while still representing all data elements in the final analysis. Data reduction techniques, such as factor analysis, are important tools available to community and social epidemiologists, but they do not always allow for full representation of the data due to exclusion of variables during the analysis (Kim & Mueller, 1978). It is critical that community and social epidemiologists have effective and innovative methods of looking at the interactive effects of social determinants of health (i.e., lack of nutritious foods, transportation, and medical care) in relation to health outcomes in a complete and comprehensive manner (Marmot & Wilkinson, 1999).

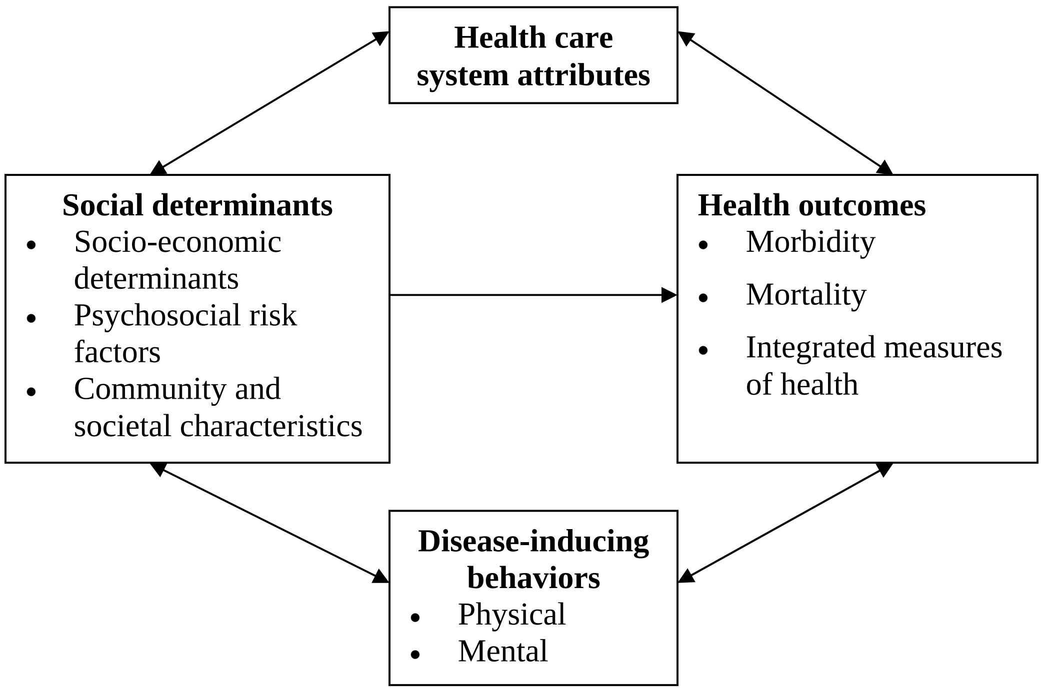
Social determinants of health data are highly complex and interconnected requiring a data reduction technique that is extremely flexible (i.e., no a priori selection of the number of resulting clusters, analyzes variables at differing levels, and handles large amounts of data records and/or variables). Self-organizing maps (SOM) address flexibility issues by processing data iteratively to allow for the best mathematical representation of data while preserving the underlying structure of the data (Erb, 1993; Kohonen, 2001). Similar to other cluster techniques, like K-Means and multidimensional scaling, SOM clusters all data elements included in the process. However, the analyst does not have to pre-select the number of resulting clusters as in K-Means (Bradley & Fayyad, 1998). Because of this, the unsupervised SOM process selects the number of valid nodes (clusters) without any a priori notions from the analyst, ultimately taking the multi-dimensional inputs and reducing them to a bi-dimensional output while preserving the integrity of the original data (Molinier, Laaksonen, & Hame, 2007; Verdu, Garcia, Senabre, Marin, & Franco, 2006). Further, the nature of SOM keeps the most similar data points together, while multidimensional scaling seeks to preserve differences. The choice between the two relies on the research question being asked; however, results from both methods are very similar (Kirt & Vainik, 2007).

Statement of the Problem

Many social environment and health models were not designed to lead researchers to specific data sources for quantification of the social problem. Instead, the models were apparently conceptually left open for broad use. While having a theoretical origin is important, knowing how to adapt a model to a specific community is vital to the accurate and consistent analysis of social determinants of health. Additionally, isolating the right variables for adequately quantifying the health problem can be difficult when the conceptual model is left so broad. Community and social epidemiologists would benefit from a SDOH model that is not only adaptable, but also points to specific variables to evaluate a variety of health problems.

The theoretical model chosen as the foundation of this study, the Public Health Model of Social Determinants of Health (PHM), addresses a pinpointed approach to social determinants of health while taking into account that access to resources and health behaviors play a varying, but contributing role, to health outcomes. The PHM (Figure 1) was developed to assist epidemiologists and public health policy makers in understanding the structure and relationship of social determinants and health, as well as indicating causal relationships that can be analyzed (Ansari, Carson, Ackland, & Vaughan, 2003). Additionally, the developers of the PHM identify variables that address each of the three social determinants categories (Table 1). The flexibility of the model allows for community and social epidemiologists to analyze disease-specific information (i.e., heart disease, cancer, and stroke) as well as conduct broad analyses across multiple health etiologies (i.e., mortality from all causes).

Figure 1: Public Health Model of the Social Determinants of Health



Note: With kind permission from Springer Science+Business Media: *Social and Preventive Medicine*, “A Public Health Model of the Social Determinants of Health,” 48, 2003, 243, Z. Ansari, N.J. Carson, M.J. Ackland, & L. Vaughan, Figure 1.

While general variables have been put forth by the developers of the PHM, the *Data Set Directory of Social Determinants of Health at the Local Level* (Hillemeier, Lynch, Harper, & Casper, 2006), developed by the Social Determinants of Health workgroup at the Centers for Disease Control (CDC), provided national data sources for most of the variables above. Although the

Directory was developed pinpointing resources that were available at the metropolitan statistical area, many of the sources within the document contain county level data, as well. See Appendix A for a detailed list of the variables included in this study.

Table 1: Identified Variables for the three PHM Social Determinants Dimensions

Socio-economic determinants	**Psychosocial risk factors**	**Community and societal characteristics**
Age	Poor social networks	Social networks and support structures
Gender	Low self-esteem	Social and community participation
Race	Self-efficacy	Civic and political involvement and empowerment
Ethnicity	Depression	Trust in people and social institutions
Education	Anxiety	Tolerance of diversity
Occupation	Insecurity	Crime rate
(Un)employment	Loss of sense of control	Poverty
Income	Isolation	Residence (urban, rural, remote)
Religion	Chronic stress	Income inequality
Housing - affordability, security of tenure, structure and maintenance of building, occupancy (including overcrowding)	High physical/psychological demand	Altruism. Philanthropy and voluntary work
	Anger/hostility	Domestic violence
	Coping	Unemployment rate
	Perception/expectations	

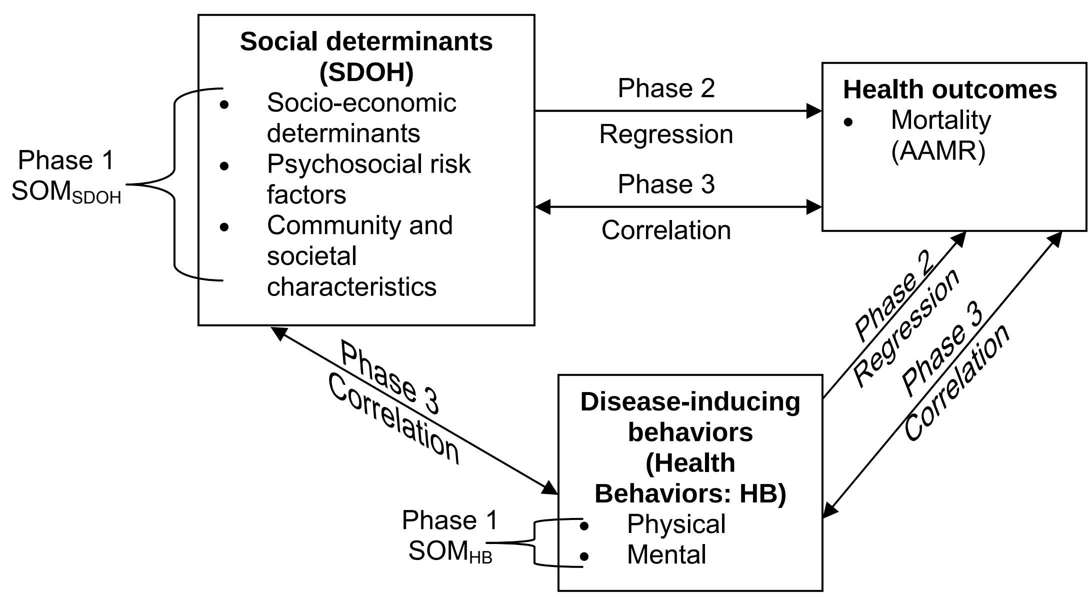
Having a theoretical model to guide variable selection is only part of the analytical process that community and social epidemiologists must address when examining social determinants of health. Matching an analytical method to this theoretical model is equally important. The Self-Organizing Map (SOM), created by Kohonen (2001), was chosen as the analytical method for this study to examine the relationships set forth in the Public Health Model of Social Determinants of Health. The SOM is a neural network technique that does not require the analyst to supervise the iterative process by matching the result of each iteration to a known target but conducts a self-learning neural network process. The SOM is categorized as unsupervised because the interaction between the analyst and the process is not required (Kohonen, 2001).

The SOM algorithm is part of several commercial statistical packages with differing graphic and output capabilities (Kohonen, 2001; SAS Institute, Inc, 2005; Viscovery, n.d.). There are also free versions of the algorithm available that can be run as a stand-alone package, SOM_pak (Kohonen, 1996), or a free extension that can be used with Matlab software (*SOM Toolbox for Matlab*, 2001). Viscovery SOMine 5.0 was used for this study because of the ability for the software to operate on a small computer and not require a server environment to operate like SAS Enterprise Miner or additional software like the Matlab extension and SOM_pak. The ability of the SOM to analyze the full extent of all variables submitted to the process remains a basic function of the SOM algorithm no matter which SOM package is chosen by a community or social epidemiologist.

Purpose of Study

The purpose of this study was to present a method for analyzing existing, nationally-available social data in a health context. This research study utilized the Public Health Model of the Social Determinants of Health (PHM) as a theoretical guide for selecting variables from existing, archival data sources to represent social determinants, health behaviors, and the health outcomes within Oklahoma at the county level, which was the unit of analysis. The PHM is a comprehensive model and links found within the model have been tested to varying degrees in differing settings (Harris, 2001; Huisman & Oldehinkel, 2008; Kopp, Skrabski, Kawachi, & Adler, 2005; Maycock & Howat, 2007). Because of the flexibility of the model, not all links were tested within this study. Figure 2 indicates the adapted model that was used. Additionally, this study introduces the Self-Organizing Map as an alternative data reduction technique for analysis of

Figure 2: Public Health Model of the Social Determinants of Health Showing Phases to be Tested and the Statistical Method



Note: Phase 1, Phase 2, and Phase 3 indicate the testable links explored within this study and correspond to the related research questions below.

With kind permission from Springer Science+Business Media: *Social and Preventive Medicine*, “A Public Health Model of the Social Determinants of Health,” 48, 2003, 243, Z. Ansari, N.J. Carson, M.J. Ackland, & L. Vaughan, Figure 1.

health and social data.

Research Questions

This research study answered the following questions:

Phase 1. What is the underlying relationship among social determinants of health (SDOH) and health behaviors (HB) within Oklahoma's counties?

Ho_1: There is no variation (one SOM cluster) in SDOH among Oklahoma counties.

Ho2: There is no variation (one SOM cluster) in HB among Oklahoma counties.

Phase 2. Is a SOM_{SDOH} cluster variable a stronger predictor of health outcome (Age-Adjusted Mortality Rate) than a SOM_{HB} cluster variable within Oklahoma counties?

H o1: The SOM_{SDOH} is a stronger predictor set of health outcome than the SOM_{HB} set.

Phase 3. What is the relationship between SDOHs, HBs, and health outcome within Oklahoma counties?

Ho1: There is no correlation between SOM_{SDOH} dummy vectors and SOM_{HB} dummy vectors.

Ho2: There is no correlation between SOM_{SDOH} dummy vectors and health outcome.

Ho3: There is no correlation between SOM_{HB} dummy vectors and health outcome.

Assumptions of Study

As with all studies, there are some assumptions that underlie this research. It is assumed that the self-reported data utilized within this study represents the population and that measurement error is randomly dispersed. Secondly, because the desired outcome is groupings or clusters of counties that display mathematically similar social determinants of health (SDOH) and health behaviors (HB), it is assumed that there are enough differences among the SDOH or HB variables at the county level that more than one Self-Organizing Map cluster can be obtained.

Significance of the Study

This study offers several advancements in the field of public health research and social determinants of health. First, a concise and well-defined model for social determinants of health is presented for community and social epidemiologists for use. Additionally, the Public Health Model for the Social Determinants of Health directs researchers to specific, nationally available data elements allowing for a broad range of analysis possibilities. Next, this study presents a statistical method that is not widely used in the public health field. Through the use of Self-Organizing Maps, community and social epidemiologists are provided a new method of analyzing large amounts of data with one comprehensive technique. The combination of an adaptable social determinants of health model and the self-organizing map technique provides community and social epidemiologists with a powerful tool to move the science of social epidemiology forward.

Definition of Terms

Age-Adjusted Mortality Rate (AAMR) – the statistic that indicates the risk of death from an event and is a good index of the severity of events within an area (Gordis, 1996). For this study, age-adjusted mortality rate was calculated from all health events resulting in a death. AAMR could also be narrowed to single events or causes of death such as heart disease, stroke, or cancer.

Community and Social Characteristics – the unique patterns of relationships and organization between the individuals within a community or society.

Compositional Approach – individual characteristics of an individual's socio-economic status (i.e., employment status, years of education completed, individual annual income)

Contextual Approach – the social network, geographic area, or community that a person inhabits that affects their ability to have wealth. Contextual variables include average house value in a geographic area, percentage of unemployed persons, and per capita income.

Ethnicity – the heritage, nationality group, lineage, or country of birth of the person, person's parents, or ancestors before their arrival to the U.S. The United States Office of Management and Budget recognize two groupings, Hispanic or Latino and Not Hispanic or Latino (United States Census Bureau, 2002).

Health Behaviors – cognitive elements such as beliefs, expectations, motives, values, perceptions; personality characteristics, including affective and emotional states and traits; and overt behavior patterns, actions, and habits that relate to health maintenance, to health restoration, and to health improvement (Gochman, 1997).

Health Inequality - the differences in health status or in the distribution of health determinants between different population groups. One example would be differences in mobility between elderly people and younger populations or differences in mortality rates between people from different social classes (Barnes & Health Department Agency, n.d.).

Health Outcomes – according to the Health Outcomes Library Core Project (AcademyHealth, 2004), health outcomes could go beyond the physiological measures of success (the absence of mortality) and could examine additional issues such as quality of life, longevity, morbidity, psychosocial functioning, cost, and complications among many others. For this study, age-adjusted mortality rate was the indicator selected to represent health outcome.

Integrated Measures of Health – integrated measures of health combine multiple types of health outcomes into one measure. An example would be the Quality-Adjusted Life Years, which is a measure that was developed for valuing states of health that reflect a person's willingness to exchange extra years of life or the risk of death for improvements in health (Dolan, 2008).

Lifestyle – the typical way of life of an individual, group, or culture (Lifestyle, 2009).

Metropolitan Statistical Area - a core geographic area containing a substantial population nucleus along with adjacent communities having a high degree of economic and social integration with that core (United State Census Bureau, 2008).

Morbidity – the presence of disease within a community.

Mortality – the occurrence of death from disease within a community.

Obese - the label for the range of weight that is greater than what is considered healthy for a given height and falls at 30 or over. It is determined by using weight and height to calculate body mass index. This is not a direct measure of body fat but is correlated (Centers for Disease Control and Prevention [CDC], 2009, January 28).

Overweight – the label for the range of weight that is greater than what is considered healthy for a given height and falls between 25 and 29.9. It is determined by using weight and height to calculate body mass index. This is not a direct measure of body fat but is correlated (CDC, 2009, January 28).

Psychosocial Risk Factors – risk factors that involve both psychological and social aspects. Psychosocial risk factors relate social conditions to mental and physical health (i.e., poor social networks, self –efficacy, chronic stress, etc.).

Race – a general social definition of race recognized in the United States consisting of the following groupings:

1. American Indian and Alaska Native
2. Asian
3. Black or African American
4. Native Hawaiian and Other Pacific Islander
5. White
6. American Indian and Alaska Native *and* White
7. Asian *and* White

8. Black or African American *and* White
9. American Indian and Alaska Native *and* Black or African American
10. >1 percent: Fill in if applicable with multiracial combinations greater than 1% of the population
11. Balance of individuals reporting more than one race
12. Total

The definition does not reflect any biological, anthropological, or genetic criteria, but it is the standard set forth by the United States Office of Management and Budget (United States Census Bureau, 2002).

Self-Organizing Map Cluster – groupings of counties that are statistically similar but not necessarily geographically arranged next to each other.

Social Determinants of Health – a broad range of social exposures that interact and cumulatively relate to a person's and a society's health (Link & Phelan, 1995; Marmot & Wilkinson, 1999). Exposures can encompass such topics as social gradients, development during early life, stress, social exclusion, work, unemployment, social support, addictions, food insecurity, transportation, and access to healthcare.

Social Gradient – the graded relationship between socio-economic status and health outcomes.

Socio-economic Status – a person's, family's, or community's relative position within a hierarchical social structure, based on their access to or control over wealth, prestige and power (Mueller & Parcel, 1981).

Chapter Organization

Chapter I provides an overview of the study, background and statement of the problem being studied, as well as the purpose of the study. The Public Health Model of Social Determinants of Health is introduced and the links being tested in this study are represented. Additionally, research questions and corresponding hypotheses are presented. The chapter ends with discussions on the assumptions and implications of the study and provides definitions used throughout the paper.

Chapter II provides a literature review of differing social determinants of health models, including the Community Guide's social environment and health model, biologic models, and disease-specific models. Additionally, the Public Health Model of the social determinants of health and a review of the components of this conceptual model are presented.

Chapter III details the design and methods used in this study. Information is provided regarding the Self-Organizing Map method and its utility for large numbers of variables.

Chapter IV presents the results of the self-organizing map analysis, as well as the results of the regression analysis used to compare the relationship between health behaviors and the health outcome.

The last chapter, Chapter V, provides a discussion of the study. The researcher's conclusions based on the results presented in Chapter IV and the research questions also are addressed within this chapter. Finally, suggestions for future research are discussed.

CHAPTER II

REVIEW OF LITERATURE

Societal influences shape many aspects of our lives, including our health. Disparities in health status have been recorded among social classes for hundreds of years, dating back to 1662, when John Graunt enumerated disparate mortality among county parishes in England. However, the impact society and social factors have on health have not become apparent until recent times. This chapter examines various models that link societal influences, termed social determinants of health, to health behaviors and health outcomes. Further, a review of current literature surrounding social determinants of health, health behaviors, and health outcome variables is provided. The chapter ends with a discussion of the differing techniques that were considered for this study.

A person's position in society is no longer thought to be the sole indicator of poor health outcomes (Berkman & Kawachi, 2000). For example, with the convergence of medical and social science research, great strides have been made in determining how chronic stress adversely affects health (Cannon, 1935). The human body is a balancing act of many systems that act simultaneously to maintain a homeostatic environment. When the delicate balance is disrupted for long periods of time or for short periods of time repeatedly (both cases classified as chronic stress), the constant, consistent internal environment of the body is altered (Marmot & Wilkerson, 1999). For some people, the alterations to the biological system may result in no adverse effects at all, but for others a variety of acute or chronic health repercussions may be triggered, such as heart disease (Orth-Gomer et al., 2000; Sawchuk et al., 2005), anxiety, and depression (Hiott, Grzywacz, Davis, Quandt, & Arcury, 2008; Orth-Gomer et al., 2000; Sawchuk et al., 2005).

In 2005, the World Health Organization (WHO) formally recognized that health inequalities were pervasive throughout the world. To address these issues and to encourage global change, the Commission on Social Determinants of Health was formed. The Commission found that not only are inequalities present among countries, but they are also present within borders (Marmot, 2007). While studying information on a global, national and local level, the Commission on Social Determinants of Health (Marmot e al., 2007) noted,

> The global context affects how societies prosper through its impact on international relations and domestic norms and policies. These in turn shape the way society, at national and local levels, organizes its affairs, giving rise to forms of social position and hierarchy. Where people are on the social hierarchy affects the conditions in which they grow, learn, live, work and age, their vulnerability to ill-health, and the consequences of ill-health (p.12).

The WHO further states the environment encapsulates the global, national and local levels and state that environmental changes have adverse and inequitable affects upon people around the world. Coastal populations, the poor, and inhabitants of arid, high mountain zones are predicted to be most affected by environmental change (World Health Organization, 2005).

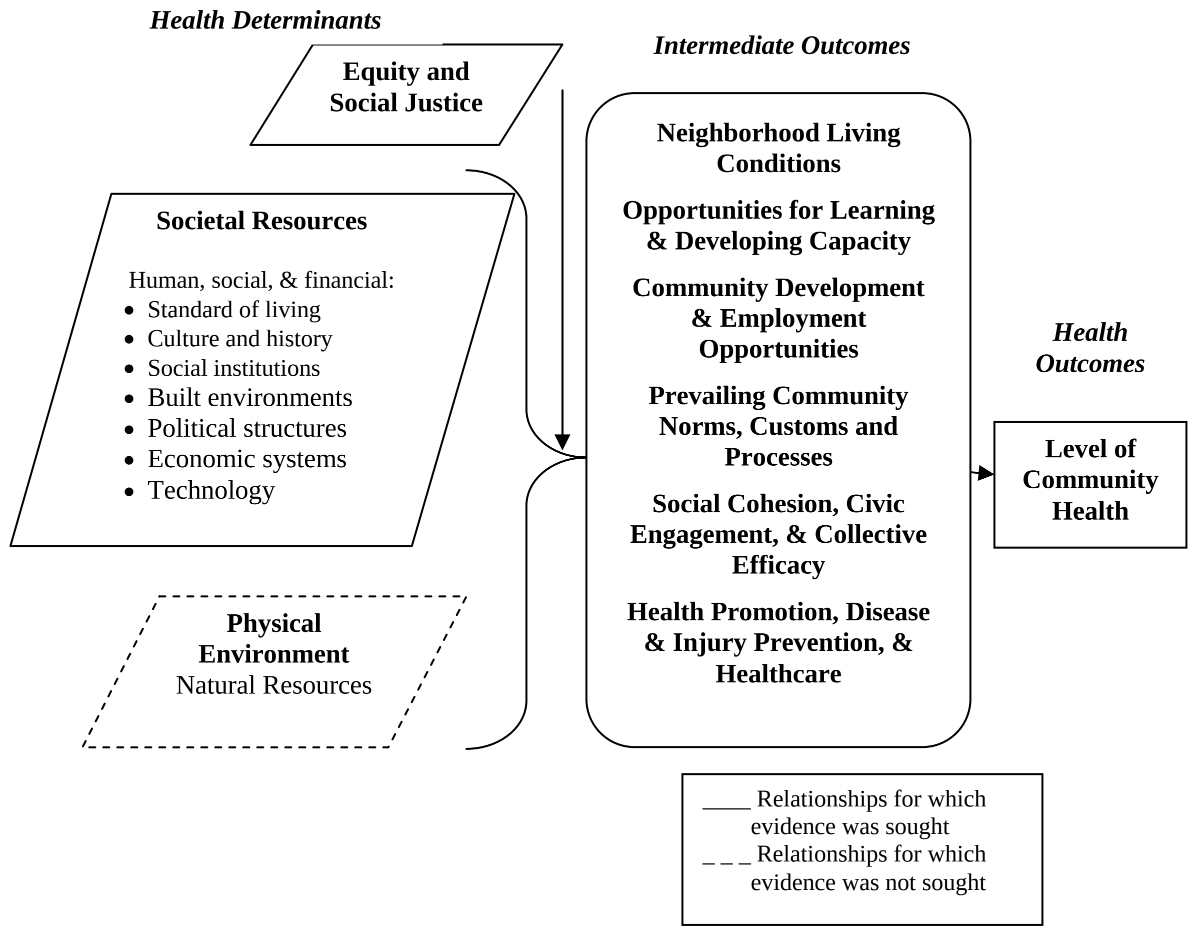
Recently, researchers have shown that social determinants of health 1) directly impact the health of individuals and populations, 2) predict individual and population health better than behaviors, 3) structure the lifestyle choices people make, and 4) interrelate to create individual and societal health (Raphael, 2003). WHO further points out that the influence of the social environment upon health is not a matter of fact or reality for all people and has subsequently identified the following areas as having a social effect upon health: 1) social gradients, 2) early life, 3) stress, 4) social exclusion, 5) work, 6) unemployment, 7) social support, 8) addiction, 9) food, and 10) transportation (Wilkinson & Marmot, 2003). The separate and unique effects, as well as combined effects, of each of these areas on health can be further studied and characterized. Because of the interconnected view of health, society, and the environment, an entirely new branch of epidemiology has formed.

Social epidemiology is a branch of epidemiology that focuses on exposures to social distribution and social determinants of health instead of a specific health outcome (Berkman & Kawachi, 2000). The term *social determinants of health* refers to a broad range of social exposures that interact and cumulatively relate to a person's health (Link & Phelan, 1995; Marmot & Wilkinson, 1999). Further, social and health policies and programs may alter these societal exposures and conditions (Anderson, Scrimshaw, Fullilove, Fielding, & the Task Force on Community Preventive, 2003). Exposures in this field are exemplified by constructs such as social networks, stressors, social gradients, exclusion, political barriers, economic forces, and social behaviors (Berkman & Kawachi, 2000; Raphael, 2006). Social determinants of health can include any of these and many other types of exposures to a person's social environment. Thus far, existing data have not been analyzed with available new techniques. Further, several theoretical models exist portraying the links between societal impacts and health outcomes.

Existing Models

Theoretical models have been developed regarding social determinants of health with various levels of specificity to health or disease processes. Some models take a broad approach to showing the links between social determinants of health and health outcomes while not specifically indicating testable or analytical variables. The Task Force on Community Preventive Services (Anderson et al., 2003) was convened in the mid-1990s and members were appointed by the Director of the CDC under the authority of the U.S. Department of Health and Human Services to study the social determinants of health from an ecological standpoint. The Task Force utilized three broad categories (social institutions, surroundings, and social relationships) as a starting point to identify six intermediate indicators of social determinants of health. The resulting model was designed to identify various aspects of the social environment that are known to affect health. Access to societal resources was the underlying principle for the development of the Community Guide's social environment and health model. According to Anderson et al. (2003, p.12), access to societal resources is what determines community health outcomes. Figure 3 graphically displays the model as a reference for the reader.

Figure 3: The Community Guide's Social Environment and Health Model



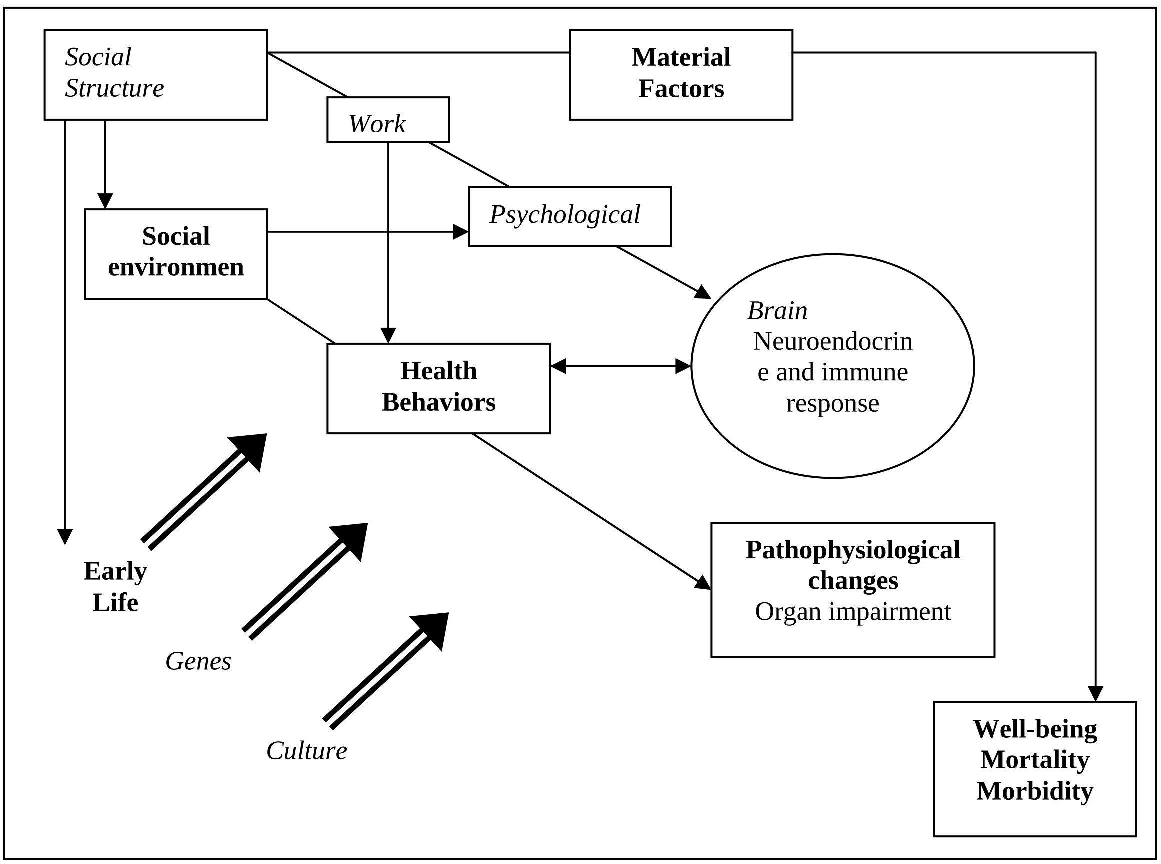
Note: With kind permission from Elsevier: *Journal of Preventive Medicine*, "The Community Guide's Model for Linking the Social Environment to Health," 24(3s), 2003, 13, L.M. Anderson, S.C. Scrimshaw, M.T. Fullilove, J.E. Fielding & the Task Force on Community Preventive Services, Figure 1.

The Task Force used the conceptual model to indicate the links between health determinants and health outcomes (Anderson et al, 2003). They further used the framework to identify interventions that would fit within each of the intermediate outcome categories (not shown in Figure 3). These interventions were developed to help drive change at all levels of society. The reader is referred to Anderson et al. (2003) for an extensive list of interventions.

Other models take a more biological approach to social determinants of health. The question these models attempt to answer is not whether our social environment affects health, but how. They seek to determine the plausible pathways from all aspects, including biology. Michael Marmot and Eric Brunner set forth such a biological model in chapter two of *Social Determinants of Health* (Marmot & Wilkerson, 1999). Their model (Figure 4) sets forth potential pathways in which one's social structure affects all aspects of life, and in turn those affected aspects begin to alter one's physical self resulting in a particular health outcome, whether good or bad. The researchers hypothesize that this biological plausibility is important to begin the discussion of causality. Establishing whether it is truly one's social environment and the interactions taking place within it that is creating poor health outcomes or the reverse is an important and necessary distinction to be made. Marmot and Brunner further point out that the science to determine the directionality of this link is far from complete, indicating that further work needs to be done surrounding data collection and analytical techniques to study the complicated links between social determinants of health and health outcomes.

Beyond conceptual and biological models, other disease-specific models regarding social determinants of health exist. These consist of a cross between conceptual models and biological processes. The University of Chicago's Center for Interdisciplinary Health Disparities Research (CIHDR) model on social determinants of health for breast cancer takes a distinct approach of downward causation. This model emphasizes the idea that upstream social and environmental determinants alter events at lower levels such as individual behavior and individual physiology all the way to the interactions that cells and genetic material have with health and disease (Gehlert et al., 2008).

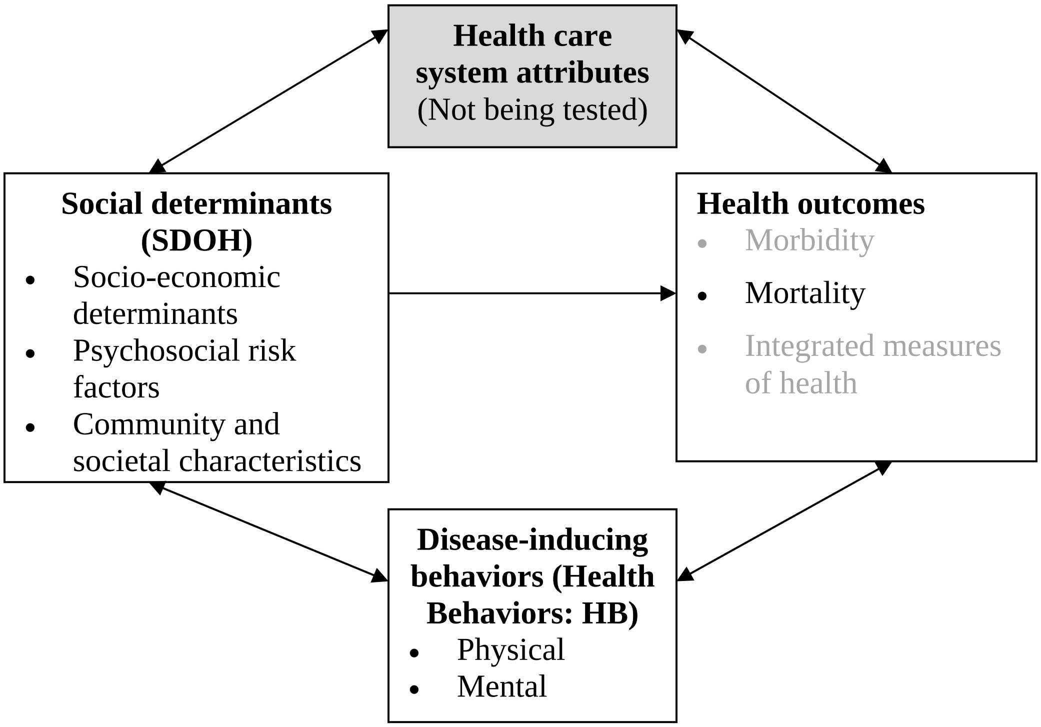
Figure 4: Marmot & Brunner's Social Determinants of Health Model



Note: With kind permission from Oxford University Press: Marmot, M. G., & Wilkinson, R. G. (1999). *Social determinants of health*. Oxford: Oxford University Press.

Given all the models available to evaluate SDOH, a community or social epidemiologist can become overwhelmed in the selection process. To assist in that endeavor, Ansari, Carson, Ackland, and Vaughan (2003) specifically targeted epidemiologists and policy makers with the development of the Public Health Model of the Social Determinants of Health (PHM). This model is a more pinpointed approach to SDOH while taking into account that access to resources and health behaviors play a varying but important contributing role to health outcomes. The full model aids in understanding the structure and relationships of SDOH by indicating potential causal pathways that can be analyzed. Figure 5, while representing the full PHM, indicates the adaptations that were made for this study. The developers of the PHM also identified specific variables (Table 1) for each of the three social determinants dimensions found within the model. Because of the adaptability and testable nature of the PHM, this was the model utilized as the foundation of this study. The remainder of this chapter explores the variables used to test a portion of the PHM. See Figure 2 and Appendix A for the specific data and sources for each variable used within this study.

Figure 5: Full Public Health Model of the Social Determinants of Health



Note: With kind permission from Springer Science+Business Media: *Social and Preventive Medicine*, "A Public Health Model of the Social Determinants of Health," 48, 2003, 243, Z. Ansari, N.J. Carson, M.J. Ackland, & L. Vaughan, Figure 1.

Socioeconomic Determinants

Our surrounding physical environment plays an important role in our health as humans. The environment provides more than just something to look at, but also provides shelter, food, air and many other required resources for daily living. Numerous researchers, both in qualitative (Muhajarine, Labonte, Williams, & Randall, 2008; Walker & Hiller, 2007) and quantitative arenas (Eberhardt & Pamuk, 2004; Fone, Dunstan, Williams, Lloyd, & Palmer, 2007; Probst, Moore, Glover, & Samuels, 2004), have reported a link between place and health. A major aspect of a person's place, or physical environment, is the neighborhood he or she lives in. People who have limited access to grocery stores, fresh fruits and vegetables, safe places for recreation, educational opportunities, income-generating opportunities, and adequate housing are at increased risk for adverse health events such as childhood obesity (Jetter & Cassady, 2006; Kipke et al., 2007; Mushi-Brunt, Haire-Joshu, Elliott, & Brownson, 2007; Sallis & Glanz, 2006) and malnourishment (Jetter & Cassady, 2006; Kipke et al., 2007; Mushi-Brunt, Haire-Joshu, Elliott, & Brownson, 2007; Sallis & Glanz, 2006). Socio-economic determinants can be categorized into three main groups: Education, Socio-economic Status, and Demographics.

Education

Researchers (Case, Fertig, & Paxson, 2005; United States Department of Justice, 2000) provide evidence that health in early life is related to educational attainment (Hack, Flannery, Schlucter, Carter, et al, 2002) and a leading indicator of adult health (Marmot & Wilkinson, 1999; Mueller & Tighe, 2007). The lasting effects of poor development during the early stage of life are not only evident with biological and physical development, but social and family factors as well - poverty, poor family cohesion, or low parental self-esteem (Marmot & Wilkinson, 1999). Giving the best start possible to children leads to future advancements in health status for entire countries.

Ensuring that babies have a healthy weight at birth is the first step to providing a healthy future. Low birth weight babies are at higher risk for numerous issues both in early and late life, such as higher mortality rates within one year of life and by age 17 (Oreopoulus, Stabile, Walld & Roos, 2008). From 2002 to 2006, 8% of the live births (over 4,000) in Oklahoma were considered low (1500 to 2499 grams) or very low (under 1500 grams) birth weight babies (Oklahoma State Department of Health, n.d.b). While there are many causes of low birth weight, some are also linked with poor cognitive development and educational attainment such as maternal tobacco use during pregnancy (Langley, Rice, van den Bree, & Thaper, 2005), maternal infection (Gay, Armstrong, Cohen, Lai, Hardy, Swales, et al., 1995), and malnutrition in utero (Horwood, Mogrdige, & Darlow, 1998). The link between low birth weight and low educational attainment is further substantiated by Hack, Flannery, Schlucter, Carter, et al (2002) in a cohort study of 242 very low birth weight infants (VLBW) compared to 233 normal birth weight controls. Fewer persons who experienced VLBW had graduated from high school when interviewed at 20 years of age compared to persons in the normal birth weight cohort ($p = 0.04$). Additionally, VLBW participant's experienced lower mean IQs ($p < 0.001$) and lower academic achievement scores ($p < 0.001$). VLBW males were also less likely to be enrolled in post-secondary education ($p = 0.002$). Because of this link and the lack of data indicating mother's nutritional status during pregnancy, low birth weight acted as a proxy variable for education in this study.

In addition to low birth weight, violence around schools has been linked to lower educational levels. Grogger (1997) found that high school graduation rates were 5.1% lower in areas of moderate violence. In addition, Grogger found that moderate violence within a neighborhood also translated into a decreased likelihood of students' attending college by 6.9%, and high school graduation levels did not dramatically decrease by more violent situations (graduation rate decreased by 5.7%). However, the likelihood of a student attending college when he or she was from a seriously violent community was reduced by 51% (Grogger).

Even with advances in national and state level violence-reporting systems, difficulties remain in obtaining county-level data addressing issues such as violence around schools (National Violence Prevention Network, 2007). Therefore, educational attainment data from the U.S. Census Bureau were used. However, violence indicators are represented within two other areas of this study, although they are not a direct reflection of community violence in or near schools.

Socio-economic Status

Socio-economic status has been a long-standing indicator for many aspects of society, including health outcomes. Marmot and Wilkinson (1999) pointed out that median income levels are less related to health than the distribution of that income at the national and state level. It is this continuum of health and income distribution that Finch (2003) defined as the social gradient. Mueller and Parcel (1981) defined socioeconomic status as a person's, family's, or community's relative position within a hierarchical social structure, based on their access to or control over wealth, prestige and power. For example, in the United States, 1% of the nation's population possesses over 30% of the nation's wealth (Wolff, 2007). This lopsided distribution of wealth creates dramatic differences in health outcomes (Marmot & Wilkinson, 1999). Additionally, Marmot, Kogevinas, and Elston (1987) stated that health benefits accumulate for persons higher up in the social gradient. Burgard, Stewart and Schwartz (2003) indicate within the United States, occupational status is a measure of social position and, in fact, may be a better indicator of income over a long time period. Within this study, occupational status was represented by the percentage of county populations represented in six categories: management, service, sales, farming and agriculture, construction and production. Further, income was represented by the percentage of county populations categorized into 12 variables. See Appendix A for specific variable information.

Income level and wealth are not the only indicators of SES, however. There are many ways to measure a person's SES and each method adds unique information to the overall picture. Beyond income, epidemiologists also examine indicators such as unemployment rates and education level as indicators of potential earning or spending power. The Townsend and Carstairs indices of social deprivation utilize characteristics as unemployment, car ownership, overcrowding or housing tenure (Berkman & Kawachi, 2000). However, Macintyre and Ellaway (2000) caution analysts about applying such variables to individuals instead of communities. For example, they point out that not all individuals may be unemployed in a community that suffers from a high unemployment rate. While this caution should be heeded while interpreting data, community-level indicators are still significant indicators of poor health outcomes.

To assist in the examining the influences of SES, two approaches have been suggested to discover where differences occur: the *compositional approach* and the *contextual approach* (Duncan, Jones, & Moon, 1998). The *compositional approach* points to individual characteristics of a person's SES. Variables such as employment status, years of education completed, individual annual income or annual household income point researchers to information that can be used to assess a person's risk for poor health outcomes at the individual or household level. Alternatively, the *contextual approach* focuses more on the social network, community, place or geographic area a person inhabits. Variables such as mean housing value, percentage of unemployed persons, and per capita income are all area-based measures (Shavers, 2007).

The U.S. Census Bureau collects contextual variables at recurring intervals to track the SES of the nation. Variables such as percent of persons employed in particular occupational groups, poverty area, working class neighborhoods, percent of owned homes and percent of households owning one or more cars are all indicators that are tracked through the decennial census and the American Community Survey (Shavers, 2007). For this study SES variables were divided into compositional and contextual categories. Compositional variables for this study included the twelve income variables discussed previously. The contextual variables used in this study were occupational status (discussed previously), the average annual unemployment rate for each county, and a variety of housing and homeownership characteristics. See Appendix A for the full list of variables with source information.

Demographics

Thisted (2003) pointed out that some socially constructed labels such as race are not considered determinants of health because of the issue of correlation versus causation. Thisted offered the following example: epidemiologic studies show a correlation between hypertension and African Americans, but the studies are not able to say that being African American causes a biologic susceptibility to hypertension. However, studies do show that race and ethnicity play an important and often confounding role in some disease patterns such as sickle cell disease (Mayfield, 1999). Therefore, such social constructs were included in this study. In fact, Probst, Moore, Glover and Samuels (2004) indicated that race/ethnicity exacerbated issues related to locality, especially in rural locations, and this relationship held true across age groups. Therefore, the following demographic variables were examined: age, gender, race, and ethnicity. While not all socioeconomic determinants were represented in this study, large amounts of available data were utilized to represent the three main groups of socioeconomic determinants (Education, Socio-economic status, and Demographics).

