

Cognitive Assessment of Geospatial Survey Data

David J. Roe, M.A.¹,
Douglas Currivan, Ph.D.¹,
James B. Holt, M.P.A., Ph.D.²,
George Ghneim, DVM, MPVM, Ph.D.¹.

¹ RTI International, ² Centers for Disease Control and Prevention (CDC)

Excerpts from a report to the CDC prepared in support of a poster to be presented at the 2006 Annual Meeting of the American Association of Public Opinion Research, May 18-21, 2006. Montreal, Quebec, Canada.

DRAFT: NOT FOR CITATION.

Abstract

Mapping survey data has become an increasingly important means of displaying survey information and highlighting differences across geographic areas. In public health, geospatial data are used by policy-makers, researchers, and other health professionals to identify and address issues of public health concern. Traditionally, geospatial data are presented in choropleth maps, where defined geographic units (i.e., political boundaries) are filled with a uniform color or pattern. Such maps are appropriate for data that have been scaled or normalized. For example, high school graduation rates displayed by county are typically represented using choropleth maps. Alternatively, geospatial data can be displayed using isopleth maps, in which the data are not aggregated to pre-defined geographic units, but rather are typically “smoothed” across adjacent geographic boundaries. There is little empirical research, however, on the differential effectiveness of choropleth versus isopleth maps. In particular, how do these two different mapping techniques affect the user’s ability to extract information from the map? Using maps of health data collected from the Behavioral Risk Factor Surveillance System (BRFSS), a widely used source of public health information, we present the results from a series of focus groups and cognitive testing sessions which were used to evaluate the way in which public health professionals and epidemiologists commonly use mapped data. Among the issues explored are usability, general preferences, ease of pattern recognition, and rate retrieval for both choropleth and isopleth maps. A clear majority of participants preferred the isopleth format, however, their assessment varied somewhat by the data retrieval task. The color scheme used also had an impact, regardless of the type of map used.

Background

The Behavioral Risk Factor Surveillance System (BRFSS), which was established in 1984 and reached national state-level data collection in 1994, was developed “to collect prevalence data on risk behaviors and preventive health practices that affect health status” (Centers for Disease Control and Prevention [CDC], 2005). BRFSS survey data are collected annually, and results are used for health promotion and disease prevention program planning, development, and evaluation. The BRFSS survey consists of a core interview (i.e., a fixed core, rotating core, and emerging core questions that reflect “late breaking” issues), optional modules, and state-added questions. RTI International conducts a variety of BRFSS activities for the CDC, including survey item appraisal and cognitive testing of items.

Many BRFSS survey data are presented in map form, allowing users to evaluate a variety of topics taking geography into consideration. While literature on best practices or methods for presenting data in map form seems scarce, it is clear that the general concept of utilizing mapping to analyze health data is a good one. Crampton notes that public health analysts focus on the spread and causes of disease not in a laboratory setting, but by combining socio-economic data or observing how social networks are linked to disease transmission (2005).

Turning to the methods used to map data, Crampton notes that there is an apparent single favorite when choosing how to map data, the choropleth map (2005). Choropleth maps, the maps currently used to present BRFSS data, are best defined as maps which present data as geographic areas shaded with intensity proportional to the data values associated with those areas (Holt, 2005).

In a literature review conducted of journal articles published between 2000 and 2004, Martin shows that choropleth maps are by far the single most frequently used method for displaying public health data in maps (2005). There is however a potential danger in the use of such maps according to Crampton (2003). Crampton warns that choropleth maps tend to hide internal variation, especially when presenting data over large areas (2003).

Alternatively, geospatial data can be displayed using isopleth maps, in which the data are not aggregated to pre-defined geographic units, but rather are typically “smoothed” across adjacent geographic boundaries. In other words these data are measured over an area but presumed to occur or be centered on point locations (Holt, 2005). There is little empirical research, however, on the differential effectiveness of choropleth versus isopleth maps. The purpose of the activities summarized in this report is to evaluate the way in which public health professionals and epidemiologists commonly use mapped data. In particular, how do these two different mapping techniques affect the user’s ability to extract information from the map?

