An exploratory spatial analysis of American adults' health literacy at county-level

By Takashi Yamashita & J. Scott Brown, Miami University, Oxford, OH, USA

Background

This study investigates the geographical distribution of limited health literacy (HL) for detecting at-risk areas at the county-level in the U.S. Whereas the national-level adult literacy surveys such as the International Adult Literacy Survey (NALS) in 1992 and the National Assessment of Adult literacy (NAAL) in 2003 showed a high prevalence of limited HL among American adults, little is known about the geographical distribution of limited HL in the U.S. The geographic information system (GIS) is used to address the understudied areas of spatial distribution of limited HL. This study provides visualized geographical distributions of limited HL data in a map format, exploratory spatial analysis, and suggestions for future HL research in the context of existing health disparities in the U.S.

Mapping HL

HL is a set of abilities to locate, understand and use health-related information to maintain and improve one's health over the life course; accumulating evidence suggests a significant association between HL and health outcomes (e.g., mortality, morbidity, health lifestyle, medication use) as well as great potential to address existing serious public health disparity issues in the U.S. (Chao, Anderson, & Hernandez, 2009; Nielsen-Bohlman, et al., 2004; U.S. Department of Health & Human Service, 2006). Recently, as the awareness of limited HL and potential public health consequences has been rising, researchers and organizations have come to realize the necessity of understanding the geographical distributions of HL. In fact, the Canadian Council on Learning (CCL) developed HL national county-level maps using the results of the International Adult Literacy and Life Skill Survey (IALSS) (Appendix 2) (Murray, et al., 2008). HL maps are clearly compelling evidence of disproportionate distribution of limited HL. Given the high prevalence of limited HL (approximately 60% of Canadian adults), mapping the proportion of limited literacy by community provided a number of significant insights for potential associations with county characteristics (e.g., proportions of immigrants, older adults, unemployed population, and others) as well as a patterned distribution (e.g., the higher proportion of limited HL in the eastern regions of Canada compared to western regions). The example of the CCL demonstrated the potential of a spatial analytical approach in an ecological study focusing on the relationship between groups of individuals (e.g., school, community, county, states) to HL research.

Interestingly, since a British epidemiologist (Snow) used a spatial analytical approach to understand the cholera epidemic in London during 1854, the benefits of mapping in large scale health studies have been widely recognized (Brownstein, et al., 2006). Goldman (1991) indicates the benefits of mapping in health studies are: 1) untangling the complex relationships between risk factors and diseases; 2) generating hypotheses based on identified geographical patterns; and 3) mining insights out of the data. Visualizing the data in a map format may display hidden associations/correlations between relevant factors. In Snow's case, findings about the residences of cholera patients around a particular public water pump provided a clue about the path of infection. Also, these spatial associations between cholera cases and a water pump are unlikely to be detected simply examining basic statistical information. Mapping is a valuable technique in public health studies and is currently used in major studies of cancer prevention, national disaster evacuation, and disability prevalence (Armstrong, et al., 2007; Cromley & McLafferty, 2002; Holt, et al., 2008; Moss, et al., 2006).

Geographic Information System (GIS) in HL research?

Geographic Information System (GIS) has been commonly utilized to visualize geographically-referenced data in a map format not only in research projects but also in practice. GIS is a computer-based technology to manage, visualize and analyze geographically-referenced data. The GIS approach is suitable for public health study because of its powerful and flexible data management, visualization, map making and spatial analysis functions (Cromley & McLafferty, 2002). As the capability of GIS has dramatically improved in accordance with significant advancements of computer technology over recent decades, spatial analysis or GIS-based analysis is now recognized as one of the most powerful analytical tools in epidemiology and public health. In health research, GIS can be used for disease surveillance purpose (Cromley, 2003); detection of high risk areas/populations (Cossman, et al., 2003; Rushton, et al., 2008); and policy and public health intervention program planning (Holt, et al., 2008; Pickle & Su, 2002; Sui & Holt, 2008). Notably, another strength of GIS-based analysis is to ensure confidentiality for sensitive personal information (e.g., income, medical condition, residential address) by using summary or aggregate data for groups of individuals (Moss, et al., 2006; Rushton, 2003).

The possibility of a GIS-based exploratory spatial analytical approach for HL research provides similar benefits. First, given the reported high prevalence of limited HL (e.g., nearly half of or 90 million American adults in 2003) (Kutner, et al., 2005), an ecological or population-level intervention is needed since in previous public health studies individual-level interventions (e.g., individual behavior change programs) are shown to be insufficient in altering health behaviors such as smoking or disease/injury prevalence in the population (Fos & Fine, 2000; Smedley & Syme, 2000). Also, ecological-level analysis is beneficial from a long-term perspective since the underlying causes of limited HL may be embedded in structural-level factors such as policy, education systems, community environments and healthcare systems (C. E. Ross & Wu, 1996). Understanding the association between HL and structural-level factors is necessary particularly for preventive purposes because potential populations at risk of limited literacy may face system-level problems (e.g., communicating with healthcare providers, misusing of medication, misreading nutrition labels) unless such structural-level factors are fundamentally improved (Ward, 2007). If, for example, the education system in a community is one of the causes of limited HL, more people are expected to have limited HL because a defective education system partially "generates" new health illiterates (Schillinger, Bindman, Wang, Stewart, & Piette, 2004; U.S. Department of Health and Human Services, 2000).

Second, the examination of geographical distribution of limited HL is important because HL is a set of context-specific skills and is anticipated to vary by region due to the differences in local healthcare systems, health problems, culture and demographics (Friedman, et al., 2009; Nutbeam, 2008; Paasche-Orlow & Wolf, 2007). Indeed, geographical differences in limited HL prevalence and its association with health outcomes including self-rated health and understanding prescription drug labels have been reported (Baker, 1997; Davis, et al., 2006; Gazmararian, et al., 1999; Willms & Human Resources Development Canada, 1999). Schillinger and Davis (2005) explain geographical differences in heath literacy through associations with disproportionately distributed disparities including income level, education system, and proportions of HL disadvantaged sub-population (e.g., immigrants, racial minorities, older adults). Further, ecological analysis of areas with high prevalence of limited literacy may inform future study and policy planning by generating hypotheses or possible etiological explanations (DeWalt & Pignone, 2005; Lee, et al., 2004).

