This is to certify that the

thesis entitled

REGIONAL PERCEPTION ON BIVARIATE CHOROPLETH MAPS

presented by

Christopher Richard Steere

has been accepted towards fulfillment
of the requirements for

M.A. degree in Geography

Richard Groop
Major professor

Date October 27, 1995

REGIONAL PERCEPTION ON BIVARIATE CHOROPLETH MAPS

by

Christopher Richard Steere

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Geography

1995

ABSTRACT

REGIONAL PERCEPTION ON BIVARIATE CHOROPLETH MAPS

by

Christopher Richard Steere

Bivariate choropleth maps are used to display the distribution of the relationship between two variables. The spectrally encoded two-variable maps published by the United States Census Bureau are considered to be the best type of bivariate choropleth map. There is little research that supports this assumption. This paper is intended to fill some of this void by comparing the spectrally encoded maps to another type of bivariate choropleth map: the complementary-color two-variable map.

A perception test is employed comparing the relative effectiveness of these two map types in conveying regional distribution information to map readers. Responses are analyzed by the consistency with which subjects defined region boundaries within the map. A standardized variance measure is used to determine the consistency of region location. Results suggest that the complementary-color two variable map is at least as effective as the spectrally encoded bivariate choropleth map in portraying regions.

ACKNOWLEDGEMENTS

I would first like to express my thanks and gratitude to my advisor, Dr. Richard E. Groop, whose guidance and editorial skills were of tremendous help to me. Although I may not have always liked his suggestions, they definitely improved the quality of my work. I would also like to thank the secretaries in the Geography Department. They were always there to answer questions and offer encouragement and I appreciate the help they were able to give me.

Many thanks go to the 72 anonymous subjects who participated in this project. Without them, there wouldn't have been much point to this paper.

To the other graduate students in the department, what can I say. We had some great times together that made life here a lot easier. Thanks for helping me keep my sanity.

Last of all, I would like to thank my wife. Through all of the trying times involved in finishing this paper, she was always there with words of encouragement. Without her love and support, none of this would have been possible.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

One of the most important activities in geographic study and analysis is the portrayal of relationships among spatial patterns. These patterns could be spatially related in any number of different ways including adjacency of regions, coincident occurrences, or regional overlap. Since these relationships are spatial in nature, most geographers would agree that thematic maps provide the best method of representing them (Rogers and Groop, 1981). Because of the complex nature of these relationships, maps displaying them can often be difficult to produce and difficult to interpret.

Mapping Spatial Relationships

Since there are many different ways that patterns can be related, there are also many ways to illustrate these spatial relationships. The simplest of these methods is used when portraying a single areal variable such as vegetation cover. The map contains clear and distinct regional boundaries and is therefore relatively easy to both produce and interpret. However, this use of distinct lines for regional boundaries is also the cause of confusion for many map readers. While the mapmaker meant the line to represent the middle of a zone of transition, many map readers see this boundary as an abrupt change from one type of vegetation to another (Jenks, 1953). Another way to represent spatial associations is to overlay different symbols in the same map space. A common application of this technique is used when mapping population distributions. The mapmaker will represent urban concentrations with graduated circles and portray rural distributions with dot symbols. Although this method is intended to aid the map reader's understanding of the distribution, many cartographers are reluctant to endorse its use. The primary reason for this is the concern that map readers will have difficulty interpreting the map due to the complexity of the map and its symbology (Jenks, 1976).

Map production and interpretation become significantly more difficult when the subject of interest is the relationship among several variables or distributions. The task of recognizing these associations has traditionally been left to the map reader. Map users are often asked to draw their conclusions by visually comparing maps showing the individual distributions. For example, a class of geography students might be given one map showing vegetation zones and one portraying climate zones and then be asked to draw conclusions about the effect of climate on vegetation. Previous research has found that unless the maps have a high positive correlation, and therefore highly similar patterns, map readers perform poorly in judging the relationship between variables (McCarty and Salisbury, 1961).

Another approach to map comparison is the mapping of statistical associations between the distributions. Robinson, for example, used isoline maps of correlation values to illustrate the relationship between rainfall and population on the Great Plains (Robinson, 1962). Although this method is useful for analytical geographers, it is questionable whether these maps are effective in communicating relationships to the average map reader. Also, due to the combining of the variables required by this method, information about the individual distributions is lost.

The final method of representing spatial associations discussed here is the use of complex regions. In short, these are regions that are defined by multiple variables. An example of this type of region would be the "Midwest" of the United States. This area can be defined by many types of variables including socio-economic, climatic, and topographic. Previous research has demonstrated that the number of features used to define a region relates to how precisely the region may be defined (Haggett, 1983). The representation of such a complex region is extremely difficult for the mapmaker because of the imprecise nature of the boundaries.

Bivariate Choropleth Mapping

This study will focus on the problem of mapping two distributions in the same map space and the use of bivariate choropleth maps to display the spatial patterns of two variables. Use of this mapping technique results in a map that displays three distributions: those of the two individual variables and that of the relationship between the variables. Consequently, this method of mapping not only eliminates the problems of visual map comparison as discussed above, but it also avoids the loss of information inherent in the statistical association mapping method.

The bivariate choropleth map gained prominence with the publication of spectrally encoded two-variable maps by the U.S. Bureau of the Census. These maps were first published in the early 1970s and are made by "crossing" two choropleth maps of different variables. This technique involves the overprinting of the tones from the individual maps to produce new tones that represent the combinations of the two distributions. Because these resultant tones are made up of the three primary colors, the maps are considered to be spectrally encoded (Olson, 1981). Due to the widespread use of Census Bureau publications, spectrally encoded maps have become somewhat familiar to the public in general and the cartographic community in particular. For this reason, these maps are considered the standard for bivariate choropleth maps.

In the years since spectrally encoded maps were first produced, cartographers have been trying to find ways to improve them. Many cartographers feel that these maps are too complex for map readers to understand. Research performed by Olson (1981) found that map readers are able to obtain some knowledge from the maps but there is a need for improvement especially with the legend arrangement. One attempt at improvement is the complementary-color two-variable map introduced by Eyton (1984). Similar to spectrally encoded maps, these maps are also produced by overprinting two individual maps. However, unlike the spectrally encoded maps, complementary-color maps utilize only two colors in the construction of their color scheme. Because these colors are complementary in nature, a gray tone is produced in those areas where the two variables have similar class values.

Eyton also attempted to improve spectrally encoded maps by simplifying the legend arrangement. The legends on Census Bureau maps utilize a four-by-four grid of cells with each cell containing one of the tones used to represent the values on the map. Complementary-color map legends also utilize a rectangular grid but the size has been reduced to three-by-three so that there are fewer tones displayed on the map. Eyton felt that map readers would be able to interpret this simplified legend arrangement more easily than that used on spectrally encoded maps (Eyton, 1984).

An important part of the information contained in any bivariate choropleth map deals with regional variations. The relationship between the two variables is not static; it changes from one geographic location to another. Since a primary focus of geographic research is to attempt to explain why these variations occur, it is essential that different map users perceive similar regions on the same map. Otherwise, the differing regional perceptions of researchers would lead to many possibly contradictory explanations.

The purpose of this research is to determine which of the two bivariate mapping techniques mentioned before, spectrally encoded and complementary-color, is more effective in representing regions. Effectiveness is defined as the consistency with which map users perceive regions on each type of map. It is reasonable to assume that the consistency of map users in perceiving similar regions will vary according to the type of map used. The technique with the higher consistency among map users will be considered more effective. It is hypothesized that the complementary-color two-variable map will be more effective at conveying regional information to map users because its symbology is less complex than that used on the spectrally encoded map.

CHAPTER 2
THE DEVELOPMENT AND USE OF BIVARIATE CHOROPLETH MAPPING

There are many types of bivariate choropleth maps available for cartographic use. These maps differ in the way the variables are represented and in how the legend is arranged. The "simplest" bivariate choropleth mapping methods are black-and-white in nature. These are considered simpler than color mapping techniques because they lack the problems inherent with the use of color in mapping. These problems will be discussed later in this chapter.