For the BRFSS mapping experiment, we conducted (1) a focus group in Research Triangle Park, North Carolina, and (2) two rounds of in-person cognitive interviews, one in Research Triangle Park and one in Atlanta, Georgia. A total of eight public health professionals participated in the focus group, nine participated in the first round of cognitive interviews, and nine participated in the second round of cognitive interviews. Both the focus group and the cognitive interviews involved six different maps of BRFSS data per interview (10 total maps were mixed in each set of cognitive interviews), and we asked four questions per map, for a total of 24 questions for each protocol.

Methodology

Focus Group

Focus groups are small groups of individuals gathered to discuss a particular topic under the direction of a moderator. The moderator promotes interaction and ensures that the discussion remains on the topic of interest. The focus group produced qualitative data on public health professionals' reactions to and perceptions of the BRFSS maps and questions. These results are solicited through semi structured group discussions during which the participants influenced one another—just as they do in real life.

The focus group protocol was designed to elicit reactions from public health professionals and epidemiologists on the different BRFSS maps and the questions asked about the maps. We expected the focus group discussion to contribute to the mapping experiment in two main ways (Fowler, 1995):

1. To help examine assumptions about the mapped data and map questions that public health professionals and epidemiologists would be asked to use.
2. To evaluate assumptions about vocabulary, the way public health professionals and epidemiologists understand key terms or concepts that will be used in the mapping experiment.

Cognitive Testing

Cognitive interviewing techniques have come into common use as a way to assess a range of problems with survey questions, including those related to comprehension, retrieval, decision, and response processes used by participants in answering questions. The most frequently used model for cognitive interviewing (Tourangeau, 1984) has four major features: (1) question comprehension, (2) retrieval of relevant information from memory, (3) decision processes, and (4) response processes. Cognitive interviewing is used to explore a person's decision-making processes in each of these areas. Willis (1999) summarized how cognitive interviewing can be used to explore these four aspects of question responding:

1. Question comprehension—studies both question intent (what does the participant believe the question to be asking?) and the specific meaning of terms in the question.
2. Retrieval from memory—examines “recallability” of information and the strategies used to retrieve that information (i.e., estimation strategies or counting of individual events).
3. Decision processes—looks at the motivation internal to the participant to thoughtfully provide an accurate response and issues related to desirable responding (or social desirability—the desire to respond in such a way as to make oneself look better, either through intentional deception or unconscious self-deception).
4. Response processes—evaluates the ability of the participant to match his or her estimation (perception, behavior, etc.) with the response options available.

The main cognitive interviewing technique we used to probe participants about each map and specific questions was *verbal probing*. Verbal probing techniques can be either concurrent probing, in which the probes are administered as the participant completes a survey item, or retrospective probing, in which the participant completes the entire survey and is then presented with specific probes about the question, retrieval, decision processes, and response processes. Each approach has advantages and disadvantages. Given that the participants’ task was to answer questions about six maps, our cognitive testing focused on retrospective verbal probing after all questions were answered. In this way we could be more certain that any discussions between the participant and interviewer would not influence the participant’s answers to subsequent questions. The goal was to elicit feedback after each participant had answered all questions about each map. We used in-depth probes to further explore each participant’s reactions to the maps and questions.

Maps and Questions

Two sets of maps were used for the focus group and cognitive interviews, six-color maps and four-color maps. Each set of maps was further divided between choropleth maps and isopleth maps, which differ in the way they present data. Maps were generated at the national level for the entire United States and at the regional level for the southeastern United States. All maps were created by the CDC and provided to RTI prior to each round of testing.

For each stage of testing, 6 of the 10 total maps were included in each protocol. The focus group employed an equal mix of 3 choropleth maps and 3 isopleth maps. For the first round of cognitive testing, 6 six-color maps were used that were also equally split between choropleth maps and isopleth maps. One variation across each cognitive protocol was the mix of national versus regional maps. All participants received at least one pair of national maps and at least one pair of southeastern maps. For the second

round of cognitive testing, 6 four-color maps were used that were split equally between choropleth maps and isopleth maps. Two additional variations were added in the second round of cognitive testing. Choropleth maps included lighter county boundaries, whereas isopleth maps included county outlines not displayed during the first round of cognitive interviews. Because this feature was added only for the second round of cognitive testing, these protocols also included some probing and discussion for comparison, but no specific questions related to these additional features.