Finally, the functions of GIS such as data visualization and spatial statistical analysis can inform future HL research and design of intervention programs (Pickle & Su, 2002). As the CCL demonstrated, geographical distribution of limited HL (See Appendix 3-1) visualized on maps helps finding clues about plausible explanations of risk factors and existing health disparities (Murray, et al., 2008). Particularly, a GIS approach can simultaneously visually examine large geographical areas unlike individual-level approaches. In addition, in this respect, visualized data on maps help target prioritizing areas for intervention. Therefore, such documentation and spatial analysis should be an immediate research concern. To sum up, the existing evidence suggests use of GIS in HL research because insights of geographical distribution would inform future HL research and large-scale interventions. Specifically, a GIS approach can provide information about regional characteristics (or possible structural-level causes of limited HL), target areas/populations with the high prevalence and easily interpreted visualized data on maps. In other words, GIS-based approach is a sensible initial step to advance HL research by providing the information about where the problems are and who are at risk of limited HL.

Research questions

Limited HL is an ongoing public health issue where many individuals face difficulties in communication, learning and decision making across a wide range of health settings. To address existing health disparities in the U.S., ecological- or population-level understanding about HL is necessary because an individual-level approach is less effective at improving the population health (Smedley & Syme, 2000). Insights regarding geographical distributions of limited HL at an ecological-level provide extremely useful information for locating high-risk areas and seeking clues of etiological associations with health outcomes. This study aims to document the geographical patterns of limited HL as well as visualized data in map format for future research and public health intervention programs to address health disparity issues in the U.S. This study addresses the following research questions:

1) What are the geographical patterns in the prevalence of limited HL at a county-level in the U.S.?

In consideration of the nation-wide prevalence of limited HL in the U.S., it is critically important to understand the patterns of limited HL distribution because it provides clues for targeting intervention locations and suggests areas/regions for further investigation. Additionally, such documentation can provide useful information for untangling complex associations between limited HL and relevant factors including health outcomes, social determinants of health, socio-demographic characteristics and SES characteristics. Moreover, if geographical patterns exist, any future HL study examining large geographical areas or populations will need to take this into account. In particular, failure to account for geographical patterns or spatial dependency could generate possible bias in statistical analysis. For instance, a linear regression model assumption would be violated if the error term was spatially autocorrelated or a function of distance between areas of inquiry (e.g., neighborhood counties have similar limited HL prevalence).

2) Where are the areas at high risk of limited HL prevalence in the U.S.?

Identifying approximate locations of areas with high prevalence of limited literacy compared to other areas suggests reasonable target areas for future research and interventions. From a research standpoint, examining/comparing areas with high limited HL rates and those with low rates is a rational approach for better understanding possible etiological mechanism. From a practice standpoint, precisely locating problematic areas or populations enables the public sector to utilize available resources more efficiently to address limited HL prevalence, and in turn, health disparities in the U.S.

Methods

Data

Data were derived from the U.S. census TIGER data (map data) (U.S. Census Bureau, 2008), and the 2003 National Assessment of Adult Literacy (NAAL) (White & Dillow, 2005). NAAL is a nationally representative assessment of English literacy among American adults aged 16 and older in households and prisons, which was conducted by the National Center for Education Statistics (NCES) in the U.S. Department of Education (White & Dillow, 2005). In this study, the prison population was excluded in order to focus on the general community-dwelling adult population. The NAAL provides statistically estimated percentages of the population with below basic prose literacy skill for every county and county-equivalent areas in the U.S. (Mohadjer et al., 2009). The NAAL 2003 collected a wide range of information including demographic and SES information in addition to the extensive literacy skills assessment from over 18,000 American adults with over-sampled African American and Hispanic sub-samples. The household sample response rate was 62%. The Survey weights for adjusting the NAAL sampling methods are provided in the dataset to generate national/state estimates for all variables.

Literacy skills including prose, document and quantitative literacy are measured using the slightly modified assessment tool from the 1992 National Adult Literacy Survey (NALS). The NALS literacy assessment tool was developed by the Educational Testing Service (ETS) (Kirsch, et al., 2002; Rudd, et al., 2004). In the NAAL, each literacy domain is defined as follows (Kutner, et al., 2006, p. 2):

- *Prose literacy:* "The knowledge and skills needed to perform prose tasks (i.e., to search, comprehend, and use information from continuous texts." Some examples are editorials, news stories, brochures, and instructional materials.
- *Document literacy:* "The knowledge and skills needed to perform document tasks (i.e., to search comprehend, and use information from noncontinuous texts in various formats." Some examples are job application, payroll forms, transportation schedule, maps, tables, and drug and food labels.
- *Quantitative literacy*: "The knowledge and skills required to perform quantitative tasks (i.e., to identify and perform computation, either alone or sequentially, using numbers embedded in printed materials)." Some examples are balancing a checkbook, figuring out a tip, completing an order form, and determining the amount of interest on a loan from an advertisement.
- *HL:* the HL skills score is derived from 28 health-related literacy tasks (12 prose; 12 document; and 4 quantitative literacy tasks in NAAL). The NAAL HL measure covers clinical, prevention, and health care system navigation aspects of skills for healthcare and services.