Psychosocial Risk Factors

In addition to socioeconomic determinants of health, psychosocial risk factors play a role in health today. The Merriam-Webster Dictionary defines *psychosocial* (2009) as “involving both psychological and social aspects” or “relating social conditions to mental health.” Martikainen, Bartley and Lahelma (2002) indicated that the interrelation of psychology and the social environment implies that community and social epidemiologists can see psychosocial factors in two ways: 1) psychosocial factors work as mediators to social factors in regard to individual health outcomes, and 2) psychosocial factors are "conditioned and modified by the social structures and contexts in which they exist (p. 1091)."

Because psychosocial risk factors result from the combination of personal psychology and a person's social environment, trying to pinpoint exact mediating or causal factors of health can be difficult. In an effort to organize all the influences of psychosocial factors the concepts of macro-, meso-, and micro-levels are used as a guide (Martikainen, Bartley, & Lahelma, 2002). Constructs such as ownership, legal and welfare structures, and distribution of income are classified as *macro-level* social structures. *Meso-level* psychosocial concepts relate to social networks, family units, how much control one has over one's work environment, the feeling of security and autonomy, or the amount of conflict a person has between one's work and family life. *Micro-level* psychosocial concepts are the individual manifestations of the other two processes. The outward representation of a loss of self-esteem when a person loses a job is one example of a micro-level psychosocial process (Martikainen, Bartley & Lahelma, 2002).

Effecting change at individual levels is important and not without merit. Health educators and medical professionals focus the majority of their careers on helping people make individual or micro-level behavior changes. The Spectrum of Prevention (Cohen & Swift, 1999) includes strengthening individual knowledge and skills as the first level of change. However, it is at the upper levels of the

Spectrum of Prevention in which the greatest change is theorized to occur. Upper levels consist of meso- and macro-level constructs ranging from fostering coalitions and networks to influencing policy and legislation. For a full discussion of the Spectrum of Prevention the reader is referred to Cohen and Swift (1999).

Macro- and Meso-level Psychosocial Risk Factors

Although changing psychosocial factors at the macro-level allow for the greatest impact across all aspects of society, actually making changes in large governmental policies ultimately prove beyond the scope of most community and social epidemiologists and, therefore, beyond the scope of this paper. It is at the meso-level of psychosocial risk factors that community and social epidemiologists can effect change while reaching a large number of people. However, because meso-level factors are measures of society and the relationships and interactions that occur within societies and communities, meso-level factors are discussed within the Community and Societal Characteristics dimension below.

Micro-level Psychosocial Risk Factors

Community and social epidemiologists can also play an important role in the micro-level psychosocial risk factors. It is at this level that variables are usually easily obtained for research. Obtaining a deep understanding of how such micro-level factors alter health is important to knowing how to best alter the outcome. Understanding emotions and how they interact with physical and mental health is one path for discovering how micro-level psychosocial risk factors might affect health.

Researchers have shown interesting links between emotions and adverse health outcomes. Denollet, Sys, Stroobant, Rombouts, Gilbert and Brutsaert (1996) found that persons with established coronary heart disease who scored high on an anxiety trait scale, as well as reporting they were socially inhibited, were four times more likely to die from both cardiac and non-cardiac related issues. A study by Mittleman et al. (1995) showed that anger was one of several triggering factors in myocardial infarctions. In addition to heart disease outcomes, weak links have been seen between emotions and cancer. According to Berkman and Kawachi (2000), links between depression and cancer have shown statistical significance, but the relative risk was much greater in persons who smoked compared to those who did not smoke, which lead researchers to hypothesize that depression and smoking interact to magnify the risk of cancer. Another prospective study (Grossarth-Maticek, Bastiaans, & Kanazir, 1985) found that persons who scored high on a rationality and antiemotionality tool, which is a measure related to the suppression of aggression, had a much higher risk of mortality from all causes of cancer (except lung cancer) over those who scored low on the tool. Additionally, all persons who died of lung cancer during the study scored in the "high" category on rationality and antiemotionality.

For this study, the micro-level emotional and psychosocial data were represented by the following variables: median number of poor mental health days experienced and days feeling nervous, number of persons being treated for depression and anxiety, and measures of isolation.

Community and Societal Characteristics

Encompassing socioeconomic determinants and psychosocial risk factors are the communities and societies in which the determinants and factors exist. The meso-level characteristics of communities and societies have been linked to health outcomes as well as health behaviors. Characteristics such as food insecurity, social networks and support, and violence all provide information about the health of a community.

Food Insecurity

Food insecurity, which is defined as “the lack of access to enough food to fully meet basic needs at all times due to lack of financial resources”, is an increasing problem across the world, including the United States (Food Research & Action Center, 2008). When food insecurity becomes a chronic issue, undernourishment and undernutrition ensue. Serious issues result from chronic food insecurity, such as severe weight loss, stunted growth, low weight, reduced cognitive ability, low productivity, or poor health status. The effects of these issues can last a lifetime (Food and Agriculture Organization of the United Nations, 1999). In the April-May 2008 editorial of the *Lancet (“Finding long-term solutions to the world food crisis”, 2008)*, food is identified as the “fundamental determinant of health (p. 1389).” With the increased use of corn for biofuels, production of staple foods has drastically decreased, resulting in skyrocketing food prices across the world. It is estimated that if prices on staple foods continue to rise, for every percentage rise in food prices an estimated 16 million people will be food insecure. This translates into 1.2 billion chronically hungry people by the year 2025 (Raswant, Hart, & Romano, 2008). The Food and Agriculture Organization (FAO) of the United Nations (2006) predict food security will not be adequately or drastically changed without “cooperation with international organizations and civil society – including both public and private sectors (p. 7).” The FAO further indicate the right policies, as well as necessary resources and political will or desire need to be in place for change to occur. Food insecurity was measured in this study through a weighted population measure. According to the Kerr Center for Sustainable Agriculture (McDermott, 2006), 15% of Oklahoma households were found to be food insecure. This percentage was applied to the number of households within each county to obtain the weighted number of households who were food insecure for this study.

Social Networks and Support

Poor social networks and a lack of social support are two other meso-level risk factors that are linked to adverse health outcomes. Berkman (1984) distinguishes between the two concepts, indicating that social networks are the ties between a person and others around them and social support is an exchange of some tangible or intangible item (i.e., emotion, goods, services, information), concluding that a social network is not necessarily a supportive environment. Several studies have shown a link between supportive social networks and decreased mortality (Eng, Rimm, Fitsmaurice, & Kawachi, 2002; Iwasaki et al., 2002; Murberg & Bru, 2001; Rutledge, Matthews, Lui, Stone, & Cauley, 2003). Social isolation repeatedly has been demonstrated to be a mediating factor within these studies. Although a social support question exists as part of the Behavioral Risk Factor Surveillance System that measures the percentage of people who receive the social and emotional support they need, data have not been collected for this variable within Oklahoma (CDC, n.d.) nor for any other state to date. Several social and community participation variables were used as proxies for social network and support, such as the amount of contributions received by charitable organizations, charitable organizations' total reported assets, the percent of registered voters by the three major political parties, the number of churches found in a community and the number of congregational members within a county. Faith-based organizations provide a unique opportunity for community members to seek support and information (Washington State Department of Social and Health Services, 2006). According to the World Health Organization, health, religion and cultural norms guide health-seeking strategies. The WHO further states that public health needs to expand collaborative relationships with these organizations in order to extend the reach of service (Haddad, Olivier, & De Gruchy, 2008). Knowing the links between social cohesion, support, and health exists presents an incomplete picture. Further work must be conducted to determine why social networks and socially supportive environments improve health. Community and social epidemiologists can prove to be strong proponents of such research.

Violence

Violence is another indicator of community and social characteristics that has an effect on health. In this study, county crime rates, the number of domestic violence reports, and the number of domestic violence services offered within a county were examined as indicators of violence. Both domestic violence and violence within the community have been linked to poor mental health outcomes in children and adults (Martinez & Richters, 1993; Osofsky, 1999), as well as developmental delays in children (Osofsky, 1999). With constant exposure to community violence a part of some people's everyday lives, focusing on physical and mental health issues is less of a priority. Researchers have shown that people in low socioeconomic communities are at higher risk for exposure to everyday violence (Martinez & Richters, 1993; Osofsky, 1999). Community and social interventions that reduce exposure to violence can potentially impact several areas of health and development.

Other indicators of community and social characteristics, as identified by the developers of the PHM are poverty, residence type, income inequality, rural and urban populations, and altruism (community-based giving). These indicators were represented within this study as well, but they have already been discussed in other sections of this paper. See Appendix A for the specific variables representing these concepts.

Social Determinants Variables for this Study

In order to assist persons interested in analyzing social issues, the University of Michigan and the CDC worked with leaders from around the world to identify data sources that represent social determinants data at a community level. The resulting *Data Set Directory of Social Determinants of Health at the Local Level* is a collection of existing data sources that represent social determinants and primarily focus on data sets that can be obtained at the metropolitan statistical area (Hillemeier, Lynch, & Casper, n.d.). The Data Set Directory guided the selection of data sources to represent the variables contained in the Public Health Model of the Social Determinants of Health as accurately as possible, while maintaining statistical stability and confidentiality. For these reasons, all data was represented at county level and not the metropolitan statistical area. Appendix A and Appendix B contains a list of all social determinants of health variables and health behavior variables to be used in this study.

Health Behaviors

According to the CDC (Kilmer et al, 2008), health behaviors are linked to the leading causes of death. Behaviors such as tobacco use, poor nutrition, lack of physical activity, and lack of vaccinations, among others, have been linked to adverse health outcomes (Holth, Wepen, Zwart, & Hagen, 2008; Kwong, Stukel, Lim, McGeer, Upshur, et al. 2008; Stewart, Cardinez, Richardson, Norman, Kaufmann, et al., 2008). Controlling these risky health behaviors may lower morbidity and mortality (Kilmer et al., 2008). Adler and Newman (2002) elude that changing behaviors will change health outcomes. When examining the underlying causes of mortality and morbidity, modifiable behaviors (Table 2) prevail as identified by Mokhdad, Marks, Stroup, and Gerberding (2004).

Table 2. Actual Causes of Death in the United States in 1990 and 2000

Actual Cause of Death	No. (%) in 1990	No. (%) in 2000
Tobacco Use	400,000 (19)	435,000 (18.1)
Poor Diet and Physical Inactivity	300,000 (14)	365,000 (15.2)
Alcohol Consumption	100,000 (5)	85,000 (3.5)
Microbial Agents	90,000 (4)	75,000 (3.1)
Toxic Agents	60,000 (3)	55,000 (2.3)
Motor Vehicle	25,000 (1)	43,000 (1.8)
Firearms	35,000 (2)	29,000 (1.2)
Sexual Behavior	30,000 (1)	20,000 (<1)
Illicit Drug Use	20,000 (<1)	17,000 (<1)
Total	**1,060,000 (50)**	**1,159,000 (48.2)**

Leading the list was tobacco use, which accounted for 18% of deaths in the United States in 2000. Smoking is related to over 30% of all cancers in the United States and 87% of lung cancer deaths (American Cancer Society [ACS], 2008). Smoking is also a major contributor to other chronic poor health outcomes such as heart disease, stroke, and respiratory diseases (ACS, 2008). Male smokers have a 23 times higher risk of developing lung cancer than non-smoking males, and female smokers have a risk of 15 times that of non-smoking females (ACS, 2008). Oddly, there is no difference in risk among smokers of “light” or “low-tar” cigarettes versus regular cigarettes (ACS, 2008). The risk of cheek and gum cancer among long-term snuff users increases nearly 50 times over non-users (ACS, 2008). Although annual cigarette consumption is decreasing in the United States, snuff manufacturing has increased by more than 75% in the past decade (ACS, 2008).

Poor nutrition and physical inactivity ranked as the second actual cause of death in the United States in 2000, accounting for approximately 365,000 deaths or 15% of the total number of deaths (Mokhdad, Marks, Stroup, Gerberding, 2005). As discussed earlier, poor nutrition has adverse effects at several levels, including physical and mental development (Horwood, Mogrdige, & Darlow, 1998). The term “poor nutrition” does not distinguish between choosing a poor diet and not having access to healthy foods, which may lead to erroneous assumptions being made about analytical results on “behavior” data. However, what can be studied are deaths due to being *overweight* and *obese* as a proxy. Mokhdad, Marks, Stroup and Gerberding (2004) pointed out that the number of overweight deaths (overweight and obese together) had the most impact on the number of deaths attributed to poor nutrition and physical inactivity. In Oklahoma, 36% of adults were overweight and 27% were obese in 2005 (Oklahoma State Department of Health [OSDH], 2007). Overweight and obese combined accounted for 63% of the adult population in Oklahoma (OSDH). The overwhelming outcome of poor nutrition and physical inactivity is diabetes, and Oklahoma ranked the 44th worst state in the nation for the percent of persons being diagnosed as diabetic (OSDH). The costs of being overweight and obesity are soaring. In 2000, obesity alone accounted for an estimated $117 billion in total costs, with over half of that (52%) being direct medical costs (CDC, 2008, September 15). Finkelstein, Fiebelkorn, and Wang (2003) estimated that costs may have reached as high as $78.5 billion, with approximately half of those costs paid by taxpayers through Medicare and Medicaid dollars. For Oklahoma, this means Medicare and Medicaid pay for approximately $390 million in medical costs associated with obesity.

Along with tobacco use and obesity, inappropriate alcohol use is associated with various medical and social problems. While researchers have shown light use of alcohol (i.e., red wine) can have beneficial effects on cardiovascular disease, researchers have also shown that overuse of alcohol can lead to serious adverse outcomes, and intake higher than 1 to 2 drinks per day is linked to increased total mortality (Goldberg, Mosca, Paina, Fisher, 2001). It is estimated that excess alcohol consumption accounted for approximately 85,000 deaths (3.5%) nationwide in 2000. Further, if previous alcohol drinkers were included in the calculations, the attributed deaths would increase to 140,000. For persons who consumed an excess amount of alcohol, Australian researchers (Ridolfo & Stevenson, 2001) showed increased relative risk (RR) of five different cancers, including breast cancer (RR 1.59), cerebrovascular disease (RR ranged from 1.06 to 7.98), hypertensive heart disease (not reported directly), and chronic liver disease and cirrhosis (9.54).

Immunizations and vaccines have done a great deal to move our country through the epidemiologic transition from infectious disease deaths to chronic health-related issues (Yusuf, Reddy, Ôunpuu, & Anand, 2001). Between 2001 and 2005, pneumonia and influenza accounted for over 300,000 deaths in the United States. A study conducted by the CDC showed that an average of 200,000 people a year are hospitalized for respiratory and heart disease complications due to influenza infections (CDC, 2004, September 22). Vaccination rates are on the rise in the United States. According to the 2007 Behavioral Risk Factor Surveillance System, 67.2% of adults aged 65 and older had received the pneumococcal vaccination and 71.9% had received an influenza vaccination. Both vaccination rates were up from 1995 where only 37.8% of adults 65 and over were vaccinated for pneumonia and 60.1% had received an influenza vaccination. For this study, both influenza and pneumonia vaccination rates were used as health behavior indicators.

Injuries are a significant source of years of potential life lost (a measure of premature death) in the United States, because the average age of death for injuries is much lower than the average age of death for other causes. Unintentional injuries are the leading cause of death for persons aged 1 to 44 and are in the top 10 leading causes of death for the remaining age groups. Overall, unintentional injuries were the 5th leading cause of death, accounting for 117,809 in 2005. Motor vehicle injuries were the number one contributor to this problem, accounting for 37% of all unintentional injuries (Centers for Disease Control and Prevention, 2009).

Lack of seatbelt use and inappropriate or improper use of child restraints are significant risk factors for injuries and death from motor vehicle incidents. The seatbelt usage rate is 75% in the United States, which is the highest it has ever been, but this rate is still much lower than other countries (Gantz & Henkle, 2002). Other industrialized countries such as Great Britain, Sweden, and Canada have seatbelt usage rates of 90% (Gantz & Henkle, 2002). Proper seatbelt use increases a person’s chance of surviving a motor vehicle crash by 45% and reduces injuries by 50% (Gantz & Henkle, 2002). Child passenger restraint usage often mirrors, and is dictated by, the adult drivers and their seatbelt use. According to the National Safe Kids Campaign, almost 40% of children who were not wearing a seatbelt were riding with adults who were not restrained themselves, while only 5% of children were unrestrained if the adult driver was wearing a seatbelt (Cody, Mickalide, Paul, & Colella, 2001).

The National Institute of Mental Health reports that mental illnesses are the leading cause of disability in the United States for persons aged 15 to 44. Additionally, suicide is listed as one of the top five leading causes of death for persons aged 10 to 54 (National Institute for Mental Health, 2008). The Youth Risk Behavior Surveillance System showed that 14.5% of high schools students had seriously thought about committing suicide in the previous 12 months and 7% stated they had actually tried to commit suicide at least once (CDC, 2008, June 6). While it is clear that mental health has a tremendous impact on morbidity and mortality in this country, it is difficult to obtain data related to mental health screening or behaviors related to mental health on a large scale that would serve as true behavior measures. However, the Behavioral Risk Factor Surveillance System does provide access to some intermediate measures related to mental health. The United Health Rankings utilize one of the intermediate measures as an indicator for poor mental health. According to the 2008 United Health Rankings, the average number of poor mental health days experienced by adults in the past 30 days was 3.4 days. Oklahoma adults, however, experienced an average of 3.9 poor mental health days, which translated to a ranking of 47th worst in the nation (United Health Foundation, 2008). While work still needs to be done on national data collection to adequately assess mental health behaviors such as screening, information related to self-reported mental health (i.e., poor mental health days, number of days that are restricted because of mental health issues, sexual violence, and intimate partner violence) was examined in this study.

It can be difficult to fully appreciate the reduction in mortality rates that preventive screenings provide. The American Cancer Society (2008) estimates that almost 1.5 million people have been diagnosed with some type of cancer and over one-third of them will die as a result. At least one- half of all new cancer cases that occur each year can be prevented by early detection and screening. Screening not only reduces mortality, but also reduces morbidity rates through earlier detection, thus allowing earlier treatment. For example, the overall 5-year survival rate for persons diagnosed with colorectal cancer is 64%. When diagnosis occurs at earlier stages (when the cancer is localized to one location in the body) the 5-year survival rate increases to 90%. Because of low screening rates, only 39% of colorectal cancer cases are caught at the early stage. However, screening for breast cancer has been tremendously beneficial. Approximately 80%-90% of all breast cancer cases are detected by mammography in women with no other symptoms of breast cancer. Visiting a doctor regularly and obtaining appropriate screenings for a variety of health issues is one way to control adverse health outcomes.

Health Outcome

The final variable to be used within this study represents the health outcome portion of the PHM. Health outcomes have been defined by the Health

Outcomes Library Core Project (AcademyHealth, 2004), as going beyond just the presence or absence of mortality, but may examine additional issues such as quality of life, longevity, morbidity, psychosocial functioning, cost, and complications, among many others. While it is encouraging to see that measures of health outcomes are broadening to allow for new types of analyses, the current study focused on one health outcome, the presence of mortality within a county. County mortality was represented by age-adjusted mortality rate (AAMR), a statistic that indicates the risk of death from an event. Mortality rates are good indices of the severity of events or problems within a community (Gordis, 1996). Mortality rate is a proportional representation of the number of people who have died from an event within a specified time period in relation to the total population of people susceptible to the event within the same time period, in this case death experienced by Oklahoma residents from any cause occurring from 2000 to 2006, and stratified by county.

Statistical Method

In order to obtain a complete picture of social determinants of health data and how they affect a community, selecting the correct variables is of critical importance. Community and Social Epidemiologists must examine innovative methods to make sense of the highly complex and interconnected social determinants of health data. A statistical method was sought to examine the interwoven nature of the social determinants of health data for this study. A functionality that needed to be present in any method selected for this study was the ability to fully and accurately represent the data in its entirety without having to pre-select or reduce the number of variables submitted. This required the statistical technique to be extremely flexible. The intent was to find a method that did not require a priori selection of the number of clusters that were discovered within the data, have the ability to analyze variables at differing levels (i.e., ratio, ordinal, nominal), and be able to handle large amounts of data records and/or variables. Self-organizing maps (SOM) addressed all of the requirements for the intended analysis and allowed for iterative processing of the data to ensure the best mathematical representation of data was achieved while preserving the underlying structure of the data (Erb, 1993; Kohonen, 2001).

Other techniques were examined that had similar functionality but were lacking in some way. Similar cluster algorithms, like K-Means and multidimensional scaling allowed for inclusion of all data elements into the process, but an analyst was required to select the number of clusters they think were in the data as in K-Means (Bradley & Fayyad, 1998). The unsupervised SOM process automatically selects the number of valid nodes (clusters) without any a priori notions from the analyst, ultimately taking the multi-dimensional inputs and reducing them to a bi-dimensional output while preserving the integrity of the original data (Molinier, Laaksonen, & Hame, 2007; Verdu, Garcia, Senabre, Marin, & Franco, 2006). Further, the nature of SOM keeps the most similar data points together, while multidimensional scaling seeks to preserve differences. The choice between the two relies on the research question being asked; however, results from both methods are very similar (Kirt & Vainik, 2007).

SOM Software

The SOM algorithm is part of several commercial statistical packages with differing graphic and output capabilities (Kohonen, 2001; SAS Institute, Inc, 2005; Viscovery, n.d.). Free versions of the algorithm are also available and can run as a stand-alone dos package, SOM_pak (Kohonen, 1996), or as an extension for Matlab software (*SOM Toolbox for Matlab*, 2001), but Matlab software must be purchased.

The ability of the SOM to analyze the full extent of all variables submitted to the process remains a basic function of the SOM algorithm no matter which SOM package is chosen by a community or social epidemiologist. However, each software package has unique options that cause the data to be processed in differing manners. For example, SAS Enterprise Miner only allows for a square or rectangular grid (SAS Institute, Inc., 2005), but this is not Kohonen's (2001) preferred method. The Viscovery SOMine 5.0 from Eudaptics utilizes a hexagonal grid (Viscovery, n.d.), the preferred method (Kohonen).

Another feature of SOM software that must be considered is cost. As mentioned previously there are several free versions of the software but they have limitations. The SOMpak runs in a dos environment and requires dos programming. The extension for Matlab, although free, requires the user to purchase the Matlab software or have access to an already purchased license (*SOM Toolbax for Matlab,* 2001). The SAS Enterprise Miner software is often found in many University computer labs but can be extremely costly to purchase, approximately $50,000 (T. Adkins, personal communication, May 13, 2008).

Viscovery SOMine 5.0 is available for a free trial with full functionality but limits the number of variables that can be processed to 100 (Viscovery, n.d.). Viscovery removed this restriction for this study.

After examing several options for analytical SOM software, the Viscovery SOMine 5.0 was used for this study because of the ability for the software to operate on a small computer and not require a server environment, the ability to create hexagonal grid, and the free trial version.

Summary

Understanding the social determinants of health and the important role they play in health outcomes is a vital step in being able to effect change. It is not surprising to see the three dimensions of the Public Health Model of the Social Determinants of Health and the variables within the dimensions overlap. The interplay between socioeconomic determinants, psychosocial risk factors, and community and social characteristics reiterates the fact that it is not just one concept or dimension that causes adverse health outcomes. Verifying the links between social determinants of health, health behaviors and the health outcome is an important step in moving forward to achieving far-reaching changes in the health of our citizens. Community and social epidemiologists have begun to discover the individual and combined effects of each social determinant of health, but further research must be conducted to confirm these pathways. New methods must be utilized to facilitate the advancement of this ever-growing research area.

It is within the scope of the Public Health Model of the Social Determinants of Health that the county-level variables for this research were selected and utilized with the Self-Organizing Map method that gives community and social epidemiologists another tool in the war against poor health.

CHAPTER III

METHOD

This study determined the utility of the Self-Organizing Map (SOM) in the analysis of social determinants of health by addressing the three research questions:

Phase 1. What is the underlying relationship among social determinants of health (SDOH) and health behaviors (HB) within Oklahoma's counties?

Phase 2. Is a SOM_{SDOH} cluster variable a stronger predictor of health outcome (Age-Adjusted Mortality Rate) than a SOM_{HB} cluster variable within Oklahoma counties?

Phase 3. What is the relationship between SDOHs, HBs, and health outcome within Oklahoma counties?

A modification of the Public Health Model of the Social Determinants of Health (PHM) provided the conceptual foundation for the variables that were used in the analysis. Variables represented three dimensions of social determinants (socio-economic determinants, psychosocial risk factors, and community and societal characteristics), as well as health behaviors (physical and mental) and health outcome (age-adjusted mortality rate). Data were obtained from existing data sources (archival data) to represent specific variables within each social determinants of health dimension at the county level. Analyses were conducted for all 77 counties in Oklahoma. Variables that were typically classified as categorical, such as demographics, were transformed to ratio-level variables by the creation of a new variable for each category which consisted of the percentage of the population that positively identified with that category. For example, race is reported by the U.S. Census Bureau within a single variable with five categorical options to choose from and a person may select as many of the categories as they identify with, allowing for a variety of combinations. For this study, the percentage of the population within each county who self-identified as white were placed into a new variable labeled Dem_Race_W. This process was repeated for every category within each categorical variable used within this study. Once all data were transformed as needed, all ratio-level variables representing social determinants of health (SDOH) were submitted to the SOM algorithm simultaneously and related mathematically.

SOM analysis is an iterative process in which vector weights are examined and the algorithm determines mathematical neighborhoods with the ultimate result being a new variable indicating cluster membership for each county. The SOM process was repeated for health behaviors (HB). The results of the individual SOMs were examined for the underlying relationships among the variables. Comparisons across SDOH and HB variables were not possible from the SOM process, but the SOM process did allow the Phase 1 research question to be answered.

The resulting SOM cluster variables for SDOH and HB were then dummy and submitted to a standard multiple regression analysis along with Age-Adjusted Mortality Rate (AAMR), which represented the health outcome (HO) using SAS 9.1 (SAS Institute, Inc., 2002). The SOM_{SDOH} dummy variables (three total variables representing the four categories from the SOM_{SDOH} variable) were entered into the model with the five SOM_{HB} dummy variables. The regression analysis tested the theoretical links of the Public Health Model of Social Determinants of Health (PHM) and answered the research question of whether SDOH or HB was a stronger predictor of the specified health outcome.

The SOM_{SDOH} vectors, SOM_{HB} vectors, and AAMR were then correlated to verify the underlying relationships. Point-biserial correlations were examined between the SOM_{SDOH} vectors and the SOM_{HB} vectors because of the dichotomous nature of the vectors. Phi correlations were used when examining the continuous AAMR variable to dichotomous SOM_{SDOH} or SOM_{HB} dummy variables. Additionally, SOM_{SDOH} and SOM_{HB} clusters and AAMR were represented spatially through ESRI's ArcGIS (geographic information system) software to aid in the final interpretation.

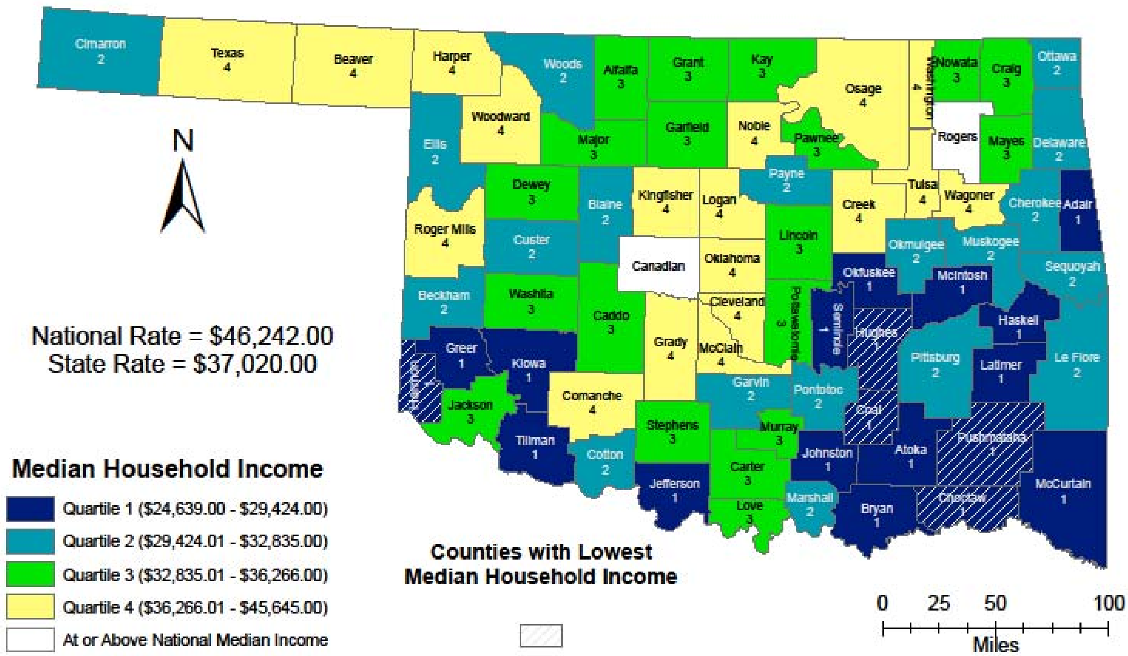
Study Site Description

The state of Oklahoma was used for this study because of the availability of a wide variety of county-level data representing social determinants of health, health behaviors, and the health outcome that could be used for model inputs. Additionally, many health behavior and health outcome variables were available at lower geographic levels, such as city, zip or census tract, in the event they should be needed for future analysis.

The U.S. Census Bureau estimated the average population of Oklahoma from 2005 to 2007 was 3,576,929 with a wide variety of ethnicities and income distribution found within the state. The majority of the population in Oklahoma was classified as white (75.4%), but Oklahoma had eight and one-half times more Native Americans and three times as many persons who identify themselves as more than one race than the national average. Oklahoma's median household income (2005 estimate) is 20% below the national average ($37,020.00 versus $46,242.00). Figure 6 shows the median household income distribution by county for the state of Oklahoma. A clear geographic distribution of income is present within Oklahoma. Four of the five counties with the lowest median household income (Choctaw, Coal, Hughes, Pushmataha), indicated by the diagonal hatching lines, are in the southeastern portion of the state.

In addition to income, other issues have visible geographic distributions at the county level within Oklahoma. As seen in figure 7, total mortality follows a very similar geographic distribution to income with most of the upper quartile counties falling in the southeastern portion of the state. While these maps do not show a causal link between low-income levels and poor health, the similarities are an indication that disparities exist within county borders.

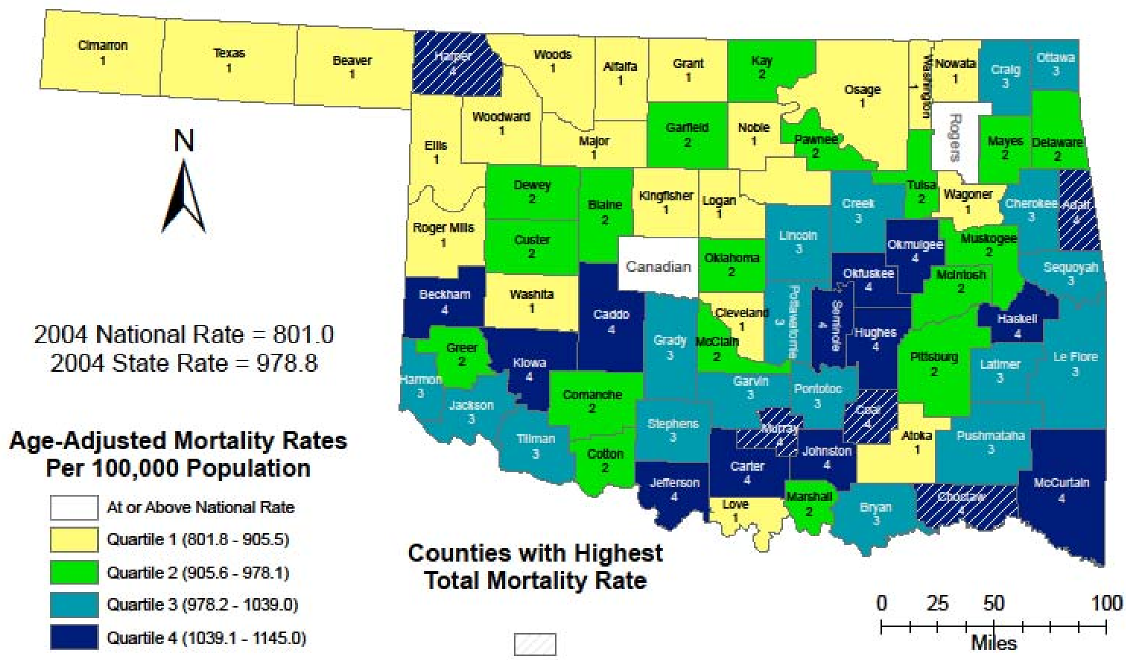
Figure 6: Median Household Income by County, Oklahoma, 2005 Estimates



Note: Quartile number is under the county name.
Source: Small Area Income & Poverty Estimates - Oklahoma Counties, U.S. Census Bureau, Data Integration Division, Small Area Estimates Branch, As obtained from http://www.census.gov/hhes/www/saipe/index.html on January 23, 2008

Obtaining readily available data at lower geographic levels such as census tracts or block groups would give better representations of community level occurrences. However, social and community epidemiologists must balance the finer detail of lower geographic levels with the sample and event sizes of the data. Accuracy and stability of statistics based on small sample sizes are questionable and maintaining confidentiality becomes difficult (Johnson, 2004). In order to assure stability and adequate sample size all analyses for this study were conducted at the county level to allow for full representation of the state of Oklahoma.

Figure 7: Total Mortality by County, 2002-2006, Age-Adjusted Mortality Rates



Note: Quartile number is under the county name.

Source: Oklahoma State Department of Health Vital Statistics 2002-2006, as obtained from http://www.health.ok.gov/ok2share on November 31, 2009.

Data Selection Criteria

Variables representing the social determinants identified by the developers of the PHM were selected from multiple sources and categorized into the three social determinants dimensions: socio-economic determinants, psychosocial risk factors, and community and societal characteristics. Additionally, variables representing health behaviors were selected to represent a wide range of personal behaviors that are associated with positive and negative health outcomes. All variables selected met the following criteria:

1. Available for most counties in Oklahoma;

2. Data timeframe must be consistent with the study timeframe (2000-2006) but not necessarily identical (i.e., the data point can represent a single year or point in time as long as it is within the study timeframe); and
3. Representative of social determinants, health behaviors or the selected health outcome (Age-Adjusted Mortality Rate)

Variables in Study

The three SDOH dimensions were represented by 114 variables distributed into 24 categories (ten categories for socio-economic determinants, four categories for psychosocial risk factors, and ten categories for community and societal characteristics). Thirty additional variables were used to represent physical and mental health behaviors within this study. Appendix A and Appendix B contain lists of all social determinants of health and health behavior variables in the study. Both appendices relay the dimension in which the variable is situated (i.e., Socio-Economic Status, Community and Societal Characteristics, Physical Health), the variable label, a description of the variable, and the source for each variable.

The dependent variable analyzed in the study was Age-Adjusted Mortality rate (AAMR) for all causes of death combined. AAMR reflects the criteria that have been outlined above for SDOH and HB variables because it was available for every county in the state and the date represented the aggregated rate of death from all causes from 2000 to 2006. The mortality rate was normalized to account for varying age-distributions within county populations in Oklahoma by using the direct method of age-adjustment. Age-adjusting by the direct method involves weighting the age-specific death rates for each county by a standard set of weights, which represent the proportion by age in a standard population. This allowed for an unbiased comparison of mortality across county borders (Gordis, 1996; Mausner & Kramer, 1985). The Oklahoma State Department of Health's (n.d.b) online queryable database called OK2Share, from which the AAMR was obtained, produces age-adjusted rates by request using the state's standard population for 2000. The general formula used by the OSDH is:

$$\text{Age Specific Mortality Rate} \times \frac{\text{Population within that age group}}{\text{OK 2000 Population}}$$

There were 77 records within the final dataset, one representing each county in Oklahoma.

Statistical Methods

The analysis for this study was conducted in three phases. Phase 1 identified the underlying links among social determinants of health and revealed the mathematical patterns that underlie the data through the use of a Self-Organizing Map (SOM). The health behaviors were also subjected to the SOM for analysis. Phase 2 tested model links between the resulting SOM clusters (SDOH and HB) and the identified health outcome, AAMR, using regression analysis. The final step (Phase 3) in the analysis tested the correlations between SDOH, health behaviors, and AAMR.

Phase 1 - Self-Organizing Map

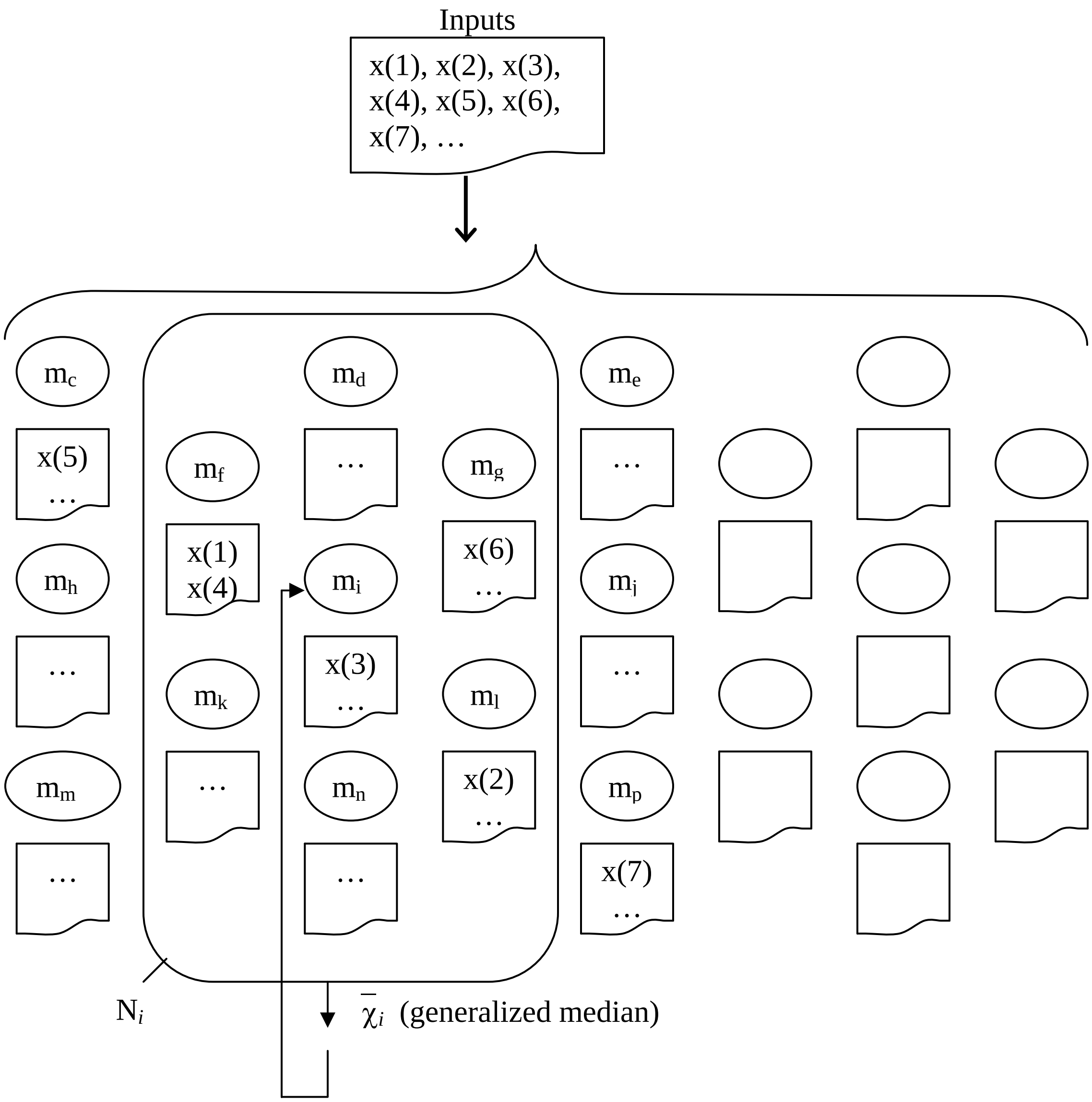
Advancements in computer technology have allowed for new methods of data analyses. One of these is the Self Organizing Map (SOM) algorithm developed by Teuvo Kohonen (2001). The SOM is a data reduction technique that allows for visualization of the underlying patterns (i.e., a mathematical map) found within the data. SOMs, and other auto neural network techniques, analyze data through nodal connections (also called neuronal connections). This is an attempt to duplicate how the human brain processes information through its neuron and synapse network (Erb, 1993; Kohonen, 2001). The SOM can be used to pre-process data in order to accomplish additional analyses while representing the full array of data in a single variable (Vesanto & Alhoniemi, 2000). Unlike traditional cluster techniques, the SOM process is nonlinear and iterative. The algorithm also awards those nodes that are mathematically near the winning node while inhibiting nodes that are farther away (Basara, 2006; Kohonen, 2001). Additionally, a target vector is not required for the unsupervised SOM to learn. This is different from other neural network techniques such as Back Propagation, which requires analyst supervision throughout the learning process to confirm target vector classification (Erb, 1993). During the self-learning process nodes are optimally categorized in order to reduce the amount of space between similar nodes in a cluster and increase the space between dissimilar clusters. The clustering of the nodes allows for the multidimensional data to be displayed in a two-dimensional map for visual representation while preserving the original topography of the data (Kohonen, 2001).

SOM Processing

Kohonen (2001) sets forth the basic process for the SOM, which is also visually displayed in Figure 8:

1. A two-dimensional ordered array of model nodes (m_i) is created and weights for each node are initialized. It is suggested that a hexagonal, triangular, or rectangular array is used to obtain a stationary state. Bacao et al. (2005) refer to this array as the output space.
2. Vector inputs (x(1), x(2),…x(t)) are submitted and compared to the model node (m_i). Vector inputs are the variables submitted to the SOM.
3. Each vector is placed in a model node sublist based on weight similarity (usually based on Euclidean distance measure $d(x, m_i)$).
4. Every vector weight is then examined to determine the Best Matching Unit (BMU), which is the vector weight that is most like the nodal weight.
5. The neighborhood of the BMU is then calculated by summing the distances within the neighborhood. The neighborhood (N_i) consists of all nodes within a set radius of the BMU.
6. The BMU and its neighbors are rewarded allowing the neighboring nodes to become more like the BMU vector. The weights that are farther away are not rewarded.
7. This process is repeated from Step 2 until a stationary state is obtained. During the iterative process, input vectors were assigned to different model nodes until the smallest distances within a node are achieved. Over time neighbors of BMU's decrease.

Figure 8: Visual Illustration of the SOM Process



Note: With kind permission of Springer Science+Business Media: *Self-Organizing Maps,* 3rd edition, 2001, p.107, "Chapter 3 - The Basic SOM", Kohonen, T. , Figure 3.1.