All map questions were constructed in a multiple-choice format with four response options per question. For each map we included a mix of questions focused on retrieval of rates, pattern recognition, and on ease of understanding. Two additional questions asked participants (1) whether they preferred the choropleth versus the isopleth map format and (2) whether they found the choropleth versus the isopleth map format easier to read. These two questions were asked separately for the national and the regional maps.

Between each round of testing, RTI reviewed the questions to determine whether revisions were necessary to clarify questions. We made few changes to the questions for the six-color map questions between the focus group and the first round of cognitive testing, but several revisions between the first and second rounds of cognitive testing. The changes between the two rounds of cognitive testing were based on the need to ensure appropriate question wording and answer choices after the shift from the six-color to four-color maps.

Participants

The focus group was comprised of public health professionals and epidemiologists working in the Raleigh/Durham area of North Carolina. Participants were employed by the North Carolina Department of Public Health and various academic departments at area universities and colleges. We recruited a total of nine public health professionals and epidemiologists for the focus group, but only eight were able to participate.

For the first round of cognitive testing in Atlanta, we recruited a group of public health professionals and epidemiologists working in the Atlanta area. Participants included public health professionals at the Georgia Department of Public Health, public health professionals and researchers at Emory University, and public health researchers with the CDC. We recruited and interviewed a total of nine participants for the first round of cognitive testing.

The second round of cognitive testing included public health professionals and epidemiologists in the Raleigh/Durham area. This group was very similar to those who participated in the earlier focus group. We recruited and interviewed a total of nine participants for the second round of cognitive testing.

Conclusions

Incorporating the results of the focus group and the two rounds of cognitive testing provides several important conclusions.

- A clear majority of participants preferred the isopleth format over the choropleth format and found the isopleth format easier to read. Results and comments indicate that this was particularly true when looking at the U.S. maps, but the majority chose the isopleth format in reference to both the U.S. and Southeast maps.
- Overall, retrieving data from U.S. maps was more difficult than retrieving data from Southeast regional maps when the maps were in choropleth format, six-color format, or both. Retrieving data from choropleth U.S. maps was especially difficult in the eastern two-thirds of the country, due to the smaller average size of states and the greater average number of counties in these states.
- Although participants generally preferred the isopleth maps, their assessment varied somewhat by the data retrieval task. For pattern recognition, participants clearly preferred the isopleth maps across all formats. For rate retrieval, participants found the choropleth maps as useful as or more useful than the isopleth maps in some situations, such as discerning county rates from a Southeast regional map or a large state on a U.S. map.
- Participants had greater difficulty distinguishing color categories with the six-color maps. This was particularly true for the middle categories as opposed to the lowest and highest categories. Participants had less difficulty distinguishing the colors with the four-color maps but did experience some difficulty distinguishing the two middle categories on these maps as well.
- On choropleth maps, discerning county boundaries was more difficult for respondents with the darker boundary outlines compared with the lighter boundaries. This was especially true when retrieving data from smaller states, states with relatively many counties, along state borders, or along coastal areas. Darker boundaries were also more difficult to discern when the color categories of adjacent counties were at the darker end of the color ramp.

Limitations

Results from this study are not intended to generalize to a larger population but rather reflect how typical participants understand and respond to the proposed maps and questions. Also, although these groups of participants included a mix of age, gender, ethnicity, and professional background, it is difficult to generalize the current cognitive interviewing with so few participants to the entire target population of BRFSS map users. We believe that the focus group and cognitive testing identified the major problems with the current four- and six-color maps but acknowledge that our results reflect only the

opinions and experiences of the 26 individuals who participated in these three rounds of testing.

Next Steps

The results of cognitive interviewing demonstrated that respondents had considerable difficulty differentiating color categories based on the orange color ramp, especially when using the six-color maps. These results supported further investigation and testing of other color ramps that provide a greater contrast between classes in a given ramp. Also, after seeing the difficulties respondents had with the dark boundary choropleth maps during the focus group and cognitive tests, we concluded both choropleth and isopleth maps should include light county boundaries. The two new maps below show the same general health data as the first two maps to the left, with a new divergent color ramp, which goes from red to blue. The new map format, designed to improve user accuracy in retrieving data from choropleth and isopleth maps, will be tested through an internet survey of approximately 400 public health researchers and professionals. Data from the internet survey will facilitate analysis of key issues such as users' accuracy of data retrieval from choropleth versus isopleth maps, differences in data retrieval from national versus regional maps and users' preferences for choropleth versus isopleth maps. These results will allow the CDC to determine the best format and features for providing online maps for BRFSS data users.