The scores (possible range: 0 – 500) for each literacy domain task are summed and classified into four levels to reflect individuals' ability to successfully complete tasks of given difficulties as follows:

- *Below basic* (HL: 0 – 184): no more than the most simple and concrete literacy skills
- *Basic* (HL: 185 -225): skills necessary to perform simple and everyday literacy activities

- *Intermediate* (HL: 226 - 309): skills necessary to perform moderately challenging literacy activities
- *Proficient* (HL: 310 - 500): skills necessary to perform more complex and challenging literacy activities

In an initial stage of this study, the restricted file of the NAAL was obtained to generate county specific estimates of the proportion of the U.S. adult population with limited HL. However, the NAAL restricted file does not include necessary variables such as county identifiers (prohibited by law) and raw individual test scores of HL (not collected for all respondents). Therefore, this study used the county-level estimate of limited (below basic level) prose literacy estimated and published by the NCES as a surrogate measure of limited HL. The prose literacy surrogate measure is a reasonable substitute for HL for several reasons. First, 28 survey items to measure HL consist of the items of prose, document and quantitative literacy. In fact, most individuals with inadequate general literacy skills have limited HL (Kutner, et al., 2006; White & Dillow, 2005). Second, the correlation between prose literacy and HL is significantly high ($r = .94$). In addition, HL and prose literacy are both highly correlated with other components of document literacy ($r_{\text{health-document}} = .99$; $r_{\text{prose-document}} = .85$, respectively) and quantitative literacy ($r_{\text{health-quantitative}} = .89$; $r_{\text{prose-quantitative}} = .87$, respectively). Thus, for the objective of the study, which is to detect at-risk areas of limited HL, the limited prose literacy estimate is the best available indicator of limited HL at the county-level.

Analysis

Data management/analyses were done using SAS 9.2 (SAS Institute Inc, 2000) and ArcGIS 9.3 (ESRI, 2009). The objectives of the analyses were twofold. The first objective is identifying geographical patterns (spatial autocorrelation) of limited HL prevalence to address the first research question (What are the geographical patterns in the prevalence of limited HL at a county-level in the U.S.?). The global Moran's I (described below) was computed to examine spatial autocorrelations or spatial patterns that are either clustered or dispersed distribution of the values (Waller & Gotway, 2004). The global Moran's I is widely used to test geography's first law, which states near things are more related than distant things (Tobler, 1970, p. 3). Therefore, the result of Moran's I indicates the geographical patterns (e.g., clustered, random or dispersed) of the limited literacy (prose literacy in this study) across the U.S.

$$\text{Moran's I} = \frac{n \sum_i \sum_j w_{ij} \, (x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j w_{ij} \, (x_i - \bar{x})^2} \qquad [\textit{Equation 3-1}]$$

Let n be the number of counties, x_i and x_j be the limited prose literacy (the number or proportion in this study) prevalence in county $_i$ and county $_j$, and w_{ij} be the binary spatial weight matrix [(1) if a county is one of the three nearest neighbors of x_i; (0) if not]. Every county $_i$ is examined for all pairs (3 nearest neighboring counties) and the sum of the results is computed (Mitchell, 2005; Waller & Gotway, 2004). Moran's I could range from -1 to 1. On one hand, if all neighboring counties had more similar values (clustered), the Moran's I coefficient becomes closer to 1. On the other hand, if the neighboring counties had more dissimilar/diverse values (dispersed), the Moran's I coefficient becomes closer to -1. If the spatial distribution is completely random, the Moran's I coefficient is 0. Also, the Z score is calculated using the difference between observed value and expected value (random spatial distribution), and the standard deviation of expected values.

The second objective is identifying the areas with high prevalence of limited prose literacy and those with low prevalence of limited prose literacy to address the second research question. Often, the local version of Moran's I called the Local Indicator of Spatial Association (LISA) is used to detect the local areas with similar values (Pfeiffer, et al., 2008). However, LISA does not show if the identified clusters of similar values are high or low but only similar or dissimilar. Therefore, in this study, the Getis and Ord Gi* (G-i-Star) statistic or hot/cold spot analysis (described below) was used because it detects areas where the significantly high (hot spot) or low (cold spot) prevalence of limited literacy is located (Getis & Ord, 1996).

$$\text{Getis} - \text{Ord Gi} * (d) = \frac{\sum_j w_{ij}(d) x_j}{\sum_j x_j} \qquad [\textit{Equation 3-2}]$$

Gi* (d) is the value of Getis-Ord Gi* for county $_i$ at a distance (d) between county $_i$ and neighboring county $_j$ (the distance between centroids of two counties $_i$ and $_j$ unless differently specified). Let w_{ij} be the binary spatial weight matrix [(1) if a county is one of the three nearest neighbors of x_i; (0) if not] for county $_i$ and its neighboring county $_j$ (Mitchell, 2005; Pfeiffer, et al., 2008). The value of each neighboring county (x_j) is multiplied by the weight and the sum of results is computed. The sum of the value of the neighboring counties is divided by the sum of values for all counties in the dataset. If the Gi* value is 0, there is not concentration or a hot/cold spot. Also, the Z score is calculated using the difference between the observed Gi* value and the expected Gi* value (the random spatial distribution), and the standard deviation of the observed Gi* value. The lower Z score indicates the existence of clusters with lower prevalence (cold spot) of limited prose literacy whereas the higher Z score indicates the existence of clusters of with higher prevalence (hot spot).

One of the decisions to make in an exploratory spatial inquiry such as Moran's I and Getis-Ord Gi* statistic is to determine an appropriate search field, which depends on theories about appropriate unit of analysis for a particular problem, and/or the size, type and distribution of spatial data (e.g., entire U.S. counties, only in the school district). That is, when comparing a county to its neighboring counties, how far or how many neighboring counties should be considered as "neighbors?" In most spatial inquiries, there is little theory guiding an appropriate field of search (Chi & Zhu, 2008). In such cases, the combination of iterative process (e.g., applying multiple definitions of neighboring counties) and a data-driven approach (e.g., the field of search with the highest Moran's I coefficient is chosen) is recommended (Anselin, 2002; Voss & Chi, 2006). In HL research, there is no theory or empirical evidence, which guides the definition of neighbors). Therefore, in this study, given the characteristics (e.g., county populations), spatial relationship, area and shape of counties are significantly diverse, the distance-based determination of neighboring counties is inappropriate, and therefore, a *k*-nearest neighbors approach was taken (Anselin, 2002; Waller & Gotway, 2004). A number of different definitions of neighboring counties considering the Brewer's mapping principal, from 3 to 15 nearest counties were examined in the preliminary analyses to compute Moran's I and Getis-Ord Gi* statistic. As a result, the 3-nearest neighbors showed the highest value (the greatest spatial-autocorrelation) both in Moran's I and Getis-Ord Gi* statistic. Thus, this study adopted 3-nearest neighbors as a criterion in spatial inquiries.