- w_{ij} = the weight vector associated with the node positioned at $column_i$ row_j
- x_k = the vector associated with pattern k

- d_{ij} = the distance between w_{ij} and a given pattern
- r = radius of neighborhood: indicates the size of the neighborhood around the winner node; it defines the topology of the SOM, must converge to 1 or 0.
- h = the neighborhood function: assumes values in [0,1]; is a function of the position of two units (the winner node and another unit) and radius (*r*); large for units close in output space (closer to 1) and small (0) for those that are far away; maximum is usually at the center of the neighborhood.
- α = the learning rate: this varies between 1 and 0, but it must converge to 0 in order to obtain a stationary state and a stable SOM; usually decreases linearly

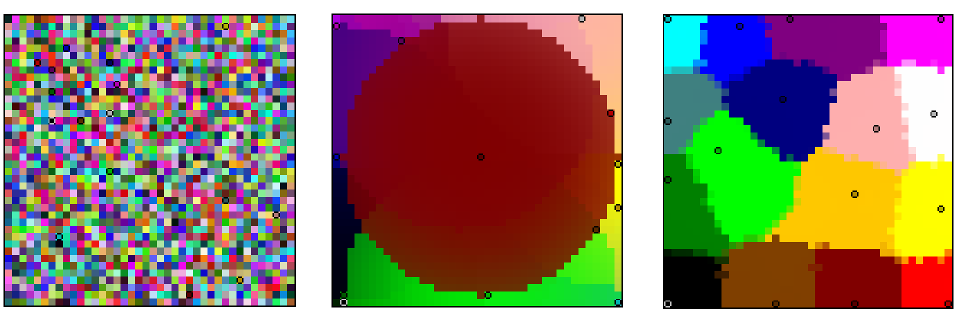
For each input pattern, four steps are needed:

1. Calculation phase: $d_{ij} = || x_k – w_{ij} ||$
   - This calculates the distance between the pattern and all nodes of the SOM
2. Voting phase: $wwinner(w_{ij} : d_{ij} = min(d_{mn}))$
   - This phase selects the nearest node as winner
3. Updating phase: $w_{ij}=w_{ij}+\alpha_h(wwinner,w_{ij})||x_k-w_{ij}||$
   - This phase updates each node of the SOM according to the update function
4. Repeat phase: The first 3 steps are repeated and the learning parameters (α and r) are updated until stopping criteria are met.

Ultimately, a successfully trained SOM occurs when the mathematical patterns that were hidden but close in the input space are mapped to the same nodes or nodes that are close in the output space (Bacao, Lobo & Painho, 2005; Kohonen, 2001).

A common example of SOM is color clustering. If one were to present a random selection of color (Figure 9a) identified by the amount of red, green, and blue that is present within each color, the iterative process of the SOM would organize all the shades of red into one corner, the shades of green into the another corner and the shades of blue into a third corner. The remaining mixed colors would organize to the most like neighbor or the color that is most like them in the remaining space. Figure 9b indicates the positions of the colors at a random point during the iterative process. Subsequent iterations of the SOM algorithm refine the cluster locations until a final solution is obtained as shown in Figure 9c (Matthews, 2004).

Figure 9. Self-Organizing Map Example Using Random Colors



a. Color randomly distributed

b. Iterative process step

c. Final SOM solution

Source: SOM example Java applet output retrieved on 10/17/08 from http://www.generation5.org/content/1999/selforganize.asp

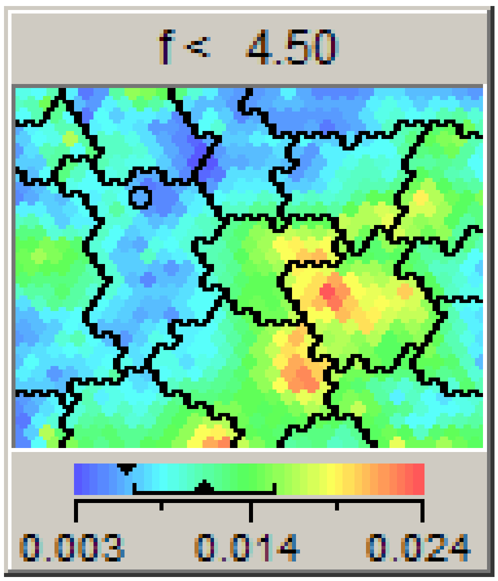
Viscovery SOMine Settings

For this study, two distinct databases were used for the SOM process. All SDOH variables, after transformation from categorical to ratio data, were combined into a single Microsoft Excel dataset and represented in columns with county being represented in the rows. A second Microsoft Excel dataset was created with all HB variables represented in columns with county in the rows. Each resultant dataset was used to create an SOM dataset within SOMine 5.0. The SOMine software automatically used the SOM batch processing method to analyze the data and arrange data on a hexagonal grid. Additional SOM parameters were set as follows (Deboeck, 1999):

- Scaling = default (scale by variance)
- Variable priority = all set to equal 1 for equal priority
- Map size = approximately 10 times the number of input vectors (1200 nodes for SDOH and 310 nodes for HB)
- Map ratio = ratio of 100:75 to force a ratio of horizontal to vertical nodes
- Map tension = set at 0.2 for greater detail in the map
- Map creation accuracy = Accurate

After the final SOM clustering was obtained, individual variable maps were examined to determine how they related to resulting SOM clusters. Figure 10 gives an example of a resulting SOM output for a single attribute variable. The black lines indicate cluster borders. A scale bar at the bottom of the attribute map indicates the range of values specific to the attribute and the corresponding color representing the values. Further, the scale bar includes an upward facing arrow that indicates the mean value of the attribute and if a node is selected in the map a downward pointing arrow indicates the value of the selected node.

Figure 10. Example Output of the SOM procedure from Viscovery SOMine 5.0



Phase 2 – Regression Analysis

A standard multiple regression analysis was conducted to test the links within the PHM. The phase also determined which was the stronger predictor, SOM_{SDOH} or SOM_{HB}. The regression procedure, PROC REG, within SAS 9.1.3 (SAS Institute, Inc., 2002) was utilized with dummy coding to accommodate the categorical nature of the resulting SOM clusters, which allowed the categorical SOM variables to be used in regression analysis. Three dummy variables were used to represent the SOM_{SDOH} clusters and five vectors were used for the SOM_{HB} clusters. Age-Adjusted Mortality Rate (AAMR) was the dependent variable used for this study in order to have the most general health outcome for regression analysis and prediction.

Regression assumptions were assessed for AAMR using appropriate techniques. Skewness and kurtosis were examined to determine if AAMR exhibited a normal distribution. Bartlett's test for homogeneity of variance was used to assess if there were differences in AAMR across the levels of each variable. Standardized residuals, representations of error, were examined for outliers that could bias the results of the regression analysis. Residuals were calculated as the difference between the observed values and the predicted values obtained during regression analysis. Standardizing residuals constrained them to a mean of zero and a standard deviation of 1. Standardized residuals greater than +/- 3 were designated as outliers (Garson, 2008; Schwab, 2006). Standardized regression coefficients were obtained to determine which vector was the stronger predictor. The standardized regression coefficients were used to present the final regression equation.

Phase 3 – Correlation Analysis

The final analytical phase answered the last research question regarding the relationship between SOM_{SDOH} variables, SOM_{HB} variables, and AAMR. Because the nominal SOM_{SDOH} and SOM_{HB} clusters were dummy for use in the regression analysis, a point-biserial correlation could be obtained for correlations between SOM_{SDOH} variables and SOM_{HB} variables. The Phi correlation is used when two dichotomous variables are used, as is the case with the dummy vectors (Garson, 2008). A Point-biserial correlation was obtained for correlations between SOM_{SDOH} or SOM_{HB} variables and AAMR because the point-biserial accommodates the dichotomous to continuous variable correlation (Garson, 2008). Determining the relationships between the different variables helped in the interpretation of the self-organizing map and regression analysis.

CHAPTER IV

RESULTS

Phase I – Self-Organizing Map

Social Determinants of Health Map

All social determinants of health variables (n =115) were submitted to the SOM algorithm via Viscovery SOMine 5.0™. The SOM algorithm identified four clusters within the SDOH data, enabling the rejection of the null hypothesis of no variation, Phase 1 – Ho_1. Figure 11 shows the clusters as they were distributed within mathematical space. The black lines within the SOM distinguish cluster divisions. Cluster 1 accounted for 38% of the counties (29 of 77). Cluster 2 accounted for 30 counties, or 39% of the total. Clusters 3 and 4 each accounted for 12% of the counties within Oklahoma. Counties that were closer together in mathematical space were grouped in the same cluster as well as grouped closer together within each cluster. For example, the counties in the bottom right hand corner of Cluster 2, Okfuskee, McIntosh, Atoka, and Pushmataha (indicated by the first three or four letters of each county name), were more mathematically similar and, therefore, displayed as a closer neighborhood (all nodes within a set radius of the best matching unit) than those that mapped near the middle of Cluster 1 (Carter, Beckham, Kay, Bryan, Pontatoc). These five counties, although similar enough to group together within a single cluster, were not mathematically similar enough to create a tight neighborhood network, but a more spread out network. Neighborhood networks do not cross cluster boundaries. Figure 12 displays the resulting clusters within geographic space for comparison. Relationships between the SDOH variables and the resulting SOM_{SDOH} clusters were examined using various SOM outputs, which will be explored in the following sections.

Figure 11: Distribution of SOM_{SDOH} in Mathematical Space with Descriptive Statistics

Cluster Color	Cluster Name	# Counties within Cluster	% of Used Data	% of Counties
1	Mid-Century Service-Oriented Communities	29	38%	38%
2	Struggling Minority Communities	30	39%	39%
3	High Income and High Education	9	12%	12%
4	Long-term Farmland	9	12%	12%
	Missing	0	0.0%	0.0%

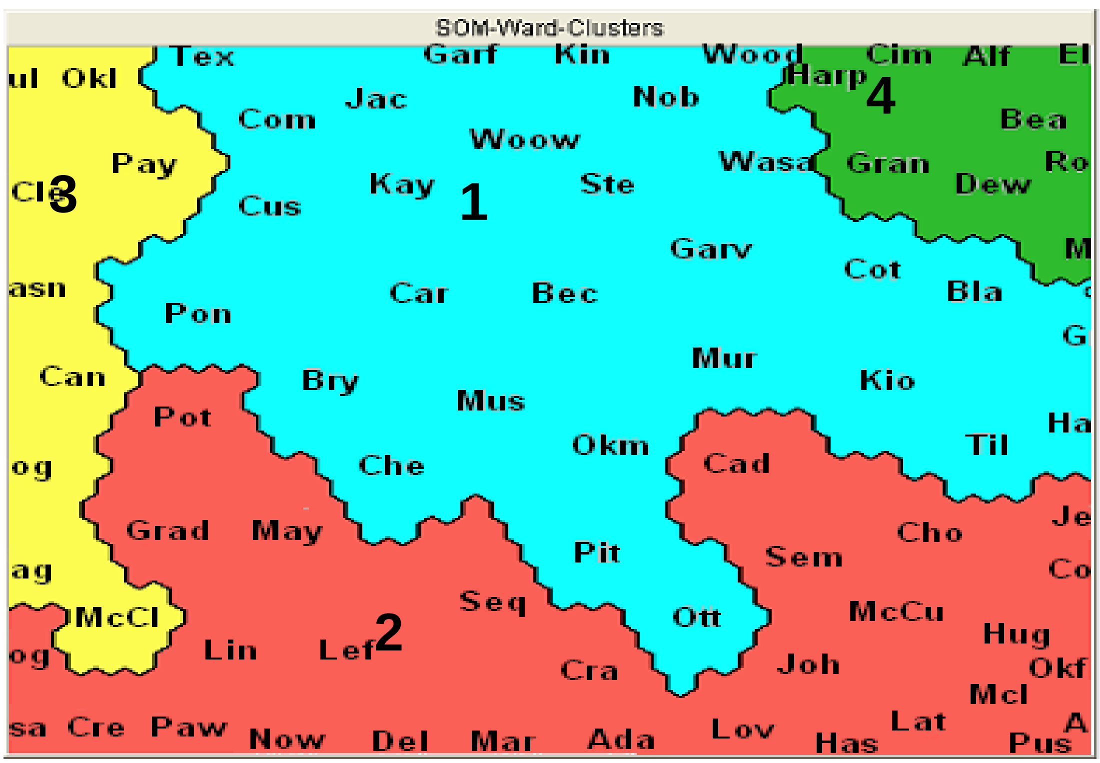
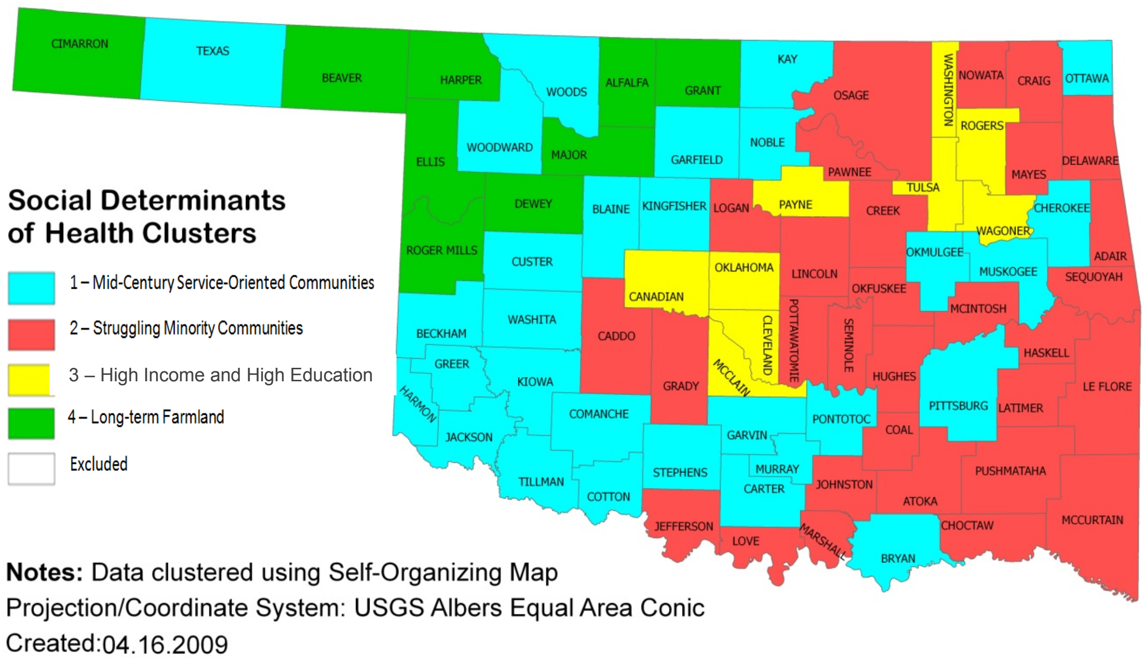


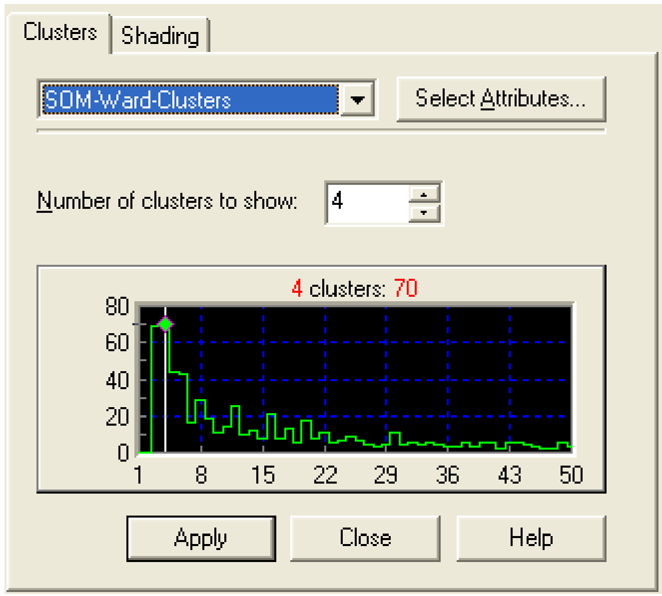
Figure 12: Geographic Distribution of SOM_{SDOH} Clusters



SDOH Cluster Verification

To verify the selection of the four clusters, the SOM-Ward-Clusters were viewed within the cluster-tuning screen. SOM-Ward-Cluster is a modification to the original Ward hierarchical agglomerative cluster algorithm but redefines the distance measure to account for the topological location of clusters (Viscovery, n.d.). Figure 13 depicts the cluster index on the Y-axis and the number of clusters on the X-axis. A high cluster index value indicates a high level of natural clustering within the data. Four clusters were determined to be the best cluster solution for this dataset with an index of 70, confirming the most natural solution for the SOM_{SDOH}.

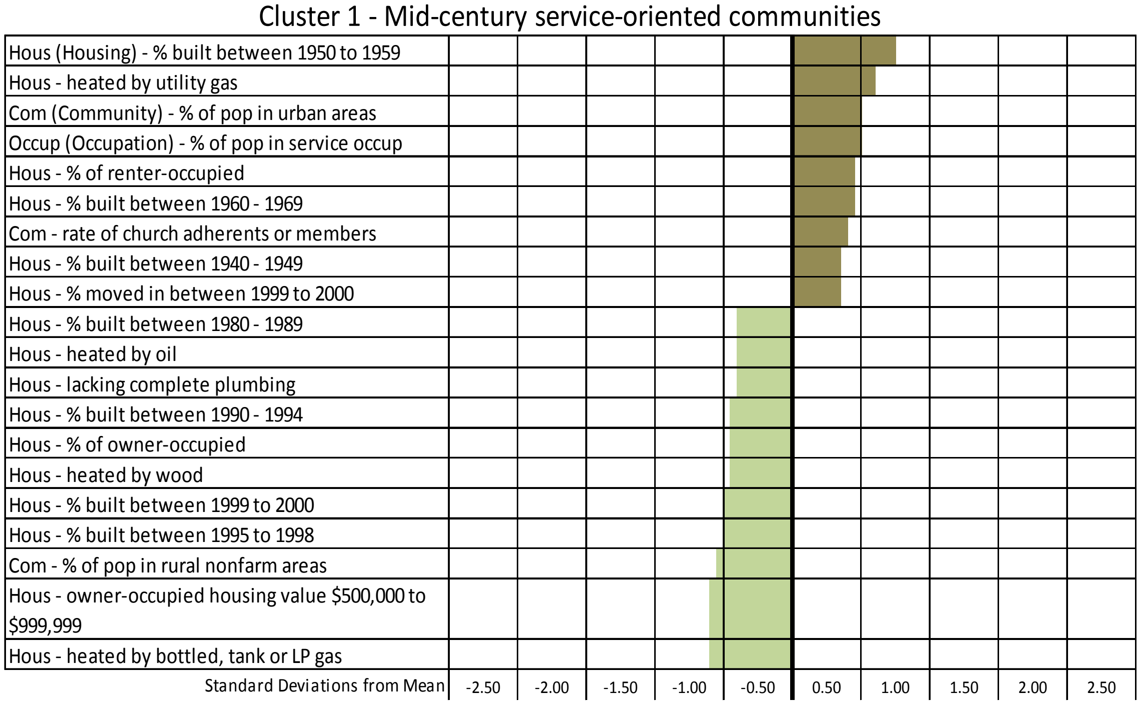
Figure 13: SOMine Cluster-Tuning Screen for SOM_{SDOH}

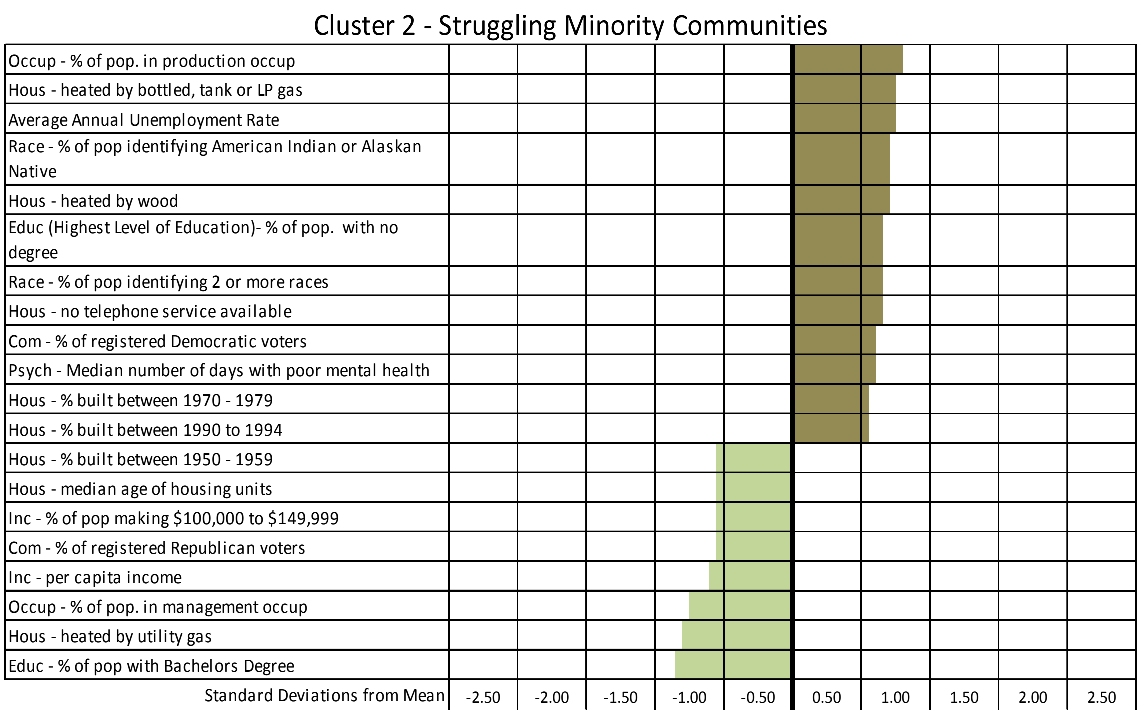


SOM_{SDOH} Cluster Descriptions

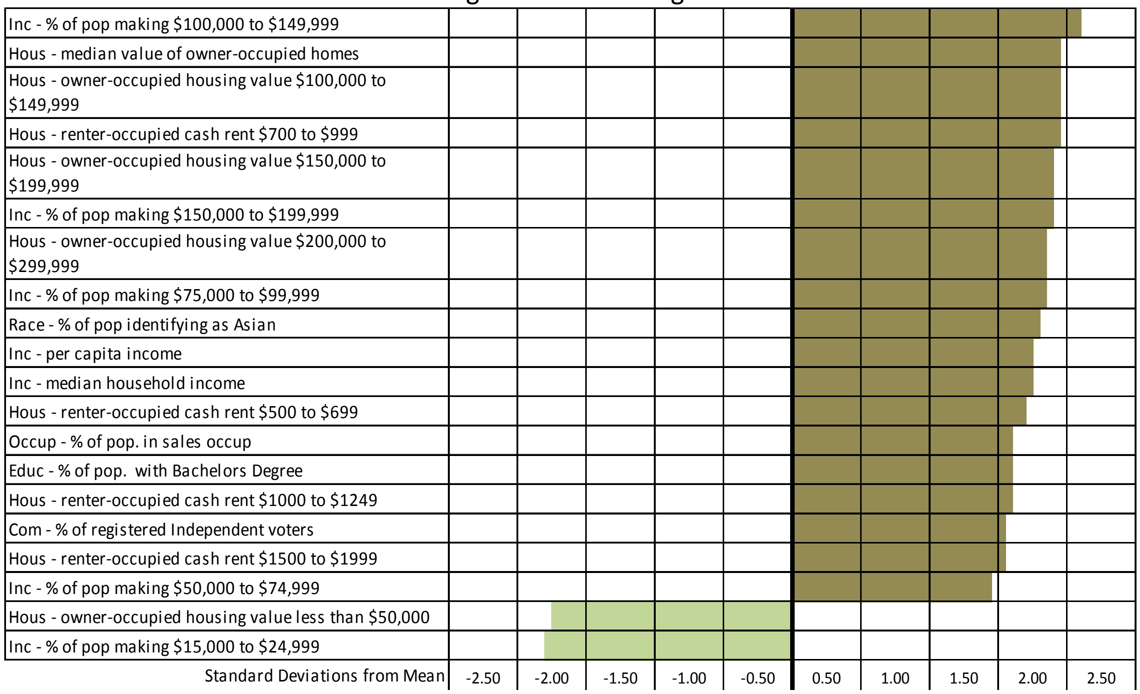
In addition to visually mapping the mathematical space among all the variables as in Figure 11, the Viscovery SOMine program allows for description of the resulting clusters by examining distinguishing variables. Figure 14 displays the most distinguishing variables for each SOM_{SDOH} cluster. The SOMine software calculates standard deviations from the grand mean and then tests for significance using a t-test. The resulting graph only displays significant variables that either positively or negatively define the cluster.

Figure 14: Distinguishing Variables by SOM_{SDOH} Cluster

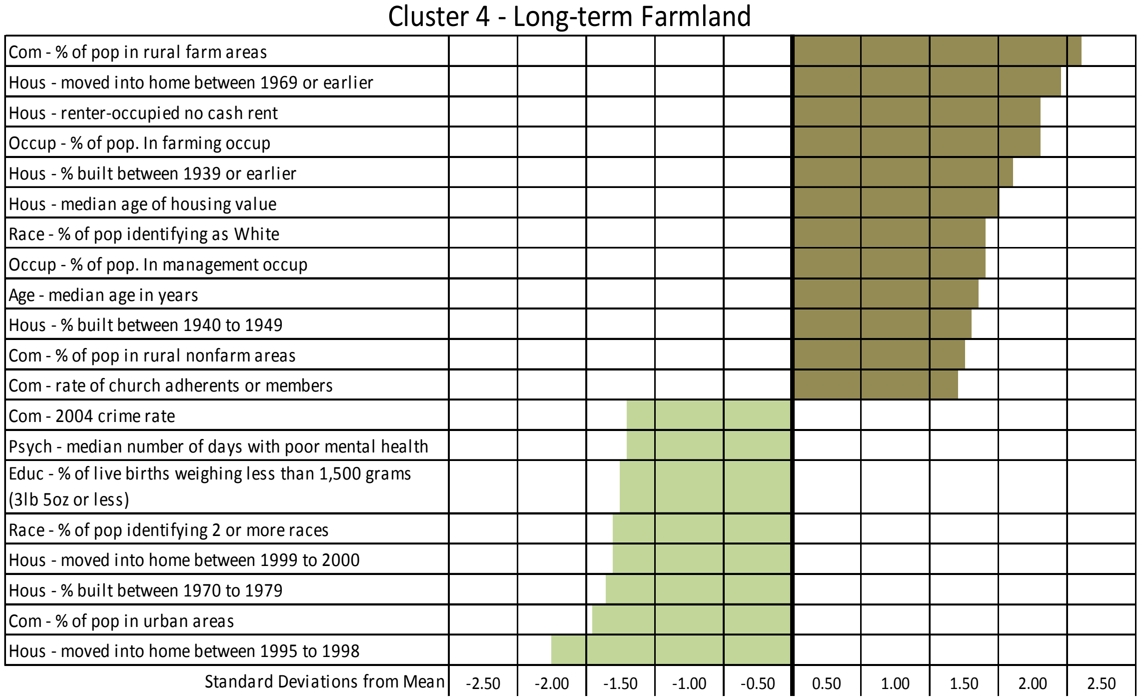




Cluster 3 - High Income and High Education



Cluster 4 - Long-term Farmland



One can quickly see from the distinguishing variables in Figure 14 that urban areas positively define Cluster 1, negatively define Cluster 4, and have no significance on Clusters 2 and 3. Other variables helped to distinguish the clusters, but to a much more varied degree. For example, education had a relatively small effect on defining clusters. Only three of the nine education variables even appeared as a distinguishing variable of clusters. The percentage of population with a bachelor's degree, the percent of population without a high school degree, and the percentage of babies born with very low birth weight were distinguishing variables, but clear patterns were difficult to identify regarding education from this information alone. However, couple this information with other distinguishing variables and interesting patterns emerge.

Mid-Century Service-Oriented Communities. Cluster 1 counties are positively defined by two of the SDOH dimensions: socio-economic status (SES) and community. SES variables had the largest impact on defining Cluster 1. Homes built in the mid-20th century with high rates of rental property and homes serviced by gas from a utility company all positively define this cluster. In addition, this cluster had the highest rate of persons who moved into homes between 1999 and 2000. Service-oriented occupations were also very prevalent in this cluster. Cluster 1 had the second largest percentage of urban population in the state and included counties such as Cherokee, Muskogee, Garfield, and Comanche. Additionally, Cluster 1 also had the second highest rate of church members in the state.

Interestingly, cluster 1 was defined by more negatively related variables than positive. Cluster 1 has low rates of newer homes and in fact some of the lowest rates of homes built after 1980. Cluster 1 also had the second to the lowest rate of rural non-farming land. Although income variables were not distinguishing variables within this cluster, both the positive and negative sides of the graph in Figure 14 indicate the lower portions of the middle class lifestyle as described by Thompson and Hickey (2005). Cluster 1 was named *Mid-Century Service-Oriented Communities* to represent the older homes and service occupations that dominate this cluster.

Struggling Minority Communities. Cluster 2 was defined by the second highest rate of production-related occupations and the lowest rate of management-related occupations and the highest rate of unemployment. In addition, Cluster 2 was defined by high rates of racial and ethnic diversity, having the highest rates of persons indicating two or more races and Native American/ American Indian population. Interestingly, Cluster 2 was also defined by high rates of houses that were heated by propane or wood and houses that did not have a home phone. This cluster was also clearly defined as having a population that most identified with the Democratic Party, while negatively associated with the Republican Party. Persons within these counties had the highest rates of poor mental health days. Houses in Cluster 2 were newer than those in Cluster 1 as evidenced by the two positive housing variables (Housing built between 1970 and 1979 and Housing built between 1990 and 1994) and the two negative housing variables (Median housing age and Housing built between 1970 and

1970). Cluster 2 was negatively associated with two income variables, both having the lowest rates: percent of population with incomes between $100,000 and $149,999 and low per capita incomes. While Cluster 2 had some of the poorest educational outcomes among the clusters, only one of the education variables was significant. The rate of persons obtaining a bachelor's degree was negatively associated with this cluster because of its low rate (7.8%). Cluster 2 was named *Struggling Minority Communities* to capture the distinguishing variables.

High Income and High Education. Cluster 3 is clearly defined by income, education, and housing variables. In fact, 80% of the distinguishing variables for Cluster 3 were income or housing related. Cluster 3, which includes counties such as Oklahoma, Tulsa and Cleveland Counties, had high rates of incomes above $50,000 a year (Inc_74, Inc_100, Inc_150 and Inc_200). Additionally, Cluster 3 had the highest median household income and per capita income. Cluster 3 also had the most significant and expensive housing among the clusters with housing values over $100,000, rent ranging from $500 to $2000 a month, and the highest median dollar value of all owner-occupied homes.

Cluster 3 was significantly defined by three additional SES variables. Cluster 3 contained the largest proportion of Asian population among the four clusters. It also accounted for the largest proportion of persons with a Bachelors degree and the most persons working in a sales-related occupation. Only one community indicator was significant for this cluster and it was positively associated with a high rate of voters registered as Independents. There were only two indicators that were negatively associated with Cluster 3: the lowest rates of housing values of less than $50,000 and incomes between $15,000 and $24,999. Cluster 3 was named *High Income and High Education*.

Long-term farmland. Within the other 3 clusters, SES variables were the top ranking significant variables. This was not the case in Cluster 4. Cluster 4 was significantly defined by community-based variables. Not only did Cluster 4 contain the largest amount of rural farmland and a significant proportion of rural non-farmland, a negative relationship existed with the urban variable because none of the population within this cluster resided in an urban area. This cluster also had much higher rates of church membership than the other clusters, a measure of community cohesion. Additionally, Cluster 4 was significantly associated with stability in its population. Houses in these counties were built in the first half of the 20^{th} century because Cluster 4 had the highest proportion of houses built before 1949. Cluster 4 also accounted for the largest proportion of persons moving into a county before 1970, and it had the oldest population (Age_Med = 41.89 years). Aligning with the farming nature, Cluster 4 was significantly associated with high rates of persons who did not have to pay anything for rent. Two occupation variables positively defined Cluster 4: farming and management (33% of which were related to farm and ranch management). This cluster is significantly associated with the Race_W variable indicating that it contained the largest proportion of white persons among the clusters (92.7%).

On the opposite end of the spectrum, Cluster 4 was associated with low rates of poor mental health days, low rates of crime, a low rate of very low birth weight babies, and a low proportion of the population who indicated 2 or more races. To support the housing stability previously mentioned, Cluster 4 counties had the lowest rates of persons who have moved into the area since 1995 and the lowest rate of houses built between 1970 and 1979. Because there is no clear alignment of this cluster with the Thompson and Hickey classes, Cluster 4 will be named for the majority of its representing variables: *Long-term Farmland*.

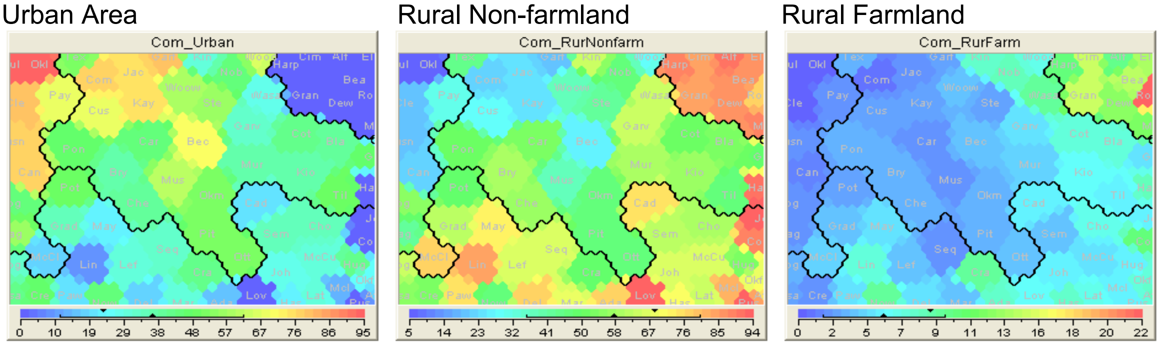
SOM$_{SDOH}$ and Individual SDOH Variables

To begin examining the distribution of the individual variables within the cluster groupings, the deviation of the group means from the grand mean were plotted by standard deviation for each of the individual SDOH variables. The lengths of the bars in Figure 14 provide a visual indicator of how the variables differ within cluster groupings. *High Income and High Education* and *Long-term Farmland* indicated large amounts of variation within the variables compared to the first two clusters. Appendix C contains a list of means and standard deviations by cluster for each SDOH variable.

Some variables showed clear patterns within the resulting clusters. For example the three variables related to measures of rural and urban land. Cluster 1, while not including the large metropolitan areas within Oklahoma (i.e., Oklahoma, Tulsa, Cleveland Counties), was associated with urban populations (Standard Deviation [SD] = +0.5) more so than rural non-farm land (SD = -0.5). This was also evident when the percentage of urban population (Com_urban) map was viewed in comparison to the resulting clusters (Figure 15). To assist in interpreting the variable maps, a scale bar at the bottom of each map indicates the data range for the subsequent variable and allows for quick reference to the map by color. In all variable maps a higher value corresponds to a white color, while low values correspond to darker colors. The variable maps allow for examination of a single variable while preserving the topological nature of the multivariate data and the resulting clusters because the counties do not change locations from the original mapped locations within the clusters. In the case of Com_Urban (the percent of designated urban area found within a county), a clear delineation in the clusters was evident. Cluster 1 consisted of a mix of more urban counties in the upper left portion of the cluster and rural non-farming land in the lower, right portion of the cluster, while Cluster 2 was comprised much more of rural non-farming land. Cluster 3 displayed a clear urban population and Cluster 4 was all rural farming, SD = +1.0, and non-farming land, SD= +2.0 (i.e. livestock). The largest portion of farmland was concentrated within Cluster 4. It was very clear that cluster 4 did not include any counties with large populations, but consisted of rural areas only.

Figure 15: Rural/ Urban Variable Maps for SOM_{SDOH}

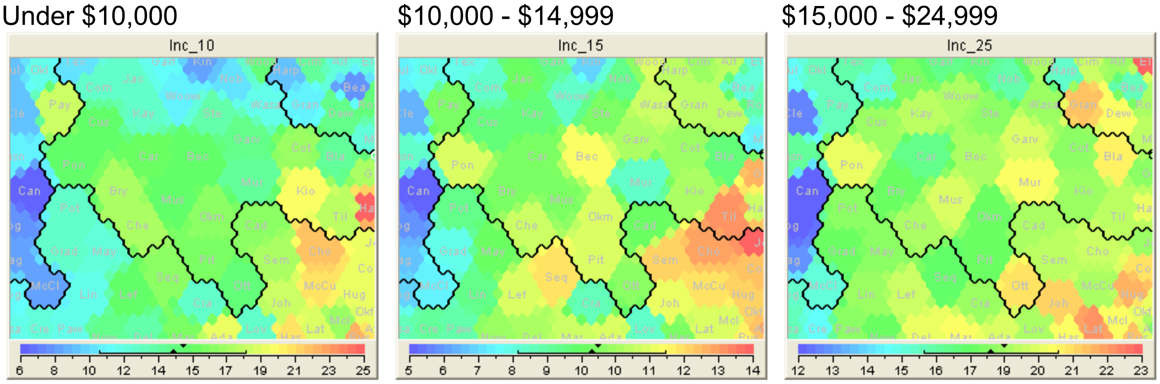
Urban Area | Rural Non-farmland | Rural Farmland



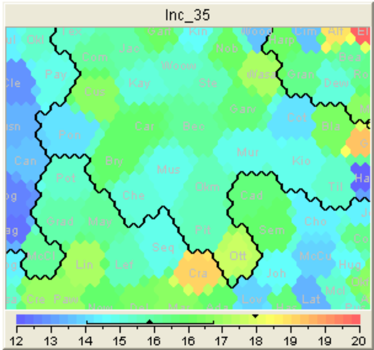
In addition to the rural/urban dichotomy that is evident within the clusters, income variables show a clear distribution among the clusters. Figure 16 displays the variable maps for the 12 income variables in order of increasing income with median household income and per capita income at the end. Lower incomes were more prevalent within the counties clustered to the lower right corner of the SOM. Incomes above $50,000 (INC_74 and above) all aggregated to the left side of the map and all corresponded to Cluster 3. Interestingly, Tillman County (right side of map in Cluster 1 - TIL) showed a dramatic disparity regarding income. While Tillman County was represented by a higher rate of persons with incomes between $10,000 and $14,999 (Inc_15), it also showed a moderately high rate of incomes over $200,000 a year. Individual variable maps for all SDOH variables can be found within Appendix D for reference.

Figure 16: Income-related Variable Maps for SOM_{SDOH}

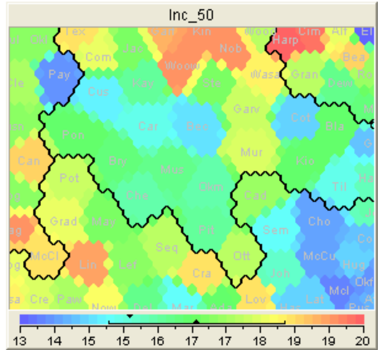
Under $10,000 | $10,000 - $14,999 | $15,000 - $24,999



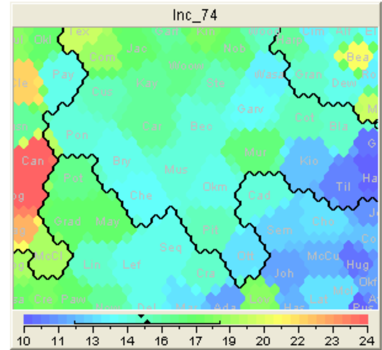
$25,000 - $34,999



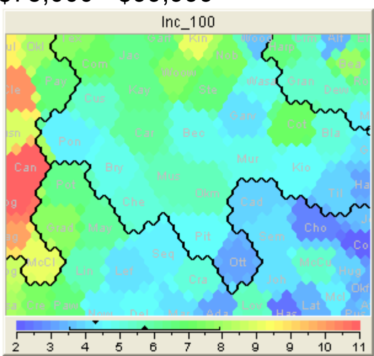
$35,000 - $49,999



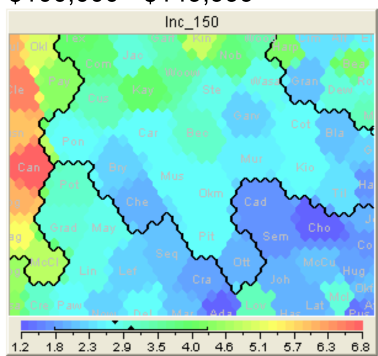
$50,000 - $74,999



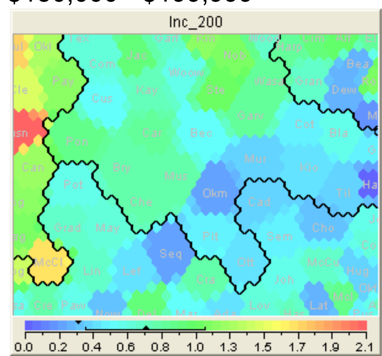
$75,000 - $99,999



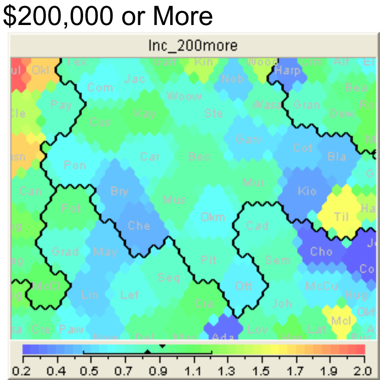
$100,000 - $149,999



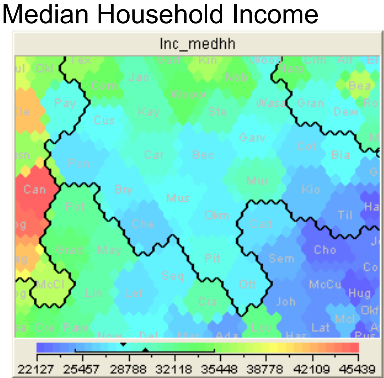
$150,000 - $199,999



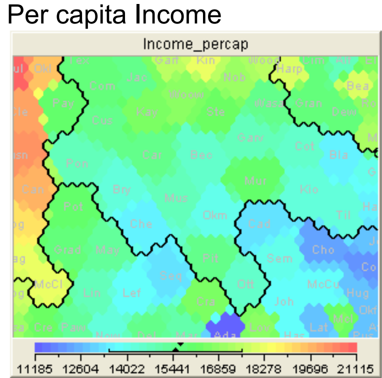
$200,000 or More



Median Household Income



Per capita Income



Health Behaviors Map

Thirty-one variables representing both physical and mental health behaviors were submitted to the self-organizing map algorithm via Viscovery SOMine 5.0™. The SOM algorithm identified six clusters within the health behavior data enabling the rejection of the null hypothesis, Phase 1 – Ho_2.

Figure 17 shows the clusters as they were distributed within mathematical space. Cluster 1, accounting for 41% of the counties (n = 31), is the largest cluster. Cluster 2, which is dark black on the left side of the map, is the second largest cluster accounting for 20% of the counties (n = 15). Cluster 3 accounted for 13.33% of counties (n = 10), followed by cluster 4 with 9% of counties (n = 7). The smallest clusters, Cluster 5 and 6 each accounted for 8% of the counties (n = 6) within Oklahoma. Figure 18 shows the geographic distribution of the resulting SOM_{HB} clusters. Both Harmon and Harper counties were deleted by the SOMine system during analysis because they were missing data for all of the HB variables. All other counties with missing data points were clustered even though the percentage of missing variables went up to 55% (17 out of 31 variables) for some counties. The flexibility of Viscovery SOMine allows for the analysis of data even with high levels of missing information. In order to represent the counties as much as possible, these counties were left in the analysis since they would be considered by the program.