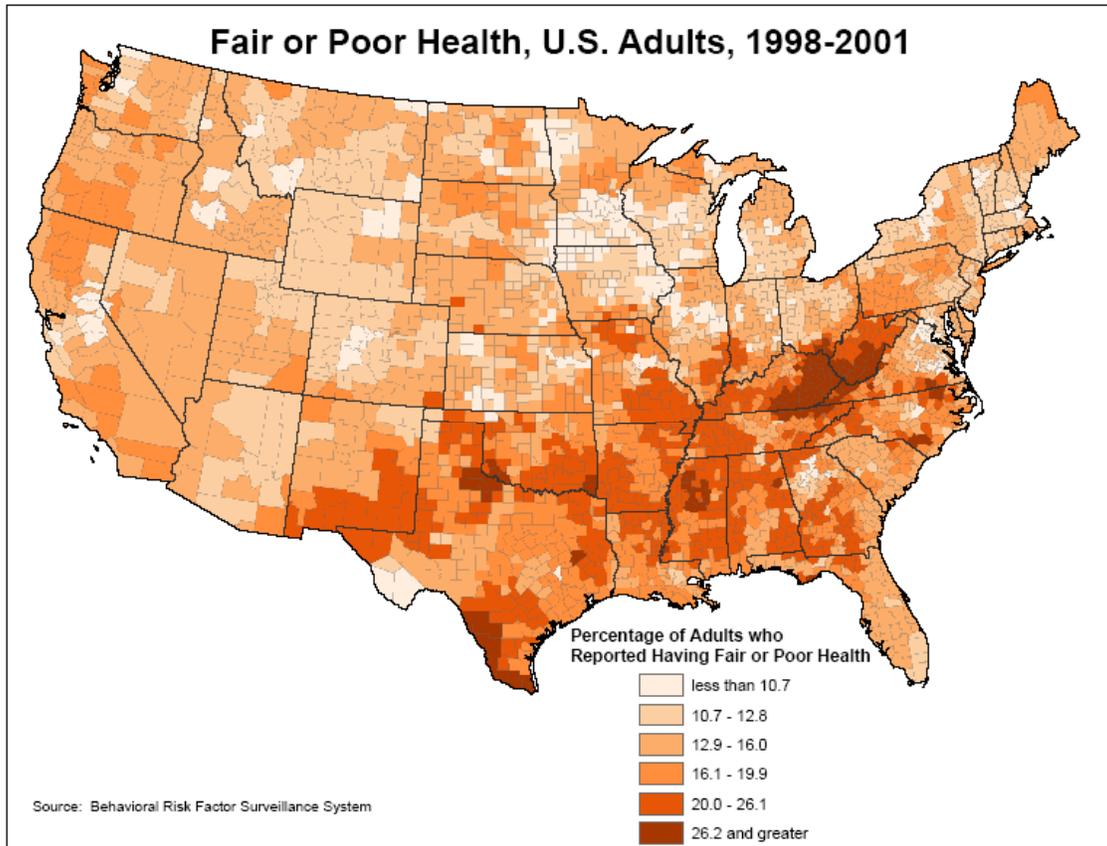
References

- Centers for Disease Control and Prevention (CDC) (2005). *Behavioral Risk Factors Surveillance System operational and user's guide version 3.0*. Atlanta, GA: National Center for Chronic Disease Prevention and Health Promotion, CDC.
- Crampton, J. 2005. *Public Health and GIS: Are We There Yet?*. Geo World. Accessed at <http://www.geospace.com/uploads/FeatureArticle/0507cgis.asp>
- Crampton, J. 2003. *Are Choropleth Maps Good for Geography?*. Geo World. Accessed at <http://www.geospace.com/gw/2003/0301/0301crit.asp>
- Fowler, F.J. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage.
- Holt, J. 2005. *BRFSS Maps: New Developments and Planned Enhancements*. Presentation made at the 2005 BRFSS Conference. Atlanta, GA. March, 2005.
- Martin, S. 2005. "Cartography, Discourse and Disease: How Maps Shape Scientific Knowledge about Disease," Master of Arts Thesis, Georgia State University.
- Tourangeau, R. 1984. "Cognitive Science and Survey Methods." In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (eds.), *Cognitive Aspects of Survey Design: Building a Bridge between Disciplines*, pp. 73-100. Washington, DC: National Academy Press.
- Willis, G.B. 1999. *Cognitive Interviewing: A "How to" Guide*. Research Triangle Park, NC: Research Triangle Institute.