Black-and-white Techniques

The first of these black-and-white maps to be discussed was developed using horizontal line patterns and gray tones to represent the two variables (Smith, 1977). Figure 2.1 is an example of this type of map. Horizontal lines with varying spaces are used to symbolize one variable while the other is symbolized by different shades of gray. The legend is a three-by-three grid of cells with the values for one variable located along the vertical axis and the values of the other variable located along the horizontal axis. Because one of the variables is represented with varying color values and the other is represented with varying pattern textures, it is relatively easy for the map reader to distinguish between the individual distributions. This allows the map user to make interpretations not only about the distribution of the relationship between the two variables but also about the distributions of the individual variables. The major drawback to using this mapping technique is

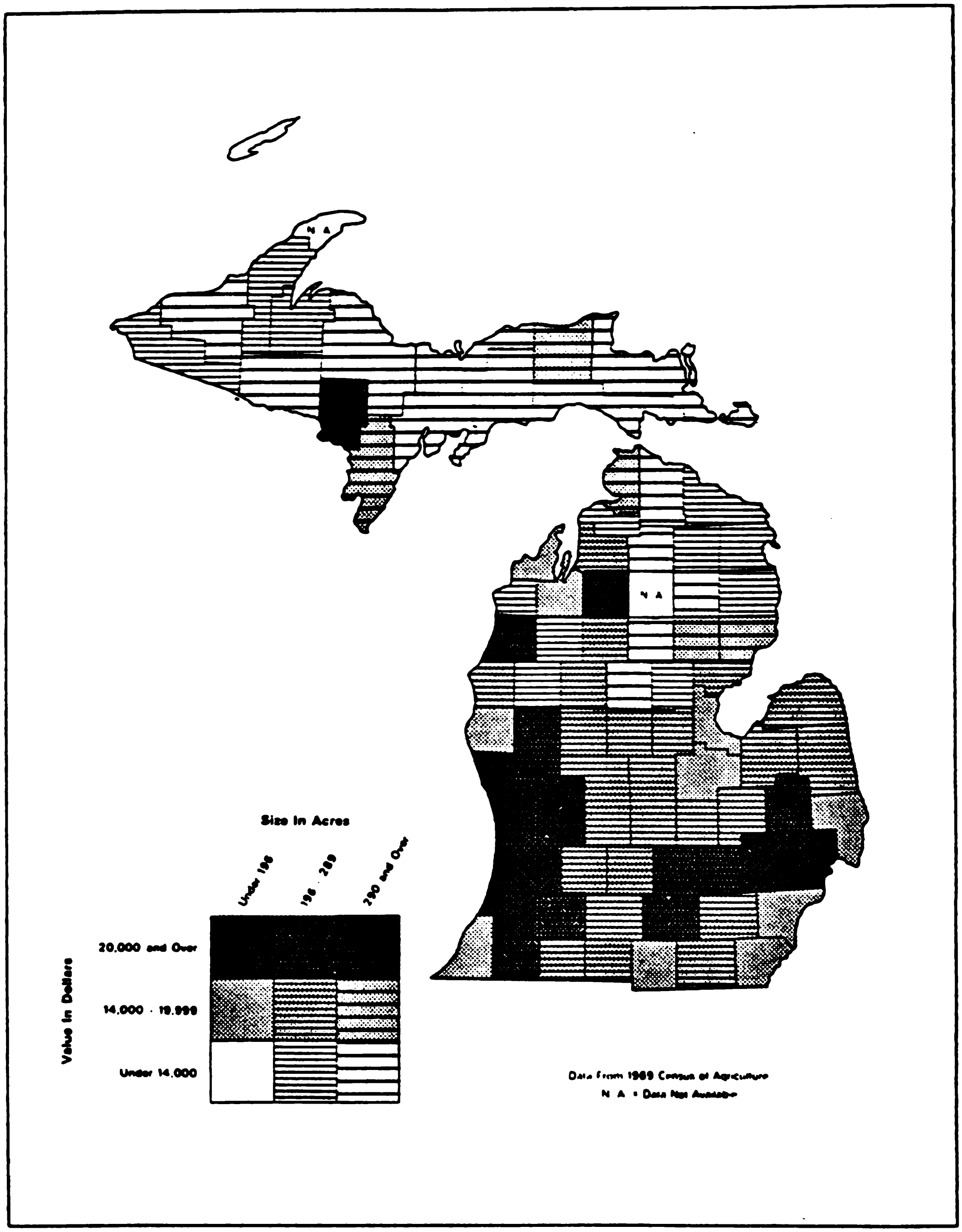


Figure 2.1 Example of Smith's mapping technique
(Source: Smith, 1977; Figure 7)

the visual noise created by the horizontal line patterns (Smith, 1977). This can be distracting to some map readers and interfere with their ability to fully understand the map's message.

Another black-and-white mapping method that has been researched uses proportionally spaced lines to represent both of the variables (Carstensen, 1982; Lavin and Archer, 1984). This is referred to as the crossed-line technique of producing bivariate choropleth maps. Figure 2.2 is an example of Carstensen's technique and Figure 2.3 illustrates Lavin and Archer's method. The distribution of one variable is displayed using proportionally spaced horizontal lines while the other variable's distribution is displayed using proportionally spaced vertical lines. The result is an unclassed bivariate choropleth map where each areal unit has a different shading pattern based on the values of the variables in that unit.

The two maps differ in the structure of their legends. Carstensen's key is constructed of "parallel lines which close together in spacing at some constant rate" (Carstensen, 1982; 57). These lines are plotted on perpendicular axes to produce a mesh appearance. The increase in line density along the two axes correspond to an increase in the variable represented. More importantly, the relationship of the variables is depicted by a shape index. If two values originate from the same relative position in their respective distributions (e.g., high and high), the shape produced by the lines crossing would be square in nature. If the two values are from different parts of their distributions, the resultant shape would be rectangular. Use of this shape index would allow map readers to determine the relationship in individual map zones or form general impressions of the relationship across the entire map (Carstensen, 1982).

Lavin and Archer's design essentially contains two keys: one for each variable. This arrangement was chosen "for the sake of clarity and to emphasize the bivariate character of the map" (Lavin and Archer, 1984; 52). The three values for each variable that are depicted in the legend correspond

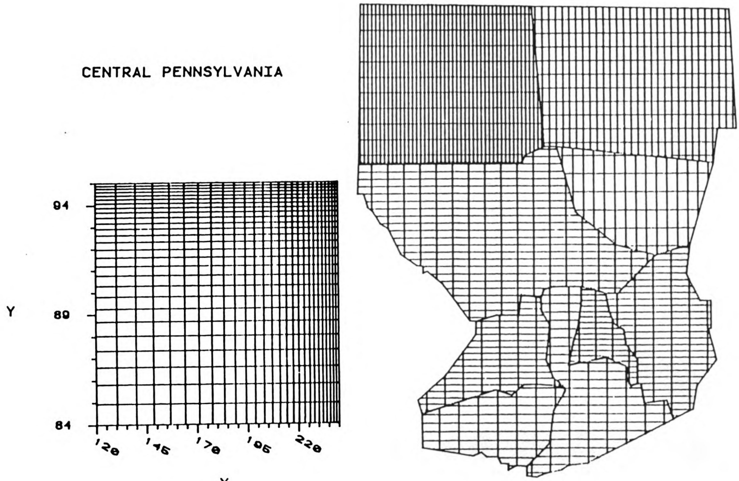


Figure 2.2 Example of Carstensen's technique (Source: Carstensen, 1982; Figure 9)

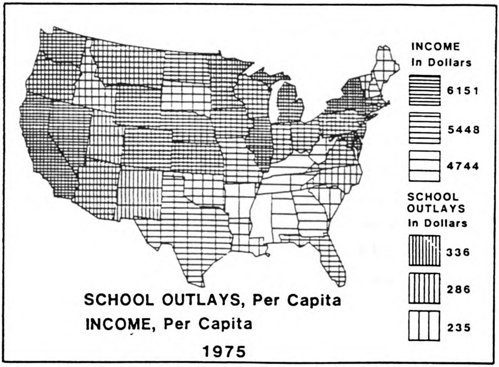


Figure 2.3 Example of Lavin and Archer's technique
(Source: Lavin and Archer, 1984; Figure 1)

to 25, 50, and 75 percent of the range of each variable. Although this legend arrangement does appear more clear than Carstensen's, it is extremely difficult to perceive the relationship between the variables using Lavin and Archer's key. In this sense, the square mesh legend used by Carstensen is the better legend for this type of bivariate choropleth map.

The crossed-line technique has a problem similar to the visual noise difficulty encountered by Smith. To many map users, the patterns of vertical and horizontal lines are visually distracting. This can inhibit their ability to use these maps in a constructive manner. Also, the zones displayed on the map must be large enough for at least two lines to cross the area in each direction so the shape index can be determined. Thus, due to the nature of the symbolization technique, this method is not suitable for maps containing small areal units (Carstensen, 1982).

Color Techniques

Other types of bivariate choropleth maps that have been developed use color to represent the variables. There are three important functions of color in any map: (1) to clarify information, (2) to increase the aesthetic appeal of the map, and (3) to encode complex information (Olson, 1975). Although the use of color in bivariate choropleth maps is strongly associated with the third function, the benefits resulting from all three functions apply to these maps. Olson's discussion of color on two-variable maps (Olson, 1975) provides a list of the considerations that must be taken into account when developing a color scheme for bivariate choropleth maps. Although her paper deals specifically with the spectrally encoded maps of the Census Bureau, the issues raised apply to any and all coding schemes that involve color.