Map creation

Based on the results from spatial analyses, the data were displayed in a map format for further visual examinations and documentation to inform future intervention/education programs for limited HL (using the surrogate measure of limited prose literacy) prevalence in the U.S. Geovisualization technique is used to explore spatial distributions on a map (MacEachren & Kraak, 2001). It is a useful approach to understand what the data contain, to locate important research areas, and to generate hypotheses for future research (Demsar, 2009). In this study, the geovisualization technique was used to visualize the spatial distribution of limited prose literacy prevalence as well as the results of hot/cold spot analyses on maps. Given the sophistication of ArcGIS 9.3 software, it is possible for a researcher to be misled depending on the use of color coding, symbols, number of classifications and areal unit. Thus, in this study, geovisualization was done according to the Brewer's basic GIS mapping principal (Brewer, 2006). According to Brewer, the number of bands or categories of variables (e.g., high prevalence: 5 to low prevalence: 1) is less than or equal to 6, and the color selection is done considering possible readers' color blindness. Also, the classifications based on the quintile of the given variables including the number as well as proportion of county adult populations with limited prose literacy for color coding were used because it is likely to be the best approach in an exploratory spatial analysis when there is little theory guiding specific classifications (Chi & Zhu, 2008). In short, the geovisualization technique was used to interpret and to further explore the results of spatial statistical analyses.

Often, a map smoothing technique such as synthetic method (Jia, Muennig, & Borawski, 2004) and head-banging method (Jia et al., 2006) is used to emphasize the geographical patterns or make a spatial distribution more visually recognizable. In this study, map smoothing was not conducted because the limited prose literacy estimate is already statistically computed with known prediction errors that are reported in the NAAL technical report (U.S. Department of Education & National Center for Education Statistics, 2009). In addition, in case a single county or a small cluster of counties with high limited prose literacy prevalence are surrounded by those with low prevalence, such areas may not be visually observable due to over-smoothing. Even a small cluster of high limited HL prevalence is important information given the objectives of this study.

Results

There are 3,143 counties and county equivalent areas in the U.S. The counties with necessary data available (n = 3,141) were included in the descriptive statistics but only the contiguous U.S. excluding Alaska, Hawaii and other distant islands/territories was analyzed for hot/cold spot analyses because of discontinuity of geographical locations. In other words, Getis-Ord Gi* statistic examines relationships between a county and its neighboring counties (either based on the distance or number of neighboring counties). Therefore, for example, including Hawaii which is substantially distant from the contiguous U.S. may result in biased results. The mean county adult population was 69,726 (SD = 224,962; range 51 – 7,206,479). The mean number of county adult population with limited prose literacy was 10,208 (SD = 58,833; range 5 – 2,413,836). The mean proportion of county adult population with limited prose literacy was 12.88% (SD = 6.17; range 3.92% - 64.40%).

With regards to the first research question about the geographical patterns of limited prose literacy, both the number (Moran's I coefficient = 0.316; $p < 0.001$) and proportion (Moran's I coefficient = 0.695; $p < 0.001$) of county adult population with limited prose literacy were highly clustered. In other words, the counties where the prevalence of limited prose literacy is either high or low were likely to be close to each other. Therefore, in the U.S., the number (See Figure 3-1) and proportion (See Figure 3-2) of county adult population with limited prose literacy are geographically patterned in such a way that neighboring areas are likely to be similar, and distant areas (from a focused area) are likely to be dissimilar.

With regards to the second research question about identifying at-risk areas, the results of hot/cold spot analysis for the number of county adult populations with limited prose literacy (Z scores of the Getis-Ord Gi* > 2.58; $p < 0.001$). The counties or county equivalent areas with limited prose literacy are in most of the highly populous areas such as the southern part of California, west part of Arizona, 2 areas including major cities (Dallas and Houston) and a few counties in the southern area in Texas, southern part of Florida, Chicago area in Illinois, Boston area in Massachusetts and Western part of New York (See Figure 3-3). However, the proportion of county adult population with limited prose literacy showed appreciably different at-risk areas. Generally, counties in the southern areas of the U.S. were more likely to have clustered areas with high proportion of limited prose literacy (See Figure 3-4). Particularly, the clusters of counties adjacent to the U.S. – Mexico borders consistently had the higher proportions (Getis-Ord Gi* Z scores > 2.58; $p < 0.001$; See Figure 3-4). Also, a few large clusters with high proportions across states of North Carolina, South Carolina, Georgia, Alabama and Mississippi were observed. Moreover, there was a comparatively large clusters of counties with high proportions in the southern part of Florida.

In the northern U.S., there were a number of clusters with lower proportions of limited prose literacy mainly from central to eastern states. Particularly, a number of statistically significant clusters with lower proportions were observed in the states of North Dakota, South Dakota, Minnesota, Iowa, Nebraska, Kansas and Wisconsin (Getis-Ord Gi* Z scores < -2.58; $p < 0.001$; See Figure 3-4). Also, several clusters of counties with lower proportion of limited prose literacy were present in the states of Michigan, Illinois and Indiana. The selected characteristics of counties in the hot spots, neutral and cold spots are reported in Table Appendix 6.