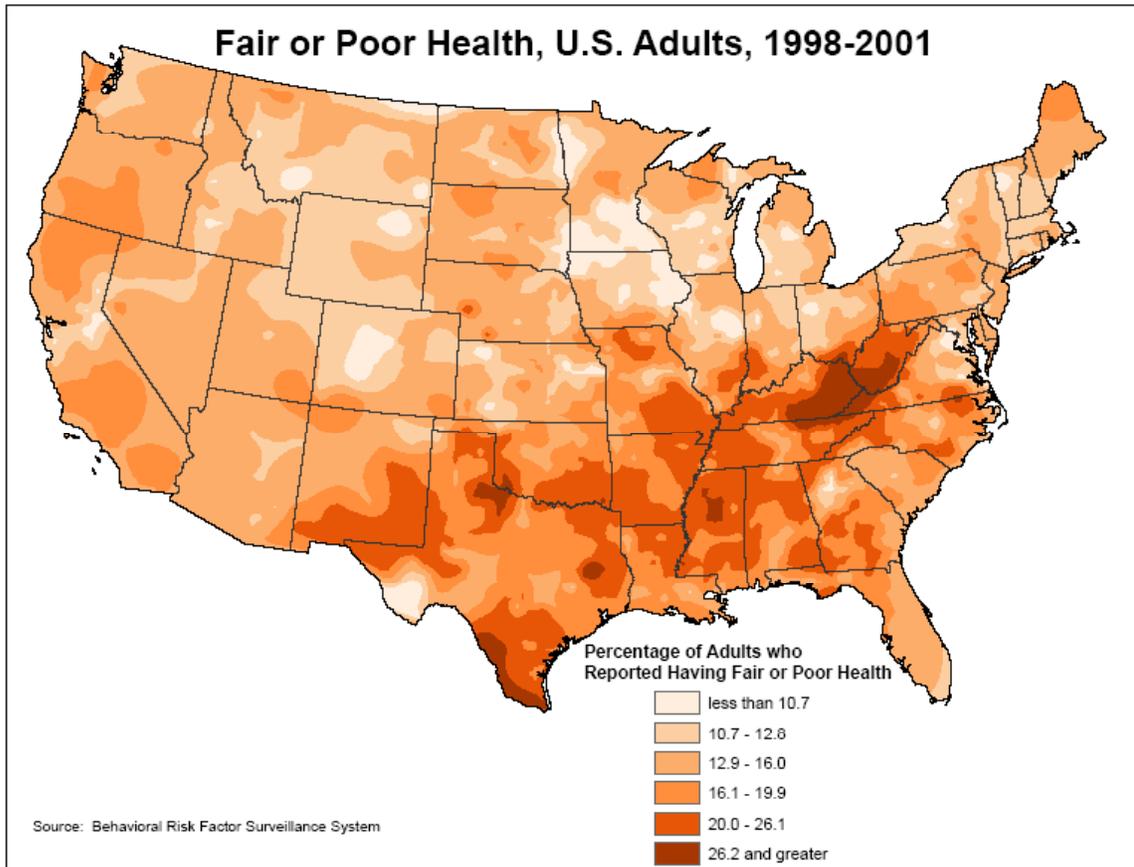
Appendix A

Sample Maps

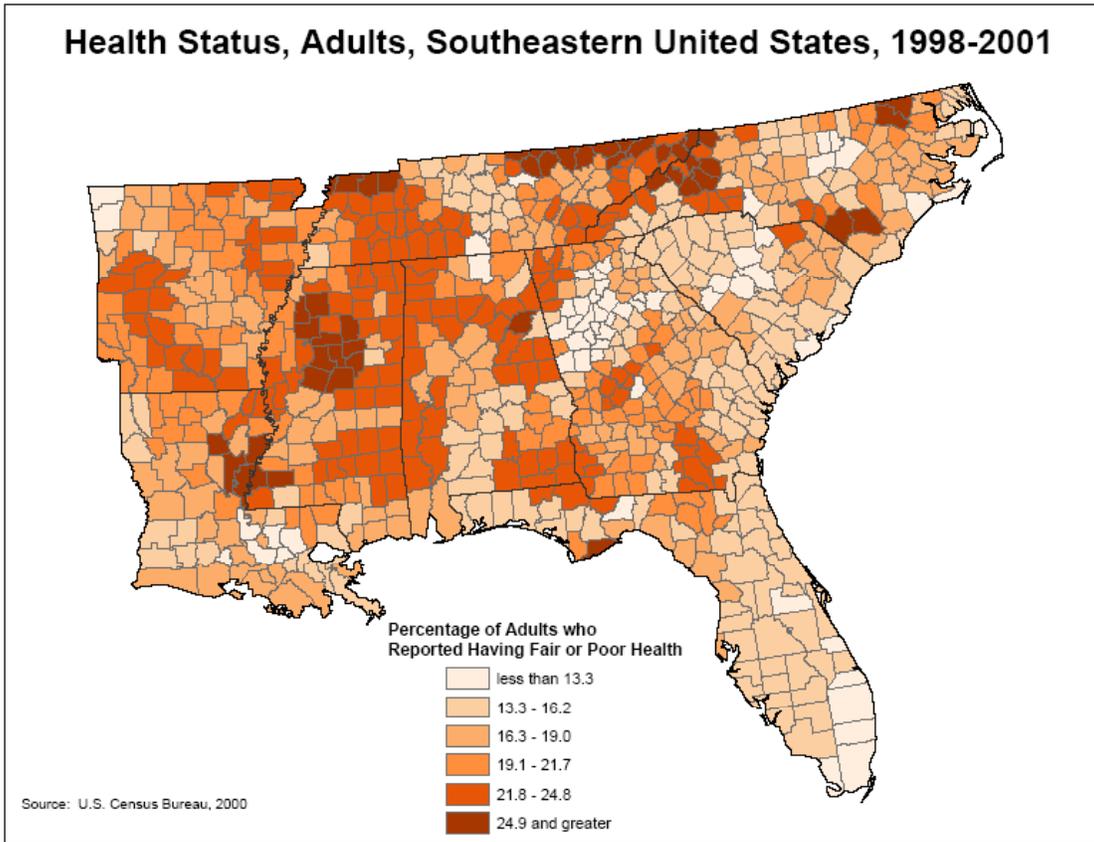
US General Health Choropleth Map



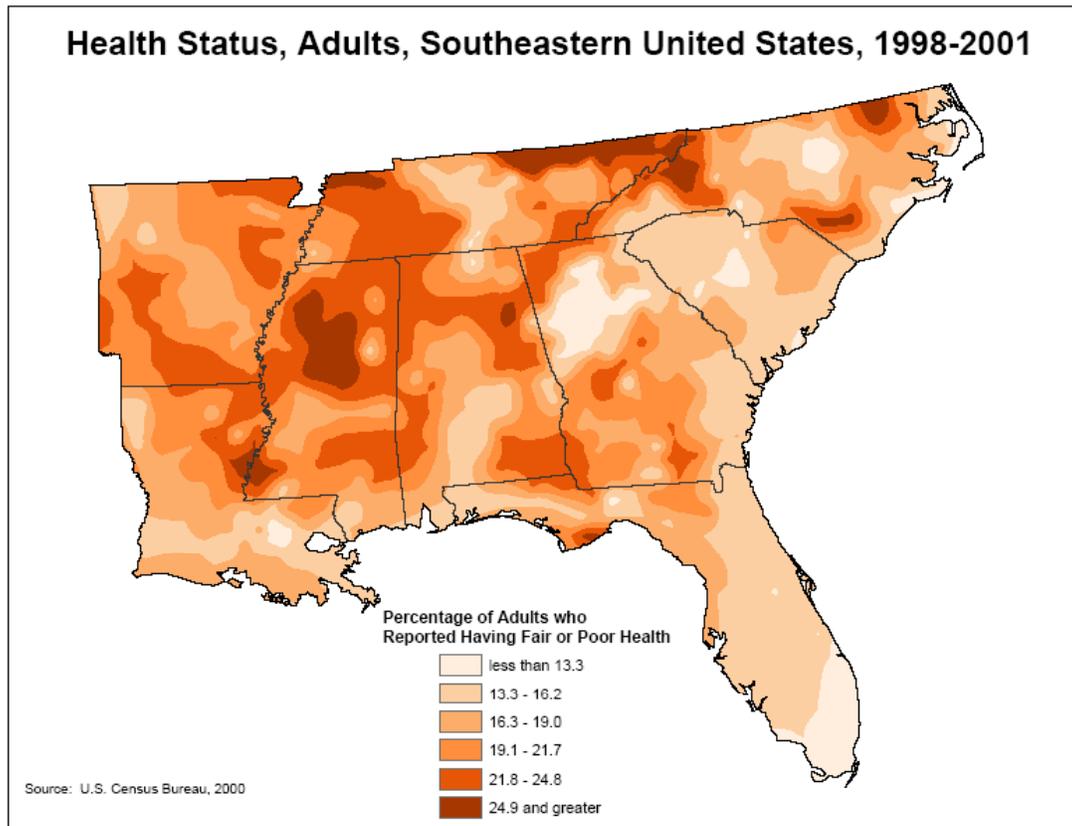
US General Health Isopleth Map



SE General Health Choropleth Map



SE General Health Isopleth Map



Appendix B
Sample Questions

US General Health Choropleth Questions

What is the predominant rate of adults reporting having fair or poor health in Nevada?

- a. Less than 13.6%
- b. 13.6%-18.4%
- c. 18.5%-24.9%
- d. Cannot assess

Which state has the largest area where less than 13.6 percent of adults report having fair or poor health?

- a. Arizona
- b. Wyoming
- c. Montana
- d. Iowa

Which two states have the largest adjacent area where 25.0 percent or more of adults report having fair or poor health?

- a. Kentucky and West Virginia
- b. Kentucky and Virginia
- c. Kentucky and Tennessee
- d. Kentucky and Ohio

How many states have at least one county where 25.0 percent of the population or greater reports having fair or poor health.

- a. 12
- b. 13
- c. 14
- d. 15

US General Health Isopleth Questions

Which 3 states have an area where there is a clear gradient (a stepwise increase) involving all 4 of the health rates?