The most common type of color bivariate choropleth map is the spectrally encoded two-variable map published by the Census Bureau. Figure 2.4 is an example of this type of map. These maps are made by overprinting the distribution maps of the individual variables. The tones that result from this process represent the combination of values of the two distributions. The term spectrally encoded is applied to these maps because "the tones representing the two distributions are composed of all three primary colors" (Olson, 1981; 259). Before being mapped, each variable's values are divided into four categories. For this reason, the legend of the spectrally encoded maps consists of a four-by-four rectangular grid with each cell containing one of the sixteen colors shown on the map. The value ranges of one variable are located along the vertical axis and the value ranges of the other variable are located along the horizontal axis.

Research performed with the spectrally encoded maps resulted in a number of conclusions about their effectiveness: (1) when given the tones used to produce the spectrally encoded maps, map users cannot spontaneously order them into the legend arrangement used on the map; (2) judging correlation on these maps is initially more difficult than performing

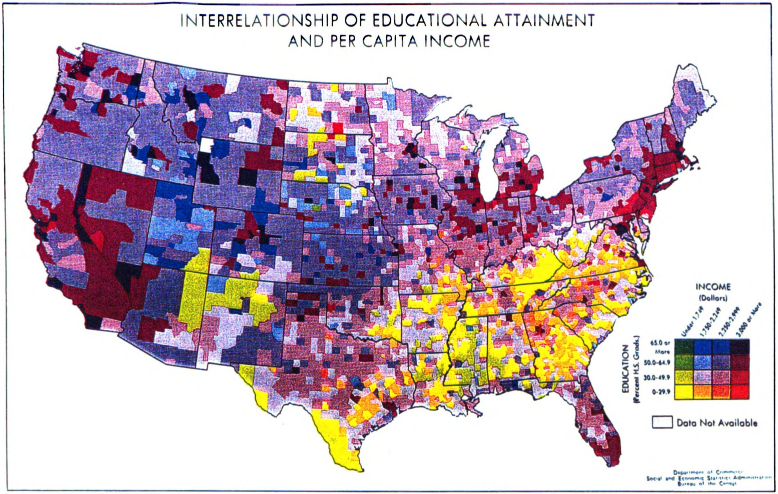


Figure 2.4 Example of spectrally encoded technique (Source: Olson, 1981; Figure 1)

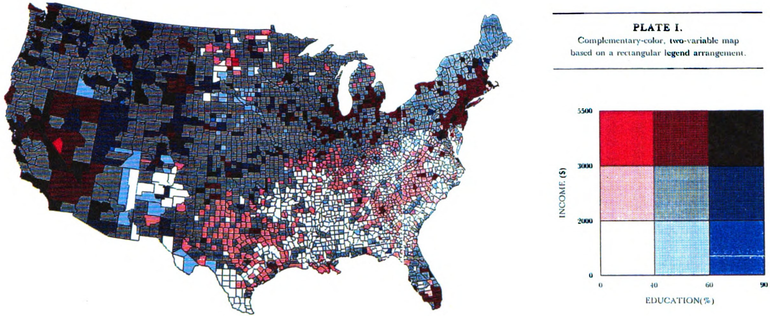


Figure 2.5 Example of complementary-color two-variable technique (Source: Eyton, 1984; Plate I)

the same task with separate distribution maps; (3) when asked to write about the information they learned from the spectrally encoded maps, map users are able to demonstrate that they gained some knowledge; and (4) map users find spectrally encoded maps interesting and appealing (Olson, 1981).

One of the main problems that Olson found with spectrally encoded maps is a lack of order in the arrangement of the legend tones. In an attempt to eliminate this difficulty, Eyton developed the complementary-color two-variable map (Eyton, 1984). An example of this map type is found in Figure 2.5. Eyton believed that map users' difficulty with the legend is a result of the unnecessary use of all three primary colors when representing the distributions. Since only two variables are displayed in the map, he proposed a color scheme using two complementary colors. The advantage of using complementary colors is that when the two distributions are overprinted, a gray tone is produced in those areas where the variables have similar class values.

The structure of the complementary-color legend is similar to that of the spectrally encoded maps. In her discussion of color on two-variable maps, Olson had the following suggestion about the legend arrangement:

> The number of categories to be used should not exceed the number that can be dealt with by the reader nor should it be so few that the map has too little information. The three-by-three matrix is both mechanically and visually more simple than the four-by-four arrangements and may actually convey more to the reader (Olson, 1975; 292).

The legend structure of complementary-color maps takes this into consideration and utilizes a three-by-three grid in order to "create easily visualized triads (light, medium, dark) of complementary colors and gray tones" (Eyton, 1984; 481).

Another method of representing two variables in the same map space was developed by intermingling small dots of only two colors on a computer screen (Groop, 1992). This produced a color unclassed two-variable map.

The two variables are represented by two different colors with the number of pixels in each color being proportional to the value of their respective variables. The intermingling of the two colors produces the appearance of intermediate colors. This results in a smooth visual change across the map and the elimination of sharply-defined regional boundaries since the colors change gradually as the variables change (Groop, 1992).

Region Perception Research

For the purposes of this study, a region is defined as "an area on the surface of the earth that is defined by a similar characteristic or characteristics that differentiate the area from surrounding areas" (Rader, 1995; 2). This definition includes both simple regions composed of one variable and complex regions made up of multiple variables.

In general, there are six ways for cartographers to portray regions: nominal maps, choropleth maps, unit grid maps, isarithmic maps, continuous tone maps, and dot maps (Rader, 1995). Each one of these different representations has a distinct way of graphically structuring the data. A good background discussion of each of these representations is provided in Rader's research on how regions are represented. The purpose of his work was to determine if each of these methods of representation adequately communicates spatial information about regions. Rader's methodology involved asking test subjects to perform several typical map use tasks on maps using the different symbolization methods for representing regions. These tasks were designed to isolate subjects' understanding of five regional concepts: location and extent, core and domain, internal structure, transitional boundaries, and map comparison.

The results of Rader's research indicate that, although spatial understanding differs with map type, "no one map is best for communicating all the spatial aspects of a region" (Rader, 1995; 64). For the specific task of comparing maps of similar regional distributions, Rader found that the response accuracy of map users did not vary significantly from map to map but that reaction times did. The maps with well-defined surface forms, such as isarithmic and continuous tone maps, were easier for map users to compare than the simpler maps.

Another study that dealt with the communication of regions was performed by Lavin (1979). He was interested in how the complexity of map patterns on choropleth maps affects map users' perceptions of regions. To examine this aspect of map communication, Lavin constructed a series of maps with increasing pattern complexities, then instructed test subjects to draw boundaries around the regions they saw on these maps. The consistency of the subjects' boundary drawing on each map was then measured. Lavin's results indicate that there is a direct relationship between pattern complexity and region perception variability (Lavin, 1979). As the complexity of the map increases, the variability of region perception among map users increases as well.

Region perception by map users using multi-variate maps has not been extensively studied by cartographers. However, several important studies have been done that establish a framework for this research. A study performed by Eilertsen examined regional perception on multi-pattern color dot maps (Eilertsen, 1980; Rogers and Groop, 1981). These maps utilize a differently colored pattern of dots to represent each distribution on the map. Eilertsen was interested in determining how effective these maps were in portraying spatial relationships. The effectiveness of the mapping technique was determined by comparing it with the traditional method of having the map reader visually compare single black-and-white dot distributions. The testing procedure involved having one group of subjects use the color dot maps and draw boundaries around regions of homogeneity and regions of mix among the distributions. A second group of subjects used the black-and-white maps to perform the same boundary drawing tasks. The responses were analyzed to measure the consistency of boundary drawing and the acc

the

con

resp

sim

det

cat

(Ca

wh

res

gro

biv

en

tw

sy

wa

co

co

accuracy of region identification. The results of the experiment indicate that the multi-pattern color dot maps are slightly more effective in communicating consistent perceptions of regions and produce more accurate responses from maps users.

Another experiment explored map users' perceptions of variable similarity on bivariate choropleth maps. The purpose of the research was to determine how the symbolism on these maps affected map readers' ability "to categorise [*sic*] variable similarity into two classes, 'similar' and 'dissimilar'" (Carstensen, 1984; 23). The term "variable similarity" refers to the degree to which pairs of values originate from the same relative positions in their respective distributions. To achieve his objective, Carstensen had two groups of test subjects produce regions of similarity and dissimilarity on a bivariate choropleth map. One group performed the task with spectrally encoded maps and the other used black-and-white crossed-line maps. The two maps were produced using the same data and differed only in their symbology. By comparing the regions produced, Carstensen found that there was a large degree of agreement between the two groups. This led him to conclude that both map designs elicit reasonable responses from map readers concerning variable similarity (Carstensen, 1984).