Discussion

This study examined publically available information from the NAAL 2003 county estimate of limited prose literacy prevalence to identify geographical patterns and to detect at-risk areas (counties) of limited HL in the U.S. The visualized data for the limited prose literacy prevalence is clearly a useful resource for informing any decision making process in a design of intervention program as well as priority area of inquiry in a future research (Demsar, 2009). Generally, identifying geographical patterns is challenging when the data are presented in a table or statistical summary. Use of geovisualization is highly recommended for a research of epidemiology and public health (e.g., disease prevalence, risk factor prevalence).

The geographical patterns of the county adult population and proportion of limited HL (based on the prose literacy as a surrogate measure; HL hereafter) had clear distribution patterns. Most adult populations with limited prose literacy were observed in the counties in highly populated states (e.g., California, Texas, New York, Florida, Illinois). This finding was consistent in the results of hot/cold spot analysis (See Figure 3-3). Such clustering counties of limited prose literacy distributions may be simply a function of population size regardless of known determinants of literacy skills (e.g., age, education) (Kutner, et al., 2006). Therefore, knowing the adult population with limited HL per se may not be sufficient for better understanding the possible mechanism of limited HL prevalence unless combined with the proportion or other covariates. On a related note, the observed county clusters suggest considerations to political boundaries (e.g., county, state) in the limited HL study since individuals' activity areas are not necessarily limited to political boundaries and an intervention to groups of neighboring counties are arguably more desirable instead of that to single county. Nonetheless, the visualized data of such are still beneficial information for planning of resource allocations targeting particular areas to reduce the number of adults with limited HL.

The proportion of limited HL distributions showed a clear distinction between the north and south regions of the U.S. The majority of counties with the high proportion of limited HL were located in the southern part of the U.S. Given the high proportion in large geographical areas, it is unrealistic to provide individual-level intervention programs to address this north-south disparity (Rose, et al., 2008). Therefore, two actions should simultaneously take place.

First, research focusing on the factors influencing the north-south disparity in HL skills needs to be conducted. Although an extensive discussion about possible determinants of HL disparity and in turn health disparity is beyond the scope of this study, structural- (either county or groups of counties) level analysis is a reasonable first step because new health illiterates would constantly increase if some structural level factors such as deficit education systems, lack of adult education opportunity and insufficient effort of public health practitioners and/or health care providers were the cause of limited HL. Second, community- and/or population-level interventions should be implemented to address the north-south disparities. Several studies noted the possibilities of large scale long-term education intervention to the individuals with limited HL skills (Devraj & Gordon, 2009; Kickbusch, 2001; Sentell & Halpin, 2006). Continuous efforts such as long-term adult education intervention to address limited HL prevalence are necessary as HL is a dynamic set of skills that changes over time in accordance with the advancement of science, technology and policy in health/medical fields (Zarcadoolas, et al., 2006).

The implications in this study include easily interpreting visualized maps with identified clustered counties at higher risk. The visualized data provide critical information about the geographical distribution of limited HL problem at the county-level in the U.S., suggest target areas for future intervention programs, and generate additional research questions that are related to limited HL prevalence. The findings from this study can be used for guiding an informed decision in designing intervention programs as well as planning policy changes aiming at the significant reduction of health disparities in the U.S. For instance, targeting the clusters of counties where the prevalence of limited HL is likely to be effective to improve HL skills in the existing adult population and preventing younger individuals from becoming health illiterates.

Additionally, the findings of this study suggest important future research questions. First, the reasons why the geographical patterns exist in the prevalence of limited prose literacy (and in turn, limited HL) need to be investigated. It is unclear whether individuals live in certain geographical areas are more likely to have limited HL skills or individuals with limited heath literacy end up living in particular geographical areas. A longitudinal data analysis considering the movement of populations across the counties is desirable. On a related note, the distribution of limited HL in the contiguous U.S. appears to follow the racial minority distributions at the county-level (Brewer & Suchan, 2001). As the hot/cold spot analysis identified in Figure 3-4, the clusters of counties with high limited HL prevalence in the southern areas visually accurately follow the Hispanics (eastern parts: e.g., California, Arizona, New Mexico, Texas) and non-Hispanic black populations (western parts: e.g., North Carolina, South Carolina, Georgia, Alabama, Mississippi). An exploration of the association between race (e.g., racial discrimination in education opportunities in north and south regions) and limited HL should be investigated in future research to disentangle complex occurrence mechanisms of limited HL.

Second, further investigations at the local level should be a high priority in future research because the number and proportion of county adult population with limited HL were mostly concentrated in populous areas in the U.S. In other words, the greater inequality between individuals with limited HL and those without may exist in populous areas. Based on the findings about the clusters of counties with limited HL prevalence in this study, examining the group of neighboring counties is recommended instead of a comparison between single counties. Third, a study that compares the areas with high prevalence of limited HL and the areas with low prevalence is useful to identify possible root causes of limited HL and in turn health disparity. Although the possible causes (e.g., education system, income inequality, poor access to adult education programs) of limited HL are mostly difficult to modify, a comparative study may provide beneficial insights more for prevention purposes rather than for the existing cases (Nielsen-Bohlman, et al., 2004). Last, while this study focused on the county-level analysis, future studies should examine the limited HL prevalence in different geographical units of analysis, including regions, states and smaller geographical areas (e.g., census tract). Particularly, multi-level analysis may have a potential to better understand factors influencing individual HL at different levels.