Figure 17: Health Behavior Cluster Segmentation with Description of Clusters

Cluster Color	Cluster Name	# Counties within Cluster	% of Used Data	% of Counties
1	Restricted	31	41%	40%
2	Health-Promoting	15	20%	20%
3	Overweight and Unsafe	10	13%	13%
4	Conflicted Intimate Partner Violence	7	9%	9%
5	Conflicted Mental & Physical Health	6	8%	8%
6	Safety Not Health-Related	6	8%	8%
	Missing	2	0%	3%

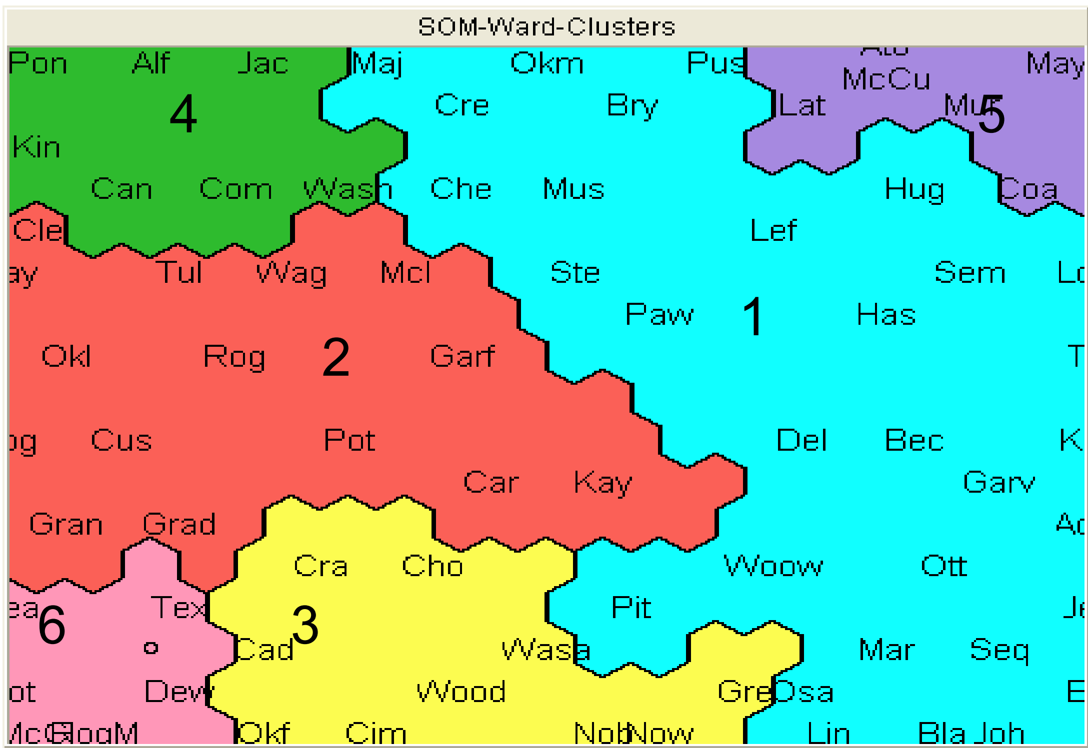
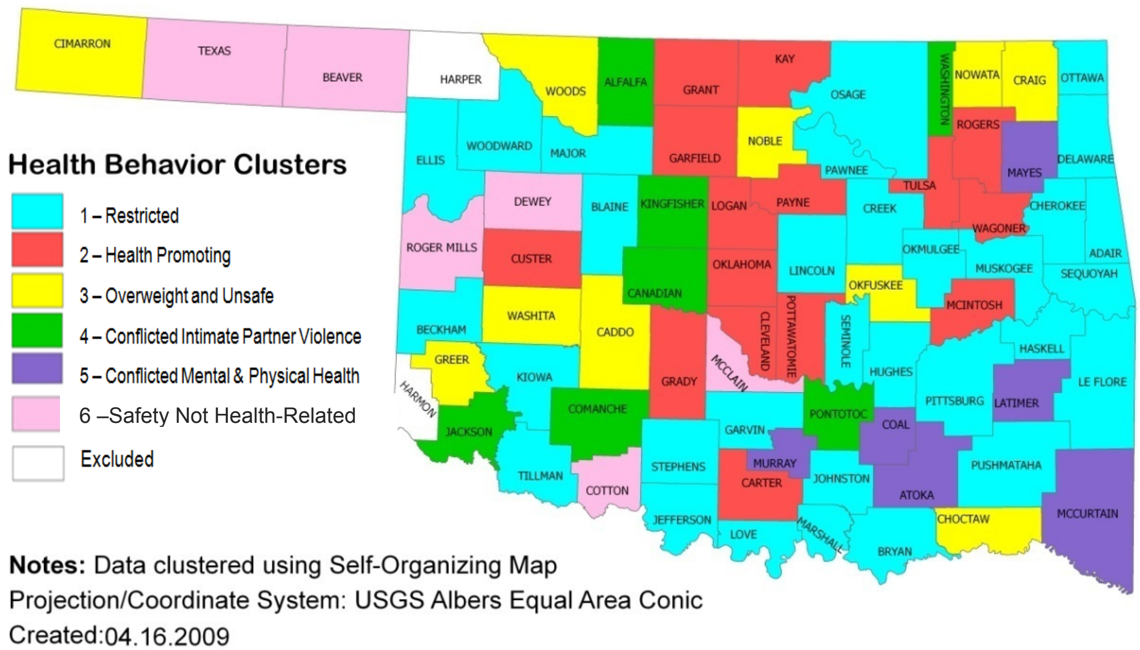


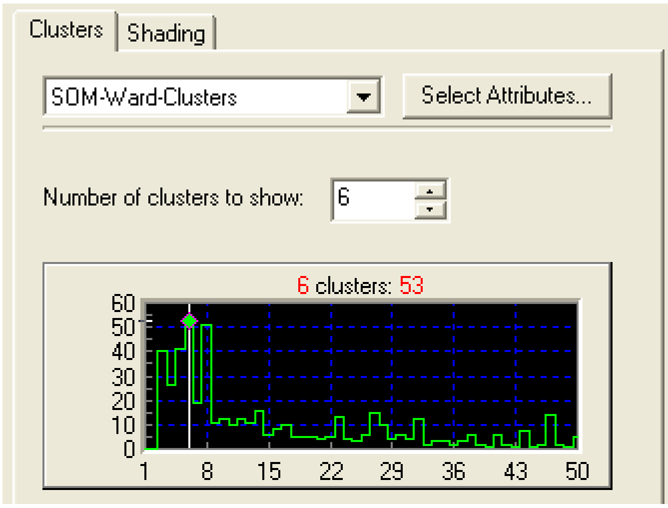
Figure 18: Geographic Distribution of Health Behavior Clusters



HB Cluster Verification

The cluster tuning screen was viewed to verify the cluster selection. While six clusters were the best solution for the HB dataset, the cluster index found in Figure 19 (53) was lower than the SDOH cluster index (70) indicating that cluster solutions within the HB data are not as natural as they are within the SDOH dataset.

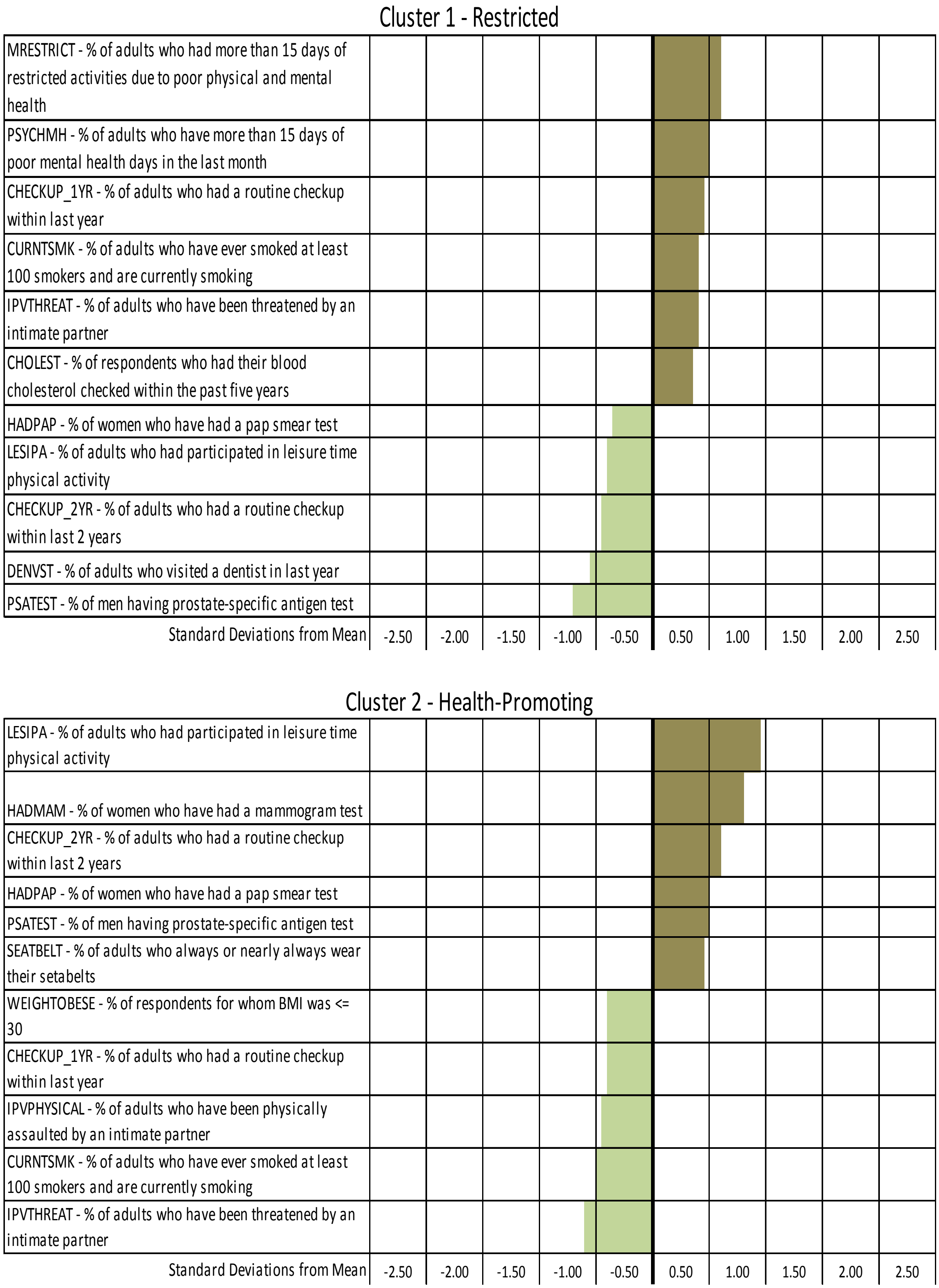
Figure 19: SOMine Cluster Tuning Screen for SOM_{HB}



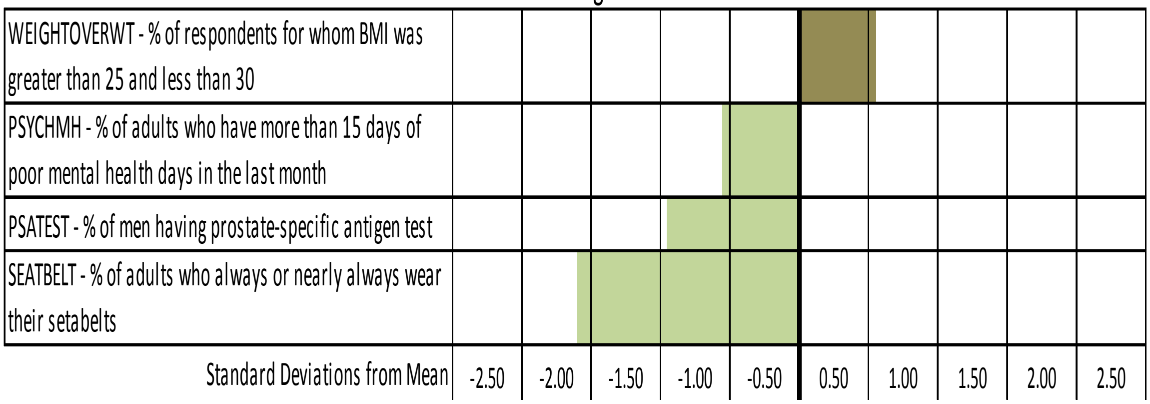
SOM_{HB} Cluster Descriptions

Though more clusters were defined during the SOM_{HB} than the SOM_{SDOH}, fewer variables significantly define each of the clusters. The number of variables that defined each cluster ranged from four to eleven (Figure 20). The small number of significant variables for each cluster made naming the clusters much more difficult.

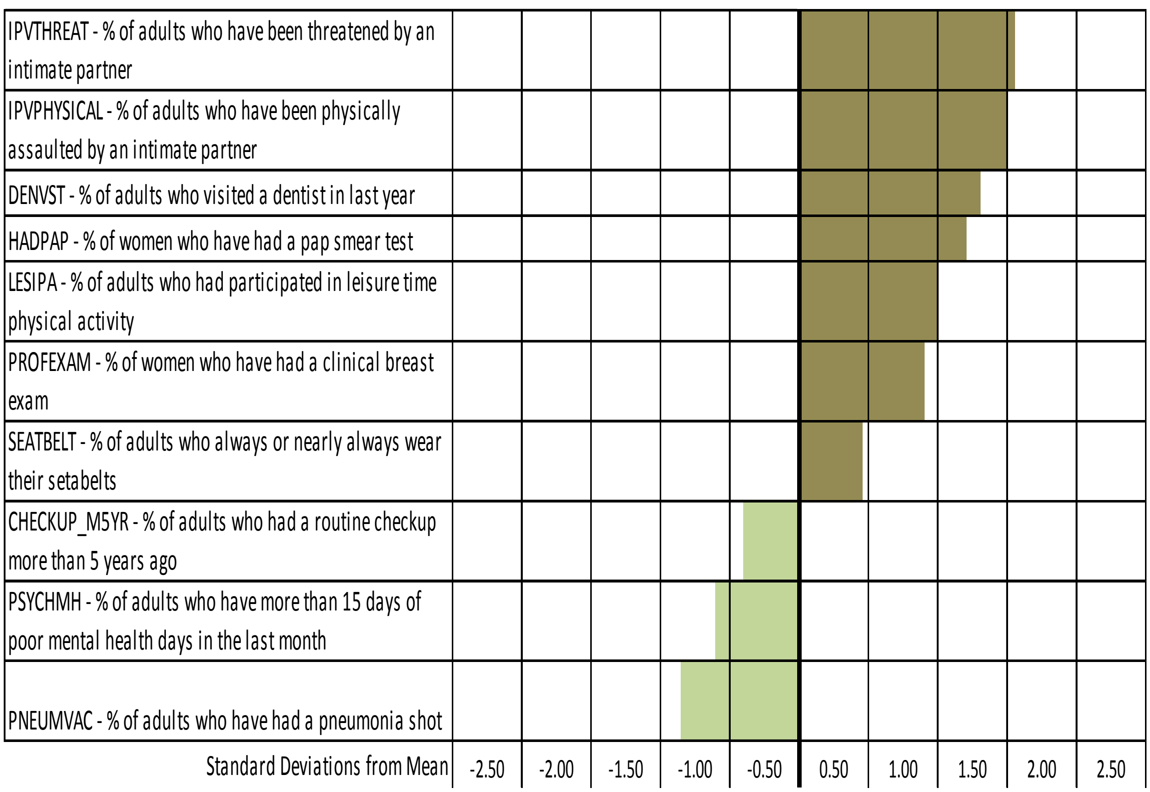
Figure 20: Distinguishing Variables by SOM_{HB} Cluster



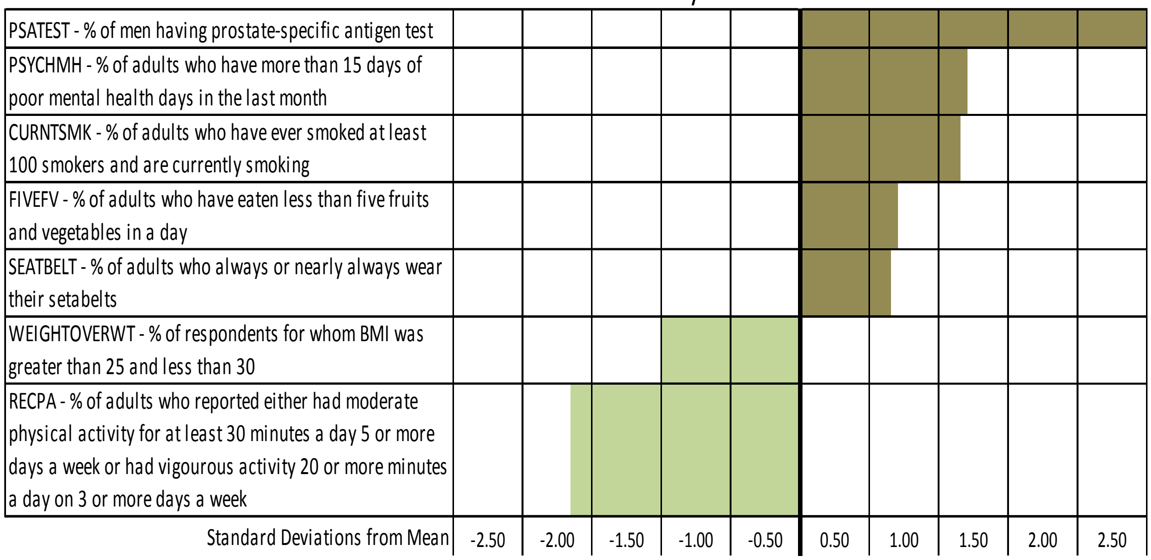
Cluster 3 - Overweight and Unsafe



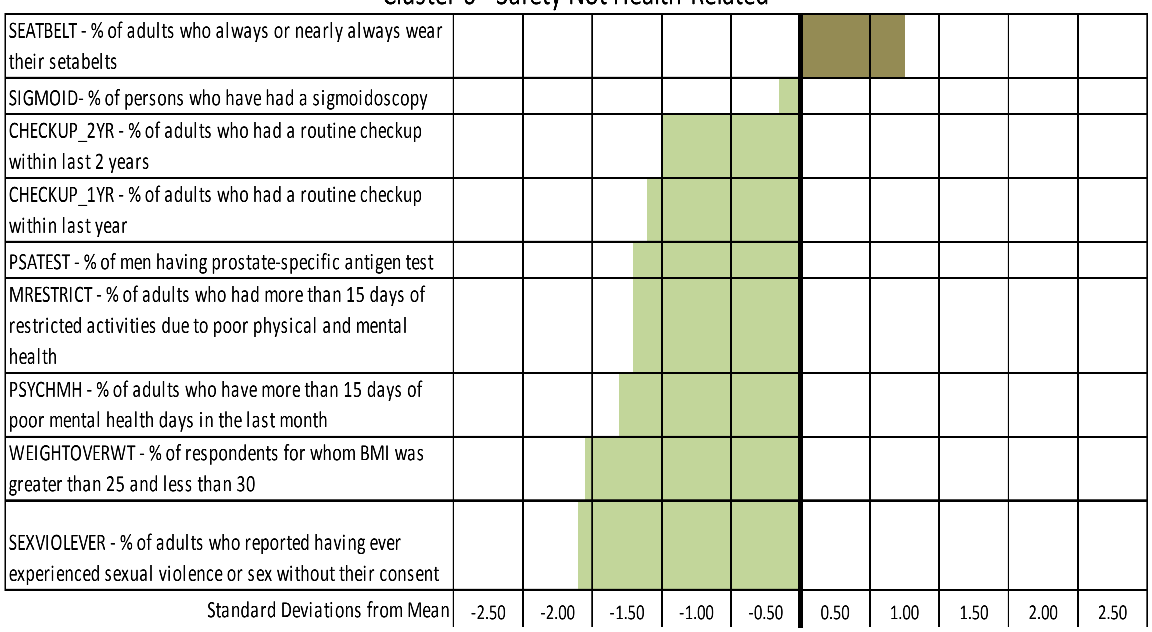
Cluster 4 - Conflicted Intimate Partner Violence



Cluster 5 - Conflicted Mental & Physical Health



Cluster 6 - Safety Not Health-Related



Restricted. Although persons living in the counties in Cluster 1 seem to receive checkups once a year, they were also more likely to be threatened by an intimate partner and be current smokers. In addition, Cluster 1 had low rates of

Pap smear exams, dental visits, checkups within the past two years, PSA tests for prostate cancer, and leisure time physical activity. Because of the potential for restricting future behaviors that goes along with a lack of screenings and the restrictive behaviors already exhibited, Cluster 1 was named *Restricted*.

Health-Promoting. Cluster 2 had very different results and variables indicated persons within these counties were Health-Promoting. High rates of leisure time physical activity, mammogram screenings, pap smears, and PSA tests indicate effort being placed in the right direction. Less significant, although still significant, seatbelt usage and physician check-ups were a sign of struggling effort. However, this cluster did have low rates of obesity, current smoking, and intimate partner violence. Cluster 2 was therefore named *Health-Promoting.*

Overweight and Unsafe. Cluster 3 only had four significant variables available to define the cluster. It had the second highest rate of adults who were overweight and the second to lowest rate of PSA test. The percentage of persons having more than 15 poor mental health days was significant, but this was a more positive relationship than the other variables (below the mean for all clusters). Seatbelt usage was Cluster 3's worst indicator. Cluster 3 was named *Overweight and Unsafe* for the two predominant defining variables: overweight and seatbelt usage.

Conflicted Intimate Partner Violence: Counties in Cluster 4 offered a bit of a conundrum. While they had high rates of healthy behaviors such as dental visits, pap smears, professional exams, leisure time physical activity and seatbelt usage, Cluster 4 also accounted for the highest rates of intimate partner violence for both threats and physical violence. To further exacerbate the issue, Cluster 4 exhibited a low percentage of persons with more than 15 poor mental health days. In addition, Cluster 4 had the lowest proportion of persons who had waited more than 5 years to have a checkup and the lowest percentage of persons receiving the pneumonia shot. Because of the extreme pattern displayed, Cluster 4 was named *Conflicted Intimate Partner Violence.*

Conflicted Mental & Physical Health. Cluster 5 was named *Conflicted Mental & Physical Health*. This was due to the interesting combination of significant variables within this cluster. Cluster 5 exhibited the highest smoking rate, a high percentage of persons reporting more than 15 poor mental health days, and a low percentage of persons getting the recommended levels of physical activity. However, the highest rate of PSA tests at 71% of the male population, high rates of fruit and vegetable consumption, high seatbelt usage, and low percentage of persons being overweight balanced the negative indicators.

Safety Not Health-Related. Cluster 6 had one significant positive indicator, seatbelt usage. In fact, it was the highest rate among the clusters. However, three other variables that were reported as negatively associated with the cluster are, in fact, supportive because they have the lowest rates among the clusters (mental restriction, more than 15 poor mental health days, and ever experiencing sexual violence). Four additional health-oriented variables (sigmoid, checkup_5yr, checkup_1yr, and psa_test) were negatively associated. Cluster 6 was named *Safety Not Health-Related.*

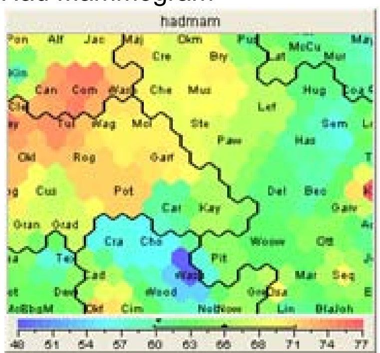
SOM_{HB} and Individual HB Variables

As with the SOM_{SDOH}, the higher numbered clusters (*Conflicted Mental & Physical Health* and *Safety Not Health-Related*) displayed greater deviations from the grand means within each variable. *Restricted* and *Health-Promoting*, which contained the largest number of counties, were mathematically closer to the grand mean within each variable than the other clusters. Figure 20 provides a set of the variables that show *Conflicted Mental & Physical Health* and *Safety Not Health-Related* have larger standard deviations away from the grand means than any of the deviations within social determinants of health clusters (3.25 versus 2).

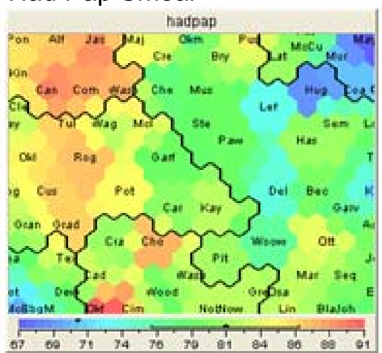
When examining the individual SOM maps for the health behavior variables, patterns begin to emerge from the data. Within the individual SOM variable maps, Figures 21 and 22, the data are again placed upon a continuum from lowest values in blue to highest values in red and each map contains a reference bar with the range of data values. As within the SOM_{SDOH}, locations of the counties did not change position within the individual maps from the overall SOM_{HB} solution found in Figure 17. Persons who tend to exhibit healthier behaviors (i.e., having mammogram or pap smear screenings, seeing a dentist, eating fruits and vegetables, having leisure time physical activity, and having normal weight) fall generally in or around Cluster 4 (top left corner of map) as in Figure 21, while Figure 22 shows this cluster as having restrictive mental health behaviors as aggregating around the variable mean or below (variable mean is indicated by upward pointing arrow on scale bar). Appendix F contains the SOM variable maps for all of the health behavior variables.

Figure 21: Screening-related Health Behavior SOM Maps

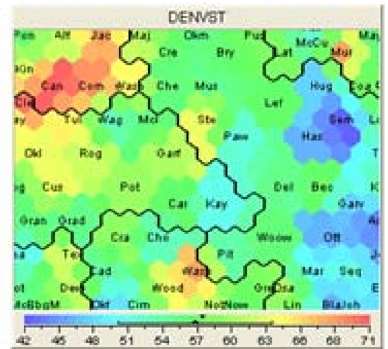
Had Mammogram



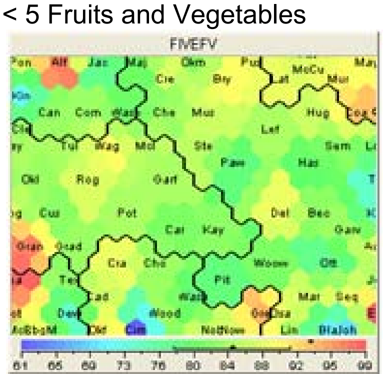
Had Pap Smear



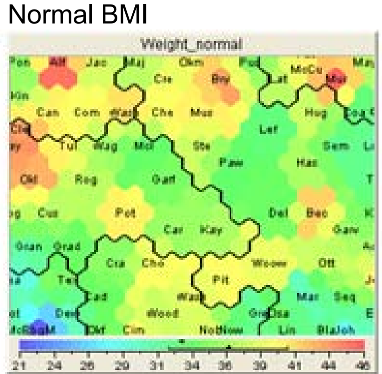
Dental Visit in Last Year



< 5 Fruits and Vegetables



Normal BMI



Leisure Time Physical Activity

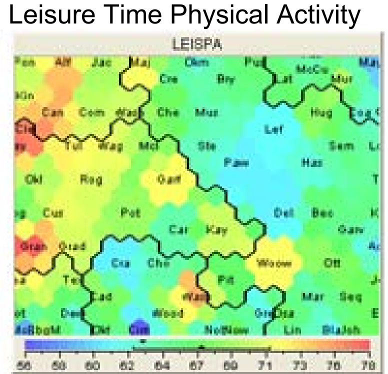
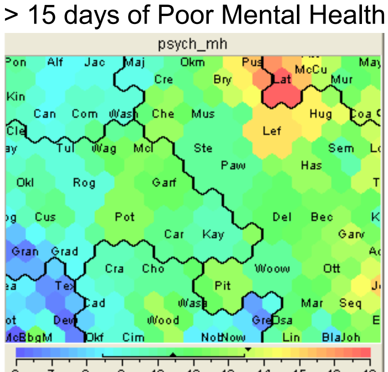
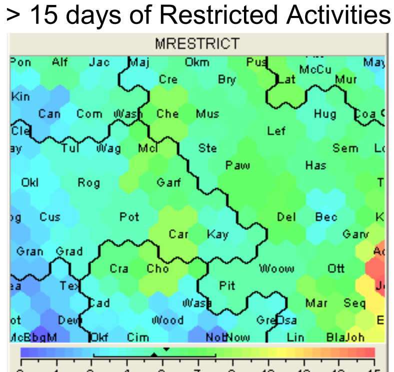


Figure 22: Restrictive Mental Health SOM Maps

> 15 days of Poor Mental Health



> 15 days of Restricted Activities



Phase 2 - Regression Analysis

After obtaining the SOM clusters for both social determinants of health (SDOH) and health behaviors (HB), each set of clusters was dummy to account for their categorical nature. The dependent variable used in the linear regression analysis was the targeted health outcome for this study, age-adjusted mortality rate (AAMR). Table 3 displays the observed AAMR by county for 2000-2006. This information is also geographically represented in Figure 23 by quartiles. Mapping data by quartiles allow for a quick visual determination of where a county lies within the range of scores

Table 3: Observed AAMR for 2000-2006 by Oklahoma County

County	AAMR*	County	AAMR*	County	AAMR*
Adair	1151.1	Grant	900.5	Nowata	932.6
Alfalfa	745.7	Greer	981.9	Okfuskee	1095.9
Atoka	972.2	Harmon	1084.2	Oklahoma	981.4
Beaver	795.3	Harper	1075.8	Okmulgee	1063
Beckham	1103.7	Haskell	1091	Osage	837.3
Blaine	1003.9	Hughes	1061.6	Ottawa	1068.8
Bryan	1019.6	Jackson	1047.8	Pawnee	1009.5
Caddo	1101	Jefferson	1126.9	Payne	818
Canadian	897.5	Johnston	1117.4	Pittsburg	1002.1
Carter	1089.8	Kay	991.3	Pontotoc	1066.5
Cherokee	1041.1	Kingfisher	940.8	Pottawatomie	1047.1
Choctaw	1126.7	Kiowa	1139.1	Pushmataha	1064.1
Cimarron	877.4	Latimer	1061	Roger Mills	875.6
Cleveland	890.2	Leflore	1063.2	Rogers	927.3
Coal	1170.7	Lincoln	1033.4	Seminole	1119.1
Comanche	979.6	Logan	928.4	Sequoyah	1020
Cotton	955.3	Love	926.3	Stephens	1039.8
Craig	1017	Major	919.5	Texas	857.9
Creek	1041.6	Marshall	959.1	Tillman	999.5
Custer	1010.9	Mayes	984	Tulsa	982.8
Delaware	939	McClain	1005	Wagoner	883.5
Dewey	1072.5	McCurtain	1148.2	Washington	892
Ellis	878.3	McIntosh	968	Washita	860.1

Garfield	984.7	Murray	1086	Woods	880.5
Garvin	1039	Muskogee	1001.4	Woodward	886.8
Grady	1041.5	Noble	900.9		

Note: * AAMR is death rate for all causes of death per 100,000 population

The mean of AAMR for all 77 counties was 996.09 with a standard deviation of 92.87 and a range of 425 (Table 4). However, when the mean and standard deviation are by cluster group, the total mean value was reduced for SOM_{HB} clusters because of the two counties (Harmon and Harper) that were not analyzed during the SOM process due to a lack of data (Table 5). The standard deviation for AAMR also increases when looking at on SOM_{HB} clusters as compared to the SOM_{SDOH} standard deviation.

Figure 23: Total Mortality in Oklahoma, 2000-2006, Age-Adjusted Mortality Rate

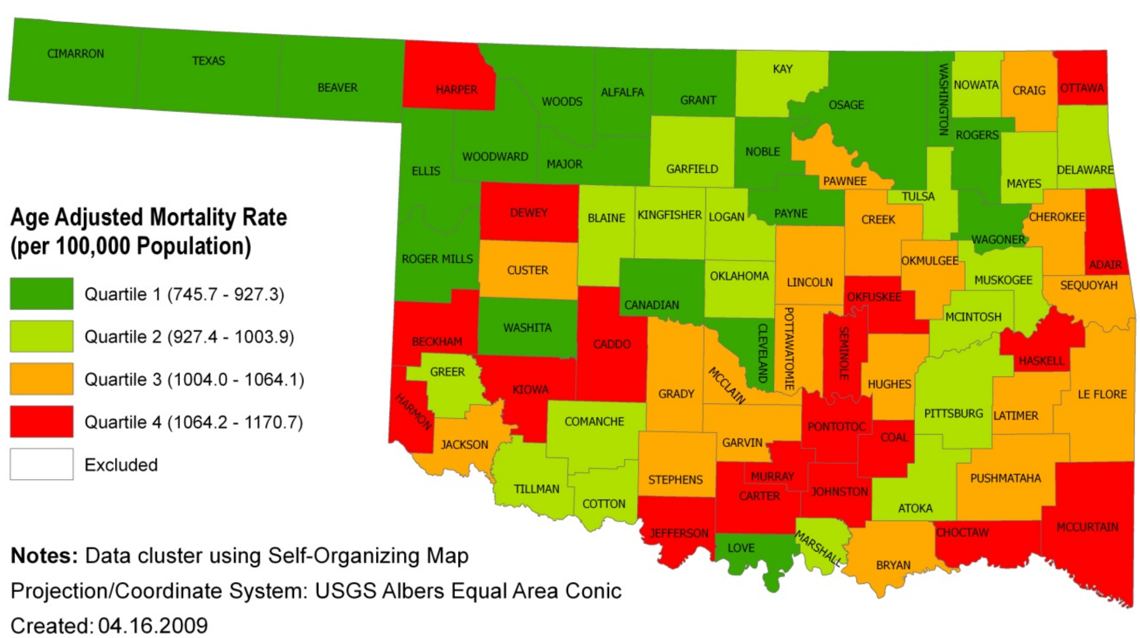


Table 4: Descriptive Statistics for Observed AAMR

AAMR Descriptive Statistics					
Mean	996.09350	Standard Deviation	92.872699	Variance	8625.33825
Median	1003.90	Skewness	-0.3558362	Kurtosis	-0.4379419
Range	425.00	Minimum	745.7	Maximum	1170.7
N	77	Quartile 1	927.3	Quartile 3	1064.1

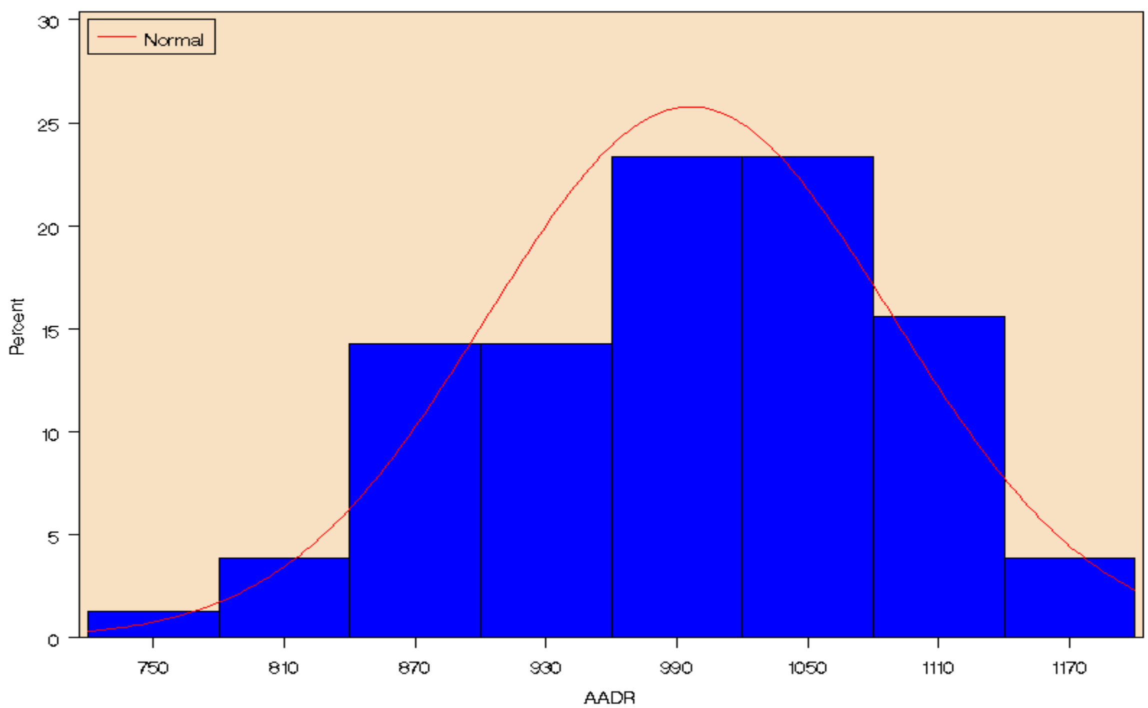
Table 5: Mean and Standard Deviation for AAMR by SOM Cluster

	AAMR			
	SDOH Cluster		HB Cluster	
	Mean	Standard Deviation	Mean	Standard Deviation
1	1004.34	74.66	1024.7	78.93
2	1038.50	80.83	963.0	72.04
3	919.74	60.11	977.4	102.31
4	904.51	110.23	938.6	108.81
5	--	--	1070.4	81.92
6	--	--	1027.8	926.9
Total	996.09	92.87	993.9	93.07

The dummy SOM_{SDOH} variables and dummy SOM_{HB} variables were entered into a standard multiple regression analysis. Regression assumptions were first assessed for violations before interpretation took place. All variable entries met the measurement level requirements (metric or dichotomous) for multiple regression analysis (Pedhazur, 1997). The SOM variables were dichotomous due to dummy coding and AAMR was an interval variable.

The sample size was adequate for multiple regression with a ratio of valid cases to independent variables of 9.375:1, which exceeded the minimum ratio of 5 to 1 (Pedhazur, 1997). AAMR was normally distributed as both skewness and kurtosis values (Table 4) fall within +/- 1.0 (Pedhazur, 1997), therefore, not requiring transformation. Figure 24 shows the roughly normal distribution of AAMR as a visual representation of the skewness and kurtosis values. Since normality cannot be assessed in dichotomous variables, normality was not tested for the vectors.

Figure 24: Histogram of AAMR with Normal Curve Indicated



Since the AAMR displayed a normal distribution, the homogeneity of variance for the dependent variable, AAMR, was assessed at each level of the vectors using Bartlett's test for homogeneity of variance (Table 6). No significant differences were found, thereby, satisfying the regression assumption of homogeneity. Additionally, standardized residuals were examined for outliers that may effect the interpretation of the data. No standardized residuals exceeded the +/- 3.0 mark, indicating no outliers existed within the AAMR data.

Table 6: Bartlett's Test for Homogeneity of Variance for AAMR

SOM Variable	Mean of Group 0	Mean of Group 1	DF	Chi-Square	Probability
SDOH1	991.11	1004.34	1	3.2987	0.0693
SDOH2	959.03	1038.50	1	0.4487	0.5209
SDOH4	1008.21	904.51	1	1.1665	0.2801
HB1	972.11	1024.72	1	1.4233	0.2329
HB3	999.68	926.93	1	0.1476	0.7009
HB4	996.39	977.40	1	0.1746	0.6761
HB5	999.55	938.56	1	0.4025	0.5258
HB6	987.20	1070.35	1	0.1003	0.7515

After assumptions were assessed, the eight variables were submitted to a standard multiple regression analysis. Table 7 displays the descriptive statistics for the model. The model was found to be significant and accounted for 37% of the variation in AAMR ($R^2 = 0.37$). The overall model was found to be significant with an F-value of 4.928 ($p < 0.0001$). Examining the dependent mean value from the regression analysis, it is evident by the reduced mean that the two counties excluded from the SOM_{HB} analysis were excluded from this analysis.

Table 7: Descriptive Statistics for Multiple Regression Analysis

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Value
Model 1	8	239,721.41	29,965.18	4.928
Error	66	401,312.94	6,080.50	
Corrected Total	74	641,034.34		
R	R^2	Adjusted R^2	Root MSE	AAMR Mean
0.612	0.374	0.298	77.98	993.86

The standardized beta weights of two of the eight variables in the model were found to be significant (Table 8). SDOH1 had a beta weight of 0.353 (p=0.049) and $SOM_{SDOH}2$ had a beta weight of 0.494 (p=0.011), while all other variables were not significant. As verification of model fit, the errors were deemed independent by a non-significant Durbin-Watson statistic of 1.989. Multicollinearity was assessed by examining the condition indices. The condition index, which should be below 30, indicating no multicollinearity, was 7.198. Variables showed tolerances ranging from 0.263 to 0.722, well above the 0.10 cut of for multicollinearity (Pedhazur, 1997). The standardized prediction equation was as follows:

$AAMR = 0.35362\ SOM_{SDOH}1 + 0.494\ SOM_{SDOH}2 + -0.14804\ SOM_{SDOH}4 + 0.14924\ SOM_{HB}1 + 0.01019\ SOM_{HB}\ 3 + -0.05483\ SOM_{HB}\ 4 + -0.05296\ SOM_{HB}\ 5 + 0.17108\ SOM_{HB}\ 6$

Table 8: Parameter Estimates for the Multiple Regression Model

Variable	DF	Unstandardized Regression Coefficients	Standard Error	t-value	Pr > |t|	Standardized Regression Coefficients	Tolerance
Intercept	1	923.099	28.021	32.94	<.0001	0	0
SDOH1	**1**	**67.590**	**33.721**	**2.00**	**0.0491**	**0.3536**	**0.3047**
SDOH2	**1**	**93.224**	**35.816**	**2.60**	**0.0114**	**0.4940**	**0.2633**
SDOH4	1	-44.338	41.257	-1.07	0.2864	-0.1480	0.4999
HB1	1	28.018	27.936	1.00	0.3196	0.1492	0.4284
HB3	1	3.473	40.596	0.09	0.9321	0.0102	0.6684
HB4	1	-14.913	34.455	-0.43	0.6666	-0.0548	0.5910
HB5	1	-16.831	36.433	-0.46	0.6456	-0.0530	0.7218
HB6	1	58.300	41.048	1.42	0.1602	0.1711	0.6538

It is from this model that the null hypothesis for research question two was answered. The SOM clusters for social determinants of health were found to be significant predictors of age-adjusted mortality rate above the SOM clusters for health behaviors. However, this was not the case for all SOM_{SDOH} variables. The dummy variable for SOM_{SDOH} cluster 4 (*Healthy No Safety-Oriented*) was not found to be significant during analysis. But the predictive abilities of SDOH exceeded the predictive abilities of health behaviors within this sample, as no SOM_{HB} variables were found to be significant.

The multiple regression analysis pinpointed which predictor set was stronger. All of the SOM variables combined accounted for 37% of the variability in the health outcome, AAMR. Only two SDOH standardized beta coefficients were significant. The research question for phase 2 was confirmed because the null hypothesis, SOM_{SDOH} is a stronger predictor set of AAMR than SOM_{HB}, was not rejected.

Phase 3 - Correlation Analysis

When the 114 social determinant and 31 health behavior variables were submitted to the two separate Self-Organizing Map (SOM) processes for analysis, the data went from being ratio or interval level data upon entry to being represented by nominal clusters upon SOM output. However, during Phase 2 of the analysis the cluster variables were dummy for entry into the multiple regression analysis, which allowed for correlations addressing dichotomous variables to be utilized. The Phi correlation was used to correlate the dichotomous variables with AAMR, while point-biserial correlations were used to correlate each of the dichotomous variables.

Interesting patterns emerged when the intra-correlations of SDOH and HB dummy variables were examined. Of the three point-biserial intra-correlations for SDOH variables, 100% were found to be significant (Table 9). The SDOH variables submitted to the SOM process (114 variables) were, thereby, validated as a cohesive construct. This was not the case for the health behavior intra-correlations, however (Table 10). Only four of the ten HB intra-correlations were found to be significant (40%). The variables representing health behavior did not exhibit the same level of cohesion as the social determinants variables, indicating that the variables were possibly measuring differing constructs or validity issues with the variables. This ambiguity among the HB variables could explain the non-significant regression coefficients obtained in Phase 2.