CHAPTER 3
REGION PERCEPTION TESTING ON BIVARIATE CHOROPLETH MAPS

The purpose of this research is to compare the effectiveness of two types of bivariate choropleth maps in representing regions. The problem is approached by having subjects perform an experimental map reading task and examining their responses. In general, this task was to draw boundaries of regions on the two-variable maps. This chapter will detail the production of the maps used in the test and the development of the testing procedure.

Map Construction

The two bivariate choropleth map types to be tested are the spectrally encoded and the complementary-color methods. These mapping techniques were chosen for the experiment for a number of reasons. The spectrally encoded maps have been published by the United States Census Bureau for more than twenty years. Figure 2.4 in Chapter 2 is an example of this type of map. Because of their widespread publication and the excellent reputation of the Census Bureau, these maps are considered the standard type of bivariate choropleth map. However, studies performed by several different researchers (Wainer and Francolini, 1980; Olson, 1981) have found that these maps still need improvement.

One of the areas identified as a problem was the lack of natural order in the legend arrangement. The development of the complementary-color two-variable map (Eyton, 1984) was an attempt to correct this problem. An example of these maps is shown in Figure 2.5 in Chapter 2. As discussed earlier, Eyton hoped to introduce logical order into the color scheme by reducing the number of colors used in constructing the maps from three to two. He also reduced the grid size of the legend from four-by-four to three-by-three in order to make the transitions from one legend category to another easier to see. These two maps were chosen for this investigation in order to determine whether or not the complementary-color maps are an improvement of the spectrally encoded maps in the area of regional perception.

A series of test maps were made using county data collected by the United States Census Bureau in 1980. Three different relationships were used in the testing in an attempt to eliminate any bias that could be introduced by the varying complexities of the distributions. Each relationship was mapped using both the spectrally encoded and the complementary-color methods. This resulted in a total of six different test maps. These maps are shown in Figures 3.1-3.6.

Since each relationship requires two variables, a total of six variables were needed to construct the maps. Their definitions are as follows: Serious Crime is the number of serious crimes (e.g., murder, rape, robbery, etc.) per one thousand persons; Unemployment is the percentage of the population that is without work; Educational Attainment is the percentage of the population that has a high school education or higher; Per Capita Income is the average annual income in dollars received by the population; Housing Value is the median value of occupied housing measured in thousands of dollars, and Cropland is the percentage of total land area that is considered cropland.

Next, each variable was categorized into three and four classes. Because the test subjects would be asked to identify regions represented by a number of different legend cells, a classification method that resulted in an approximately equal map area for each category was necessary. For this reason, the quantile break method was used. The employment of this classification method resulted in the variable ranges shown in Figures 3.1-3.6. For the complementary-color maps, the observations of the variables

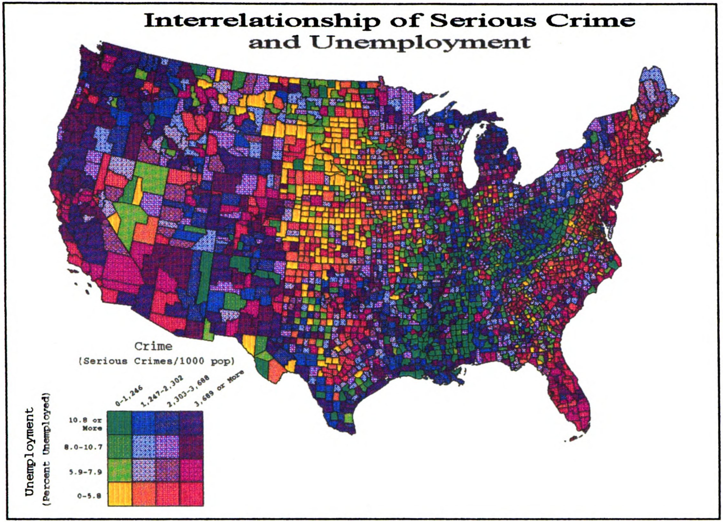


Figure 3.1 Spectrally encoded test map of Serious Crime vs. Unemployment (Reduced 60%)

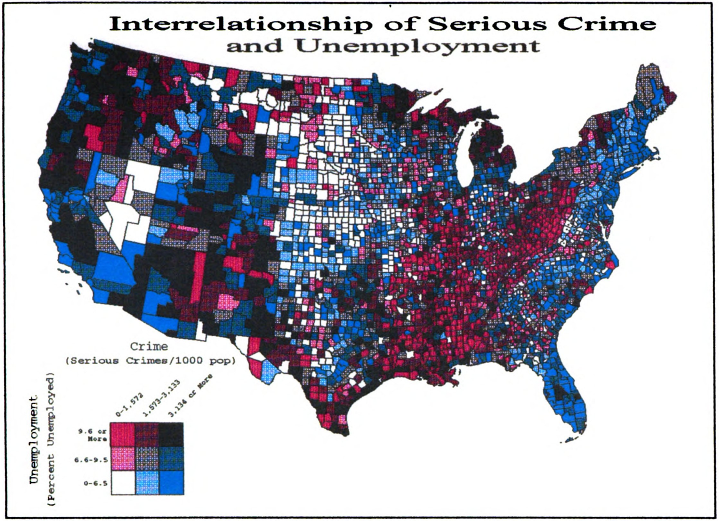


Figure 3.2 Complementary-color test map of Serious Crime vs. Unemployment (Reduced 60%)

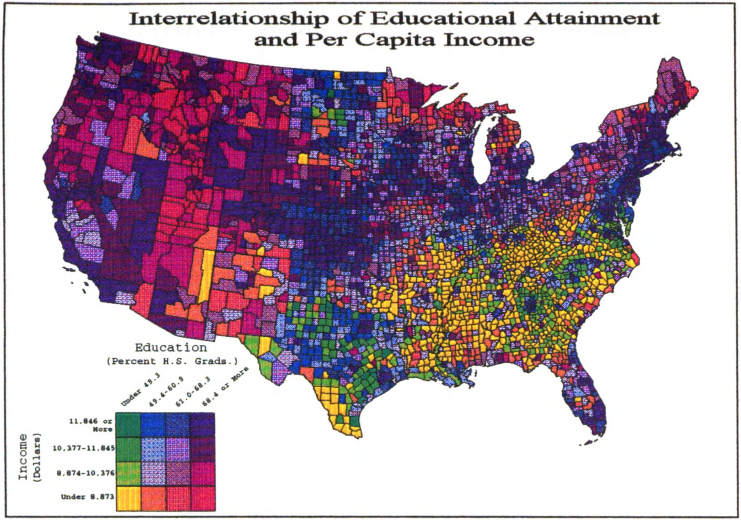


Figure 3.3 Spectrally encoded test map of Educational Attainment vs. Per Capita Income (Reduced 60%)

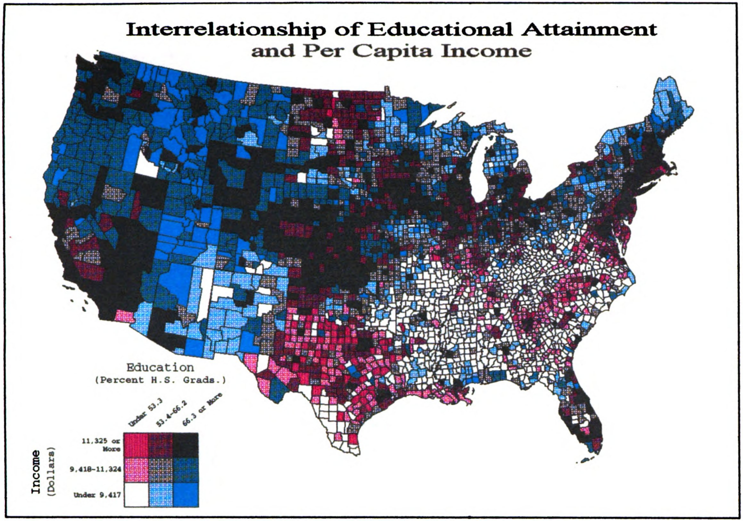


Figure 3.4 Complementary-color test map of Educational Attainment vs. Per Capita Income (Reduced 60%)

Interrelationship of

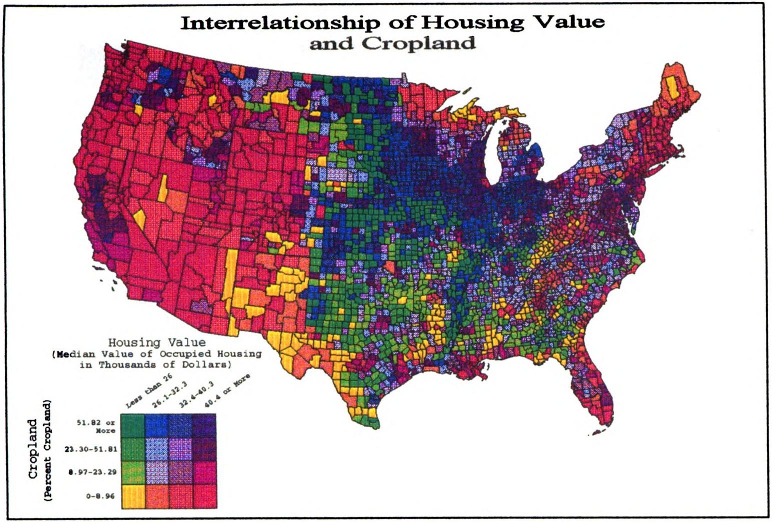


Figure 3.5 Spectrally encoded test map of Housing Value vs. Cropland (Reduced 60%)