- a. Ohio, North Carolina, and Virginia
- b. Ohio, North Carolina, and Kentucky
- c. Ohio, North Carolina, and Georgia
- d. Ohio, Kentucky, and Virginia

How many states have at least one county where 25.0 percent of the population or greater reports having fair or poor health.

- a. 12
- b. 13
- c. 14
- d. 15

Which two states have the largest adjacent area where less than 13.6 percent of adults report having fair or poor health?

- a. Minnesota and Iowa
- b. Montana and Wyoming
- c. Montana and Idaho
- d. Colorado and Utah

What is the predominant rate of adults reporting having fair or poor health in Washington state?

- a. Less than 13.6%
- b. 13.6%-18.4%
- c. 18.5%-24.9%
- d. Cannot assess

Southeast General Health Choropleth Questions

Which of these states has the largest number of counties along a border where 24.0% or more adults report having fair or poor health?

- a. Louisiana
- b. Mississippi
- c. North Carolina
- d. Tennessee

What is the predominant rate of adults reporting fair or poor health in Alabama along its border with Mississippi?

- a. Less than 15.1%
- b. 15.1%-19.2%
- c. 19.3%-23.9%
- d. Cannot assess

Which of these states has the largest number of counties where less than 15.1% of adults report having fair or poor health?

- a. Florida
- b. Georgia
- c. Louisiana
- d. South Carolina

Which two Southeast states do not have at least one county where 24.0% or more adults report having fair or poor health?

- a. Alabama and Florida
- b. Alabama and Georgia
- c. Arkansas and Florida
- d. Georgia and South Carolina

Southeast General Health Isopleth Questions

What is the predominant rate of adults with fair or poor health in Florida?

- a. Less than 15.1%
- b. 15.1%-19.2%
- c. 19.3%-23.9%
- d. Cannot assess

Which state has three distinct areas where there is a clear stepwise increase across all four of the categories of adults with fair or poor health?

- a. Alabama
- b. Arkansas
- c. North Carolina
- d. Tennessee

How many Southeast states have at least one area where less than 13.3% of adults report fair or poor health?

- a. 5
- b. 7
- c. 8
- d. 9

Which of these states has the largest number of counties where less than 15.1% of adults report having fair or poor health?

- a. Florida
- b. Georgia
- c. Louisiana
- d. South Carolina