Table 9: Point-biserial Intra-correlations for SDOH Dummy Variables

Variable	$SOM_{SDOH}1$	$SOM_{SDOH}2$	$SOM_{SDOH}4$
$SOM_{SDOH}1$ p-value	1	**	**
$SOM_{SDOH}2$ p-value	**-0.62** **<.0001**	1	**
$SOM_{SDOH}4$ p-value	**-0.28** **0.01**	**-0.29** **0.01**	1

Table 10: Point-biserial Intra-correlations for HB Dummy Variables

Variable	$SOM_{HB}1$	$SOM_{HB}3$	$SOM_{HB}4$	$SOM_{HB}5$	$SOM_{HB}6$
$SOM_{HB}1$	1	**	**	**	**
p-value					
$SOM_{HB}3$	**-0.25**	1	**	**	**
p-value	**0.03**				
$SOM_{HB}4$	0	-0.12	1	**	**
p-value	-0.33	0.32			
$SOM_{HB}5$	**-0.27**	-0.09	-0.13	1	
p-value	**0.02**	0.41	0.28		
$SOM_{HB}6$	**-0.25**	**-0.25**	-0.12	-0.09	1
p-value	**0.03**	**0.03**	0.32	0.41	

The inter-correlations between SDOH and HB dummy variables were examined to assess the first null hypothesis for research question 3. Only 27% of the inter-correlations were found to be significant (Table 11). Even though the null hypothesis of no correlation between the constructs could not be rejected, the minimal correlation between the variables supports the notion that social determinants of health and health behaviors are two separate constructs as the Public Health Model of the Social Determinants of Health indicates.

Table 11: Point-biserial Inter-correlations among Dummy SDOH and HB Variables

	SOM_{HB}									
SOM_{SDOH}	1	p-value	3	p-value	4	p-value	5	p-value	6	p-value
1	0.08	0.49	-0.02	0.83	0.02	0.85	0.13	0.26	-0.13	0.28
2	0.2	0.08	**-0.24**	**0.04**	0.08	0.49	**-0.26**	**0.02**	**0.26**	**0.02**
4	-0.11	0.32	**0.38**	**0**	-0.01	0.94	0.04	0.74	-0.1	0.38

To address the last two null hypotheses within Research Question 3, the eight Phi correlations were examined (Table 12). Of the three SDOH correlations with AAMR, two were significant (66.67%). Alternatively, only 40% of the Phi correlations between HB and AAMR were significant. Both sets of variables were correlated to age-adjusted mortality rate, which caused a failure to reject the first two null hypotheses for research question 3. However, the SOM_{SDOH} variables produced more significant correlations, confirming the stronger predictive abilities of the social determinants of health variables over the health behavior variables.

Table 12: Phi Correlation Matrix for AAMR to SOM_{SDOH} and SOM_{HB}

Variable	AAMR
$SOM_{SDOH}1$	0.07
p-value	0.55
$SOM_{SDOH}2$	**0.37**
p-value	**0.00**
$SOM_{SDOH}4$	**-0.36**
p-value	**0.00**

Variable	AAMR
$SOM_{HB}1$	**0.28**
p-value	**0.01**
$SOM_{HB}3$	-0.21
p-value	0.07
$SOM_{HB}4$	-0.07
p-value	0.55
$SOM_{HB}5$	-0.19
p-value	0.10
$SOM_{HB}6$	**0.24**
p-value	**0.03**

Summary

The self-organizing map was able to handle the large amounts of data during one process and put forth interpretable information. The SOM_{SDOH} had a more natural fitting map than the SOM_{HB} as evidenced by the cluster index (70 to 53). The resulting number of clusters differed between the two analyses with more clusters resulting from the analysis with the smaller number of variables (health behaviors). The SOM_{SDOH} resulted in four clusters being named *Working Class, Lower Class, Upper-Middle Class- Urban, and Long-term Farmland.* The SOM_{HB} resulted six clusters being named, *Restricted, Health-Promoting, Unhealthy, Conflicted, Conflicted Mental & Physical Health, Safety not Health Oriented.* When comparing the geographical distribution of both sets of clusters, social determinants of health clusters tend to visually align better with totally mortality (AAMR) than the health behavior clusters. Additionally, social determinants of health dummy variables were found to be stronger predictors of age-adjusted mortality rate than health behavior dummy variables through standard multiple regression analysis and correlation analysis. Relationships that were found during the SOM process were confirmed during correlation analysis.

CHAPTER V

DISCUSSION

Research on how the environment and society interact with health continues to grow, and addressing policy and social issues to alter adverse health outcomes have been highlighted (Marmot & Wilkinson, 1999; Raphael, 2003). The ability for social indicators to predict health outcomes opens more avenues to prevention than changing individual health behaviors alone. Social determinants of health, although not a new concept in public health, have been constructs of interest of late. One result of this much-needed attention was the creation of a new field of science in order to study the complex phenomenon of social determinants - Social Epidemiology (Berkman & Kawachi, 2000). Additionally, an increase in the understanding and the value for community-level health indicators and flexible analytical methods to study the relationships between social indicators and health outcomes have emerged (Marmot & Wilkinson, 1999).

The intent of this study was to present a method for analyzing existing, nationally available social data in a health context to further elucidate the links that are found within the Public Health Model of the Social Determinants of Health (PHM). The PHM is a comprehensive model with testable links, which was used as the theoretical guide for selecting and categorizing variables from archival data sources that were publicly available. Such sources include the Behavioral Risk Factor Surveillance System, Oklahoma Vital Records, U.S. Census Bureau data, and Bureau of Labor Statistics, as well as others. Variables represented social determinants (socio-economic determinants, psychosocial risk factors, and community and societal characteristics), health behaviors (physical and mental behaviors), and health outcome (age-adjusted mortality rate, AAMR) within Oklahoma at the county level. Although smaller community levels could be defined within some datasets, county was the smallest geographic level that was available across all datasets in this study. Finally, this study introduced the Self-Organizing Map (SOM) as an alternate data reduction technique for analysis of health and social data. To examine the testable links of the PHM, the research questions and analyses for this study were divided into three phases: SOM for initial data reduction, regression analysis for prediction, and correlation analysis to determine relationships among the variables.

SOM Analysis

Phase 1 sought to determine the underlying relationships between the vast amounts of data that were utilized within this study. Two hypotheses were posed for the first phase, which sought to establish the amount of variation present within social determinants of health variables or health behavior variables. If no variations were present, then a single cluster would have resulted for each set of variables, while multiple clusters would have surfaced if variation existed. After submitting variables to individual SOM processes, both null hypotheses were subsequently rejected because significant variations existed within the data as evidenced by the number of resulting clusters and the cluster means for each of the variables (Appendix C & E).

SOM_{SDOH} Clusters

While all variables were represented to varying degrees within each cluster, the clusters were defined by a reduced set of distinguishing variables that were unique to each. The SOM_{SDOH} resulted in a four-cluster solution with clusters being named: *Mid-Century Service-Oriented Communities, Struggling Minority Communities, High Income and High Education, and Long-term Farmland.* The SOM_{HB} resulted in six clusters being named: *Restricted, Health-Promoting, Overweight and Unsafe, Conflicted Intimate Partner Violence, Conflicted Mental & Physical Health, Safety Not Health-Related.*

Thompson and Hickey (2005) defined working class as manual or service oriented workers with low job security, having common household incomes ranging from $16,000 to $30,000, and possessing a high school education. The resulting *Mid-Century Service-Oriented Communities* cluster from the current SOM analysis significantly identified with service-oriented occupations but was not significantly defined by education or income. However, the per capita income and median household income fell within the range indicated by Thompson and Hickey. Other studies included housing characteristics as an indicator of social class and deprivation (Kreiger, Williams, & Moss, 1997; Langhout, Rosselli, & Feinstein, 2007). The Townsend index (Berkman & Kawachi, 2000; Kreiger, Williams & moss, 1997) indicates high rates of rental property as a measure of social deprivation for a community. The *Mid-Century Service-Oriented Communities* cluster found within the current study indicated high rates of renter-occupied housing, among several other indicators of older homes that are often used for transitory rental property (Kemeny, 1978). The Gamaliel Foundation (2006) noted that white collar and professional families have been moving out of the cities and older suburban areas and leaving them to middle and working class families. The *Mid-Century Service-Oriented Communities* found within the current study identified with and upheld this statement as a significant indicator of the *Mid-Century Service-Oriented Communities* cluster was urban areas.

Thompson and Hickey (2005) defined lower class as a group of persons who have poorly paid positions or rely on governmental assistance and have low levels of education. The resulting *Struggling Minority Communities* cluster from the current SOM analysis significantly identified with production occupations and was negatively associated with management positions. Counties within this cluster also had high unemployment rates. While low-income variables were not significant distinguishing variables for this cluster, high-income variables were negatively associated with the variable. These variables combined to conform to the Thompson and Hickey category of lower class.

The third cluster, *High Income and High Education,* closely aligned with Thompson and Hickey's (2005) upper middle class designation: 1) highly educated; 2) professionals and managers; and 3) household incomes varying from the high 5-figure range to above $100,000. Distinguishing variables found within the cluster from the current study were mainly around income (between

$75,000 and $200,000) and housing characteristics (high rent and high home values). As with the Thompson and Hickey definition, the *High Income and High Education,* cluster from this study was significantly associated with sales occupations and high rates of college education.

The final SDOH cluster obtained during the SOM process was labeled *Long-term Farmland.* Hunt (2002) related that income was not an accurate measure of farmers of today because they either invest assets in farming corporations or reinvest their assets into their own farms. Additionally, farmers, or farm owners more specifically, are much less migratory than those who work on farms. According to the 2002 Agriculture Census (United States Department of Agriculture, 2004), 72% of principal operators have worked on the same farm for 10 years or more and the average age of principal operators was 55.3 years. The resulting *Long-term Farmland* cluster was named such based on the distinguishing variables obtained. Rural farming and agriculture land were the most significant variables for this cluster. Similar to the agricultural census data (United States Department of Agriculture, 2004), the *Long-term Farmland* cluster had the largest median age group among the clusters (41.9 years), although it was slightly lower than the national rate. Additionally, farming and management occupations were significantly related to *Long-term Farmland*. As for lack of mobility, many housing characteristics related to older homes and long-term residence were all found to be significant distinguishing variables of *Long-term Farmland*. Finally, the *Long-term Farmland* cluster was significantly defined by the large percentage of persons within these counties who were white as well as the small proportion of the population who would self-identify with more than two races. This finding coincides with the agricultural census where 97% of principal operators were white (United States Department of Agriculture, 2004).

SOM_{HB} Clusters

As with the SOM_{SDOH} clusters, all health behavior variables were represented within each cluster but they were defined by a reduced set of distinguishing variables. The SOM_{HB} resulted in six clusters being named: *Restricted, Health-Promoting, Overweight and Unsafe, Conflicted Intimate Partner Violence, Conflicted Mental & Physical Health, Safety Not Health-Related.*

Cluster 1, called *Restricted,* was dominated with significant variables that restricted current and future activities. Variables included threats of violence from intimate partners, tobacco use, high levels of poor mental health days, and restricted activities due to poor mental or physical health. In addition, *Restricted* had low rates screening exams, health care provider visits, and low physical activity levels. These variables were similar to those found within a study conducted by Vest, Catlin, Chen, and Brownson (2002) that found persons experiencing intimate partner violence were also under- or uninsured, were currently smoking, had self-reported fair/poor health, and had frequent mental distress issues.

A cluster emerged from the SOM analysis of health behavior data within the current study that represented health promoting behaviors. Counties within Cluster 2, named *Health-Promoting,* were found to have significant distinguishing variables dealing with health screenings, safety, and physical activity. In addition, low rates of obesity, current smoking, and intimate partner violence were found within this cluster. This finding corresponds to other cluster analysis studies conducted on health behaviors in which health-promoting clusters were found, one from Israel (Hagoel, Ore, Neter, Silman, & Rennart, 2002) and one from Germany (Schneider, Huy, Schussler, Diehl, & Schwarz, 2009). Variables included in both of these analyses were similar to the current study: regular tobacco use, unhealthy diet, and physical inactivity (Hagoel et al.; Schneider et al.). In addition, the Schneider et al. study included excessive alcohol consumption, while the Hagoel et al. study included periodic medical checkups. While the health-promoting clusters accounted for the largest proportion of participants in both studies (Israel = 44%; Germany = 25%), *Health-Promoting* within the current study accounted for a slightly smaller percentage of counties (20%) and was the second largest cluster. This difference in proportions could be a result of examining counties instead of persons as in the other studies or a general reflection of attitudes toward health promoting behaviors within country borders.

Cluster 3, *Overweight and Unsafe,* was only defined by four variables, but all of those variables had generally unhealthy connotations to them. Low rates of PSA test and seatbelt usage, mediocre mental health, and high rate of proportion overweight combined to make the *Overweight and Unsafe* cluster. This cluster was similar to one discovered by Schlundt et al. (2003) in which they identified an “overweight and unhealthy” cluster.

In contrast to Cluster 1, *Conflicted Intimate Partner Violence* (Cluster 4) provided interesting information regarding intimate partner violence. When actual physical violence inflicted by an intimate partner is significant, as it is in Cluster 4, counties exhibit higher rates of health screenings, health provider exams, seatbelt usage, and physical activity. This result is not only contradictory to Cluster 1 but to other studies that have examined intimate partner violence (Coker et al., 2002; Vest et al., 2002). The conflicted nature of these variables may suggest that analyses of the past may not have captured the full range of information regarding intimate partner violence and other related variables were missing from analysis.

Cluster 5 was named *Conflicted Mental & Physical Health* because of the combination of significant variables obtained. *Conflicted Mental & Physical Health* exhibited the highest smoking rate, poor mental health days, and low levels of physical activity. In contrast, *Conflicted Mental & Physical Health* also exhibited the highest rate of PSA testing, high rates of fruit and vegetable consumption and seatbelt usage, and low percentage of overweight. Hagoel et al. (2002) found a cluster that exhibited similar contrasting variables, which they called “ambivalent.” The “ambivalent” cluster consisted of high rates of overeating and alcohol consumption but displayed higher rates of medical visits and physical activity than another cluster within the study. While the variables are not the same between the current study and the Hagoel et al. study, the concept of balancing healthy and unhealthy behaviors is the same.

The use of a seatbelt while in a vehicle that experiences a crash has been shown to dramatically reduce fatal injuries by more than 50% depending on the type of vehicle (Dissanayake & Ratnayake, 2009). A study of youth risk behaviors conducted by Cox, Larkin, and Scott (n.d.) found that students who never wore seatbelts also had greater levels of feeling sad or hopeless (OR = 1.57), considered suicide (OR = 1.88), or had attempted suicide (OR = 2.11) compared to those who always wore seatbelts. The *Safety Not Health-Related* cluster, Cluster 6, within the current study found a similar relationship between seatbelt usage and mental health. However, the cluster was also defined by a significant negative relationship with other health behavior indicators. The addition of negative health behaviors indicates that some persons are much more safety-conscious than they are concerned with physical health. Another possibility for the differences could be the fact that the Cox et al. study was conducted with youth while the data represented in the current study focuses on adults. Competing priorities may force adults to focus on one aspect of health - in this case safety - versus a comprehensive view.

Regression Analysis

Raphael (2003) stated that social determinants of health predict individual and population health better than health behaviors. This study sought to examine which group of variables were stronger predictors of health outcome by analyzing the reduced SOM output for both social determinants of health and health behaviors through regression analysis. The findings from this study supported the notion that social determinants of health were stronger predictors of health outcome than health behaviors. When looking at the predictive abilities of the social determinants clusters, *Mid-Century Service-Oriented Communities and Struggling Minority Communities* were both found to significantly predict age-adjusted mortality rate. This finding was similar to a study conducted by Regidor et al. (2005), in which mortality risk for male workers (skilled and un-skilled) was greater than that of men who were in management or professional positions. As a result of the significant findings related to social determinants of health within the current study, community and social epidemiologists can begin to use social-related data in a more effective and reliable manner by utilizing the SOM algorithm. They are no longer solely dependent upon health behavior data that may not provide accurate reflections of population health.

Correlation Analysis

In addition to looking for mathematical patterns and predictive abilities of social determinants of health data, relationships between the variables were also sought. Information about what the structure of the data after the SOM process occurred was assessed for construct validity. Link and Phelan (1995) and Marmot and Wilkinson (1999) all indicate that social determinants of health data interact and create a cumulative effect on a person’s health. The interactive nature of social determinants of health was once again displayed within this study. By compiling a naturally fitting SOM, significant standardized regression coefficients, and significant intra-correlations (correlations among the three SDOH dummy variables), it was determined that social determinants data were representative of a single construct.

Use of SOM Clusters to Examine Individual Variables

Beyond the large concepts of the analysis, individual variable results supported and added to existing theory. For example, Oreopoulus, Stabile, Walld and Roos (2008) found a relationship between low birth weight babies and higher mortality rates. Within this study, the clusters with the highest percentage of low birth weight and very low birth weight babies (*Mid-Century Service-Oriented Communities and Struggling Minority Communities)* also had the highest AAMR based on the SOM. Additionally, similar to the study conducted by Hack, Flannery, Schlucter, Carter, et al (2002), the clusters with the highest rates of very low birth weight babies also had the highest percentage of adults with no high school degree.

Marmot and Wilkinson (1999) suggested that the distribution of wealth creates differences in health outcomes. Analyzing individual characteristics of socio-economic status, the compositional approach (Duncan, Jones, & Moon, 1998), within the current study resulted in vague and conflicting information. The SOM clusters with high income levels related to age-adjusted mortality rates near the overall mean. Further, the cluster with the highest median household income had the second lowest AAMR, but the lowest median household income had the highest AAMR. While high-income levels did not translate into exceptionally better health, the current study showed an inequitable relationship between low income and poor health.

Examining data on a larger scale using the contextual approach (social networks, community, and geographic area) set forth by Duncan, Jones, and

Moon, (1998), the current study continued to expand upon existing theory. The Townsend and Carstairs indices of social deprivation (Berkman & Kawachi, 2000) utilized unemployment as part of the measure of social deprivation. Thus, it was not a surprise to see that the cluster with the highest unemployment rate also had the highest AAMR. However, what was surprising was to discover that when examining the housing characteristics (also used in the Townsend and Carstairs indices [Berkman & Kawachi, 2000]) no clear pattern prevailed for home owners within Oklahoma, but the cluster with the highest rate of low-income rental property did have the highest rates of AAMR, which supports findings from other studies that examined public housing (Fertig & Reingold, 2007) and rental property (Dun, 2002) in relation to health outcomes. Information such as this could allow community and social epidemiologists to focus on the most vulnerable populations who are truly in need of public health intervention and resources by targeting those counties with high rates of low-income rental property.

Conclusions

Several conclusions in relation to the research questions can be drawn based upon the results of this study. In general, results indicate that while health behavior variables consisted of more than one cluster, the geographic distribution of the social determinants of health clusters obtained visually aligned with the geographic distribution of total mortality (AAMR) better than the geographic distribution of the health behavior clusters. The regression analysis further substantiated the predictive abilities of social determinants of health variables over health behaviors and the intra-correlations among the SOM_{SDOH} dummy variables provided construct validity. Self-organizing maps (SOM) were found to be useful in preserving the mathematical nature of the data while reducing the data to useable and testable levels. Link and Phelan (2005) and Marmot and Wilkinson (1999) defined social determinants of health as the broad range of social exposures that interact and cumulatively relate to health. By finding a natural mathematical relationship among the SDOH variables and significant correlations, the current study provided additional evidence of the interwoven nature of social determinants of health data and supported their definition. The significant predictive ability of the SDOH clusters also provided evidence toward the impact of social determinants of health on health outcomes. Below you will find a series of specific conclusions based on the analysis of this study.

SOM and Resource Allocation

The SOM provided a new and interesting look into the nature of the social determinants data. Categorizing the SDOH variables provided an organized way to examine the impact of each individual variable upon the cluster distribution allowing one to quickly determine outlying counties within the data. Viewing information in this manner may allow for better alignment of resources and efforts to target specific issues within a community. For example, one could quickly identify the extreme disparity in income found within Tillman County. Additional in-depth community assessments could be conducted to determine why this disparity exists in the first place and where the disparities are truly occurring within the county borders. Income information available from the U.S. Census

Bureau could narrow down if the disparities are occurring across city limits or concentrated within one location. Tillman County is ranked 36th in the State for total mortality. While this rate in itself is not the worst rate among Oklahoma counties, examining mortality data at lower geographic levels would provide information to determine if and where the true health disparities exist (Kreiger, Williams, & Moss, 1997). Programmatic and community efforts could then be designed to bring the persons in the bottom rung of the income scale upward to eliminate the extreme income disparity that exists within this county.

Oklahoma's Turning Point Initiative is an example of such efforts. Oklahoma has an extensive network of Turning Point community coalitions whose purpose is to link all aspects of a community together to address the varying issues that affect health (Oklahoma State Department of Health, 2009). Turning Point communities across the state focus on projects related to increasing educational levels, creating healthier workforces in order to decrease health insurance expenditures, fostering economic development within a community to attract higher paying jobs, and promoting many other public health-related activities. Efforts such as these have benefits beyond increasing income; they translate into sustainable communities and healthier populations in the long run (Strong and Healthy Oklahoma, n.d.). Coupling community level efforts with appropriate alignment of state resources could allow for maximal impact on health outcomes.

Archival Data

Additionally, examining social data in a health context maximizes the analysis of existing data that were collected for other purposes, therefore, decreasing the collection of redundant, expensive data. By using these existing, publicly available data sources and the self-organizing map algorithm, community and social epidemiologists can examine health issues on limited budgets with minimal collection efforts. There is great benefit in knowing which social determinants of health affect certain areas (Marmot & Wilkinson, 1999). Some social determinants variables cannot and shouldn't be changed. For example, the rural-urban dichotomy of land use is needed in order to maintain a sustainable society (Forster, 2009). However, knowing which areas are affected by a particular health outcome or other changeable social determinants could allow community and social epidemiologists to work smarter and not harder. Having the ability to divert needed resources to a particular area while not burdening others with unnecessary items, tasks, or issues would allow precious resources to be used wisely.

Practical Application of SOM to Target Public Health Interventions

In the light of numerous explanatory variables that need to be analyzed to understand health outcomes, having a method that allows for the discovery of patterns within the data and guides an analyst to specific variables is a powerful tool (Basara & Yuan, 2008). The SOM algorithm pinpoints variables that play active roles in the formation of clusters. Patterns within the individual attribute maps and the distinguishing variable charts could be examined for the types of services that are generally needed within a county (i.e., mental health versus physical activity or nutrition or employment versus education). Further, community and social epidemiologists could select only the most important variables to a specific cluster group on which to focus additional analysis or programmatic efforts. For example, when the distinguishing variables were examined in the SOM_{HB} (Figure 23) having a prostate-specific antigen test (PSATEST) was a very significant variable in the formation of Cluster 5, *Conflicted Mental & Physical Health*. However, in Clusters 1, 3, or 6 (*Restricted, Overweight and Unsafe,* and *Safety Not Health-Related)*, this variable, although significant in the formation of the cluster, was in the negative direction. When an epidemiologist's efforts are targeted at increasing prostate screenings, focusing on the *Conflicted Mental & Physical Health* cluster would prove those efforts to be wasted because men in these counties already have very high rates of prostate screening. Focusing screening efforts on men living in a county within the other three clusters could create significant health benefits. However, this conjecture cannot be confirmed through this analysis because the health outcome was not entered into the same SOM as health behaviors or social determinants of health. One can only visually examine how AAMR clusters within health behaviors.

Construct Considerations

While it was encouraging to visually see the mathematical structure of the social determinants of health and health behavior data, several conclusions can be drawn about the construct development of both sets of data. For the social determinants of health data, the number of variables was dominated by socio-economic determinants. Even within that construct, variables were heavily influenced by or related to income. One could argue that just because you have a high level of education or income doesn't mean you couldn't choose to live well below what those means could afford. The opposite is true for some persons that have low income as well; they could choose to live well above their means, albeit probably for a limited time. However, when considering aggregated data, as in this study, those issues tend to even out. This is especially true at such a large geographic level as county (Eisgruber & Schuman, 1963). This heavy weighting of income-related variables within the SDOH dataset translated into clusters being driven by these variables.

Similarly, it can be equally concluded that construct development issues appeared in two other areas: the psychosocial risk factors for SDOH and the health behavior variables. Due to a lack of available psychosocial data, variables were limited to only five in a pool of 114 for SDOH. This could account for the low appearance of psychosocial risk factors within the distinguishing variables. In fact, psychosocial risk factors were only found to be a defining variable twice, and it was the same variable (Median number of poor mental health days) appearing in *Struggling Minority Communities* and *Long-term Farmland.*

In addition, the construct development issues could have affected the results of the regression analysis and the correlations. Social and Community Epidemiologist need to be aware of such issues when developing studies for

SOM analysis and attempt to balance the variables so one construct does not have the ability to overpower the results.

Limitations

There are several limitations to this study that must be considered. First, because health behavior data are not as readily available as most of the social determinants data, trying to obtain data for all counties within Oklahoma proved to be difficult. The Behavioral Risk Factor Surveillance System (BRFSS) imposed automatic cell suppression on variables with less than three observations leaving records with missing data points (Oklahoma State Department of Health, n.d.a). Some of the health behavior questions also reside within the optional modules of the BRFSS and, therefore, are not asked every year by every state (CDC, n.d.). This may require the data to be aggregated by non-consecutive years. For example 2002, 2004, and 2006 might be aggregated instead of 2002 through 2006. This could potentially impact the data by introducing bias, because the values may look very different in 2003 and 2005 than they do in the included years (Hartman, Forsen, Wallace, & Neely, 2002).

Secondly, the number of social determinants of health variables out-numbered the health behavior variables by almost four to one. This could have impacted some results obtained during analysis. Because the SOM is an iterative process, the number of variables entered into the process dictated the number of iterations (Deboeck, 1999). If there were not enough health behavior variables to adequately represent the health behavior construct, then the SOM would have difficulty finding a stable resolution or exit the iterative process too soon (Deboeck, 1999). Balancing the number of variables could be one method for determining if this was a limitation within this study.

Thirdly, the selection of the unit of analysis, county, could be considered a limitation to this study. While county-level data allowed for the analysis of the entire state of Oklahoma in SDOH data and most of the state in HB data, having a smaller geographic area to analyze would allow for finer detail within the results. Aggregated data ultimately assigns information to persons that may not be truly reflective of their reality (Eisgruber & Schuman, 1963). Understanding this phenomenon and considering it carefully before conclusions are drawn about a specific area is criticial for Community and Social Epidemologists.

SOM Process

The Self-Organizing Map algorithm developed by Teuvo Kohonen (2001) was designed to analyze complex data structures while preserving their natural mathematical relationships. The SOM process has been used in various industries including computer gaming (Wu, Liu, Thomas, & Huang, 2000), genetic research (Wang, Delabie, Aasheim, Smeland, & Myklebost, 2002), and clinical medicine (Oyana Boppidi, Yan, & Lwebuga-Mukasa, 2008), but to date, use of this method in public health has been limited (Basara & Yuan, 2008). Within this study, the SOM algorithm was used to analyze 145 variables in two separate analyses (114 SDOH and 31 HB) with 77 records in each. As Wang, Delabie, Aasheim, Smeland, and Myklebost (2002) found, patterns within the data can be quickly assessed and summarized.

Since the SOM process was found to be of significant benefit to analyze large amounts of data, the following steps are provided for community and social epidemiologists who wish to employ the same technique with their large data sets. These steps are laid out for use with the Viscovery SOMine software.

Dataset creation phase:

1. Ensure data are in the proper format
    a. If individual raw data points are available, categorical data can be used.
    b. If aggregated data points are available, transform categorical data to percentages of the population included in the variable (i.e., sex = male to % of male in population).
    c. Include a naming variable but make it around three characters for labeling.
2. Put the data in a Microsoft Excel table or text format (.txt).
3. Open a new SOM project, name the project and select the location to be saved.
4. Import the formatted data from its location.
    a. Adjust the format of any variables if needed
    b. Indicate if the first row is data or variable names by checking the box.
5. Select the key attributes of the data. This is usually the ID or naming variable.
6. Define any nominal variables included in the dataset through the wizard.

7. Tune (i.e., transform, replace values, etc.) the variables that were entered into in the system.
    a. Records can be deleted (i.e., outlier cases)
    b. Replacements defined for missing variables (i.e., mean, median, constant)
    c. Transformations can be performed, either sigmoid or logarithmic

    Note: All data were cleaned and adjusted before entry into the SOM process in this study but could have been easily accomplished in this step.
8. Write the dataset (either full or sample) to the new data mart.

Modeling Phase:

1. Select the data mart created in the last phase
2. Prioritize the attributes (variables). All variables within this study were set to 1 because they all entered the SOM process equally. If variables theoretically have more weight than others the priority should be adjusted to reflect that. Priority settings cannot exceed 1.
3. Creating the actual SOM is the next step. Several options are available to define the map parameters:
    a. Map Size
        i. Number of Nodes: should be set to roughly **ten times** the number of variables entered into the process
        ii. Shape: three choices exist
            1. Automatic - the system selects the best shape

2. Square Map – creates a square map

   Note: Kohonen (2001) cautions against a square map as it could create convergence issues)

   Ratio – Preferred method as it gives the control to the analyst

   Note: This was the method used within this analysis. The ratio was set to 100:75.

b. Training Schedule: set to accurate for a more detailed map.

   Note: The larger the dataset the longer it will take to create a more detailed map so the normal setting may be preferred.

c. Tension: smaller tension values allow the map to adapt to the data space and cause less averaging of the data. Values can range from 0-2.

   Note: The value for this study was 0.2 to allow the data space to be represented fully while still coming to convergence

4. The final step in this process is defining the segments. This is where the bulk of the analysis for this study took place. The following items were used to define the clusters for this study

   a. SOM – Ward clusters – the original cluster solution

   b. Attribute maps – individual maps for every variable entered into the process. Original cluster locations are displayed with a black line and the data range is displayed by varying shades of color.

c. Segment tables – segment tables enable examination of all the variables in a table format by cluster.
d. Deviation Charts – display the standard deviations from mean for each cluster
e. Statistics – includes Descriptives, Correlations, Principal Components, Histograms, Frequency tables, Box plots and Scatter plots at several different levels: individual records (nodes), cluster, neighborhood, specific selection, or the entire map.

Model Application Phase:

Note: This phase was not conducted within this study because it was not within the purpose of the SOM. Explanation of the process is only cursory, but the steps are similar to those listed above and follow general data mining processes.

1. The model created in the previous phase can be applied to similar data or a subset of the data used in the model creation phase.
2. A data mart must be created as above
3. A model must be chosen that was created as above
4. Segmentations must be examined for model fit
5. Cluster segments can be exported for additional analysis.

Implications of Findings

This study provided several implications regarding theoretical and practical research of social determinants of health and self-organizing maps. First, this study provided empirical evidence in support of the links within the

Public Health Model of the Social Determinants of Health (PHM). However, the full model was not tested using the self-organizing map (SOM). Future work needs to be done regarding health care access variables. Such work could provide needed information to help settle the long debate of what is more important to good health outcomes: health care access, health behaviors, or social determinants. Although this study has provided evidence that social determinants of health are stronger predictors of mortality than health behaviors, adding another layer of information to the analysis could change this outcome.

While the PHM indicated the variables that could represent social determinants of health, health behaviors, and health outcomes, this study brings out several considerations and implications for future studies. The PHM implies more weight is given to social determinants of health variables than health behaviors or health access variables as evidenced by the number of variables that are needed to represent social issues over health access or behaviors. This study points out that work may need to be done to balance the number of variables representing each dimension, but the value of the dimensions and their relationship to health outcomes are still related. Further, if social determinants of health variables continue to predict health outcomes to a higher degree, then the Public Health Model of the Social Determinants of Health should be altered to accurately reflect this weighting.

This study also provided practical implications for community and social epidemiologists. It showed the power of combining an adaptable public health model with a flexible analytical tool. The SOM gives the epidemiologist the ability to represent large amounts of data without having to manually adjust the number of clusters until the best solution is achieved. The SOM also fully represents the mathematical nature of the data and reduces the data to useable levels. Further, community and social epidemiologists can utilize the methods set forth in this study to verify associations between other subsets of public health data. One possible direction for future research would be to combine all of the variables used within this study into a single SOM analysis. Questions such as the prostate example above could be assessed for impacts on health outcome and determine what societal factors are associated with low screening rates.

This study is only a starting point in public health research using the PHM and self-organizing maps. This was an exploratory study that can be used to help further refine social determinants of health models to accurately reflect the individual and society. Confirmatory studies need to be conducted to verify the results.

Community and social epidemiologists can apply the SOM techniques to help flesh out the real world problems that may not be evident when just examining health data. Once narrowed they can more appropriately devise strategies to resolve the underlying issues of adverse health outcomes and not continue to perpetuate the bandaging of the intermediate issues of health behaviors.

REFERENCES

AcademyHealth (2004). Health Outcomes Core Library Project. Retrieved August 15, 2008 from http://www.nlm.nih.gov/nichsr/corelib/houtcomes.pdf.

Adler, N.E., Newman, K. (2002). Socioeconomic disparities in health: Pathways and policies. *Health Affairs, 21,* 60-76.

American Cancer Society. *Cancer facts & figures 2008*. Atlanta: American Cancer Society; 2008.Retrieved January 30, 2009, from http://www.cancer.org/docroot/STT/content/STT_1x_Cancer_Facts_and_Figures_2008.asp?from=fast.

Anderson, L. M., Scrimshaw, S. C., Fullilove, M. T., Fielding, J. E., & the Task Force on Community Preventive Services. (2004). The Community Guide's model for linking the social environment to health. *American*

Ansari, Z., Carson, N.J., Ackland, M.J., & Vaughan, L. (2003). A public health model of the social determinants of health. *Social and Preventive Medicine, 48*, 242-251.*Journal of Preventive Medicine, 24*(3s), 12-20.

Bacao, F., Lobo, V., & Painho, M. (2005). The self-organizing map, the Geo-SOM, and relevant variants for geosciences. *Computers & Geosciences, 31,* 155-163.

Barnes, R., & Health Development Agency (n.d.). Health Impact Assessment Glossary of Terms. Retrieved November 20, 2007, from http://www.who.int/hia/about/glos/en/index1.html.

Basara, H.G. (2006). *Development of an Ecologically Derived Environmental Health Model Using Geographic Information Systems.* Unpublished Dissertation.

Basara, H.G. & Yuan, M. (2008). Community health assessment using self-organizing maps and geographic information systems. *International Journal of Health Geographics, 7,*67.

Berkman, L.F. (1984). Assessing the physical health effects of social networks and social support. *Annual Review of Public Health, 5,* 413-432.

Berkman, L. F., & Kawachi, I. (Eds.). (2000). *Social epidemiology*. Oxford; New York: Oxford University Press.

Bradley, P.S., & Fayyad, U.M. (1998). *Refining initial points for K-means clustering.* Retrieved on October 30, 2008 from Microsoft Research Website: ftp://ftp.research.microsoft.com/pub/tr/tr-98-36.pdf.

Burgard, S., Stewart, J., & Schwartz, J. (2003). *Occupational status.* Retrieved February 10, 2009, from http://www.macses.ucsf.edu/Research/Social%20Environment/notebook/occupation.html.

Cannon, W. B. (1935). Stresses and strains of homeostasis. *American Journal of Medical Science, 189*, 1-14.

Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstances. *Journal of Health Economics, 24,* 365-389.

Centers for Disease Control and Prevention (n.d.). *Behavioral Risk Factor Surveillance System: Questionnaires.* Retrieved May 31, 2009, from http://apps.nccd.cdc.gov/BRFSSModules/ModByState.asp?Yr=2008.

Centers for Disease Control and Prevention (2004, September 22). *Seasonal Influenza-Associated Hospitalizations in the United States.* Retrieved January 30, 2009, from http://www.cdc.gov/flu/about/qa/hospital.htm.

Centers for Disease Control and Prevention (2008, June 6). Youth Risk Behavior Surveillance—United States, 2007. *Morbidity & Mortality Weekly Report, 57*, SS-4.

Centers for Disease Control and Prevention (2008, September 15). *Preventing Obesity and Chronic Diseases Through Good Nutrition and Physical Activity.* Retrieved May 5, 2009, from http://www.cdc.gov/nccdphp/publications/factsheets/prevention/obesity.htm.

Centers for Disease Control and Prevention (2009, January 28). *Defining Overweight and Obesity.* Retrieved May 5, 2009, from http://www.cdc.gov/nccdphp/dnpa/obesity/defining.htm.

Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. (2009, February 1). Web-based Injury Statistics Query and Reporting System (WISQARS). Accessed on February 1, 2009 from http://www.cdc.gov/ncipc/wisqars/.

Cody, B.E., Mickalide, A.D., Pau,l H.P., Colella, J.M. (2002). Child passengers at risk in America: A national study of restraint use. Washington (DC):

National SAFE KIDS Campaign. Retrieved on January 30, 2009, from http://www.usa.safekids.org/content_documents/ACFD6C.pdf.

Cohen, L. & Swift, S. (1999). The spectrum of prevention: Developing a comprehensive approach to injury prevention. *Injury Prevention*, *5*, 203-207.

Cohn, D. V. (2007). The Growing Global Chronic Disease Epidemic. *Population Reference Bureau.* Retrieved November 19, 2006, from http://www.prb.org/Articles/2007/GrowingGlobalChronicDiseaseEpidemic.aspx.

Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., Brandt, H. M., et al. (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine, 23,* 260-268.

Cox, M., Larkin, E. G., & Frank, S. (n.d.). *Relationship of pro-health safety behaviors with health risk behaviors* [PowerPoint slides]. Retrieved July 16, 2009, from http://www.case.edu/med/adolescenthealth/publications/cox_relationship_of_pro-health.ppt.

Deboeck, G. (1999). Data mining with self-organizing maps: Part II: A practical application. Dokus Publishing. Available online at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.26.4690.

Denollet, J., Sys, S. U., Stroobant, N., Rombouts, H., Gilbert, T. C., & Brutsaert, D. L. (1996). Personality as an independent predictor of long-term mortality in patients with coronary heart disease. *The Lancet, 347,* 417-421.

Dissanayake, S., Ratnayake, I. (2009, January 11-15). *Estimating economic benefits due to increased seatbelt use: A case study.* Paper presented at the 88th Annual Meeting of the Transportation Research Board. Abstract retrieved July 16, 2009, from http://pubsindex.trb.org/document/view/default.asp?lbid=881673.

Dolan, P., (2008). Developing methods that really do value the 'Q' in the QALY. *Health Economics, Policy and Law,* 3, 69-77.

Duncan, C., Jones, K., & Moon, G. (1998). Context, composition and heterogeneity: Using multilevel models in health research. *Social Science & Medicine, 46,* 97-117.

Dunn, J. R. (2002). Housing and inequalities in health: a study of socioeconomic dimensions of housing and self reported health from a survey of Vancouver residents. *Journal of Epidemiology and Community Health, 56,* 671-681.

Eberhardt, M. S., & Pamuk, E. R. (2004). The importance of place of residence: examining health in rural and nonrural areas. *American Journal of PublicHealth, 94*, 1682-1686.

Eisgruber, L.M. & Schuman, L.S. (1963). The usefulness of aggregated data in the analysis of farm income variability and resource allocation. *Journal of Farm Economics,45*, 587-591.

Eng, P. M., Rimm, E. B., Fitsmaurice, G., & Kawachi, I. (2002). Social ties and changes in social ties in relation to subsequent total and cause-specific mortality and coronary heart disease incidence in men. *American Journal of Epidemiology, 155,* 700-709.

Erb, R. J. (1993). Introduction to backpropagation neural network computation. *Pharmaceutical Research, 10*, 165*.*

Fertig, A.R. & Reingold, D.A. (2007). Public housing, health, and health behaviors: Is there a connection? *Journal of Policy Analysis and Management, 26,* 831-859.

Finch, B. K. (2003). Early origins of the gradient: The relationship between socioeconomic status and infant mortality in the United States. *Demography, 40*, 675-699.

"Finding long-term solutions to the world food crisis." (2008, 26 April-2 May) February 12) [Editorial]. *The Lancet.* Retrieved April 27, 2008, from http://www.sciencedirect.com/science/article/B6T1B-4SC0YX5-1/2/fc03d9853f12c8dfc1cf7a8a09221a80

Finkelstein, E.A., Fiebelkorn, I.C., Wang, G. (2003). National medical spending attributable to overweight and obesity: How much, and who's paying? *Health Affairs*, *W3,*219–226.

Fone, D., Dunstan, F., Williams, G., Lloyd, K., & Palmer, S. (2007). Places, people and mental health: A multilevel analysis of economic inactivity. *Social Science & Medicine, 64*, 633-645.

Food and Agriculture Organization of the United Nations (1999). The state of food insecurity in the world 1999. Rome, Italy: Author.

Food and Agriculture Organization of the United Nations (2006). The state of food insecurity in the world 2006. Rome, Italy: Author.

Food Research & Action Center. (2008, November 24). *Hunger in the U.S.* Retrieved on February 2, 2009, from http://www.frac.org/html/hunger_in_the_us/hunger_index.html.

Forster, T. (2009). *Eating and conserving biodiversity at the same time: Re-linking urban and rural sectors to achieve sustainable food security and resilient communities* [PowerPoint slides]. Retrieved July 1, 2009, from http://www.iclei.org/fileadmin/template/project_templates/localactionbiodiversity/user_upload/Images/Urban_Nature_2009/Talks/Thomas_Forster.pdf.

Gamaliel Foundation (2006). *Gamaliel Foundation strategic plan 2000 through 2010.* Retrieved July 13, 2009, from http://www.gamaliel.org/Foundation/goals.htm.

Gantz, T. & Henkle, G. (2002, October). *Seatbelts: Current issues.* Retrieved January 30, 2009, from http://www.preventioninstitute.org/traffic_seatbelt.html.

Garson, G. D. (2008). “Correlation”, from *Statnotes: Topics in Multivariate Analysis.* Retrieved June 20, 2009 from http://faculty.chass.ncsu.edu/garson/PA765/correl.htm#ordinal.

Gay, C.L., Armstrong, D., Cohen, D., Lai, S., Hardy, M.D., Swales, T.P., et al. (1995). The effects of HIC on cognitive and motor development in children born to HIV-seropositive women with no reported drug use: Birth to 24 months. *Pediatrics, 96,* 1078-1082.

Gehlert, S., Sohmer, D., Sacks, T., Mininger, C., McClintock, M., & Olopade, O. (2008). Targeting health disparities: A model linking upstream determinants to downstream interventions. *Health Affairs, 27*, 339-349.

Gochman, D.S., Ed. (1997). *Handbook of health behavior research*; New York, Plenum.

Goldberg, I.J., Mosca, L., Paina, M.R., Fisher, E.A., (2001). Wine and your heart: A science advisory for healthcare professionals from the Nutrition Committee, Council on Epidemiology and Prevention, and Council on Cardiovascular Nursing of the American Heart Association. *Stroke, 31,* 591-594.

Gordis, L. (1996). *Epidemiology.* Philadelphia, PA: W.B. Saunders Company.

Graunt, J. (1662). *Natural and political observations mentioned in a following index, and made upon the bills of mortality*. Retrieved February 22, 2008, from Western Washington University Website: http://www.ac.wwu.edu/~stephan/Graunt/10.html.

Grogger, J. (1997). Local violence and educational attainment. *The Journal of Human Resources, 32,* 659-682.

Grossarth-Maticek, R., Bastiaans, J., & Kanazir, D. T. (1985). Psychosocial factors as strong predictors of mortality from cancer, ischaemic heart disease and stroke: The Yugoslav prospective study. *Journal of Psychosomatic Research, 29,* 167-176.