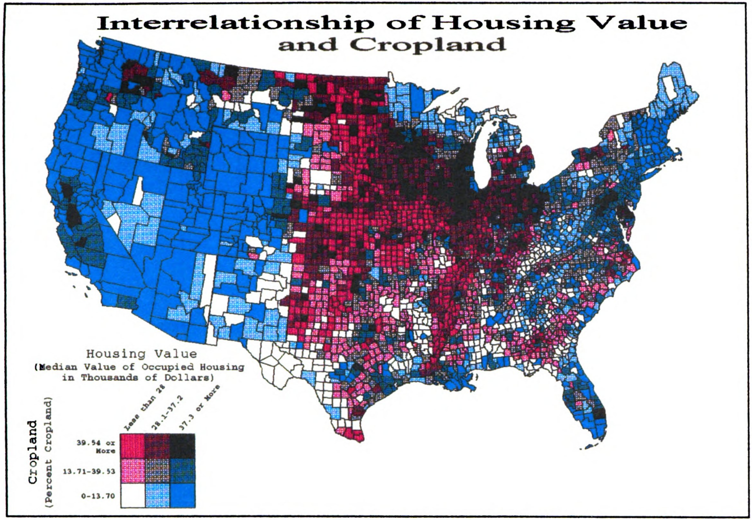


Figure 3.6 Complementary-color test map of Housing Value vs. Cropland (Reduced 60%)

that made up each map pair had to be combined to create the nine categories displayed in the legend. The map pair of Serious Crime and Unemployment will be used as an example to clarify the explanation. Two numbers were assigned to each observation; the first number represented the observation's position in the distribution of Serious Crime and the second number signified its position in the Unemployment distribution. The first number was assigned in the following manner: observations in the lowest quantile of Serious Crime were assigned a one, those in the next quantile a two, and those in the highest quantile were assigned a three. For the observations of Unemployment, a zero was given to those in the lowest quantile, a three for the next quantile, and a six for the highest quantile.

When the assigned values for an observation are added together, the result is the number of the appropriate legend cell for that observation. For example, if Ingham County, Michigan fell into the lowest quantile of Serious Crime and the highest quantile of Unemployment, it would be assigned a one and a six. The sum of these two numbers is seven indicating that Ingham County should be shaded according to the tone in the upper left cell of the legend. Figure 3.7 shows the relationship between the assigned values and

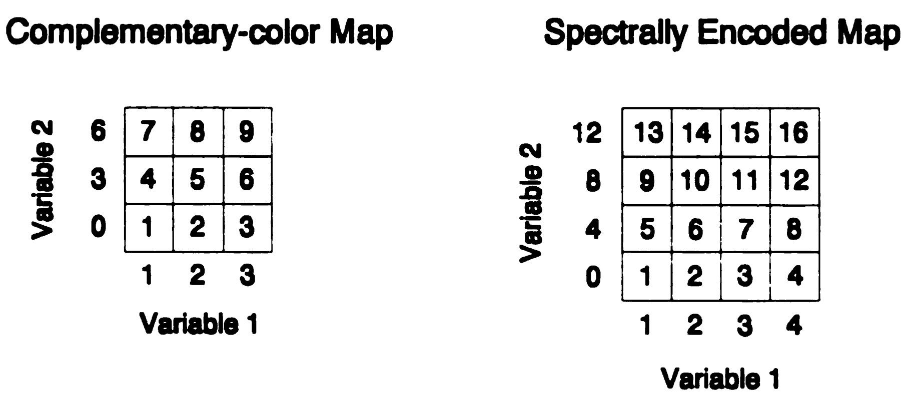


Figure 3.7 Relationship between assigned values and legend cells

the l
othe

orde
wer
stru
quar
a on
assig
dete
ther

dete
liter
Chap
map
not a
gree
were
the l
map
hard
com
and
acce

Map
files
polyg
categ

the legend cells. The same procedure was used to categorize the data for the other two map pairs.

Slight modifications were made to the method described above in order to categorize the data for the spectrally encoded maps. These changes were required because the spectrally encoded maps use a four-by-four legend structure. The observations for each variable were divided into four quantiles. Observations for the first variable in each map pair were assigned a one, two, three, or four and observations for the second variable were assigned a zero, four, eight, or twelve. Again, the values were added to determine the correct legend cell. When the entire process was finished, there were six different classification schemes; one for each test map.

The appropriate color for the legend cells on each type of map was determined by mixing colors to match those of the sample maps in the literature. These maps have been reproduced in Figure 2.4 and Figure 2.5 in Chapter 2. The general guidelines for producing the colors on both types of maps are known, but the exact tones used in the construction of the maps are not available. An attempt was made to replicate these colors by mixing red, green, and blue tones on a computer screen. Many different combinations were tried before satisfactory color schemes were found. Figure 3.8 shows the RGB percentages that were used in the construction of both types of maps. There is one note of caution; because of differences in software and hardware, these percentages should not be perceived as the "correct" combinations. These color combinations are specific to the mapping package and printer used in this experiment and are not guaranteed to produce acceptable color schemes in other configurations.

The data was mapped utilizing a thematic mapping package called MapViewer (MapViewer Version 1.1). This package has the ability to link files containing polygon boundaries with data files associated with those polygons and create choropleth maps based on that information. Each set of categorized data was matched to a file containing the boundaries of the

Complementary-color Map

100 R 0 G 0 B	80 R 0 G 0 B	0 R 0 G 0 B
100 R 80 G 80 B	80 R 80 G 80 B	0 R 80 G 80 B
100 R 100 G 100 B	80 R 100 G 100 B	0 R 100 G 100 B

Spectrally Encoded Map

0 R 100 G 0 B	55 R 75 G 100 B	65 R 65 G 90 B	65 R 25 G 85 B
50 R 100 G 0 B	85 R 85 G 100 B	90 R 80 G 100 B	80 R 15 G 70 B
75 R 100 G 0 B	90 R 85 G 95 B	90 R 75 G 85 B	95 R 20 G 75 B
100 R 100 G 0 B	100 R 75 G 0 B	100 R 55 G 0 B	100 R 20 G 0 B

Figure 3.8 RGB percentages used in the legends

counties in the contiguous United States. The appropriate color was assigned to each category and the legend and title were placed on each map. The six test maps were then printed on sheets of 8½" x 11" paper using an HP Paintjet printer. The resultant maps are shown in Figures 3.1-3.6.

Testing Procedure

The design of the testing procedure is quite similar to the procedure used by Lavin in his doctoral research (Lavin, 1979). He was interested in the relationship between map pattern complexity and regional perception. To examine this relationship, Lavin conducted an experiment in which map users drew regional boundaries on maps with varying complexities of pattern. He then analyzed these boundaries to determine the consistency among the map users. Since this research also involves the analysis of regional boundaries, Lavin's testing procedure was used for this experiment.

A basic assumption of this experiment is that the process of region perception involves the mental construction of lines around those areas that are seen as internally homogeneous. It follows logically that the experimental task readers will be asked to perform is that of reproducing those boundaries on the maps themselves. In order for the test to be statistically relevant, a minimum of twelve subjects had to be tested with each test map. Since there were six test maps, the test was given to a group of 72 graduate and undergraduate students enrolled in geography courses at Michigan State University.

Although all of the subjects were familiar with maps through their geography courses, it was unclear as to how much they knew about the concept of regions. Since the test subjects' differing levels of knowledge could introduce bias into the results, a training session was designed to ensure that all of the subjects were acquainted with the regional concept.

In previous studies of region drawing, two styles of drawing have emerged and are referred to as atomistic and generalistic (Lavin, 1979).

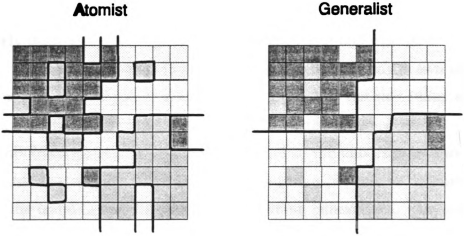


Figure 3.9 Examples of atomistic and generalistic styles of drawing regions (Source: Lavin, 1979; Figure 5.1)

Figure 3.9 illustrates the differences between these two styles of boundary drawing. The atomist draws boundaries wherever there is a change in color. The generalist, however, draws boundaries that encompass only the major trends and ignores the minor variations. Since the overall spatial trends were of more interest in this research, the training session was designed to produce test subjects who drew boundaries in the generalistic style.