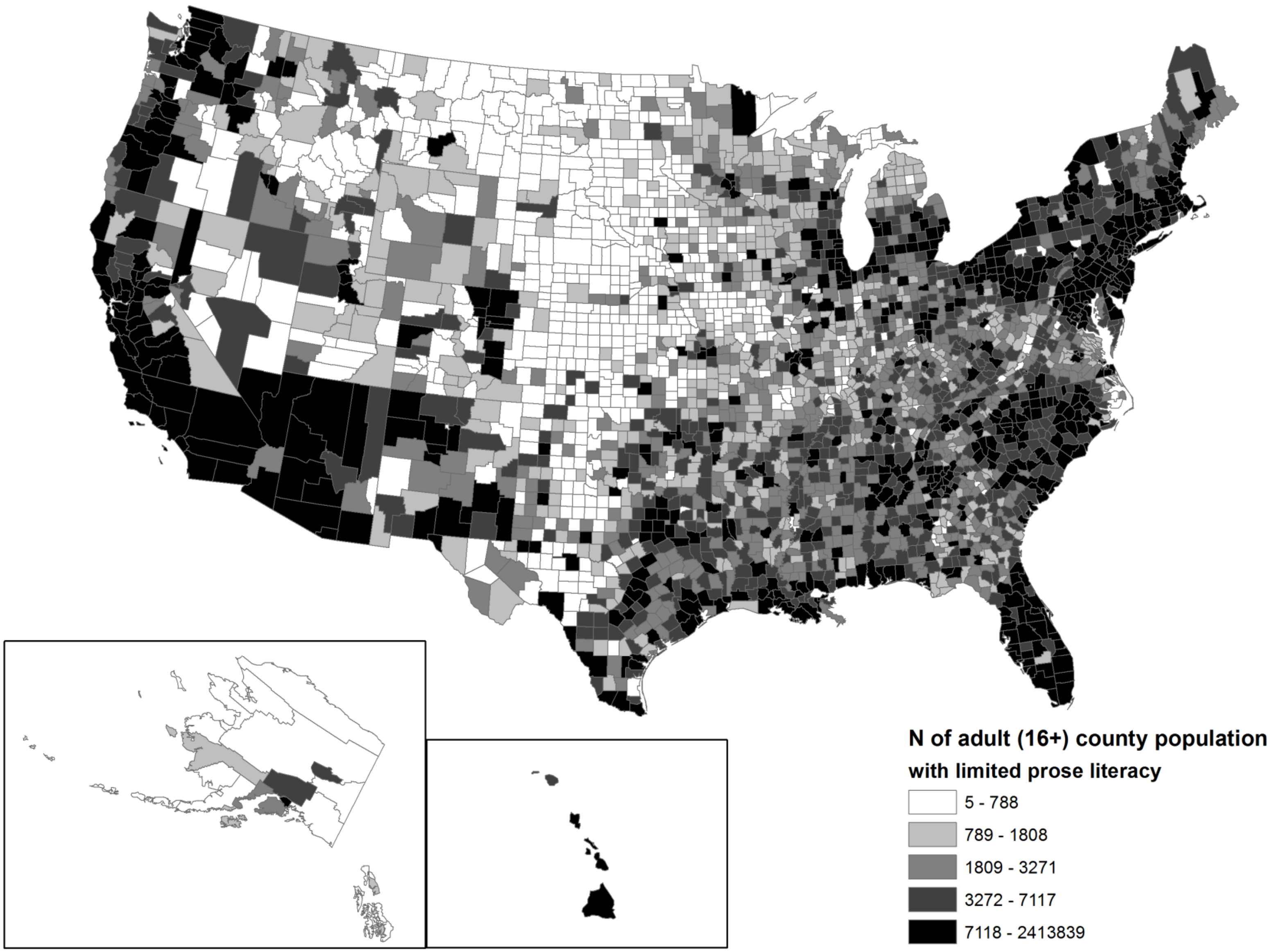
Limitations

There are several potential limitations in this study. First, some data limitations including the use of prose literacy estimate at county-level with known prediction errors are not ideal in a HL study. However, given the information available and the high correlation between prose and HL, the prose literacy estimate is a reasonable (and perhaps the best available) surrogate measure to HL for all U.S. counties. Second, the results of spatial analyses would require caution due to the shape and areas of each county. Some counties are significantly larger than some others. In addition, counties facing coast lines (i.e., edge effect) may have different associations with neighboring counties because of their unique spatial relationships (e.g., facing the ocean).

Although a *k*-nearest neighbors approach was taken to address such issues, there might have been some unobserved influence in spatial analysis. On a related note, the findings from this study are limited to the county-level since use of different analysis unit (e.g., census tracts) may result in different findings (a.k.a., MAUP: modifiable areal unit problem). Issues with potential influence of terrain/unit of analysis are virtually always present in spatial analysis. However, given the countries are most common political units in the U.S., county-level analysis is sensible for generating practical information in HL study from a public health standpoint. Finally, the visualized data on a map format can be comparatively easily manipulated to mislead researchers although this study followed the existing map creation guideline to avoid potential bias (Brewer, 2006).