Hack, M., Flannery, D.J., Schluchter, M., Carter, L., Borawski, E., & Klein, N. (2002). Outcomes in young adulthood for very-low-birth-weight infants. *The New England Journal of Medicine, 346*, 149-157. Retrieved February 10, 2009, from Research Library database. (Document ID: 100030154).

Haddad B., Olivier, J., De Gruchy, S. (2008). *The potential and perils of partnership: Christian religious entities and collaborative stakeholders responding to HIV and AIDS in Kenya, Malawi and the DRC.* Study commissioned by Tearfund and UNAIDS. Interim report. ARHAP. Retrieved February 17, 2009, from http://www.arhap.uct.ac.za/downloads/TFUNAIDS_full_June2008.pdf.

Hagoel, L., Ore, L., Neter, E., Silman, Z., & Rennert, G. (2002). Clustering women's health behaviors. *Health Education & Behavior, 29*, 170-182.

Harris, M.I. (2001). Racial and ethnic differences in health care access and health outcomes for adults with Type 2 diabetes. *Diabetes Care, 24,* 454-459.

Hartman, J.M., Forsen, J.W., Wallace, M.S., & Neely, G. (2002). Tutorials in clinical research: Part IV: Recognizing and controlling bias. *The Laryngoscope, 112,* 23-31.

Hill, J. O., & Peters, J. C. (1998). Environmental contributions to the obesity epidemic. *Science, 280*(5368), 1371-1374.

Hillemeier, M., Lynch, J., Harper, S., & Casper, M. (n.d.). *Data set directory of social determinants of health at the local level.* Retrieved December 4, 2008 from the Centers for Disease Control Website: http://www.cdc.gov/dhdsp/library/data_set_directory/pdfs/data_set_directory.pdf.

Hiott, A. E., Grzywacz, J. G., Davis, S. W., Quandt, S. A., & Arcury, T. A. (2008). Migrant farmworker stress: Mental health implications. *The Journal Of Rural Health, 24*, 32-39.

Holth, H.S., Werpen, H.K., Zwart, J.A., & Hagen, K. (2008). Physical inactivity is associated with chronic musculoskeletal complaints 11 years later: Results from the Nord-Trøndelag Health Study. *BMC Musculoskeletal Disorders, 9,* 159.

Horwood, L. J., Mogridge, N., & Darlow, B. A. (1998). Cognitive, educational, and behavioural outcomes at 7 to 8 years in a national very low birthweight cohort. *Archives of Diseases in Childhood- Fetal and Neonatal Edition, 79,* F12-20.

Huisman, M. & Oldehinkel, A.J. (2008). Income inequality, social capital and self-inflicted injury and violence-related mortality. *Journal of Epidemiology and Community Health, 63,* 31-37.

Iwasaki, M., Otani, T., Sunaga, R., Miyazaki, H., Xiao, L. Wang, N., et al. (2002). Social networks and mortality based on the Komo-Iso cohort study in Japan. *International Journal of Epidemiology, 31,* 1208-1218.

Jetter, K. M., & Cassady, D. L. (2006). The availability and cost of healthier food alternatives. *American Journal of Preventive Medicine, 30*, 38-44.

Johnson, G. D. (2004). Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modeling. *International Journal of Health Geographics, 3,* 29.

Kemeny, J. (1978). Forms of tenure and social structure: A comparison of owning and renting in Australia and Sweden. *British Journal of Sociology, 29,* 41-56.

Kilmer, G., Roberts, H., Hughes, E., Li, Y., Valluru, B. Fan, A., et al. (2008). Surveillance of certain health behaviors and conditions among states and selected local areas – Behavioral risk factor surveillance system (BRFSS), United States, 2006. *Morbidity & Mortality Weekly Report*, *57* (SS07), 1-188.

Kim, J.O., & Mueller, C.W., (1978). *Introduction to factor analysis: What is it and how to do it.* Beverly Hills, CA: Sage Publications.

Kipke, M. D., Iverson, E., Moore, D., Booker, C., Ruelas, V., Peters, A. L., et al. (2007). Food and park environments: Neighborhood-level risks for childhood obesity in East Los Angeles. *Journal of Adolescent Health, 40*, 325-333.

Kirt, T. & Vainik, E. (2007). Comparison of the methods of self-organizing maps and multidimensional scaling in analysis of Estonian emotion concepts. *NODALIDA Conference Proceedings.* Retrieved May 4, 2009 from http://dspace.utlib.ee/dspace/bitstream/10062/2561/1/reg-Kirt-19.pdf.

Kohonen, T. (2001). *Self-Organizing Maps.* 3rd edition, New York, NY: Springer.

Kopp, M.S., Skrabski, A., Kawachi, I., & Adler, N.E. (2005). Low socioeconomic status of the opposite sex is a risk factor for middle aged mortality. *Journal of Epidemiology and Community Health, 59,* 675-678.

Kreiger, N., Williams, D.R., & Moss, N.E. (1997). Measuring social class in US public health research: Concepts, methodologies and guidelines. *Annual Review of Public Health, 18,* 341-378.

Kwong, J.C., Stukel, T.A., Lim, J., McGeer, A.J., Upshur, R.E., et al. (2008, October). The effect of universal influenza immunization on mortality and health care use. *Public Library of Science Medicine, 5*, 1440- 1452. doi:10.1371/journal.pmed.0050211

Langhout, R.D., Rosselli, F., & Feinstein, J. (2007). Assessing classism in academic settings. *The Review of Higher Education, 30,* 145-184.

Langley, K., Rice, F., van den Bree, M.B., & Thapar, A. (2005). Maternal smoking during pregnancy as an environmental risk factor for attention deficit hyperactivity disorder behaviour: A review [Abstract]. *Minerva Pediatrica, 57*, 359-361.

Lifestyle, (2009). In *Merriam-Webster Online Dictionary*. Retrieved May 5, 2009, from http://www.merriam-webster.com/dictionary/lifestyle.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior, 35*(Extra Issue), 80-94.

Macintyre, S. & Ellaway, A. (2000). Ecological Approaches: The rediscovery of the role of the physical and social environment. In L. F. Berkman & I.

Kawachi (Eds.), *Social Epidemiology* (pp. 332-348). Oxford: Oxford University Press.

Marmot, M. (2007). Achieving health equity: From root causes to fair outcomes. *The Lancet, 370*, 1153-1163.

Marmot, M. G., Baum, F., Begin, M., Berlinguer, G., Chatterjee, M., Foege, W. H., et al. (2007). *Achieving health equity: From root causes to fair outcomes*. Retrieved. from World Health Organization Website: http://whqlibdoc.who.int/publications/2007/interim_statement_eng.pdf.

Marmot, M.G., Kogevinas, M., & Elston, M.A. (1987). Social/economic status and disease. *Annual Review of Public Health, 8,* 111-135.

Marmot, M. G., & Wilkinson, R. G. (1999). *Social determinants of health*. New York, NY: Oxford University Press.

Martikainen, P., Bartley, M., & Lahelma, E. (2002). Psychosocial determinants of health in social epidemiology. *International Journal of Epidemiology, 31,* 1091-1093.

Martinez, P. & Richters, J. E. (1993). The NIMH community violence project II: Children's distress symptoms associated with violence exposure. In D. Reiss, Reichters, J. E., Radke-Yarrow, M. & Scharff, D. (Eds.), *Children and violence* (pp. 22-35). New York, NY: Guilford Press.

Matthews, J. (2004). Self-Organizing Nets. Retrieved October 17, 2008 from http://www.generation5.org/content/1999/selforganize.asp.

Mausner, J. S. & Kramer, S. (1985). *Epidemiology: An introductory text* (2nd ed.). Philadelphia, PA: W.B. Saunders Company.

Maycock, B.R., & Howat, P. (2007). Social capital: Implications from an investigation of illegal anabolic steroid networks. *Health Education Research, 22*, 854-863.

Mayfield, E. (1999). *New hope for people with sickle cell anemia.* Retrieved October 30, 2008 from U.S. Food and Drug Administration Website: http://www.fda.gov/fdac/features/496_sick.html

McDermott, M. (Ed.). (2006). *Closer to home: Healthier food, farms and families in Oklahoma.* Poteau, OK: Kerr Center for Sustainable Agriculture*.*

Mittleman, M. A., Maclure, M., Sherwood, J. B., Mulry, R. P., Tofler, G. H., Jacobs, S. C., et al. (1995). Triggering of acute myocardial infarction onset by episodes of anger. *Circulation, 92*, 1720-1725.

Mokdad AH, Marks JS, Stroup DF, Gerberding JL., (2004). Actual causes of death in the United States, 2000. *JAMA, 291*, 1238-1245.

Mokdad AH, Marks JS, Stroup DF, Gerberding JL., (2005). Correction: Actual causes of death in the United States, 2000. *JAMA, 293*, 293-294.

Molinier, M.m Laaksonen, J. & Hame, T. (2007). Detecting man-made structures and changes in satellite imagery with a content-based information retrieval system built on self-organizing maps. *IEEE Transactions on Geoscience and Remote Sensing, 45,* 861-874.

Mueller, C. W., & Parcel, T.L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development, 52*, 13-30.

Mueller, C.W., & Tighe, J.R. (2007). Making the case for affordable housing: Connecting housing with health and education outcomes. *Journal of Planning Literature, 27*, 371-385.

Muhajarine, N., Labonte, R., Williams, A., & Randall, J. (2008). Person, perception, and place: What matters to health and quality of life. *Social Indicators Research, 85*, 53-80.

Murberg, T. A., & Bru, E. (2001). Social relationships and mortality in patients with congestive heart failure. *Journal of Psychosomatic Research, 51,* 521-527.

Mushi-Brunt, C., Haire-Joshu, D., Elliott, M., & Brownson, R. (2007). Fruit and vegetable intake and obesity in preadolescent children: The role of neighborhood poverty and grocery store access. *American Journal of Health Education, 38*, 258-265.

National Institute for Mental Health (2008, September 8). *Statistics.* Retrieved on January 31, 2009, from http://www.nimh.nih.gov/health/statistics/.

National SAFE KIDS Campaign (NSKC) (2004). *Unintentional firearm injury fact sheet*. Washington (DC): NSKC. Retrieved January 30, 2009, from http://www.usa.safekids.org/tier3_cd.cfm?folder_id=540&content_item_id=1131.

National Violence Prevention Network (2007). History of NVDRS. Retrieved February 10, 2009, from http://www.preventviolence.net/winvdrs/history.html.

Oklahoma State Department of Health (2007). *State of the state's health report, 2007*. Retrieved on January 30, 2009, from http://www.ok.gov/health/pub/boh/state/.

Oklahoma State Department of Health (2009). *Oklahoma turning point initiative.* Accessed July 14, 2009, from http://www.ok.gov/health/Community_Health/Community_Development_Service/Turning_Point/index.html.

Oklahoma State Department of Health (n.d.a). Behavioral Risk Factor Surveillance System. Accessed July 1, 2009, from http://www.ok.gov/health/pub/wrapper/ok2share.html.

Oklahoma State Department of Health (n.d.b). Vital Statitics Birth & Death Data. Retrieved March 23, 2008, from www.health.ok.gov/OK2SHARE.

Oreopoulos, P., Stabile, W., Walld, R., & Roos, L. L. (2008). Short-, medium-, and long-term consequences of poor infant health - An analysis using siblings and twins. *Journal of Human Resources, 43,* 88-138.

Orth-Gomer, K., Wamala, S. P., Horsten, M., Schenck-Gustafsson, K., Schneiderman, N., & Mittleman, M. A. (2000). Marital stress worsens prognosis in women with coronary heart disease: The Stockholm female coronary risk study. *JAMA, 284*, 3008-3014.

Osofsky, J. D. (1999). The impact of violence on children. *The Future of Children, 9*, 33-49.

Oyana, T.J., Boppidi, D., Yan, J., & Lwebuga-Mukasa (2008). Exploration of geographic information systems (GIS)-based medical databases with self-organizing maps (SOM): A case study of adult asthma. In B. L. Friis-Christensen A, H. Pundt (Eds.), *The European Information Society.* Berlin, Germany: Springer.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction.* Orlando, Fl: Holt, Rinehart & Winston, Inc.

Probst, J. C., Moore, C. G., Glover, S. H., & Samuels, M. E. (2004). Person and Place: The Compounding Effects of Race/Ethnicity and Rurality on Health. *American journal of Public Health, 94*(10), 1695-1703.

Psychosocial, (2009). In *Merriam-Webster Online Dictionary*. Retrieved May 5, 2009, from http://www.merriam-webster.com/dictionary/psychosocial.

Raphael, D. (2003). Bridging the gap between knowledge and action on the societal determinants of cardiovascular disease: How one Canadian community effort hit - and hurdled - the lifestyle wall. . *Health Education*, *103*, 13, from www.emeraldinsight.com/0965-4283.htm

Raphael, D. (2006). Social determinants of health: Present status, unanswered questions, and future directions. *International Journal of Health Services, 36*(4), 651-677.

Raswant, V., Hart, N, & Romano, M. (2008). *Biofuel expansion: Challenges, risks and opportunities for rural people.* Unpublished manuscript.

Regidor, E., Ronda, E., Martinez, D., Calle, M. E., Navarro, P., & Dominguez, V. (2005). Occupational social class and mortality in a population of men economically active: The contribution of education and employment situation. *European Journal of Epidemiology, 20,* 501-508.

Ridolfo, B. & Stevenson, C. (2001). *The quantification of drug-caused mortality and morbidity in Australia, 1998.* Canberra: Australian Institute of Health and Welfare, Commonwealth Department of Human Services and Health; 2001. Category No. PHE 29.

Rutledge, T., Matthews, K., Lui, L. Y., Stone, K. L., Cauley, L. A. (2003). Social networks and marital status predict mortality in older women: Prospective evidence from the study of osteoporotic fractures (SOF). *Psychosomatic Medicine, 65,* 688-694.

Sallis, J. F., & Glanz, K. (2006). The role of built environments in physical activity, eating, and obesity in childhood. *Future of Children, 16*, 89-108.

SAS Institute, Inc. (2002). SAS Help and Documentation, 2002-*2004.* Cary, N.C.

SAS Institute, Inc. (2005). *What's new in SAS Enterprise Miner 5.2*. Cary, N.C.

Sawchuk, C. N., Roy-Byrne, P., Goldberg, J., Manson, S., Noonan, C., Beals, J., et al. (2005). The relationship between post-traumatic stress disorder, depression and cardiovascular disease in an American Indian tribe. *Psychological Medicine, 35*, 1785-1794.

Schlundt, D. G., Larson, C., Ahmed, N. U., Keith, H., McClellan, L., Marrs, M. (2003, November 18). *Mapping healthy and unhealthy neighborhoods using cluster analysis and GIS: Analysis of the Nashville REACH 2010 community survey.* Paper presented at the 131st Annual Meeting of APHA. Abstract retrieved July 15, 2009, from http://apha.confex.com/apha/131am/techprogram/paper_70428.htm.

Schneider, S., Huy, C., Schussler, M., Diehl, K., & Schwarz, S. (2009). Optimising lifestyle interventions: identification of health behaviour patterns by cluster analysis in a German 50+ survey. *European Journal of Public Health*, 19, 271-277.

Schwab, A. J. (2006). *Hierarchical multiple regression* [PowerPoint slides]. Retrieved June 21, 2009 from http://www.utexas.edu/courses/schwab/sw388r7_spring_2006/SolvingProblems/MultipleRegression_CompleteHierarchicalProblems_spring2006.ppt

Shavers, V. L. (2007) Measurement of socioeconomic status in health disparitites research. *Journal of the National Medical Association, 99,* 1013-1023.

Shults, R.A. (2004). Child passenger deaths involving drinking drivers—United States, 1997−2002 [published erratum appears in Morbidity & Mortality Weekly Report, 53,109]. *Morbidity & Mortality Weekly Report, 53*, 77–9.

SOM toolbox for Matlab (2001). Retrieved August 4, 2009 from http://www.cis.hut.fi/projects/somtoolbox/links/somsoftware.shtml.

Stewart, S.L., Cardinez, C.J., Richardson, L.C., Norman, L., Kaufmann, R.K., et al. (2008). Surveillance for cancers associated with tobacco use: United States, 1999-2004. *Morbidity & Mortality Weekly Report, 57*, SS08, 1-33.

Strong and Healthy Oklahoma. (n.d.). *Make it your business for a strong and healthy oklahoma* [PowerPoint slides]. Retrieved July 14, 2009, from http://www.ok.gov/strongandhealthy/documents/Make%20it%20Your%20Business%2010302008.ppt.

Thisted, R. A., & Hiller, J. E. (2003). Are there social determinants of health? *Perspectives in Biology and Medicine, 46* (3 supplement), S65-S73.

Thompson, W., & Hickey, J. (2005). *Society in focus.* Boston, MA: Pearson.

United Health Foundation (2008). *America's health rankings: A call to action for individuals * their communities, 2008 edition.* Minnetonka, MN: United Health Foundation. Retrieved January 31, 2009 from http://www.americashealthrankings.org/2008/pdfs/2008.pdf.

United States Census Bureau (2002). *Questions and answers for Census 2000 data on race.* Retrieved May 6, 2009, from http://www.census.gov/Press-Release/www/2001/raceqandas.html.

United States Census Bureau (2008). *Metropolitan and micropolitan statistical areas.* Retrieved May 28, 2009, from http://www.census.gov/population/www/metroareas/aboutmetro.html.

United States Department of Agriculture (2004). *2002 Census of agriculture.* Retrieved July 14, 2009, from http://www.agcensus.usda.gov/Publications/2002/Volume_1,_Chapter_1_US/USVolume104.pdf.

United States Department of Justice (2000). *Juvenile justice bulletin: The High/Scope Perry preschool project.* Retrieved May 5, 2009, from http://www.ncjrs.gov/pdffiles1/ojjdp/181725.pdf.

Verdu, S.V., Garcia, M. O., Senabre, C., Marin, A.G., & Franco, F.J.G. (2006). Classification, filtering, and identificatin of electrical customer load patterns through the use of self-organizing maps. I*EEE Transactions on Power Systems, 21*, 1672-1682.

Vesanto, J. & Alhoniemi (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks, 11,* 586-600.

Vest, J. R., Catlin, T. K., Chen, J.J., & Brownson, R. C. (2002), Multistate analysis of factors associated with intimate partner violence. *American Journal of Preventive Medicine, 22,* 156-163.

Viscovery (n.d.). *Viscovery SOMine 5.0.* Retrieved August 4, 2009 from http://www.viscovery.net/somine/.

Walker, R. B., & Hiller, J. E. (2007). Places and health: A qualitative study to explore how older women living alone perceive the social and physical dimensions of their neighbourhoods. *Social Science & Medicine, 65*, 1154-1165.

Wang, J., Delabie, J., Aasheim, H.C., Smeland, E. & Myklebost, O. (2002). Clustering of the SOM easily reveals distinct gene expression patterns: Results of a reanalysis of lymphoma study. *BMC Bioinformatics, 3*,

Washington State Department of Social & Health Services (2006, September). *Faith-based organizations and chemical dependency recovery support services legislative report.* Retrieved August 5, 2009 from http://www.dshs.wa.gov/pdf/ea/GovRel/Leg0307/FaithBasedAll0307.pdf.

Watt, G. (2002). The inverse care law today. *The Lancet, 360*, 9328, 252-254.

Wilkinson, R.G., & Marmot, M.G. (Eds.) (2003). *Social determinants of health: The solid facts*. Copenhagen, Denmark: World Health Organization. Retrieved from http://www.euro.who.int/document/e81384.pdf.

Wolff, E.N. (2007). *Recent trends in household wealth in the United States: Rising debt and the middle-class squeeze.* Unpublished manuscript, New York University.

World Health Organization (2008). *The world health report 2008: Primary health care – Now more than ever.* Retrieved May 4, 2009 from http://www.who.int/whr/2008/whr08_en.pdf.

World Health Organization. (2005). Climate and Health Fact Sheet. Retrieved November 20, 2007, 2007, from http://www.who.int/globalchange/news/fsclimandhealth/en/index.html.

Wu, Y., Liu, Q., & Huang, T.S. (2000, Jan). An Adaptive Self-Organizing Color Segmentation Algorithm with Application to Robust Real-time Human Hand Localization. *In Proc. IEEE Asian Conf. on Computer Vision,* pp. 1106-1111, Taiwan. Retrieved July 1, 2009, from http://www.ece.northwestern.edu/~yingwu/.

Yusuf, S., Reddy, S., Ôunpuu, S., & Anand, S. (2001). Global burden of cardiovascular diseases: Part I: General considerations, the epidemiologic transition, risk factors, and impact of urbanization**.** *Circulation, 104,* 2746-2753.

APPENDICES

Appendix A

Social Determinants of Health Variables: Determinant Dimension and Category,

Variable Label, Description & Source

Determinant Category	Variable Label	Description	Source
SOCIO-ECONOMIC DETERMINANTS			
Education			
Low Birth Weight	VLBW	A live birth weighing less than 1,500 grams (3 lb. 5 oz. or less).	Oklahoma Vital Records
	LBW	A live birth weighing less than 2,500 grams (5 lb. 8 oz. or less).	Oklahoma Vital Records
Education	Educ_Nodegree	Percent population: no high school diploma or GED	U.S. Census Bureau
	Educ_HS	Percent population: high school diploma or GED only	U.S. Census Bureau
	Educ_AD	Percent population: Associates degree	U.S. Census Bureau
	Educ_BD	Percent population: Bachelors degree	U.S. Census Bureau
	Educ_MD	Percent population: Masters degree	U.S. Census Bureau
	Educ_PD	Percent population: Professional school degree	U.S. Census Bureau
	Educ_DD	Percent population: Doctoral degree	U.S. Census Bureau
Socio-Economic Status			
Occupation	Occup_Mangt	Percent population: Management, professional, and related occupations	U.S. Census Bureau
	Occup_Service	Percent population: Service occupations	U.S. Census Bureau
	Occup_Sales	Percent population: Sales and office occupations	U.S. Census Bureau
	Occup_Farm	Percent population: Farming, fishing, and	U.S. Census Bureau

		forestry occupations	
	Occup_Const	Percent population: Construction, extraction, and maintenance occupations	U.S. Census Bureau
	Occup_Prod	Percent population: Production, transportation, and material moving occupations	U.S. Census Bureau
Income	Inc_10	Household Income: Less than $10,000	U.S. Census Bureau
	Inc_15	Household Income: $10,000 to $14,999	U.S. Census Bureau
	Inc_25	Household Income: $15,000 to $24,999	U.S. Census Bureau
	Inc_35	Household Income: $25,000 to $34,999	U.S. Census Bureau
	Inc_50	Household Income: $35,000 to $49,999	U.S. Census Bureau
	Inc_74	Household Income: $50,000 to $74,999	U.S. Census Bureau
	Inc_100	Household Income: $75,000 to $99,999	U.S. Census Bureau
	Inc_150	Household Income: $100,000 to $149,999	U.S. Census Bureau
	Inc_200	Household Income: $150,000 to $199,999	U.S. Census Bureau
	Inc_200more	Household Income: $200,000 or more	U.S. Census Bureau
	Inc_medhh	Median household income (dollars)	U.S. Census Bureau
	Income_percap	Per capita Income	U.S. Census Bureau
(Un)employment	Unemp	Average Annual Unemployment Rate	U.S. Bureau of Labor Statistics
Housing	Hous_occ	Occupied housing units	U.S. Census Bureau
	Hous_vac	Vacant housing units	U.S. Census Bureau
	Hous_age	Median Age of housing units	U.S. Census Bureau
	Hous_50	Owner-occupied housing value: Less than $50,000	U.S. Census Bureau

	Hous_100	Owner-occupied housing value: $50,000 to $99,999	U.S. Census Bureau
	Hous_150	Owner-occupied housing value: $100,000 to $149,999	U.S. Census Bureau
	Hous_200	Owner-occupied housing value: $150,000 to $199,999	U.S. Census Bureau
	Hous_300	Owner-occupied housing value: $200,000 to $299,999	U.S. Census Bureau
	Hous_500	Owner-occupied housing value: $300,000 to $499,999	U.S. Census Bureau
	Hous_mill	Owner-occupied housing value: $500,000 to $999,999	U.S. Census Bureau
	Hous_overmill	Owner-occupied housing value: $1,000,000 or more	U.S. Census Bureau
	Hous_medown	Median (dollars) value of owner-occupied homes	U.S. Census Bureau
	Hous_tenown	Tenure: Owner-occupied	U.S. Census Bureau
	Hous_tenrent	Tenure: Renter-occupied	U.S. Census Bureau
	Hous_utgas	Utility gas	U.S. Census Bureau
	Hous_bgas	Bottled, tank, or LP gas	U.S. Census Bureau
	Hous_elec	Electricity	U.S. Census Bureau
	Hous_oil	Fuel oil, kerosene, etc.	U.S. Census Bureau
	Hous_coal	Coal or coke	U.S. Census Bureau
	Hous_wood	Wood	U.S. Census Bureau
	Hous_solar	Solar energy	U.S. Census Bureau
	Hous_otherfuel	Other fuel	U.S. Census Bureau
	Hous_nofuel	No fuel used	U.S. Census Bureau
	Hous_noplum	Lacking complete plumbing facilities	U.S. Census Bureau
	Hous_nokitch	Lacking complete kitchen facilities	U.S. Census Bureau
	Hous_nophone	No telephone service available	U.S. Census Bureau

	Hous_1occup	0 to 1 occupants per sleeping quarters	U.S. Census Bureau
	Hous_1_5occup	1.01 to 1.50 occupants per sleeping quarter	U.S. Census Bureau
	Hous_1_5moccup	1.51 or more occupants per sleeping quarter	U.S. Census Bureau
	Hous_M2000	Moved in 1999 to 2000	U.S. Census Bureau
	Hous_M1995	Moved in 1995 to 1998	U.S. Census Bureau
	Hous_M1990	Moved in 1990 to 1994	U.S. Census Bureau
	Hous_M1980	Moved in 1980 to 1989	U.S. Census Bureau
	Hous_M1970	Moved in 1970 to 1979	U.S. Census Bureau
	Hous_M1969	Moved in 1969 or earlier	U.S. Census Bureau
	Hous_B2000	Built 1999 or 2000	U.S. Census Bureau
	Hous_B1995	Built 1995 to 1998	U.S. Census Bureau
	Hous_B1990	Built 1990 to 1994	U.S. Census Bureau
	Hous_B1980	Built 1980 to 1989	U.S. Census Bureau
	Hous_B1970	Built 1970 to 1979	U.S. Census Bureau
	Hous_B1960	Built 1960 to 1969	U.S. Census Bureau
	Hous_B1950	Built 1950 to 1959	U.S. Census Bureau
	Hous_B1940	Built 1940 to 1949	U.S. Census Bureau
	Hous_B1939	Built 1939 or earlier	U.S. Census Bureau
	Hous_rent99	Renter-occupied housing units: With cash rent: under $100	U.S. Census Bureau
	Hous_rent100	Renter-occupied housing units: With cash rent:$100 to $299	U.S. Census Bureau
	Hous_rent300	Renter-occupied housing units: With cash rent:$300 to $499	U.S. Census Bureau
	Hous_rent500	Renter-occupied housing units: With cash rent:$500 to $699	U.S. Census Bureau
	Hous_rent700	Renter-occupied housing units: With cash rent:$700 to $999	U.S. Census Bureau
	Hous_rent1000	Renter-occupied - housing units: With cash rent:$1000 to $1249	U.S. Census Bureau
	Hous_rent1250	Renter-occupied housing units: With	U.S. Census Bureau

		cash rent:$1250 to $1499	
	Hous_rent1500	Renter-occupied housing units: With cash rent:$1500 to $1999	U.S. Census Bureau
	Hous_rent2000	Renter-occupied housing units: With cash rent:$2000 and over	U.S. Census Bureau
	Hous_rentnocash	Renter-occupied housing units: With no cash rent	U.S. Census Bureau
Demographics			
Age	Age_Med	Median age in years for the population of the county	U.S. Census Bureau
Gender	Gender_F	Percent population: female	U.S. Census Bureau
Race	Race_W	Percent population: as white only	U.S. Census Bureau
	Race_B	Percent population: as black only	U.S. Census Bureau
	Race_I	Percent population: as American Indian or Alaskan Native only	U.S. Census Bureau
	Race_A	Percent population: as Asian only	U.S. Census Bureau
	Race_PI	Percent population: as Native Hawaiian and other Pacific Islander only	U.S. Census Bureau
	Race_O	Percent population: as some other race only	U.S. Census Bureau
	Race_2races	Percent population: as 2 or more races	U.S. Census Bureau
Ethnicity	Hispanic	Percent population: as being of Hispanic origin	U.S. Census Bureau
PSYCHOSOCIAL RISK FACTORS			
Low self-esteem	Psyc_MH	Median number of days with poor mental health	Behavioral Risk Factor Surveillance System
Depression	Psyc_Depression	Number of persons	Oklahoma Department

		treated for depressive disorders	of Mental Health and Substance Abuse Services
Anxiety	Psyc_anxiety	Number of persons treated for anxiety disorders	Oklahoma Department of Mental Health and Substance Abuse Services
Isolation	Psyc_Ling	Percent Population: Linguistically Isolated	U.S. Census Bureau
	Psyc_alone	Number of persons treated for mental health disorders that live alone	Oklahoma Department of Mental Health and Substance Abuse Services
COMMUNITY AND SOCIAL CHARACTERISTICS			
Insecurity	Psyc_fdins_calc	Calculated: 15.2% of population in Oklahoma are food insecure	http://www.kerrcenter.com/publications/closer_to_home/chapter02.pdf
Social and community participation	Com_AvgRevenue	Average revenue for charitable organizations	http://nccsdataweb.urban.org/NCCS/Public/
	Com_AvgAssets	Average amount of assets for charitable organizations	http://nccsdataweb.urban.org/NCCS/Public/
Civic and political involvement and empowerment	Com_Rep	Percent of Registered Voters: Republican	Oklahoma State Election Board
	Com_Dem	Percent of Registered Voters: Democrat	Oklahoma State Election Board
	Com_Ind	Percent of Registered Voters: Independent	Oklahoma State Election Board
Crime rate	Com_Crime	Oklahoma 2004 Crime rate by county	Oklahoma State Bureau of Investigation
Domestic violence	Com_Dvserv	Number of Domestic Violence Services by County	http://www.okdhs.org/okdhslocal/docs/domestic_violence.htm or http://www.aardvarc.org/dv/states/okdv.shtml
	Com_Dvrep	Rate of Domestic Violence Reports by County	http://www.coph.ouhsc.edu/coph/HealthPolicyCenter/Pubs/2005/GTF/justice.pdf
Poverty	Com_Povfam	Percent of Families	U.S. Census Bureau

		below Federal Poverty Level	
	Com_PovInd	Percent of Individuals below Federal Poverty Level	U.S. Census Bureau
Residence (urban, rural, remote)	Com_Urban	Percent Population: Urban	U.S. Census Bureau
	Com_RurFarm	Percent Population: Rural Farm Area	U.S. Census Bureau
	Com_RurNonfarm	Percent Population: Rural Nonfarm Area	U.S. Census Bureau
Income inequality	Com_Gini (Gini)	Gini Coefficient of Income Inequality	Burkey, Mark L. "Gini Coefficients for the 2000 Census", March 2006. www.ncat.edu/~burkeym/Gini.htm
Altruism, Philanthropy and voluntary work	Com_church_type	Number of different types of churches	http://www.glenmary.org/grc/default.htm
	Com_rateofadh	Rate of church adherents or membership for each county	http://www.glenmary.org/grc/default.htm

Appendix B

Health Behavior Variables: Variable Label, Description & Source

Variable Label	Description	Source
PHYSICAL HEALTH		
CHECKUP_1yr	About how long has it been since you last visited a doctor for a routine checkup? Within past year	Behavioral Risk Factor Surveillance System
CHECKUP_2yr	About how long has it been since you last visited a doctor for a routine checkup? Within past 2 years	Behavioral Risk Factor Surveillance System
CHECKUP_5yr	About how long has it been since you last visited a doctor for a routine checkup? Within past 5 years	Behavioral Risk Factor Surveillance System
CHECKUP_m5yr	About how long has it been since you last visited a doctor for a routine checkup? 5 or more years ago	Behavioral Risk Factor Surveillance System
CHECKUP_never	About how long has it been since you last visited a doctor for a routine checkup? Never	Behavioral Risk Factor Surveillance System
LEISPA	During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise? (Had PA)	Behavioral Risk Factor Surveillance System
RECPA	Adults reported in either moderate physical act: 30+ min/day for 5+ days/week, or vigorous act: 20+ min/ day on 3+ days (Not at risk)	Behavioral Risk Factor Surveillance System
FIVEFV	Do you eat the recommended number of fruits and vegetables in a day? (Less than 5 servings)	Behavioral Risk Factor Surveillance System
WEIGHT_Normal	Recomputed variable: Normal = Respondents for whom _BMI < 25.00.	Behavioral Risk Factor Surveillance System
WEIGHT_Overwt	Recomputed variable: Overweight = Respondents for whom 25.00 <= _BMI < 30.00.	Behavioral Risk Factor Surveillance System
WEIGHT_Obese	Recomputed variable: Obese = Respondents for whom 30.00 <= _BMI < 99.102.	Behavioral Risk Factor Surveillance System
LASTDEN3	How long has it been since you last visited a dentist or a dental clinic for any reason? Include visits to dental	Behavioral Risk Factor Surveillance System

	specialists, such as orthodontists.	
DENVST	Percent of adults who visited a dentist in the last year	Behavioral Risk Factor Surveillance System
CURNTSMK	This variable combines results from two questions: Ever smoked at least 100 cigarettes in their lifetime and currently smokes every day or some days.	Behavioral Risk Factor Surveillance System
STOPSMK2	Percent of adult smokers who have stopped smoking for one day or longer in the past 12 months because they were trying to quit smoking.	Behavioral Risk Factor Surveillance System
BINGEDRK	Adults having five or more drinks on one occasion	Behavioral Risk Factor Surveillance System
HEAVYDRK	Adult men having more than two drinks per day and adult women having more than one drink per day	Behavioral Risk Factor Surveillance System
FLUSHOT3	Percent of adults who have had a flu shot in the past 12 months.	Behavioral Risk Factor Surveillance System
PNEUVAC3	Percent of adults who have had a pneumonia shot.	Behavioral Risk Factor Surveillance System
SEATBELT	Percent of adults who always or nearly always wear their seatbelts.	Behavioral Risk Factor Surveillance System
HADMAM	Percent of women who have had a mammogram.	Behavioral Risk Factor Surveillance System
PROFEXAM	Percent of women who have had a clinical breast exam.	Behavioral Risk Factor Surveillance System
HADPAP2	Percent of women who have had a pap smear test.	Behavioral Risk Factor Surveillance System
PSATEST	Percent of men who have had a prostate-specific antigen test.	Behavioral Risk Factor Surveillance System
CHOLEST	Percent of respondents who had their blood cholesterol checked within the past five year	Behavioral Risk Factor Surveillance System

SIGMOID	Percent of persons who have had a sigmoidoscopy.	Behavioral Risk Factor Surveillance System
MENTAL HEALTH		
Psych_MH	Percent of adults who had more than 15 poor mental health days in the last month.	Behavioral Risk Factor Surveillance System
MRESTRICT	Percent of adults who had more than 15 days of restricted activities due to poor physical and mental health.	Behavioral Risk Factor Surveillance System
IPVTHREAT	Has an intimate partner EVER THREATENED you with physical violence?	Behavioral Risk Factor Surveillance System
IPVATTEMPT	Has an intimate partner EVER ATTEMPTED physical violence against you? (Yes)	Behavioral Risk Factor Surveillance System
IPVPHYSICAL	Has an intimate partner EVER hit, slapped, pushed, kicked, or physically hurt you in any way? (Yes)	Behavioral Risk Factor Surveillance System
SEXVIOLEVER	Has anyone EVER had sex with you after you said or showed that you didn't want them to or without your consent? (Yes)	Behavioral Risk Factor Surveillance System

Appendix C

Table of Means and Standard Deviations by SDOH Clusters

	Total		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Attribute	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Age_ Median	37.79	3.09	37.12	2.85	38.1	2.09	34.8	3.46	41.89	1.88
Com_ Char_ avg_ assets	961,061	1,399,517	894,434	924,340	803,092	1,002,250	2,427,378	2,955,530	235,998	346,860
Com_ Char_ avg_ revenue	691,910	819,343	697,282	804,374	634,866	584,492	1,044,209	1,079,748	512,448	1,233,115
Com_ church_ type	19.35	9.41	20.59	6.76	16.67	4.82	32.33	17.18	11.33	2.74
Com_ Crime	27.72	12.84	30.7	11.66	26.21	7.88	38.52	19.64	12.33	6.42
Com_ Dem_ 2009	35.76	10.12	34.76	9.76	41.48	8.23	28.78	4.54	26.85	10.12
Com_ Dvrep	527	302	670	387	473	159.7	520.8	237.5	253.4	160.7
Com_ Dvserv	2.39	0.861	2.241	0.511	2.367	0.615	3	2	2.333	0.5
Com_ fdins_ calc	6812	14946	4489	3615	3841	2575	30274	36589	737	222
Com_ Ind_ 2009	5.538	1.794	5.551	1.513	5.036	1.663	8.308	0.808	4.403	0.921
Com_ Povfam	12.73	3.74	13.03	3.01	14.51	3.77	8.2	2.04	10.4	2.02
Com_ PovInd	16.56	4.33	17.04	3.6	18.43	4.13	11.73	3.87	13.63	2.34
Com_ rateofadh	685	176	761	148	593.5	113.9	538	138	895	155
Com_ Rep_ 2009	20.54	10.21	18.69	8.35	14.99	7.85	33.24	3.92	32.32	6.52
Com_ RurFarm	5.64	3.97	4.411	2.176	5.51	2.14	1.68	1.306	13.99	3.63
Com_ RurNonfarm	57.97	22.72	45.98	16.32	69.34	13.64	30.7	23.66	86.01	3.63
Com_ Urban	36.39	25.43	49.61	17.7	25.15	14.27	67.62	24.86	0	0

Educ_ AD	4.456	1.721	4.464	1.946	4.418	1.456	5.867	1.393	3.142	1.038
Educ_ BD	10.37	3.12	10.91	2.2	7.78	1.42	15.38	2.95	12.22	1.18
Educ_ DD	0.443	0.598	0.419	0.346	0.2824	0.1957	1.204	1.428	0.2996	0.1646
Educ_ HS	35.68	3.99	34.89	3.65	37.27	2.95	30.6	4.21	38.01	2.97
Educ_ MD	3.667	1.328	3.923	1.301	3.082	0.732	5.162	2.011	3.295	0.812
Educ_ Nodegr ee	23.22	5.47	23.06	4.52	26.8	4.36	15.67	2.93	19.35	2.57
Educ_ PD	0.962	0.399	0.945	0.355	0.795	0.215	1.499	0.558	1.037	0.393
Gender _F	50.43	1.64	50.34	1.84	50.62	1.27	50.65	0.91	49.85	2.55
Gini	0.4484	0.026	0.4506	0.0236	0.4547	0.0248	0.4325	0.035	0.4366	0.0208
Hispani c	4.71	4.86	6.75	6.64	2.727	1.934	4.058	2.201	5.39	4.64
Hous_ 1_ 5moccu p	0.947	0.451	0.941	0.553	0.967	0.338	1.042	0.454	0.805	0.458
Hous_ 1_ 5occup	2.364	0.813	2.325	0.882	2.76	0.564	2.089	0.471	1.451	0.752
Hous_ 100	35.68	6.27	35.97	6.93	34.23	4.58	43.05	4.82	32.17	4.79
Hous_ 150	10.31	4.63	9.55	2.99	9.38	3.21	19.28	4.5	6.89	2.37
Hous_ 1occup	96.69	1.14	96.73	1.37	96.27	0.74	96.87	0.85	97.74	1.07
Hous_ 200	3.772	1.722	3.469	1.036	3.438	1.494	7.045	1.105	2.591	0.853
Hous_ 300	2.122	1.069	1.775	0.735	2.05	0.868	4.118	0.568	1.483	0.717
Hous_ 50	46.86	12.67	48.35	11.07	49.52	8.71	24.31	7.77	55.79	7.93
Hous_ 500	0.774	0.544	0.618	0.315	0.783	0.556	1.523	0.557	0.5	0.473
Hous_ age	31.68	7.76	34.41	6.75	27.47	4.3	25.44	4.93	43.11	6.05
Hous_ B1939	15.06	7.79	16.11	6.63	12.45	5.47	7.87	3.11	27.53	6.47
Hous_ B1940	9.16	3.65	10.65	3.1	7.33	2.42	5.71	2.61	13.94	2.23
Hous_ B1950	12.18	3.84	14.76	3.87	10.09	2.01	10.7	5.2	12.29	1.76
Hous_ B1960	13.75	2.47	14.87	2.81	12.83	1.62	13.34	2.78	13.63	2.3
Hous_ B1970	21.53	4.29	20.11	4.12	23.82	2.13	24.13	3.22	15.84	4.23
Hous_	16.41	4.97	14.51	4.64	18.54	3.53	20.9	4.37	10.95	2.9

B1980										
Hous_ B1990	4.39	2.2	3.408	1.972	5.57	1.96	5.694	1.635	2.328	0.589
Hous_ B1995	5.58	2.79	4.205	2.236	6.95	1.89	8.36	3.17	2.663	1.07
Hous_ B2000	1.947	1.138	1.382	0.848	2.421	0.796	3.307	1.328	0.825	0.478
Hous_ bgas	20.1	9.95	14.17	5.79	27.42	8.14	9.68	8.03	25.25	6.17
Hous_ coal	0.00387	0.01218	0.0014	0.00558	0.00766	0.01793	0.00303	0.00631	0	0
Hous_ elec	23.32	8.21	22.28	7.81	26.04	8.66	24.85	5.62	16.08	5.39
Hous_ M1969	10.65	3.71	11.1	2.8	9.25	1.94	6.78	2.08	17.78	1.76
Hous_ M1970	12.09	1.95	11.63	1.81	12.56	1.49	10.07	1.88	14.03	1.59
Hous_ M1980	16.92	1.91	16.2	1.5	17.5	1.25	15.36	2.32	18.81	2.45
Hous_ M1990	15.68	1.49	15.3	1.64	16.31	1.19	15.61	1.73	14.86	0.89
Hous_ M1995	26.07	2.82	25.66	2.1	26.96	1.92	29.35	1.22	21.15	1.52
Hous_ M2000	18.59	3.98	20.11	3.66	17.41	1.51	22.83	5.35	13.38	1.62
Hous_ medown	53905	13047	52028	10000	51230	8755	78867	8317	43911	8195
Hous_ mill	0.314	0.286	0.1465	0.1305	0.376	0.286	0.481	0.285	0.476	0.403
Hous_ nofuel	0.1472	0.1145	0.1578	0.102	0.1732	0.1365	0.1276	0.0487	0.046	0.0573
Hous_ nokitch	2.822	1.353	2.553	1.169	3.136	1.127	1.319	0.536	4.148	1.601
Hous_ nophone	6.38	2.64	6.38	1.81	8.02	2.64	3.065	0.887	4.281	1.207
Hous_ noplum	2.163	1.151	1.71	0.568	2.743	1.149	0.975	0.389	2.881	1.535
Hous_ occ	84.97	5.77	84.98	3.89	83.98	7.08	91.97	1.27	81.26	2.38
Hous_ oil	0.1368	0.1199	0.0915	0.1024	0.1695	0.108	0.0707	0.0365	0.24	0.1673
Hous_ otherfuel	0.506	0.364	0.414	0.281	0.593	0.399	0.2758	0.1467	0.741	0.451
Hous_ overmill	0.1673	0.1506	0.1282	0.1089	0.2158	0.1796	0.2025	0.1032	0.096	0.1579
Hous_ rent100	24.41	9.35	24.27	8.35	27.38	8.68	10.83	4.02	28.53	6.12
Hous_ rent1000	0.618	0.74	0.544	0.503	0.421	0.469	1.804	1.171	0.326	0.555
Hous_ rent125	0.184	0.293	0.208	0.341	0.0863	0.1487	0.4267	0.2708	0.192	0.388

0										
Hous_ rent1500	0.172	0.323	0.1488	0.2223	0.095	0.247	0.675	0.507	0	0
Hous_ rent2000	0.0721	0.1301	0.0669	0.1032	0.0478	0.1078	0.2423	0.2018	0	0
Hous_ rent300	38.9	6.54	41.21	6.22	38.73	5.76	39.31	5.46	31.58	6.4
Hous_ rent500	13.93	7.5	14.96	5.6	11.42	5.62	26.42	5.38	6.53	4.09
Hous_ rent700	3.49	3.55	3.16	2.35	2.442	1.735	10.4	4.92	1.16	0.909
Hous_ rent99	1.635	1.209	1.295	1.017	2.086	1.27	1.035	0.86	1.827	1.434
Hous_ rentnocash	16.59	7.35	14.13	4.78	17.29	3.93	8.85	5.3	29.86	7.7
Hous_ solar	0.0191	0.0362	0.0178	0.0358	0.014	0.0369	0.03667	0.0224	0.0225	0.045
Hous_ tenown	74.3	5.53	71.78	4.83	76.26	2.76	71.27	10.14	78.92	2.64
Hous_ tenrent	25.7	5.53	28.22	4.83	23.74	2.76	28.73	10.14	21.08	2.64
Hous_ utgas	52.64	15.78	61.13	11.45	40.33	13.65	63.85	10.43	55.13	12.03
Hous_ vac	15.03	5.77	15.02	3.89	16.02	7.08	8.027	1.273	18.74	2.38
Hous_ wood	3.12	3.02	1.738	1.622	5.26	3.51	1.104	0.892	2.491	1.636
Inc_ 10	14.39	3.92	14.67	3.35	16.18	3.61	9.97	3.55	11.91	2.33
Inc_ 100	5.63	1.94	5.442	1.289	4.744	1.524	9.23	1.57	5.555	1.137
Inc_ 15	9.8	1.94	9.98	1.51	10.57	1.86	7.058	1.392	9.39	1.54
Inc_ 150	3.02	1.25	2.895	0.713	2.36	0.801	5.648	0.81	2.994	0.778
Inc_ 200	0.739	0.37	0.667	0.266	0.611	0.238	1.432	0.324	0.702	0.353
Inc_ 200more	0.85	0.352	0.811	0.301	0.746	0.306	1.301	0.428	0.872	0.252
Inc_ 25	18.06	2.16	18.24	1.31	18.8	1.64	14.15	1.84	18.95	2.28
Inc_ 35	15.47	1.42	15.46	1.27	15.62	1.23	14.02	0.99	16.41	1.93
Inc_ 50	16.87	1.73	16.95	1.53	16.38	1.77	17.5	1.58	17.62	2.06
Inc_ 74	15.17	3.08	14.87	2.23	13.99	2.5	19.68	3.72	15.59	2.72
Inc_ medhh	29943	5039	29297	3294	27664	4034	38733	5256	30826	2798
Income_ percap	15526	2067	15316	1279	14357	1693	19134	1510	16494	1110
LBW_ p	6.358	0.985	6.645	1.045	6.383	0.781	6.067	0.712	5.644	1.302
Occup_ Const	12.49	1.81	12.25	1.41	13.07	1.64	11.43	2.66	12.4	2.17

Occup_ Farm	2.319	1.859	2.117	1.35	2.066	1.161	0.464	0.375	5.671	2.129
Occup_ Mangt	28.04	3.65	28.04	2.15	25.32	2.24	32.02	3.72	33.14	1.95
Occup_ Prod	17.8	4.36	16.89	2.9	21.21	3.89	13.99	2.14	13.18	2.18
Occup_ Sales	23	2.97	23.24	2.22	21.98	2.63	27.77	1.65	20.81	1.92
Occup_ Service	16.35	2.25	17.47	2.36	16.35	1.68	14.32	1.21	14.8	2.21
Psyc_ alone	439	1092	312	285	215.5	164.9	1995	2792	38.78	20.55
Psyc_ anxiety	125	281	96.5	77.1	62	46.6	467	687	14.33	4.23
Psyc_ Depression	567	1217	438	339	295	227	2376	3030	77.1	50.6
Psyc_ Ling	1.148	1.131	1.387	1.463	0.755	0.48	1.35	0.829	1.482	1.48
psych_ mh	10.98	2.38	10.7	1.94	12.28	2.28	9.7	0.93	8.11	2.23
Race_ 2races	4.34	2.01	3.743	1.62	5.63	1.85	4.516	1.205	1.747	0.556
Race_ A	0.484	0.636	0.488	0.389	0.256	0.192	1.614	1.179	0.1042	0.0288
Race_ B	3.79	4.03	4.72	4.42	3.55	3.5	4.77	4.9	0.646	1.341
Race_ I	9.64	7.64	7.62	6.68	14.78	7.42	6.39	2.93	2.288	1.715
Race_ O	2.26	3.09	3.55	4.27	1.128	1.237	1.688	1.235	2.501	2.994
Race_ PI	0.0532	0.1129	0.0954	0.1756	0.027	0.0227	0.03941	0.019	0.01835	0.01709
Race_ W	79.42	9.42	79.78	8.86	74.63	8.11	80.98	5.51	92.7	3.44
Unemp	5.15	1.59	4.724	1.223	6.32	1.51	4.5	0.474	3.244	0.364
VLBW_ p	1.206	0.408	1.255	0.268	1.307	0.404	1.222	0.249	0.7	0.583

Appendix D

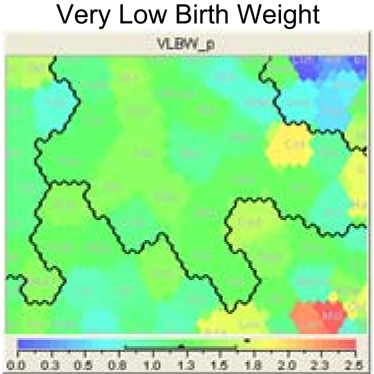
SOM Maps for Social Determinants of Health

Variable names appearing at the top of each map correspond to the variable names found in Appendix A. The scale bar at the bottom of each map indicates the range of data points. The arrow pointing upward indicates the mean value for variable.