The training session involved the use of a prepared script and map so that all test subjects were presented with the same information. Both the script and map can be found in Appendix A. In general, subjects were read a short introductory paragraph explaining the purpose of the experiment and the basic concept of regions. They were then shown a sample bivariate choropleth map and told that this map was similar to the ones that they would be working with during the test. The same map with a few regional boundaries drawn on it was then shown to the subjects. It was explained to them that these regions were not the "right" regions but just one person's perceptions. It was also made clear to the subjects that the drawing of regional boundaries is a purely subjective process and that different people will likely draw different regions. Additionally, the concept of generalization was introduced by explaining that the regions drawn on the sample map contained minor areas of different colors but that these areas were not significant enough to be considered regions themselves.

This training session is almost identical to the one used by Lavin. Before administering his main test, he conducted a pre-test to determine whether the training was successful in producing generalists. The results of this pre-test indicate that trained subjects are much more likely to draw boundaries in the generalistic style (Lavin, 1979). Therefore, the training session can be used as an effective means of producing subjects who draw boundaries in a generalistic manner.

After the training session was complete, each test subject was given a map and a set of instructions (Appendix B). Because of the large number of test subjects, it was not economically feasible to produce a separate test map for each subject. To solve this problem, each test map had a sheet of clear transparency laid over and registered to it. This resulted in the subjects drawing boundaries on the transparency so that each individual test map could be used by multiple test subjects.

The first part of the instructions asked the subjects to take a moment to familiarize themselves with the distributions of the two variables shown on the map and the map legend itself. This was necessary because it was unlikely that the subjects had much experience with bivariate choropleth maps and they needed to feel comfortable with the map before they started drawing region boundaries. The subjects were then instructed to draw lines around a specific type of region. For example, one of the instructions was as follows:

Carefully draw lines around areas that you see as predominantly **high**

S

w

In a sin

bounda

purpose

introdu

The ins

variabl

variabl

corresp

T

questio

intende

and sug

were in

were co

follow

color th

I

Serious Crime and **high** Unemployment. Label each of these areas with the letter A.

In a similar manner to the first region type, subjects were asked to draw boundaries around four other regions and label them as B, C, D, or E. The purpose of asking for several different region types was to eliminate any bias introduced by test subjects having difficulty with the colors used on the map. The instructions for all of the test maps asked for the same categories of variables using the same labelling scheme regardless of the map type or variable pair. Table 3.1 shows the category combinations and their corresponding labels.

The subjects were then asked to answer a series of subjective questions about the region types. The answers to these questions were intended to assist with the explanation of their region drawing performance and suggest possible improvements to the maps. For example, if subjects were inconsistent in their drawing of boundaries around Region A but they were consistent in naming Region A as the most difficult to see, it would follow logically that the color representing Region A should be changed to a color that is easier for map readers to perceive.

In the instructions, the regions in question were specified by

Table 3.1 Region labelling scheme used in the test instructions

Region Label	Variable 1	Variable 2
A	high	high
B	low	high
C	low	low
D	high	low
E	average	average

categ
subje
abilit
deter
on th
short
This
instru
In th
obtai

Data

resea
Eiler
frequ
numb
data
becau
meth

be th
analy
exper
subje
bound
thus e
respon
respon

categories of the two variables rather than by color. This required that subjects demonstrate an understanding of the legend used on the map. The ability of map users to correctly utilize the legend is an important part of determining whether or not these maps are useful tools. To collect more data on this aspect, the final part of the instructions asked subjects to write a short paragraph explaining what the map told them about the two variables. This part of the instructions was designed in a similar manner to Olson's instructions in her experiment with spectrally encoded maps (Olson, 1981). In this way, the results of this experiment could be compared to the results obtained by Olson.

Data Collection Method

The data collection method was similar to that used by Eilertsen in her research with multi-pattern color dot maps (Eilertsen, 1980). In her research, Eilertsen manually registered a grid to her response maps then compiled a frequency count for each grid cell. This frequency count indicated the number of test subjects who included that cell in a particular region. The data collection in this research followed the same general process but, because of the large number of test maps that needed to be analyzed, a method of automating Eilertsen's procedure was devised.

The use of a geographic information system (GIS) was determined to be the most efficient method of compiling the frequency counts needed to analyze the data that had been collected. The specific GIS used in this experiment was IDRISI (IDRISI Version 4.0). Because the regions test subjects drew were spatial entities, a digitizer was used for data entry. The boundaries of each region on the response maps were traced on the digitizer thus entering each region into the GIS as a polygon. Each test subject's response map was digitized into five separate files; one for each type of response region.

With the subjects' responses now in digital form, each map had be

sectioned in a manner similar to Eilertsen's use of a grid. In a GIS, this is accomplished through a process called rasterization. The system divides the study area into a grid with a user-specified cell size and then assigns each grid cell a value according to the attribute of the area in that cell. Each file was rasterized with a 0.125 inch grid and, if more than half of a grid cell was within a polygon, that cell was assigned the ID of that polygon. Since all of the polygons had an ID of one, every grid cell within a polygon was given a value of one and every grid cell not within a polygon was given a value of zero.

The next step was to produce a frequency count for each grid cell indicating how many test subjects placed that cell in a polygon. Since there were six different test maps and five different region types, a total of thirty frequency counts were calculated. These frequency counts were compiled by combining the response files through addition. The value of a specific grid cell in one file was added to the value of the same grid cell in another file. This was done for every grid cell in the study area. Since every cell that was in a polygon has a value of one, adding the values of cells in different files produces a frequency count that indicates the number of test subjects who included that cell in a region. Figure 3.10 illustrates this procedure. An example of one of the frequency counts is displayed in Figure 3.11. The other frequency counts can be found in Appendix C.

Consistency of region location on the maps was determined by calculating the variance of responses for each region type. Variance measures the variability of responses and, in most cases, a low variance is desired. In this experiment, however, a low variance indicates high variability in regional perception where test subjects generally disagree in drawing regions. Conversely, if the test subjects locate similar regions, the variance will be high (Rogers and Groop, 1981). This relationship is illustrated in Figure 3.12.

Although the raw variance is easy to calculate, it has one major

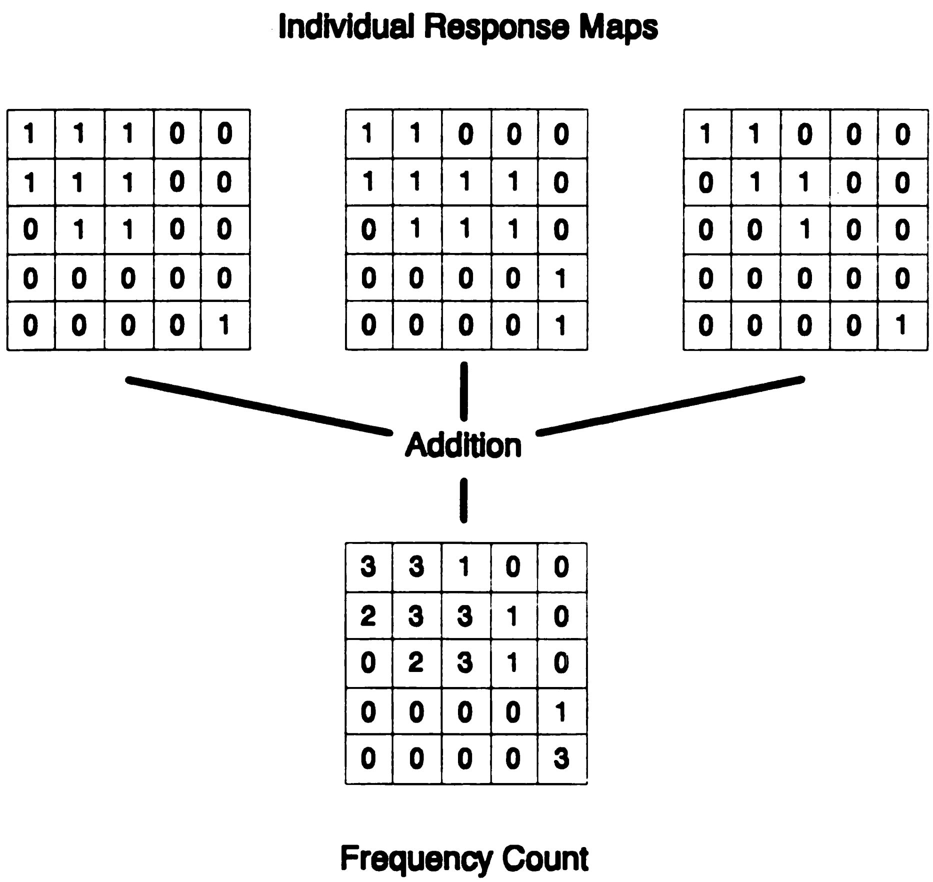


Figure 3.10 Procedure used to compile frequency counts

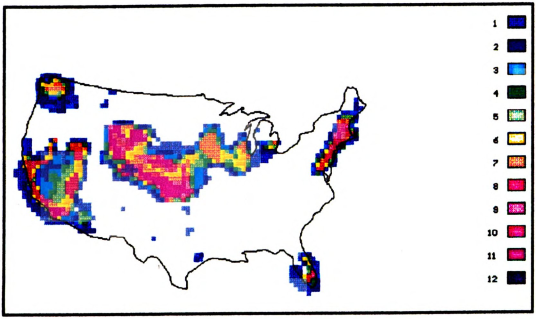


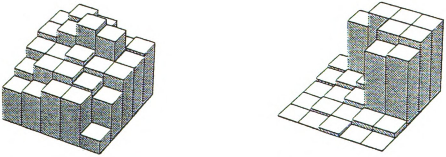
Figure 3.11 Example of a frequency count