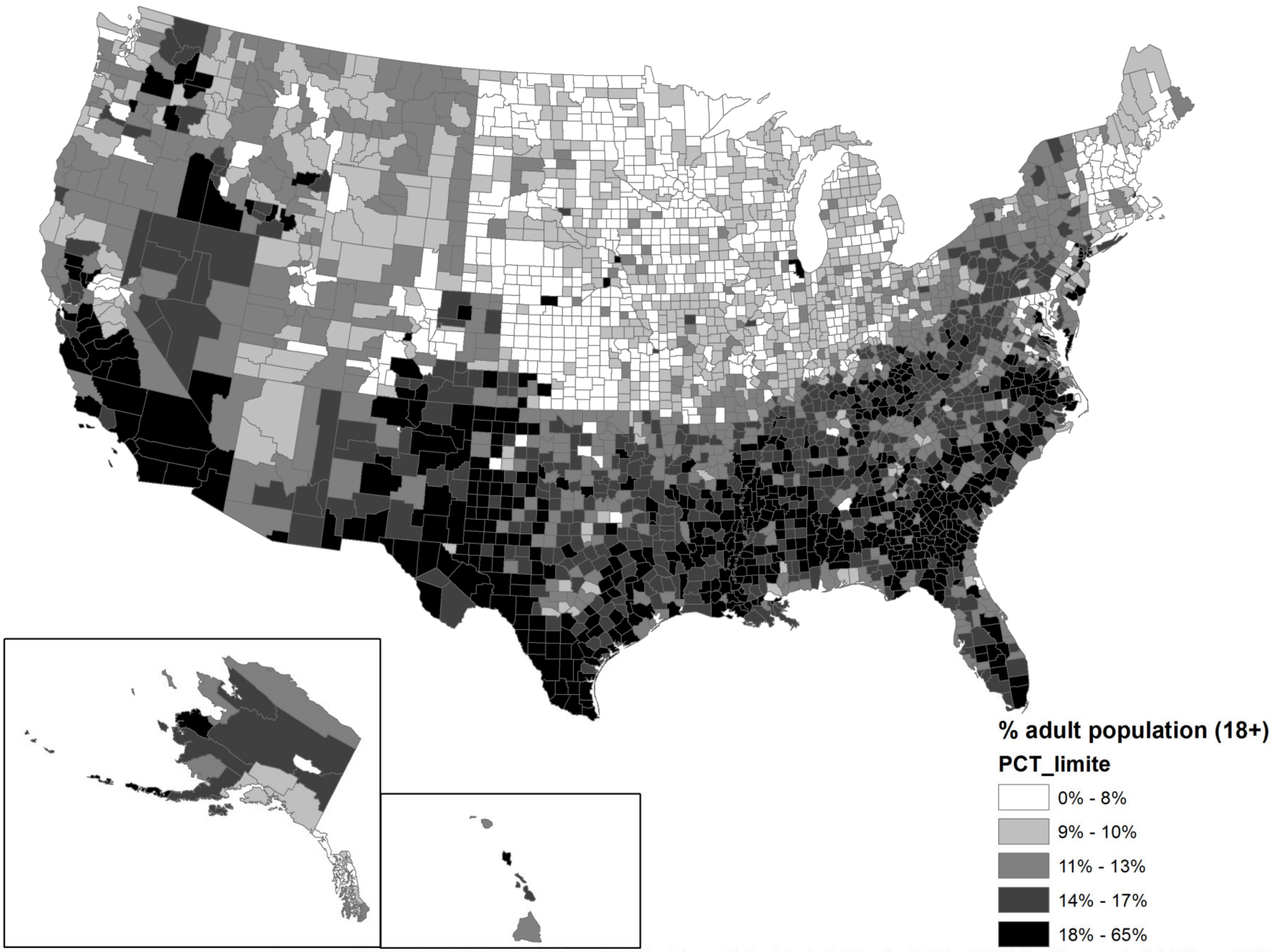
In conclusion, the GIS-based visualized data as well as exploratory spatial analyses detected and confirmed the geographical patterns and high risk areas of limited HL using the surrogate measure of limited prose literacy skills in the U.S. adult populations. The large numbers of American adults with limited HL are likely to reside in the U.S. counties in populated states such as California, New York, Texas and so forth, and their surrounding areas arguably because of the outstandingly greater population sizes compared to other areas/regions. However, the distinctive disparities in the proportion of county adult populations with limited HL were observed between the northern and southern parts of the U.S. That is, most counties with higher prevalence of limited HL were located in the southern parts, particularly around the U.S. – Mexico borders. Such identifications of geographical patterns as well as at-risk areas can inform design of possible intervention programs, decision making in policy planning (e.g., resource allocation) and future health disparity research focusing on HL.

Figure 3-1: the estimated number of county adult populations with limited prose literacy skills in the U.S. by the quintile



Data source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy (NAAL)

Figure 3-2: The proportion of county adult populations with limited prose literacy skills in the U.S. by the quintile points



Data source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy (NAAL)

Figure 3-3: Hot/Cold spot (Getis-Ord Gi) analysis for the number of county adult populations with limited prose literacy.*