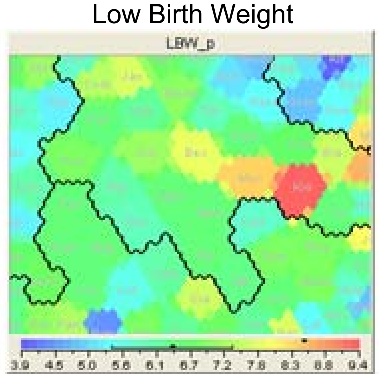
Social Economic Determinants

Education

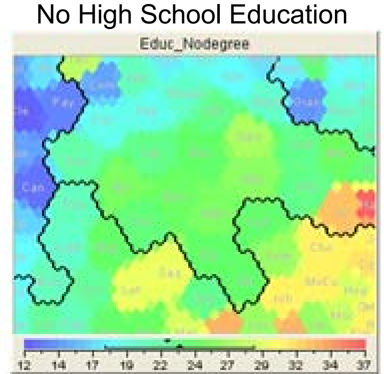
Very Low Birth Weight



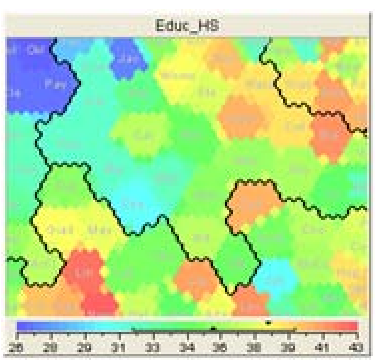
Low Birth Weight



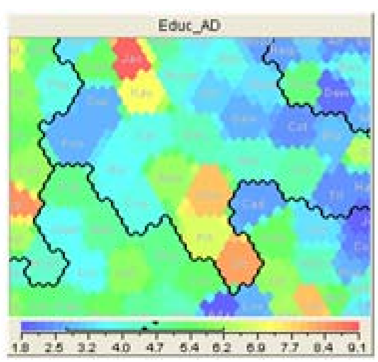
No High School Education



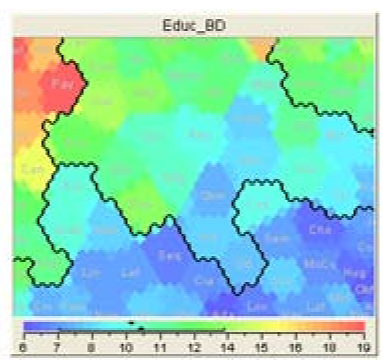
Percent population: high school diploma or GED only



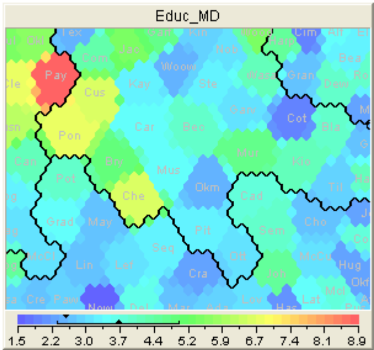
Percent population: Associates degree



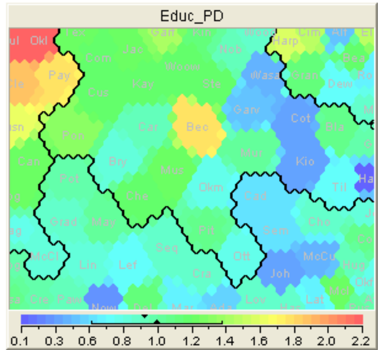
Percent population: Bachelors degree



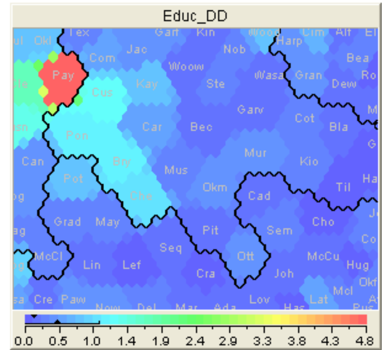
Percent population: Masters degree



Percent population: Professional school degree

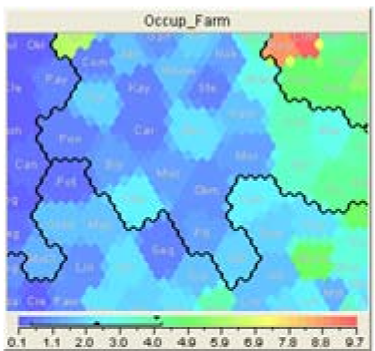


Percent population: Doctoral degree

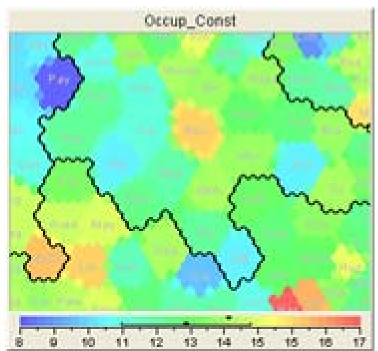


Socio-Economic Status: Occupation

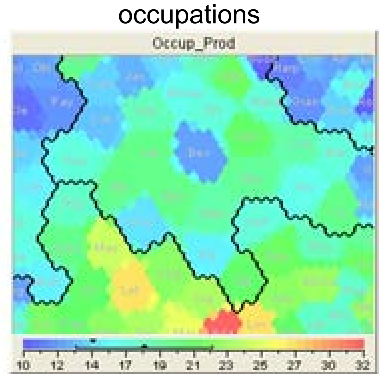
Percent population: Farming, fishing, and forestry occupations



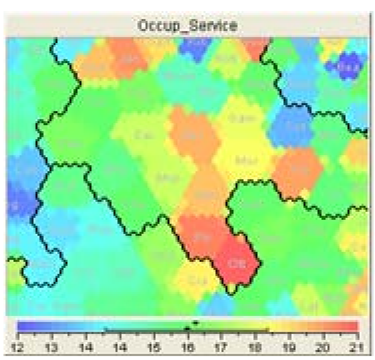
Percent population: Construction, extraction, and maintenance operations



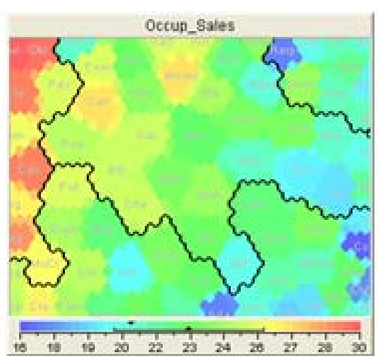
Percent population: Production, transportation, and material moving occupations



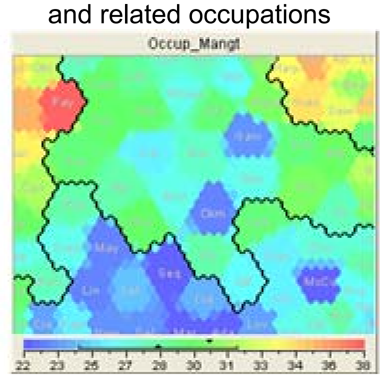
Percent population: Service occupations



Percent population: Sales and office occupations

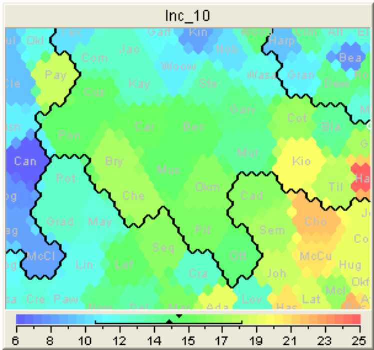


Percent population: Management, professional, and related occupations

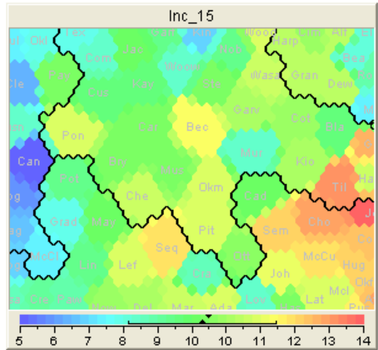


Socio-Economic Status: Income

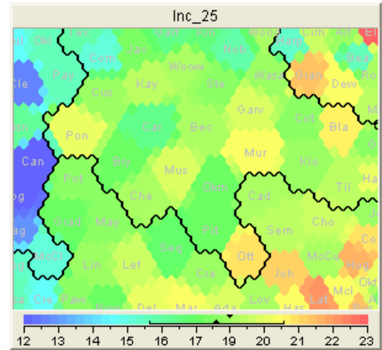
Household Income:
Less than $10,000



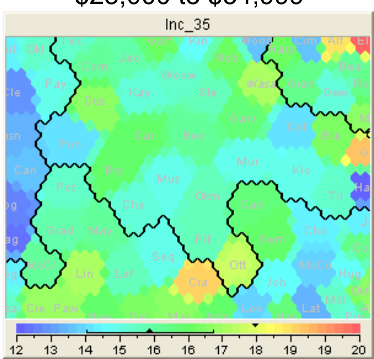
Household Income:
$10,000-$14,999



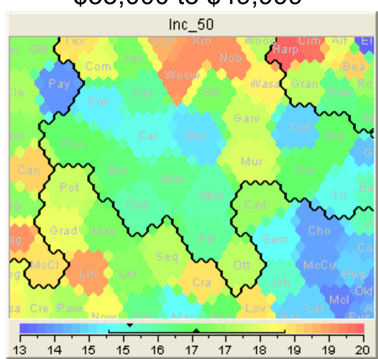
Household Income:
$15,000 to $24,999



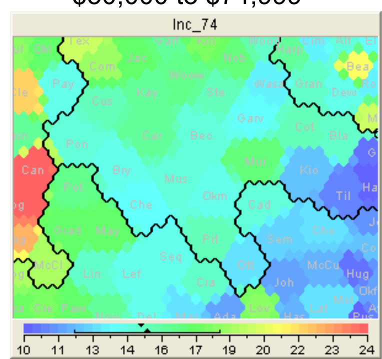
Household Income:
$25,000 to $34,999



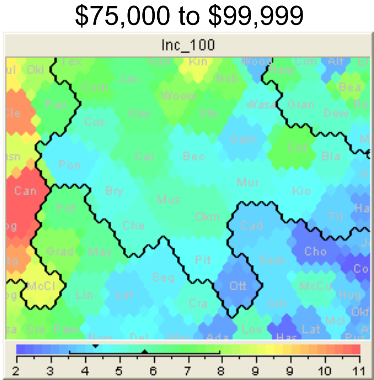
Household Income:
$35,000 to $49,999



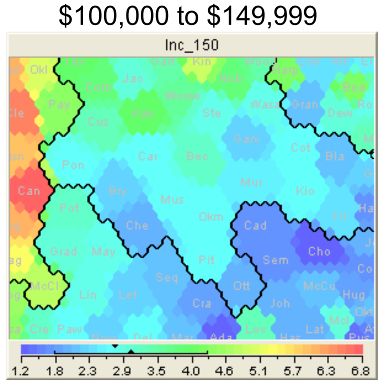
Household Income:
$50,000 to $74,999



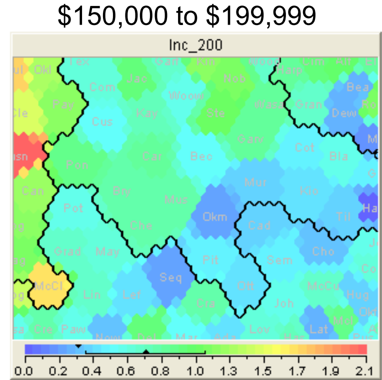
Household Income:
$75,000 to $99,999



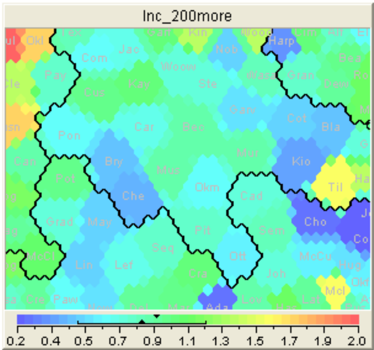
Household Income:
$100,000 to $149,999



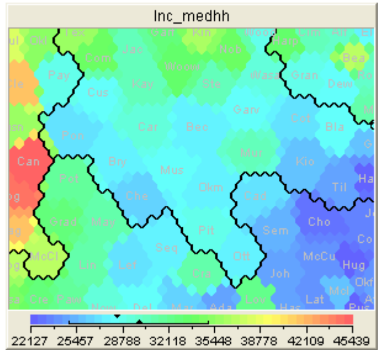
Household Income:
$150,000 to $199,999



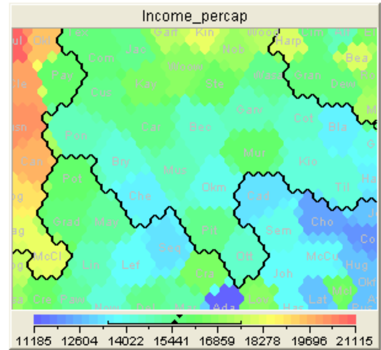
Household Income: $200,000 or more



Median household income (dollars)

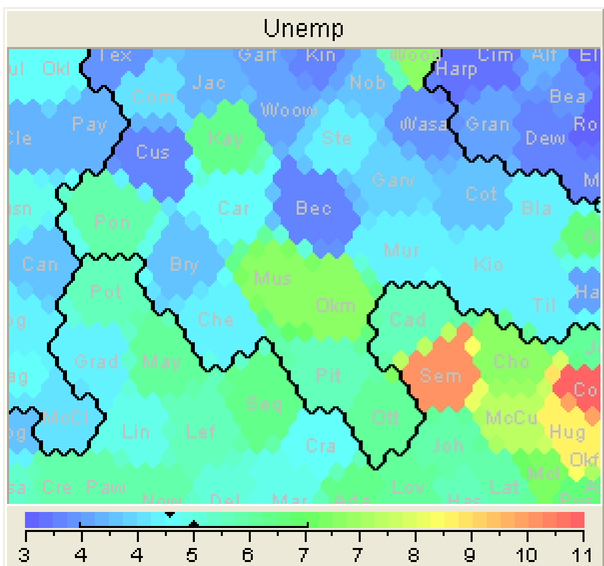


Per capita income in 1999 dollars



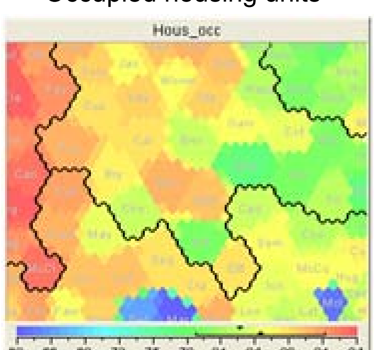
Socio-Economic Status: Unemployment

Average Annual Unemployment Rate

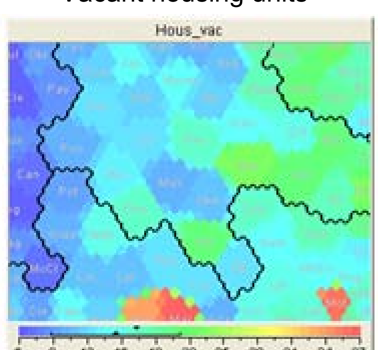


Socio-Economic Status: Housing

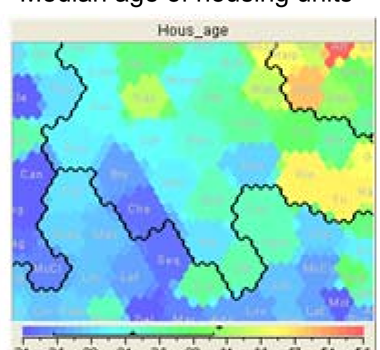
Occupied housing units



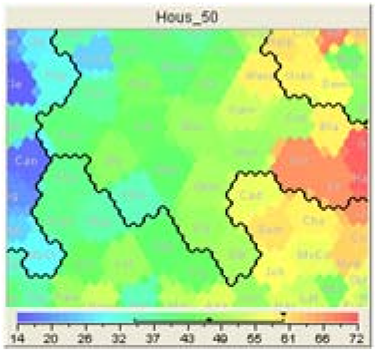
Vacant housing units



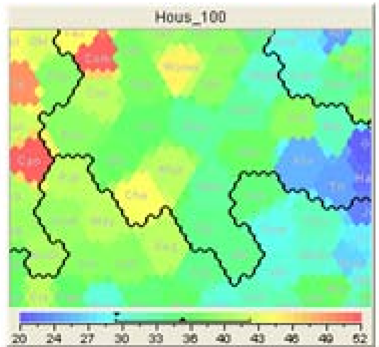
Median age of housing units



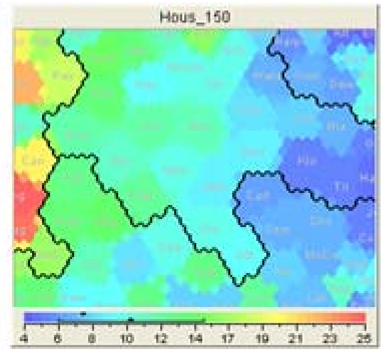
Owner-occupied housing value: less than $50,000



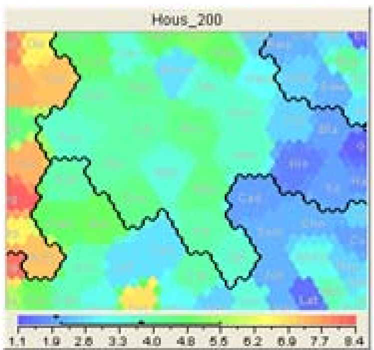
Owner-occupied housing value: $50,000 to $99,999



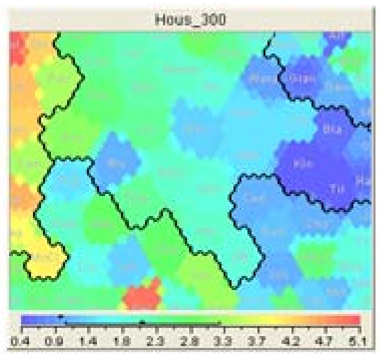
Owner-occupied housing value: $100,000 to $149,999



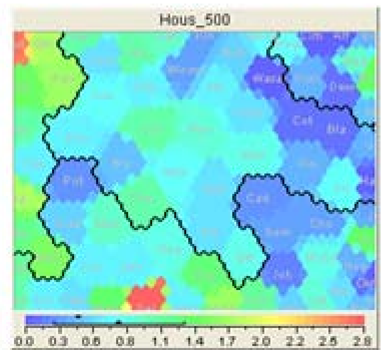
Owner-occupied housing value: $150,000 to $199,999



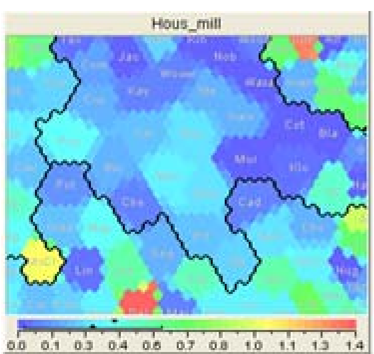
Owner-occupied housing value: $200,000 to $299,999



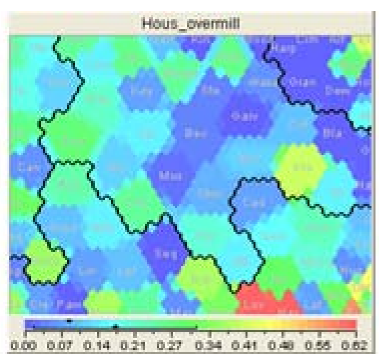
Owner-occupied housing value: $300,000 to $499,999



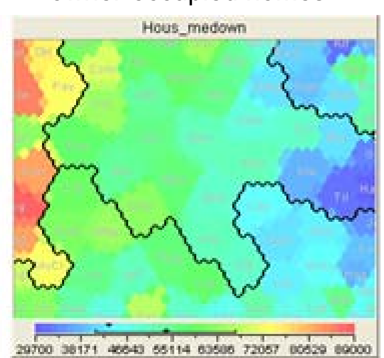
Owner-occupied housing value: $500,000 to $999,999



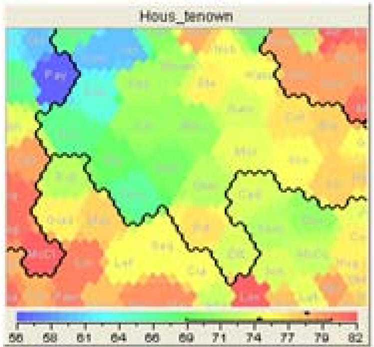
Owner-occupied housing value: $1,000,000 or more



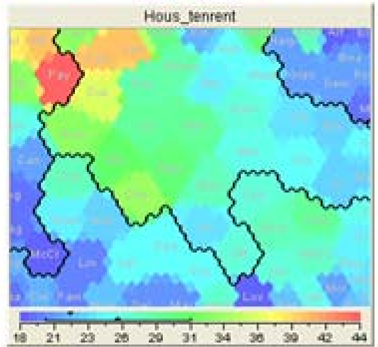
Median (dollars) value of owner-occupied homes



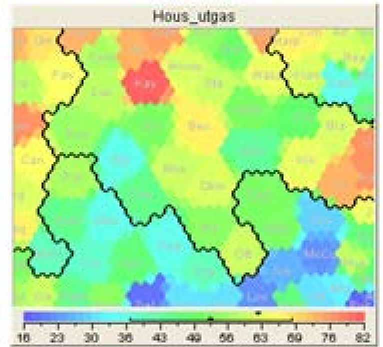
Tenure: Owner-occupied



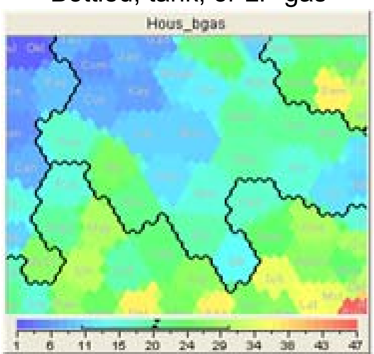
Tenure: Renter-occupied



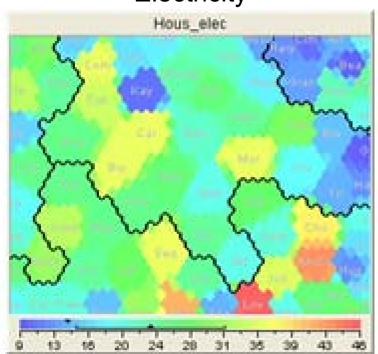
House Heating Fuel:
Utility gas



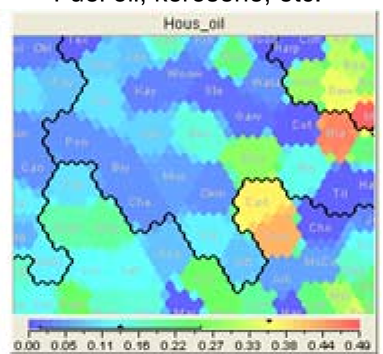
House Heating Fuel:
Bottled, tank, or LP gas



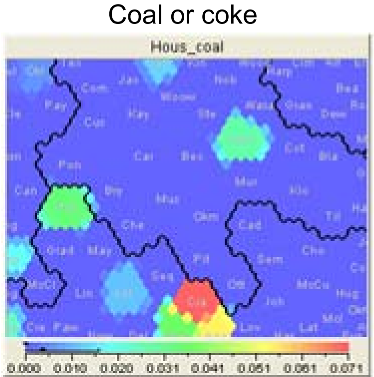
House Heating Fuel:
Electricity



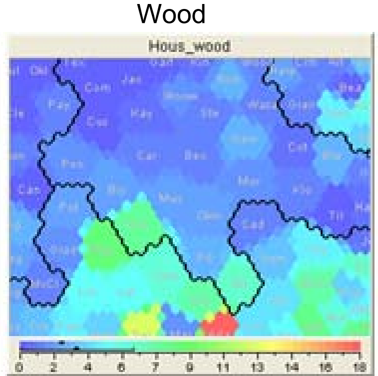
House Heating Fuel:
Fuel oil, kerosene, etc.



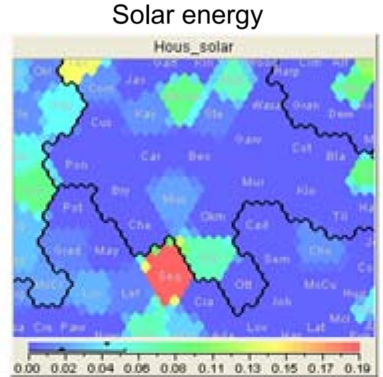
House Heating Fuel:
Coal or coke



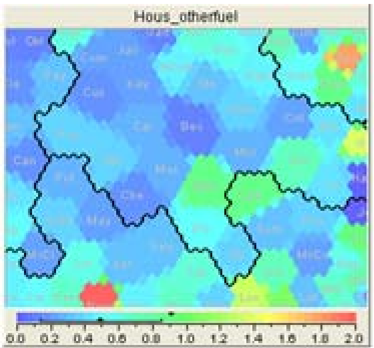
House Heating Fuel:
Wood



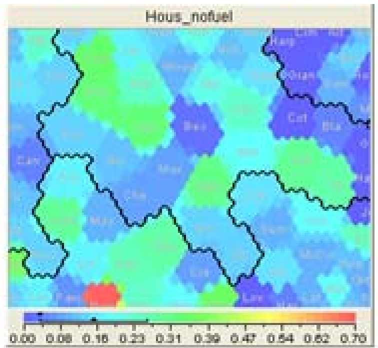
House Heating Fuel:
Solar energy



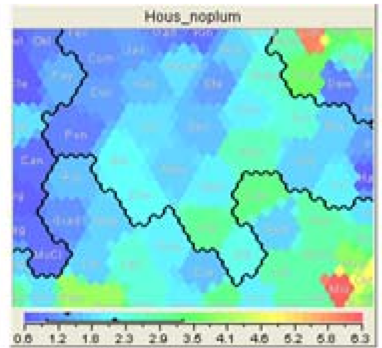
House Heating Fuel:
Other fuel



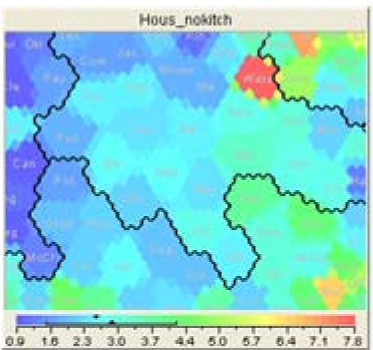
House Heating Fuel:
No fuel used



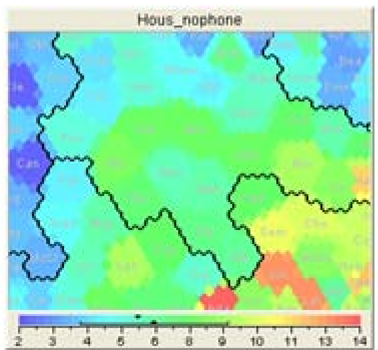
Lacking complete plumbing
facilities



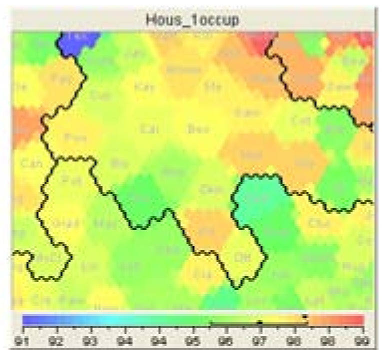
Lacking complete kitchen
facilities



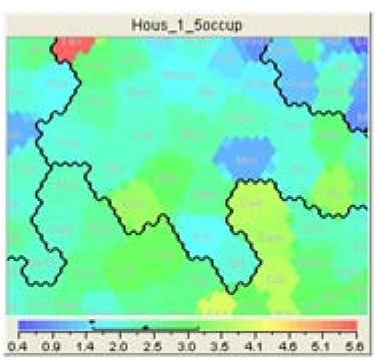
No telephone service available



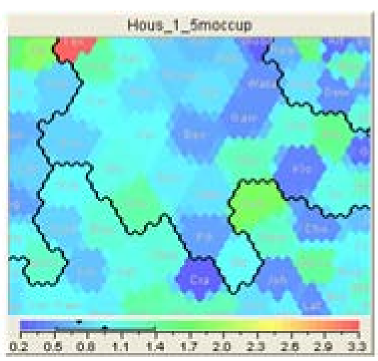
1.00 or less persons per
room



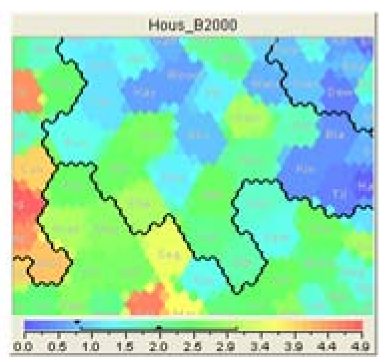
1.01 to 1.50 persons per
room



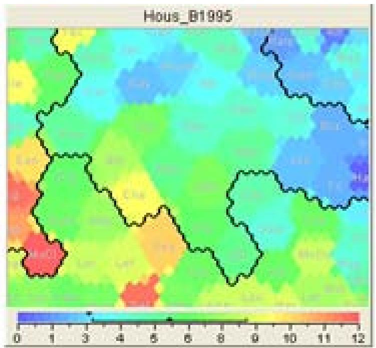
1.51 or more persons per
room



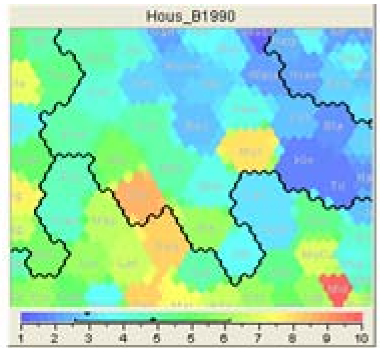
Built 1999 to 2000



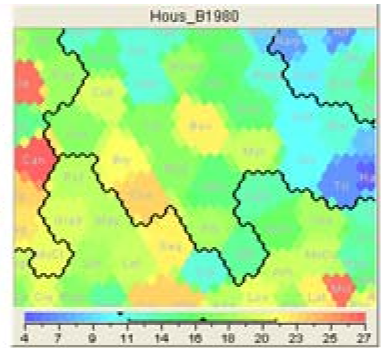
House Built 1995 to 1998



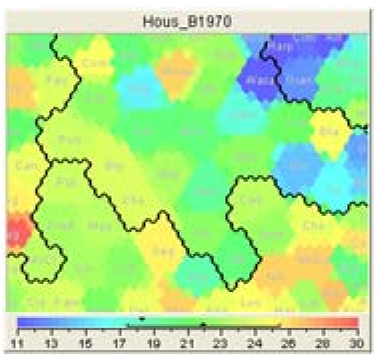
House Built 1990 to 1994



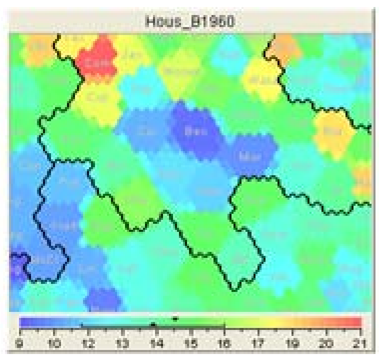
House Built 1980 to 1989



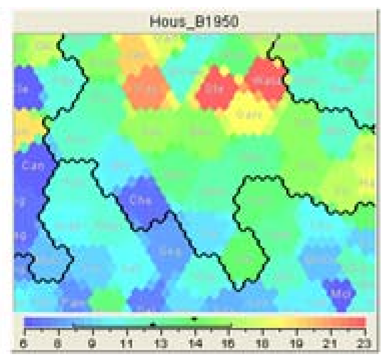
House Built 1970 to 1979



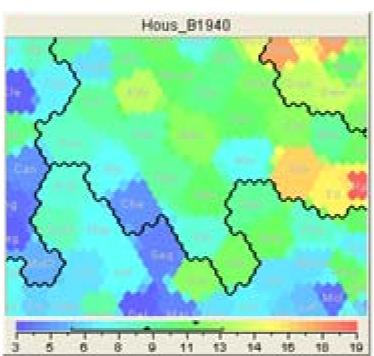
House Built 1960 to 1969



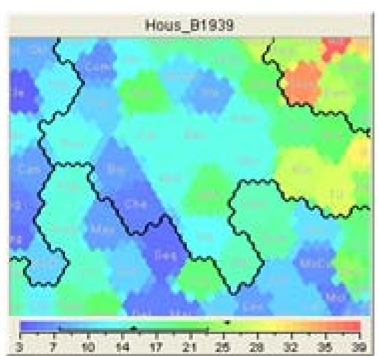
House Built 1950 to 1959



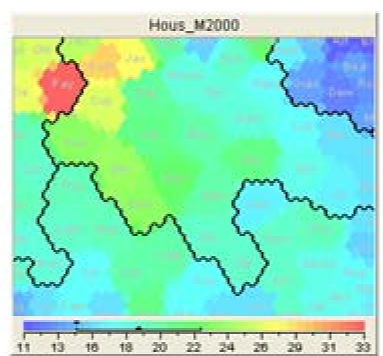
House Built 1940 to 1949



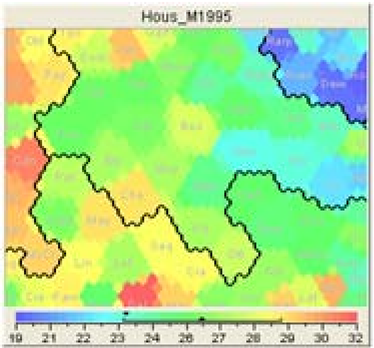
House Built 1939 or earlier



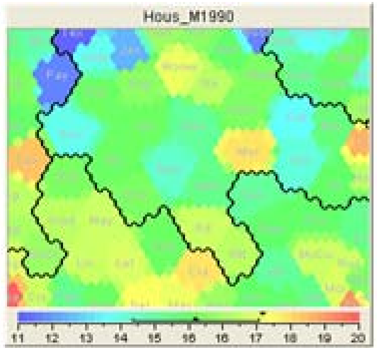
Moved in 1999 to 2000



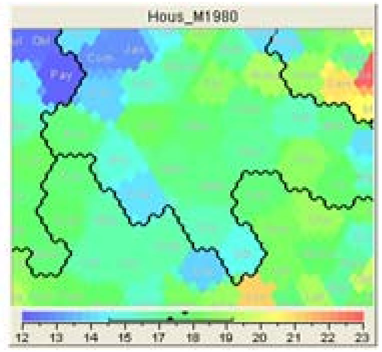
Moved in 1995 to 1998



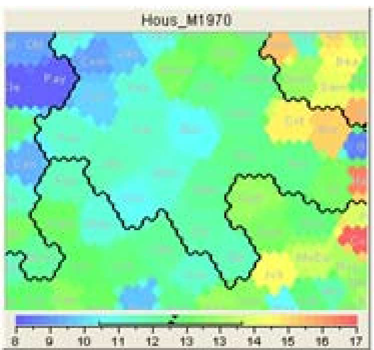
Moved in 1990 to 1994



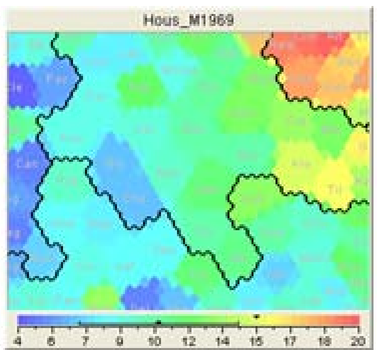
Moved in 1980 to 1989



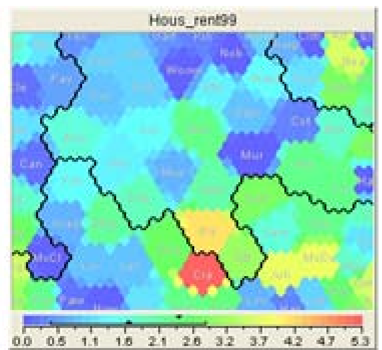
Moved in 1970 to 1979



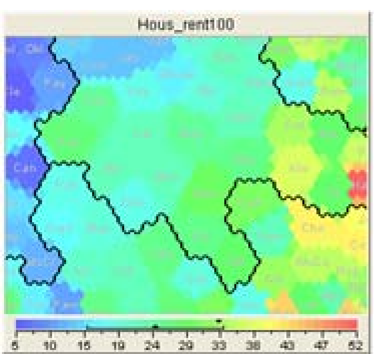
Moved in 1969 or earlier



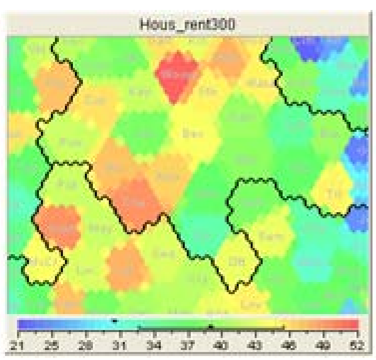
Renter-occupied housing units: With cash rent: under $100



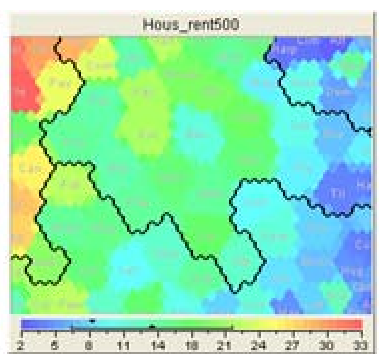
Renter-occupied housing units: With cash rent: $100 to $299