22	27	25	20	21
22	25	30	28	20
20	20	22	20	20
16	20	20	18	15
13	14	17	17	5

Low Variance

2	24	30	30	30
1	3	5	30	30
0	1	3	26	27
0	1	1	2	2
0	0	1	0	0

High Variance



Inconsistent Agreement

Consistent Agreement

Figure 3.12 Relationship between variance and agreement among map readers (Source: Rogers and Groop 1981, Figure 5) disadvanta

calculated

another ty

frequency

The diagr

perfect ag

Sir

measure

research

attempt t

called a

maximur

Th

regional

where X

is the nu

the expec

Th

variances

frequency

group of s

drawing d

V_{ratio} range

among tes

disadvantage as a measure of region perception variability. The raw variance calculated for one type of region is not comparable to the raw variance of another type of region. This is the result of the varying number of zero-frequency cells and the differing total response frequencies (Lavin, 1979). The diagram in Figure 3.13 demonstrates this difficulty. The subjects are in perfect agreement about the location of the regions yet the variances differ.

Since the raw variances cannot be compared, another consistency measure must be devised. Lavin encountered the same difficulty in his research with pattern complexity on choropleth maps (Lavin, 1979). In an attempt to solve the problem, he developed a standardized variance measure called a V_{ratio}. The V_{ratio} is the ratio of the observed frequency variance to the maximum variance possible:

$$V_{ratio} = \frac{V_{observed}}{V_{max}}$$

The value of V_{max} occurs when all subjects are in agreement about regional location. It is calculated with the formula:

$$V_{max} = \frac{\frac{\Sigma X}{N}\left(N - \frac{\Sigma X}{n}\right)^2 + \left(n - \frac{\Sigma X}{N}\right)\left(\frac{\Sigma X}{n}\right)^2}{n - 1}$$

where X is the frequency response to a cell, n is the total number of cells, N is the number of subjects, $\Sigma X/n$ is the mean response per cell, and $\Sigma X/N$ is the expected number of cells.

There are two major advantages to using the V_{ratio} instead of the raw variances. The first is that the V_{ratio} standardizes the raw variances of each frequency count. This makes it possible to compare the consistency of one group of subjects drawing regions with the consistency of another group drawing different regions. The second advantage is that the values of the V_{ratio} range from 0.0 to 1.0 with higher values indicating stronger agreement among test subjects. This makes interpretation of the results easier because

Figu

they
64%
197

ratio
subje
diffe
V_{ratio}
deper
comp
has no
maxim

15	15	15	0
15	15	0	0
15	15	0	0
0	0	0	0

0	0	0	0
0	0	0	15
0	0	15	15
0	0	0	0

105	Total Response Frequency	45
59.06	Variance	36.56

Figure 3.13 Relationship between variance and number of zero-frequency cells on two maps of complete agreement (Source: Rogers and Groop, 1981; Figure 6)

they can be treated as percentages. For example, a V_{ratio} of 0.64 means that 64% of the maximum agreement between subjects was achieved (Lavin, 1979).

Lavin also issues some cautionary notes about the use of the variance ratio. The calculation of the V_{ratio} is dependent on the number of test subjects and the total response frequency. If the test were given to a different group of test subjects, V_{max} would have to be recomputed and the V_{ratio} would change. Therefore, this consistency measure is experiment dependent and values calculated from different experiments should not be compared. Also, Lavin's derivation of V_{max} was performed empirically and it has not been mathematically proven that his computation results in the maximum possible variance (Lavin, 1979).

CHAPTER 4
EVALUATING BIVARIATE CHOROPLETH MAPS

The first part of this chapter presents the variance ratios that were calculated from the boundaries drawn by subjects during the experiment. This section compares the values of different region types and also of different maps. The second part of this chapter examines the paragraphs written in response to the open-ended question found in the instructions. Assessment of the subjects' understanding of the map message is based mainly on subjective reasoning.

Examination of Variance Ratios

As stated earlier, the research question addressed in this experiment was which type of bivariate choropleth map is better at communicating regions to the map reader. This question was answered by comparing the consistency with which map users draw regions on the spectrally encoded maps with their consistency when drawing on the complementary-color maps. To perform this analysis a method of evaluating the consistency of boundary drawing was developed. As discussed in the previous chapter, the analysis involved the use of a consistency measure called the V_{ratio}. Despite the reservations mentioned earlier, the V_{ratio} still provides an acceptable relative measure of the consistency of region location.

For ease of discussion, each of the test maps will be referred to by a number. Table 4.1 shows how these numbers correspond to the test maps. Several test subjects did not follow the instructions or drew boundaries that were too ambiguous to judge accurately. These tests were discarded leaving

Table 4.1 Numbering scheme for the test maps

Mapped Relationship	Spectrally Encoded	Complementary-color
Serious Crime vs. Unemployment	1	2
Educational Attainment vs. Per Capita Income	3	4
Housing Value vs. Cropland	5	6

Table 4.2 Number of test map responses

Test Map	# of Responses
1	10
2	11
3	12
4	11
5	10
6	11

Table 4.3 Variance ratios for response regions

Region	Map 1	Map 2	Map 3	Map 4	Map 5	Map 6
A	52.14	56.53	56.44	66.93	44.23	56.95
B	61.99	68.85	35.71	44.14	50.84	50.94
C	66.85	64.51	78.35	60.39	53.74	51.40
D	49.65	55.59	43.91	47.19	72.72	75.51
E	40.93	27.50	48.57	29.78	38.31	30.75

65 response maps from the testing procedure. Table 4.2 shows how these responses were distributed among the six test maps.

The variance ratios that were calculated for each of the five response regions are listed in Table 4.3. These values are interpreted as the percentage of the maximum possible variance. A value of one hundred would indicate total agreement among the test subjects and a value of zero would signify absolute disagreement. Because of the reservations expressed earlier, the variance ratios of only those maps displaying the same distribution can be compared. Therefore, the results for Map 1 can only be compared to those of Map 2, Map 3 to Map 4, and Map 5 to Map 6. For the definitions of each region type refer to Table 3.1 in Chapter 3 which shows how the region labels correspond to categories of each variable. From the results, it can be seen that the test subjects were more consistent drawing Regions A, B, and D on the complementary-color maps and Regions C and E on the spectrally encoded maps. This disparity among the subjects' consistency could be caused by the relative difficulty the subjects had perceiving each type of region.

In the second part of the instructions, the subjects were asked to decide which region type was the easiest to see. For the spectrally encoded maps, the total number of subjects who chose a certain region type was calculated. These totals were then converted into the percentage of subjects using spectrally encoded maps who thought that region was the easiest to see. The same calculation was done with the answers from those subjects using the complementary-color maps. The results are shown in Table 4.4.

A large majority of the test subjects who used the spectrally encoded maps chose Region C as the easiest to see. This could explain why their boundaries around Region C were drawn more consistently than the subjects using the complementary-color maps. It is reasonable to relate these results to the color used to represent this particular region. Region C is used to label those regions where both variables displayed are low in value. In the

Table 4.4 Percentage of subjects rating each region easiest to perceive

Region	Spectrally Encoded	Complementary-color
A	6	23
B	0	17
C	69	20
D	23	37
E	2	3

spectrally encoded legend scheme, these regions are represented by a yellow tone. This tone is significantly brighter than the other tones used in the color scheme and seems to cause the counties that fall into this category to stand out from the rest of the map. Region C on the complementary-color maps is represented by white and this appears to cause the counties in this region to fade into the background of the map. These two effects resulted in the subjects using the spectrally encoded maps drawing boundaries around region C more consistently than the subjects using the complementary-color maps.