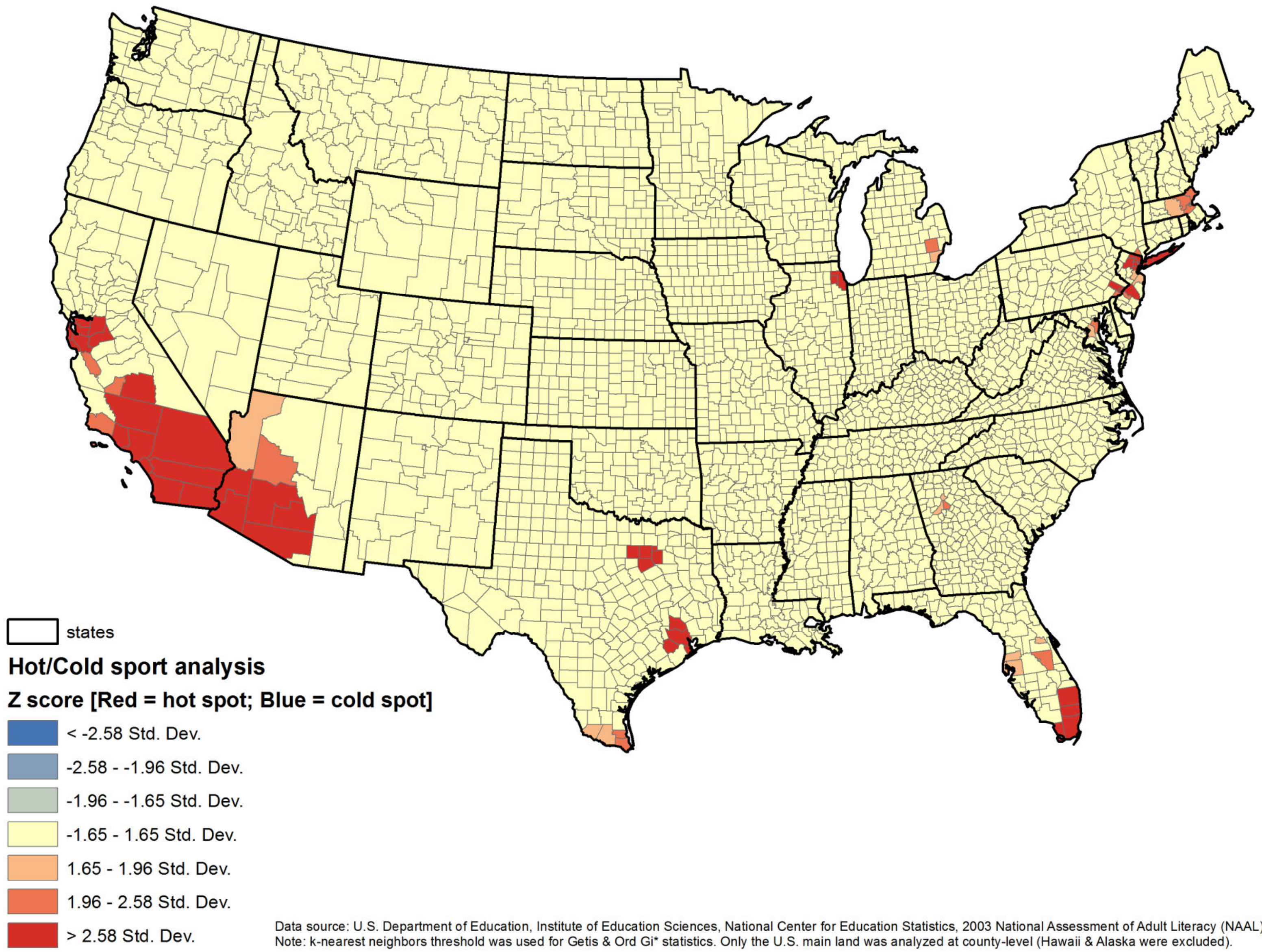
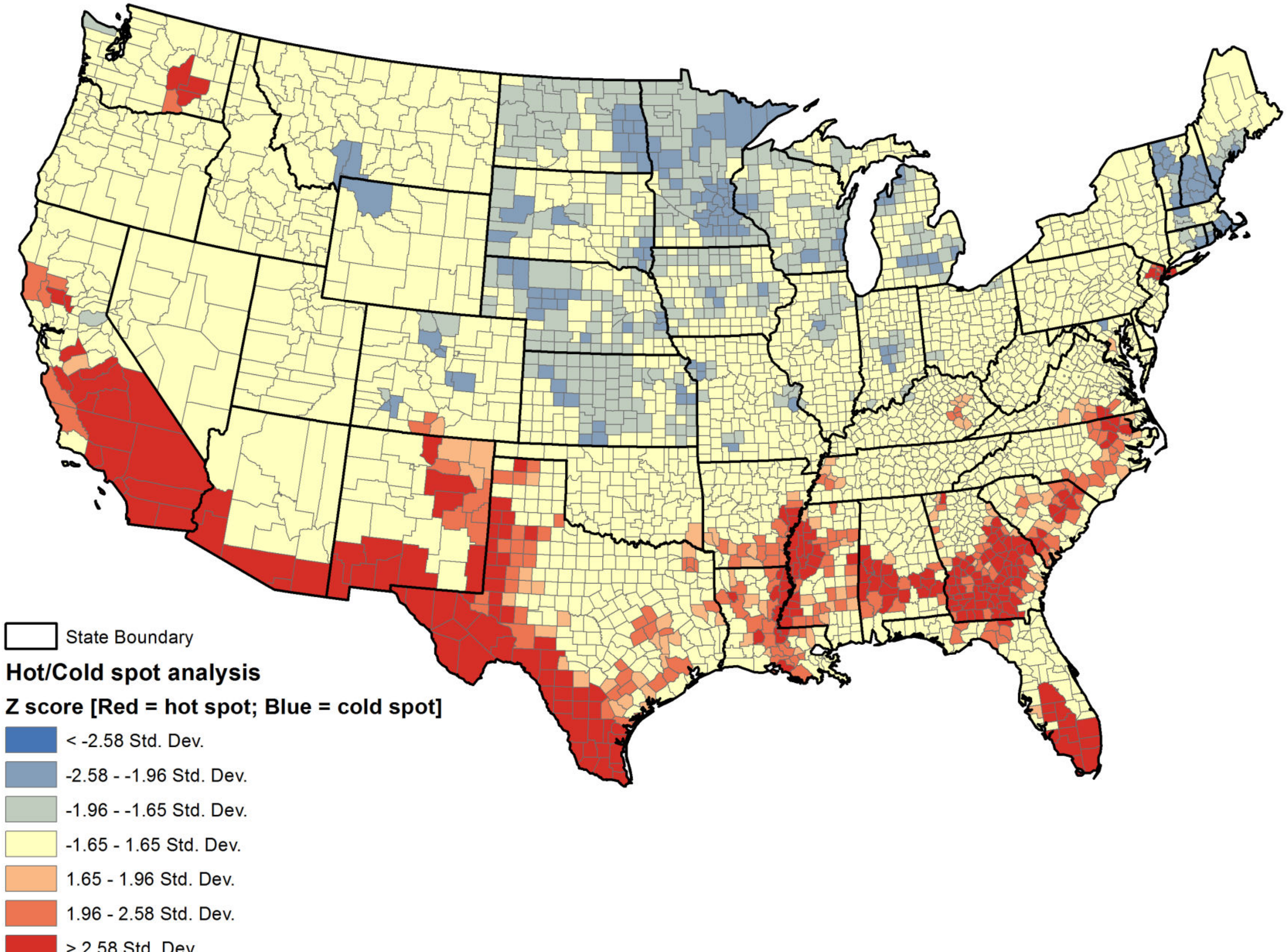


Figure 3-4: Hot/Cold spot (Getis-Ord Gi) analysis for the proportion of county adult populations with limited prose literacy*



Data source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy (NAAL).
Note: k-nearest neighbors threshold was used for Getis & Ord Gi* statistics. Only the U.S. main land was analyzed at county-level (Hawaii & Alaska were excluded).