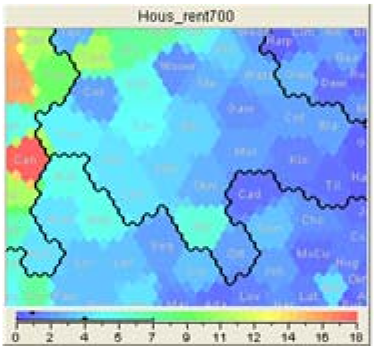
Renter-occupied housing units: With cash rent: $300 to $499



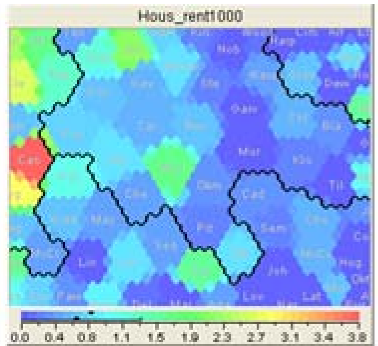
Renter-occupied housing units: With cash rent: $500 to $699



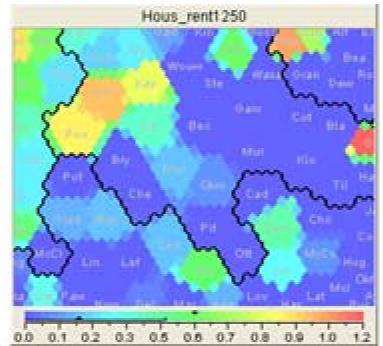
Renter-occupied housing units: With cash rent: $700 to $999



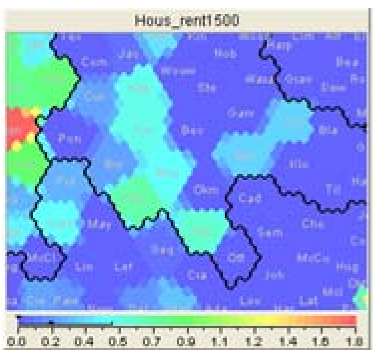
Renter-occupied housing units: With cash rent: $1000 to $1249



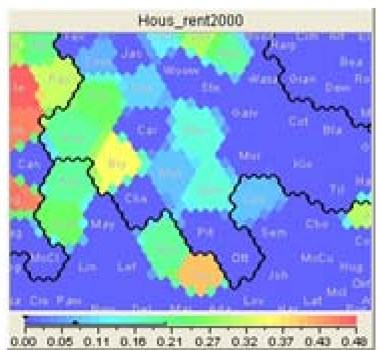
Renter-occupied housing units: With cash rent: $1250 to $1499



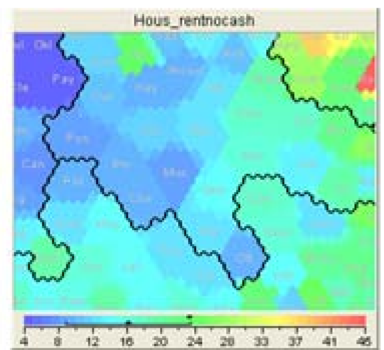
Renter-occupied housing units: With cash rent: $1500 to $1999



Renter-occupied housing units: With cash rent: $2000 and over

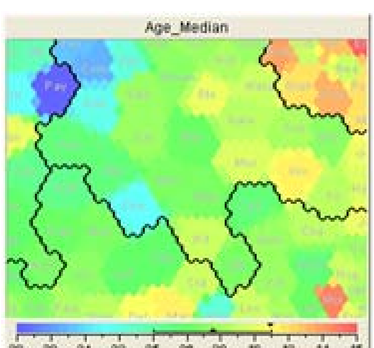


Renter-occupied housing units: With cash rent: $100 to $299

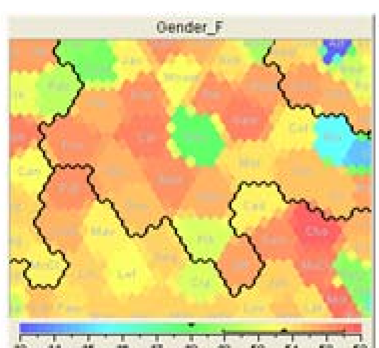


Demographics

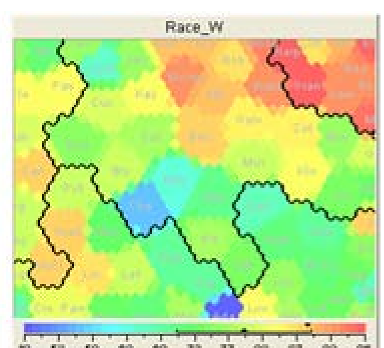
Median age in years for the population of the county



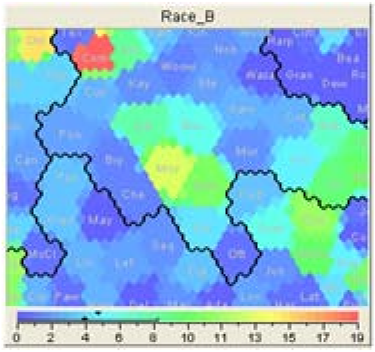
Percent population: female



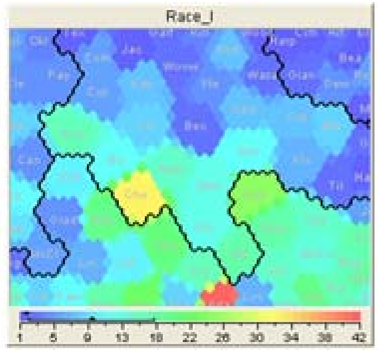
Percent population: as white only



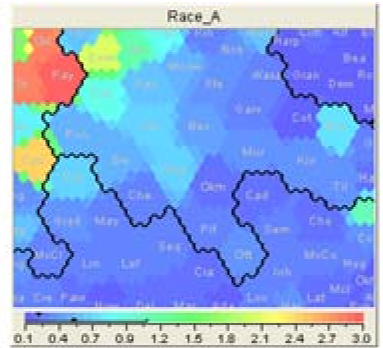
Percent population: as black only



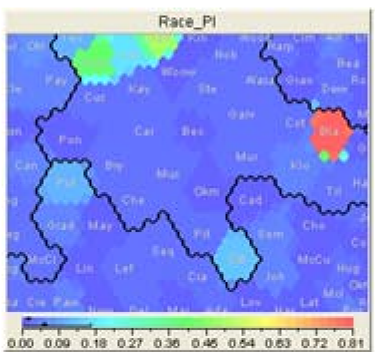
Percent population: as American Indian or Alaskan Native only



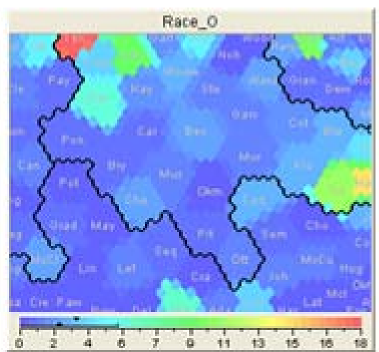
Percent population: as Asian only



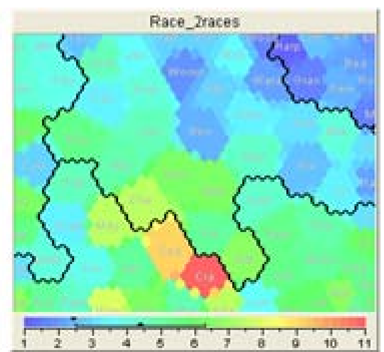
Percent population: as Native Hawaiian and other Pacific Islander only



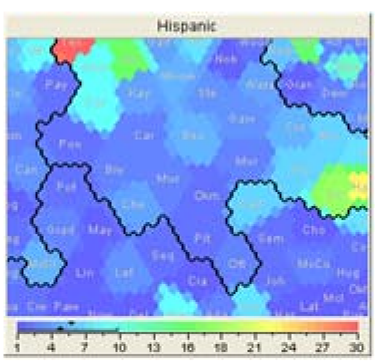
Percent population: as some other race only



Percent population: as 2 or more races

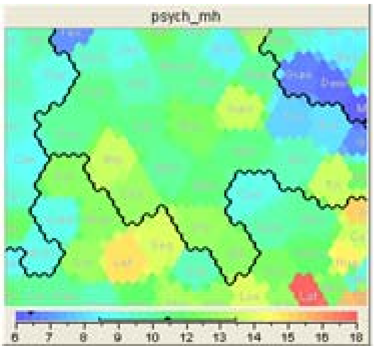


Percent population: as being of Hispanic origin

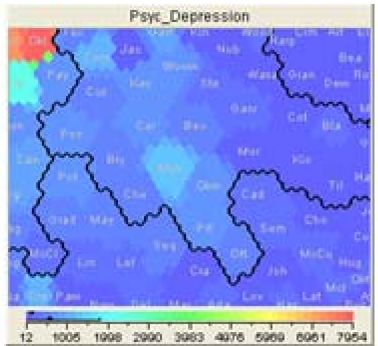


Psychosocial Risk Factors

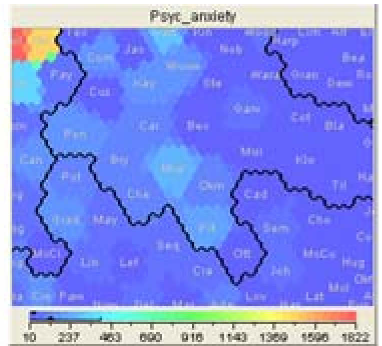
Median number of days with poor mental health (% of >15 poor mh days)



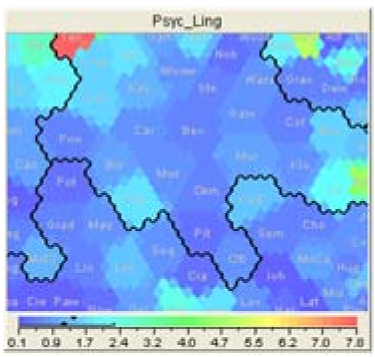
Number of persons treated for depressive disorders (<10 were set to missing)



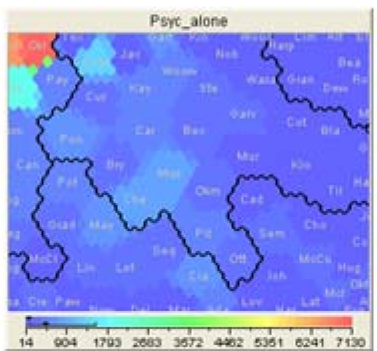
Number of persons treated for anxiety disorders



Percent population: Linguistically isolated

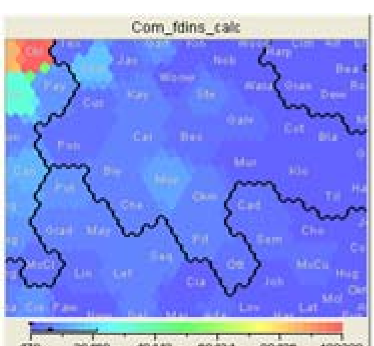


Number of persons treated for mental health disorders that live alone

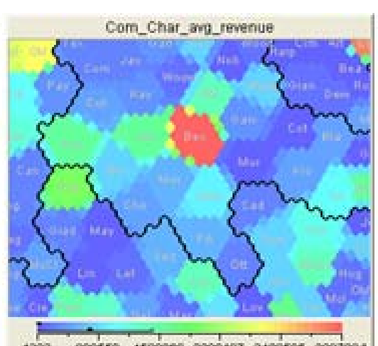


Community and Social Characteristics

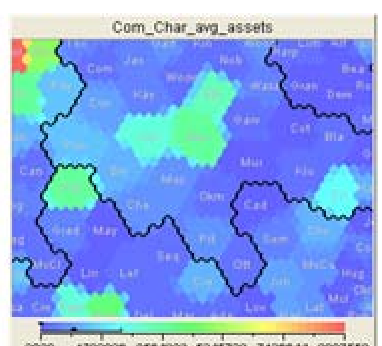
Calculated: 15.2% of population in Oklahoma are food insecure



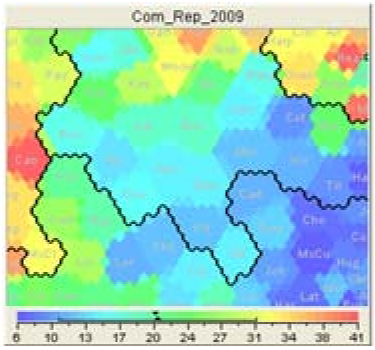
Average revenue for charitable organizations



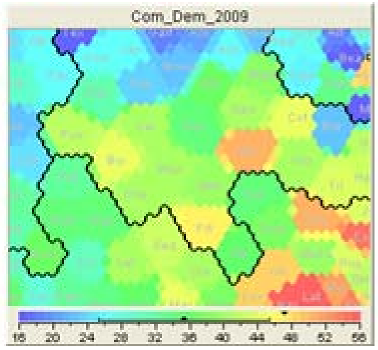
Average amount of assets for charitable organizations



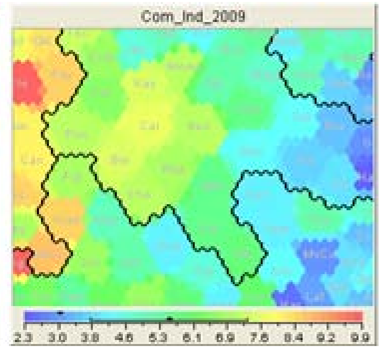
Percent of Registered Voters: Republican



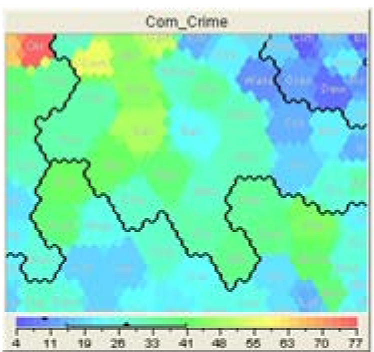
Percent of Registered Voters: Democrat



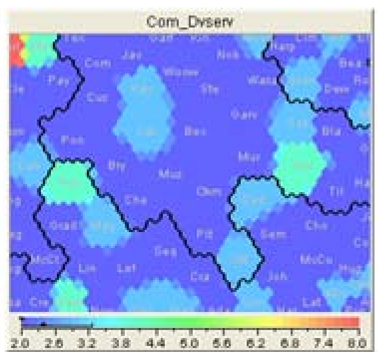
Percent of Register Voters: Independent



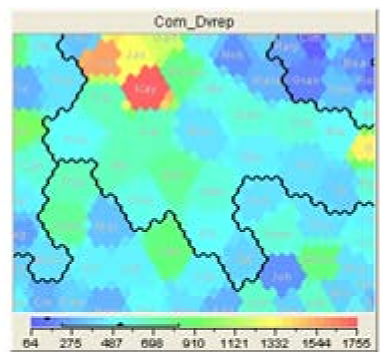
Oklahoma 2004 crime rate by county (index crime rate per 1,000 population)



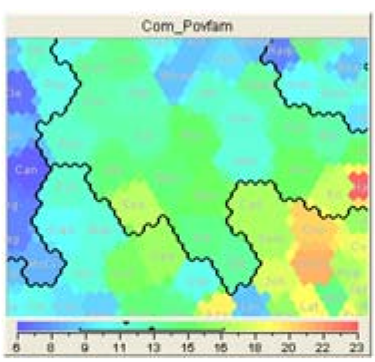
Number of Domestic Violence Services by County



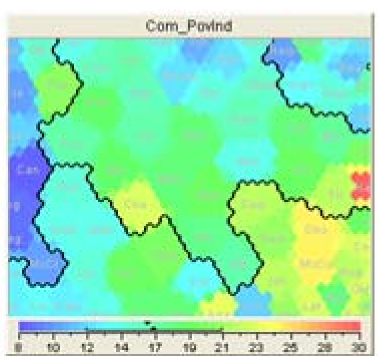
Rate of Domestic Violence Reports by County



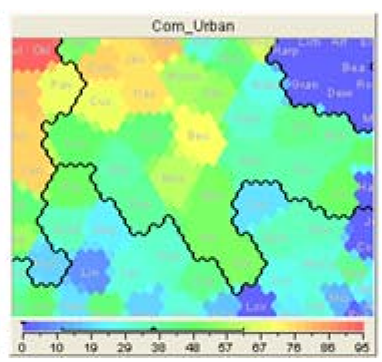
Percent of Families below Federal Poverty Level



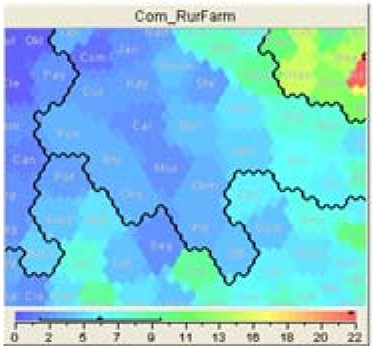
Percent of Individuals below Federal Poverty Level



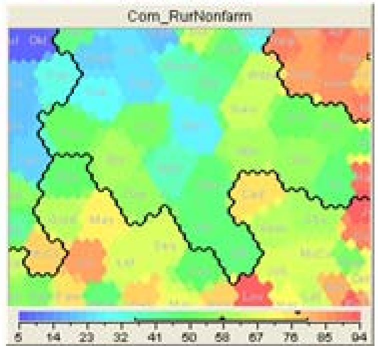
Percent Population: Urban



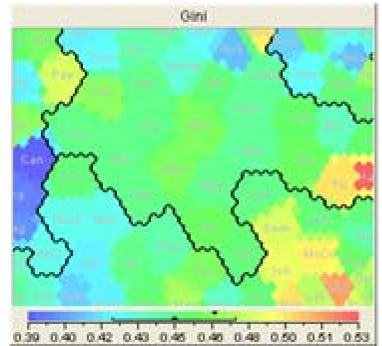
Percent Population: Rural Farm Area



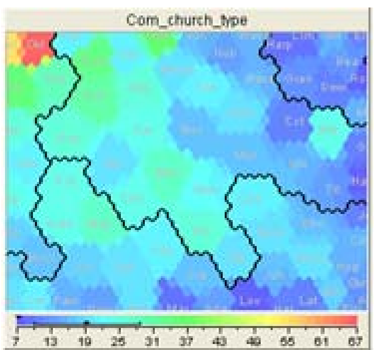
Percent Population: Rural Nonfarm Area



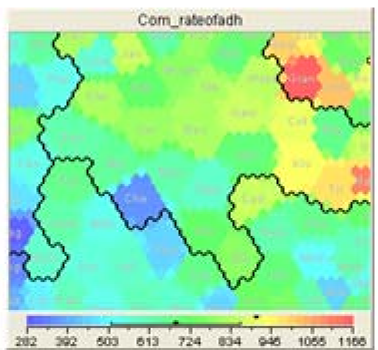
Gini Coefficient of Income Inequality



Number of Different Types of Churches



Rates of church adherence per 1000 population



Appendix E

Mean and Standard Deviation for Health Behavior Variables by Cluster

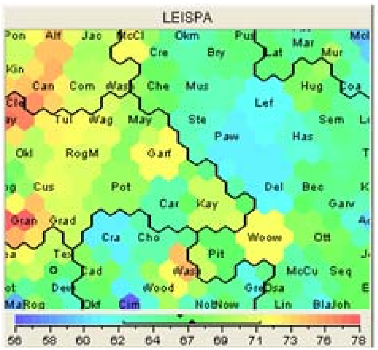
Attribute	Total Mean	Total Std. Dev.	Cluster 1 Mean	Cluster 1 Std. Dev.	Cluster 2 Mean	Cluster 2 Std. Dev.	Cluster 3 Mean	Cluster 3 Std. Dev.	Cluster 4 Mean	Cluster 4 Std. Dev.	Cluster 5 Mean	Cluster 5 Std. Dev.	Cluster 6 Mean	Cluster 6 Std. Dev.
Bingedrk	88.2	3.4	88.7	2.9	87.6	2.5	89.8	3.7	87.3	3.2	89.8	2.6	84.5	5.6
Checkup_1Yr	60.2	5.8	62.9	6.0	58.1	3.1	60.3	4.3	61.7	4.2	56.4	7.1	53.9	4.7
Checkup_2Yr	14.4	4.7	12.2	4.0	17.2	2.5	11.5	4.8	15.9	3.1	16.7	5.7	19.0	5.5
Checkup_5Yr	9.9	3.4	11.2	3.4	9.9	1.9	9.1	4.5	9.5	3.6	9.1	2.0	6.5	2.6
Checkup_M5Yr	12.2	4.8	10.9	4.1	11.9	2.4	15.9	6.2	10.1	2.1	14.0	6.4	14.2	7.9
Checkup_ Never	3.3	2.5	2.9	1.8	3.0	1.7	3.2	3.0	2.8	1.5	3.9	3.4	6.4	4.2
Cholest	34.3	5.2	36.1	4.5	32.9	3.3	29.6	7.5	32.6	4.4	37.2	5.0	33.9	6.9
Curntsmk	26.9	5.4	28.9	4.2	24.3	3.7	25.3	4.3	26.1	2.6	33.1	2.5	20.1	9.0
Denvst	56.6	6.5	53.0	5.3	59.5	5.4	57.8	6.2	65.0	5.4	57.2	6.5	57.0	5.3
Fivefv	84.5	6.4	83.1	5.9	85.5	3.9	84.1	9.0	83.0	7.9	89.0	2.4	87.3	8.7
Flushot3	61.1	5.0	61.1	3.9	62.3	2.6	57.4	5.7	60.8	6.2	62.4	7.9	63.7	6.9
Hadmam	65.6	5.9	65.1	4.6	70.2	4.0	60.3	7.3	69.9	5.8	61.4	4.0	60.2	3.9
Hadpap	81.2	5.2	79.2	4.0	83.7	2.5	84.1	4.9	87.5	1.4	74.2	7.4	76.5	9.2
Heavydrk	96.2	1.5	96.4	1.1	96.5	0.8	96.7	0.7	96.3	1.2	95.4	2.3	92.4	5.2
Ipvattempt	13.8	4.4	13.6	2.4	12.5	2.9	10.6	5.8	22.3	5.4	15.6	1.3	10.7	6.3
Ipvphysical	15.6	4.3	16.9	2.5	13.6	3.0	10.8	4.4	21.9	2.6	19.4	2.9	10.6	6.9
Ipvthreat	15.0	3.3	16.3	2.2	13.1	2.4	10.3	2.2	19.9	2.1	16.3	1.4	14.6	2.7
Leispa	67.0	4.4	65.2	2.8	71.1	3.8	65.1	5.3	71.3	2.3	64.3	3.3	66.4	4.8
Mrestrict	5.5	2.5	7.0	2.5	4.7	1.9	4.5	2.0	4.5	1.4	5.6	2.3	2.5	1.5
Pneuvac3	32.5	7.4	31.8	6.9	31.8	3.8	39.1	8.1	26.3	4.0	44.4	12.7	30.4	1.6
Profexam	90.0	3.3	89.5	3.1	90.7	1.5	91.5	4.9	93.0	1.9	86.2	3.7	88.8	3.1
Psatest	49.3	6.7	44.8	4.0	52.5	3.6	42.9	0.0	53.1	5.6	70.9	0.0	41.4	0.0
Psych_Mh	11.0	2.4	12.2	1.8	10.2	1.6	9.7	1.6	9.5	0.7	13.9	2.5	7.8	2.3
Recpa	40.6	4.9	41.9	5.3	40.3	1.8	43.7	3.7	39.6	4.2	32.7	3.3	37.8	1.7
Seatbelt	89.1	5.7	88.9	4.7	91.6	3.2	80.0	7.9	91.5	1.3	92.8	1.1	93.4	0.2
Sexviolever	6.5	2.4	6.6	1.3	6.5	1.9	4.0	2.7	8.7	4.3	7.6	1.7	2.7	0.0
Sigmoid	42.7	6.0	41.3	6.0	44.8	4.5	38.2	6.4	46.1	5.5	47.9	6.6	41.9	0.0
Stopsmk2	23.0	3.6	23.3	3.7	22.6	2.1	22.6	6.1	22.6	1.3	22.1	1.8	25.2	5.5
Weight_Normal	36.3	4.3	35.9	3.7	37.6	3.8	36.2	2.9	39.3	3.3	39.0	5.0	29.6	4.5
Weight_Obese	26.1	4.5	27.3	3.7	24.3	2.4	23.6	5.4	25.8	5.2	28.3	6.6	27.1	6.3
Weight_Overwt	37.6	5.0	36.9	3.9	38.1	2.9	40.2	3.6	34.8	7.2	32.7	3.2	43.3	7.4

Appendix F

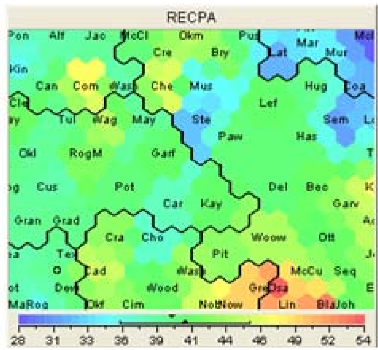
SOM Maps for the Health Behavior Variables

Variable names appearing at the top of each map correspond to the variable names found Appendix B.

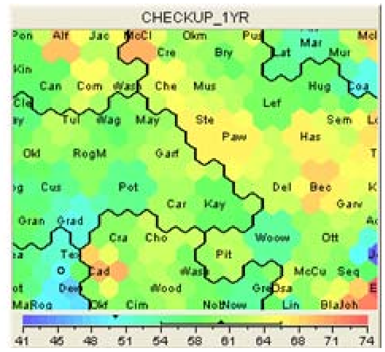
Leisure Time Physical Activity



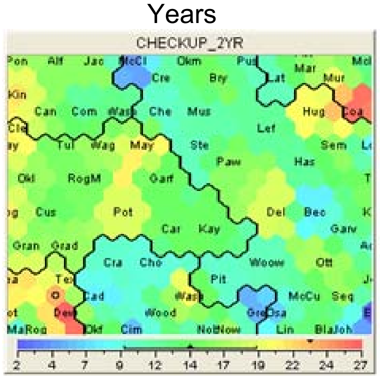
Recommended Amounts Physical Activity



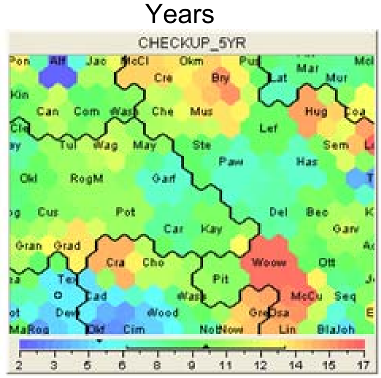
Doctors Visit Within Past Year



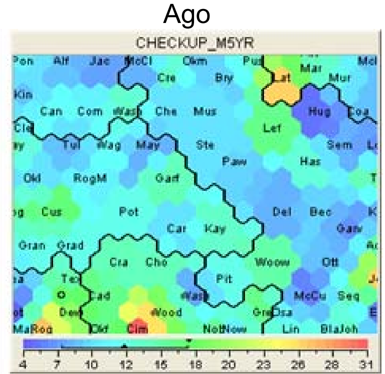
Doctor Visit - Within Past 2 Years



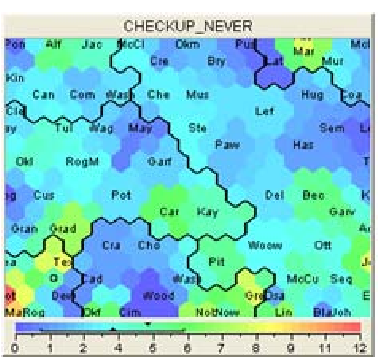
Doctor Visit - Within Past 5 Years



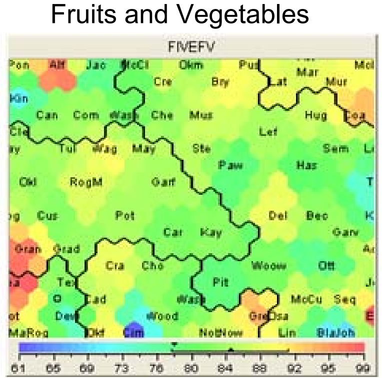
Doctor Visit - 5 or More Years Ago



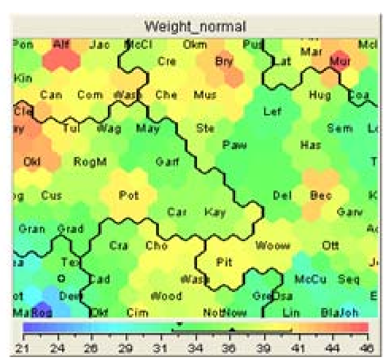
Doctor Visit - Never



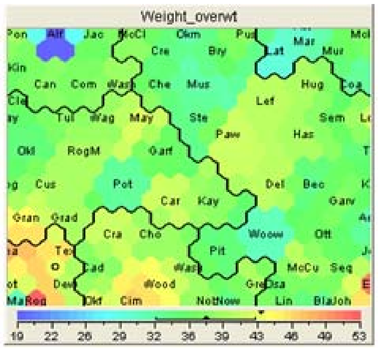
Recommended Number of Fruits and Vegetables



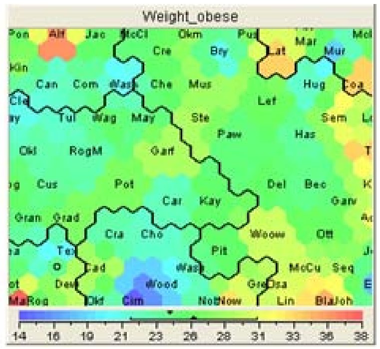
Normal Weight - BMI < 25



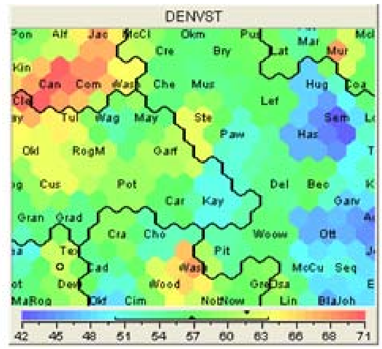
Overweight - BMI 25 - 30



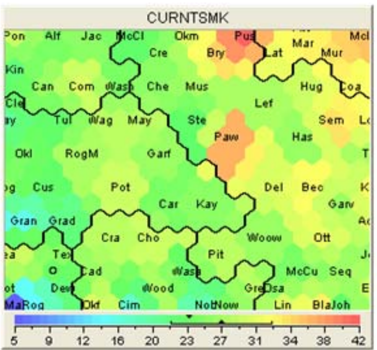
Obese - BMI >= 30



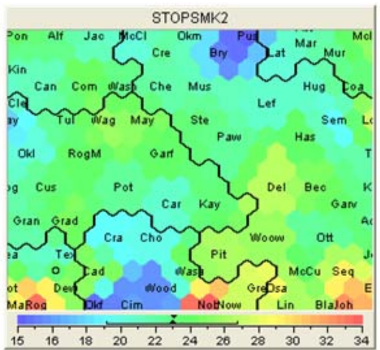
Dental Visit Within Past Year



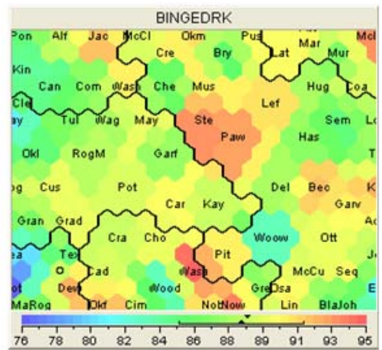
Current Smokers



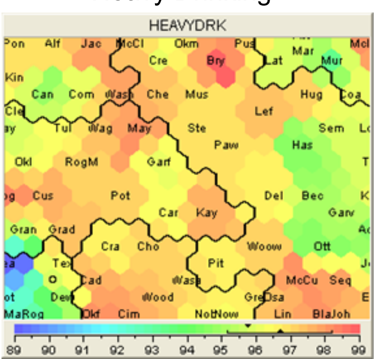
Stopped Smoking for Day or Longer



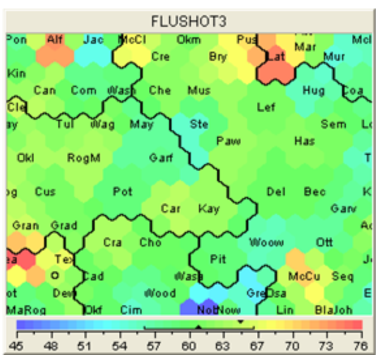
Binge Drinking



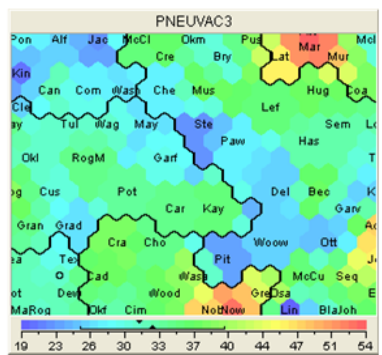
Heavy Drinking



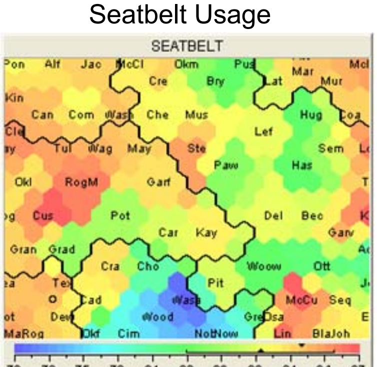
Influenza Vaccination



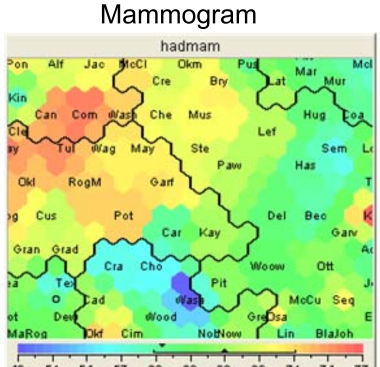
Pneumonia Vaccination



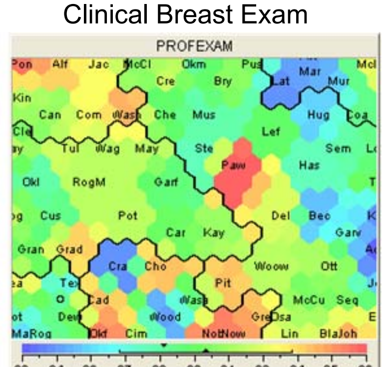
Seatbelt Usage



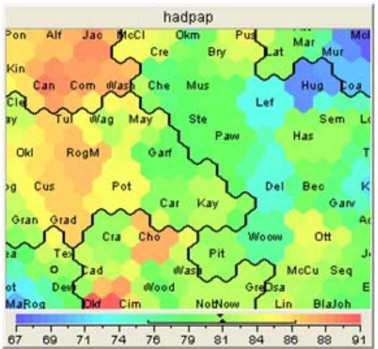
Mammogram



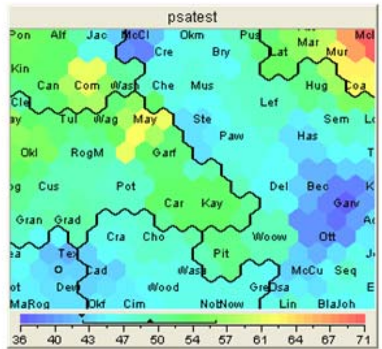
Clinical Breast Exam



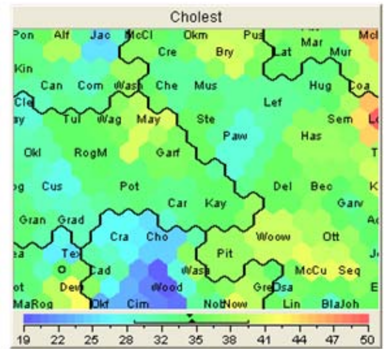
Pap Smear Test



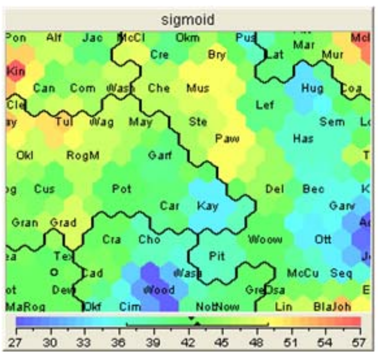
Prostate-Specific Antigen Test



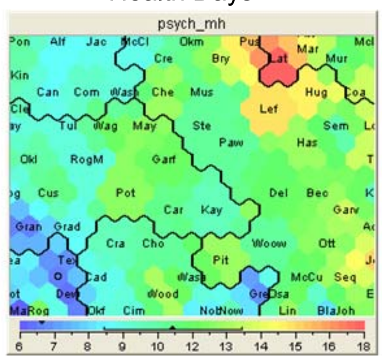
Cholesterol Check - Within The Past Five Years



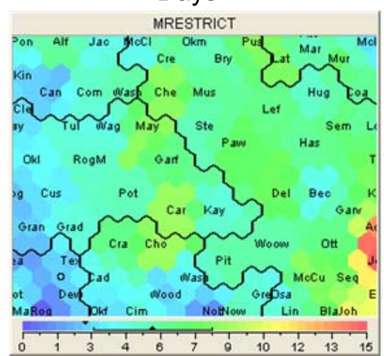
Sigmoidoscopy



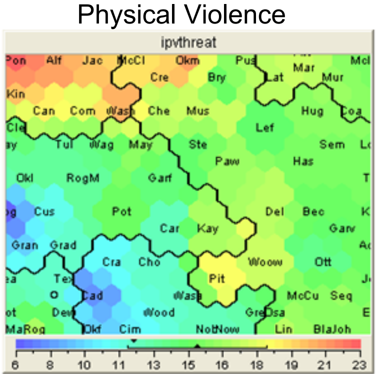
More Than 15 Poor Mental Health Days



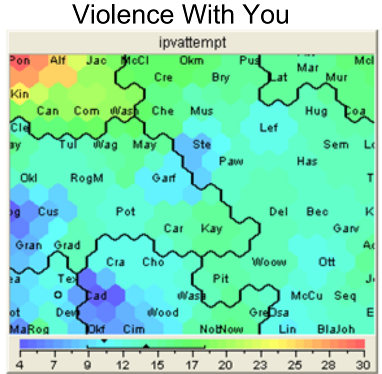
More Than 15 Poor Health Days



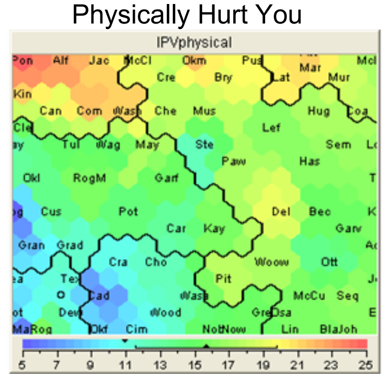
Intimate Partner EVER THREATENED You With Physical Violence



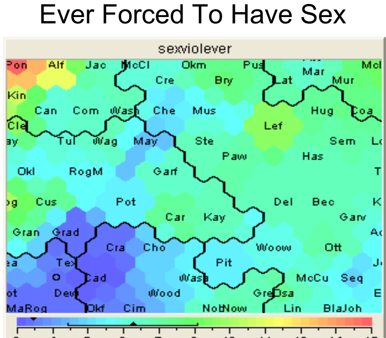
Intimate Partner EVER ATTEMPTED Physical Violence With You



Intimate Partner EVER Hit, Slapped, Pushed, Kicked, or Physically Hurt You



Ever Forced To Have Sex



Appendix G

OSU Institutional Review Board Letter

Oklahoma State University Institutional Review Board

Date:	Wednesday, March 18, 2009
IRB Application No	ED0957
Proposal Title:	The Utility of Self-Organizing Maps in the Analysis of Social Determinants of Health
Reviewed and Processed as:	Exempt

Status Recommended by Reviewer(s): Approved Protocol Expires: 3/17/2010

Principal Investigator(s):

Miriam McGaugh
5401 Colfax Place
Oklahoma City, OK 73112

Janice Miller
313 Willard
Stillwater, OK 74078

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

☐ The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval.
2. Submit a request for continuation if the study extends beyond the approval period of one calendar year. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of this research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved protocols are subject to monitoring by the IRB and that the IRB office has the authority to inspect research records associated with this protocol at any time. If you have questions about the IRB procedures or need any assistance from the Board, please contact Beth McTernan in 219 Cordell North (phone: 405-744-5700, beth.mcternan@okstate.edu).

Sincerely,

Shelia Kennison, Chair
Institutional Review Board

VITA

Miriam Jane McGaugh

Candidate for the Degree of

Doctor of Philosophy

Dissertation: THE UTILITY OF SELF-ORGANIZING MAPS IN THE ANALYSIS OF SOCIAL DETERMINANTS OF HEALTH

Major Field: Educational Psychology

Biographical:

Education:
Completed the requirements for the Doctor of Philosophy in Educational Psychology at Oklahoma State University, Stillwater, Oklahoma in December, 2009.

Completed the requirements for the Master of Science in Epidemiology at University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma in August, 2001.

Completed the requirements for the Bachelors of Science in Biology at Oklahoma City University, Oklahoma City, Oklahoma in May, 1998.

Experience:
Epidemiologist, Oklahoma State Department of Health, Community Development Service, Oklahoma City, OK, 2005-present.

Epidemiologist, Oklahoma State Department of Health, Injury Prevention Service, Oklahoma City, OK 2001-2005.

Professional Memberships:
Oklahoma Public Health Association; State and Territorial Injury Prevention Directors Association; Beta Beta Beta Biological Honor Society

Name: Miriam McGaugh Date of Degree: December, 2009

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: THE UTILITY OF SELF-ORGANIZING MAPS IN THE ANALYSIS OF SOCIAL DETERMINANTS OF HEALTH

Pages in Study: 185 Candidate for the Degree of Doctor of Philosophy

Major Field: Educational Psychology

Scope and Method of Study: The purpose of this study was to present a method for analyzing existing, nationally-available social data in a health context. This research study utilized the Public Health Model of the Social Determinants of Health (PHM) as a theoretical guide for selecting variables from existing, archival data sources to represent social determinants, health behaviors, and the health outcomes within Oklahoma at the county level, which was the unit of analysis. Additionally, this study introduced the Self-Organizing Map algorithm as an alternative data reduction technique for analysis of health and social data. The study further analyzed which of these sets of health variables were stronger predictors of the health outcome, age-adjusted mortality rate. Three phases of research were conducted: self-organizing map analyses to determine the underlying mathematical structures of social determinants of health variables and health behavior variables, a multiple regression analysis on the two resulting SOM solutions for determination of the stronger predictor set, and correlation analysis among the SOM variables and the health outcome to determine construct development.

Findings and Conclusions: The overall results support the use of the Self-Organizing Map algorithm in the analysis of public health data. The SOM analysis of social determinants of health variables revealed four clusters: *Mid-Century Service-Oriented Communities*, *Struggling Minority Communities*, *High Income and High Education*, and *Long-term Farmland.* The health behavior SOM analysis identified six clusters: *Restricted, Health Promoting, Overweight and Obese, Conflicted Intimate Partner Violence, Conflicted Mental and Physical Health,* and *Safety Not Health-Related.* The multiple regression analysis resulted in a significant model with two significant parameters: *Mid-Century Service-Oriented Communities* and *Struggling Minority Communities.* Correlation analysis identified a cohesive construct for the social determinants of health SOM variables, but a less coherent health behavior SOM construct. Further research is needed to define this area better and extend these findings.

ADVISER'S APPROVAL: Dr. Janice Miller