In the second part of the instructions, the subjects were also asked which region type was the hardest to see. Their answers were analyzed in the same manner as described above and the results are shown in Table 4.5. A large percentage of the complementary-color subjects chose Region E as the hardest to see. This could indicate a possible explanation for their difficulty in locating Region E consistently. In the complementary-color legend, Region E is represented by the gray tone in the center of the legend grid. In all three of the distributions, the counties shaded this color are dispersed throughout the map with no large concentrations. Lavin's research (Lavin, 1979) showed that a high pattern complexity such as this results in a low variance ratio. This could explain why subjects using the

Table 4.5 Percentage of subjects rating each region hardest to perceive

Region	Spectrally Encoded	Complementary-color
A	26	3
B	14	29
C	6	3
D	17	8
E	37	57

complementary-color maps had a more difficult time perceiving Region E than those who used the spectrally encoded maps.

Another possible reason the subjects using the complementary-color maps performed less consistently when drawing boundaries around Region E is related to the legend structure of the two maps. In the three-by-three legend used on the complementary-color maps, Region E is represented by one color, the gray tone in the center cell. However, on the spectrally encoded maps, the four-by-four legend structure causes Region E to be represented by four colors, the four light tones located around the center of the structure. This difference in the number of tones representing Region E results in more counties, and therefore larger map areas, being classified in Region E on the spectrally encoded maps than on the complementary-color maps. As a consequence, Region E is easier for the subjects to perceive on the spectrally encoded maps and they perform more consistently in their region location.

None of the subject groups performed well when drawing boundaries around Region E. The variance ratios for this region type were, in most cases, far below the variance ratios calculated for the other four region types. There are two possible explanations for this; both dealing with the legend arrangement. On both maps, Regions A, B, C, and D are located at the four corners of the legend grid. It is possible that this positioning makes these categories appear more prominent than the other categories. These legend cells also could stand out because of the colors used to represent them. The method used to construct the color schemes on both types of maps results in the corner legend cells being assigned pure hues. This causes these cells, and the categories they represent, to catch the reader's attention more than the other cells.

An interesting pattern emerged from the examination of the variance ratios for each region type. By looking at the concentration and areal extent of each region type on each test map, a correlation was found between the variance ratio and the distribution of the region. As the concentration and areal extent of a particular region type increased, the variance ratio for that region type increased also. For example, Region D on Map 6 has a variance ratio of 75.51. On the map, this region type is represented by the blue tones and has three distinct and reasonably large areas. On the other hand, Region B on Map 3 has a variance ratio of 35.71. This region type is represented by the green tones and does not have any significant areas displayed on the map.

Evaluation of Paragraph Answers

In the final part of the instructions, test subjects were asked to write a short paragraph describing the information they extracted from the map. The results of this part of the experiment cannot be treated in a statistical manner because of the subjective nature of the question. However, these results do make it possible to determine whether the subjects were able to obtain information from these maps or if the basic structure of the maps was too complex to allow basic understanding.

The answers were read to determine whether the subjects demonstrated an understanding of the relationship being displayed. The answers were then divided into three categories: those written by subjects showing understanding, those by subjects not showing understanding, and those left

Table 4.6 Categorization of paragraph answers from spectrally encoded maps

Category	**# of Subjects**
Demonstrated understanding	20
Did not demonstrate understanding	10
Left answer blank	6

blank by subjects. These results are displayed in Table 4.6 and Table 4.7. Approximately two-thirds of the subjects in both groups wrote answers that demonstrated some level of understanding about the distribution being displayed. This clearly indicates that both of these types of bivariate choropleth maps are able to communicate information to map readers.

The answers by the subjects who demonstrated understanding were then divided into groups according to the relationship that was displayed. This was done to determine whether one distribution was being understood more than the others. For both types of maps, each distribution was understood by approximately the same number of subjects. These results show that, at least for the distributions used in this experiment, the comprehension of the map message by map readers is not dependent on the relationship being mapped.

Table 4.7 Categorization of paragraph answers from complementary-color maps

Category	**# of Subjects**
Demonstrated understanding	21
Did not demonstrate understanding	10
Left answer blank	5

CHAPTER 5
CONCLUSIONS

In this thesis, the literature concerning bivariate choropleth maps has been discussed and an experiment has been developed that tested the consistency of regional perception among map readers using these maps. The purpose of this research was to determine which of two bivariate choropleth map types, spectrally encoded or complementary-color, was more effective in communicating regional information to the map user. In this final chapter, the significance of the results and suggestions for future research will be addressed.

Significance of Results

The values of the calculated variance ratios show that, of the five regions subjects were asked to identify, three were found more consistently on the complementary-color maps and two on the spectrally encoded maps. Although the subjects using the complementary-color maps performed slightly better, these results are not significant enough to indicate that one type of map is more effective at regional communication than the other. Even though the results of this experiment were not able to determine which map type was better, there are several other conclusions that can be drawn from the information gathered.

The first of these findings is that the type of bivariate choropleth map on which readers are more consistent in finding regions is not dependent on the distribution that is being displayed. This conclusion is supported by the variance ratios calculated for each region type. No matter which distribution was displayed, the test subjects located the regions more consistently on the same type of map. Although the absolute values of the variance ratios were influenced by the distribution, the relative value of one map type to the other was not affected. This would seem to indicate that the mapmaker's decision about which type of map to use should not be dependent on the distribution that is being displayed.

Another conclusion that can be made from the results of this experiment is that map readers perceive regions of extreme values - highest and lowest - better than they recognize regions of intermediate values. As discussed in the previous chapter, this is related to the legend structure and the arrangement of colors used in the legend. According to Olson, this is desired because "the correspondence of extreme values weighs most heavily on the relationship" (Olson, 1975; 290). From this research, it would appear that both of these types of bivariate choropleth maps are successful in producing the desired effect of emphasizing the extreme values.

Previous research has shown that the pattern complexity of a map is an important factor in how consistent map readers are in drawing regional boundaries (Lavin, 1979). This experiment has demonstrated that this relationship applies to individual region types as well. The pattern complexity of a region type is related to the concentration and relative size of the areas representing that region type. For example, a region type composed of many small areas scattered across the map space has a higher pattern complexity than a region type with one or two large concentrations located on the map. In this experiment, the region types with higher pattern complexities had much lower variance ratios than the region types with low pattern complexities. Therefore, this would appear to support Lavin's conclusion that pattern complexity and consistency of boundary drawing have an inverse relationship.

The final conclusion that can be made from this research concerns map readers' comprehension of the information presented on bivariate choropleth maps. A significant majority of subjects were able to understand and discuss the relationships that were being shown on the test maps. Although many cartographers have criticized these maps for being too complicated for map readers to understand, the results of this experiment seem to indicate that this is not true. This agrees with the conclusions Olson reached in her research with spectrally encoded maps (Olson, 1981).

Suggestions for Future Research

Previous studies have shown that bivariate choropleth maps can be valuable tools for both geographic researchers and the casual map reader. However, there has been very little research done to determine which type of these maps is best to use in any particular situation. More testing needs to be done with other types of two-variable choropleth maps to determine how well map readers perform region location tasks on those maps. The map types used in this experiment are only two of the many options available to cartographers. Also, this research dealt only with the consistency of map readers in locating regions. Although this is an important consideration, another important factor in a map's effectiveness is the accuracy with which map readers identify regions. Are map readers able to more accurately identify regions on one type of map than on any other? Before one particular type of bivariate choropleth map can be considered the best option, both of these areas need to be explored.

It has been suggested in previous research (Olson, 1975; Eyton, 1984) that a four-by-four legend structure may present too many colors for map readers to sort and adequately comprehend. This was the reason Eyton developed the complementary-color maps using a three-by-three legend. However, since the color scheme used on the spectrally encoded maps has been found to be visually attractive to map readers, it would be interesting to see if a three-by-three legend structure as suggested by Olson (1981) would make these maps more effective. Therefore, a possible area of future research would be the construction of a three-by-three legend using the color scheme presently utilized on spectrally encoded maps and then testing its effectiveness at representing regions.

It is important that the research of bivariate choropleth maps be continued. Despite any reservations critics may have, these maps have been shown to be effective at displaying the relationship between two variables. With further research and development, these maps could become even more valuable tools for both the geographic researcher and the casual map reader.

APPENDICES

APPENDIX A

TRAINING SESSION SCRIPT AND MAP

The reason for this experiment is to find out information about how people see regions on maps. A region is an area of sameness. It is an area that has some characteristic that makes it different from surrounding areas or regions.

(SHOW EXAMPLE MAP)

If I asked you to look at this map and show me the regions that you see, you would see regions which would be a little different from those that others would see. If I asked you to draw lines around regions you see, you would draw the lines somewhat differently from other people.

As an example, I can show you how I draw lines around some of the regions I see.

(DRAW REGIONAL BOUNDARIES)

Notice that most of my regions have more than one shade within them but they are predominantly one shade. Also notice that some of my regions overlap each other.

Even though these are the regions I see, I can fully understand how anyone else might see different regions. If you had to draw the regions on this map I am sure your regions would be different from mine. There is no absolutely correct way of drawing regions. What I am interested in is differences in the way people see and draw regions.

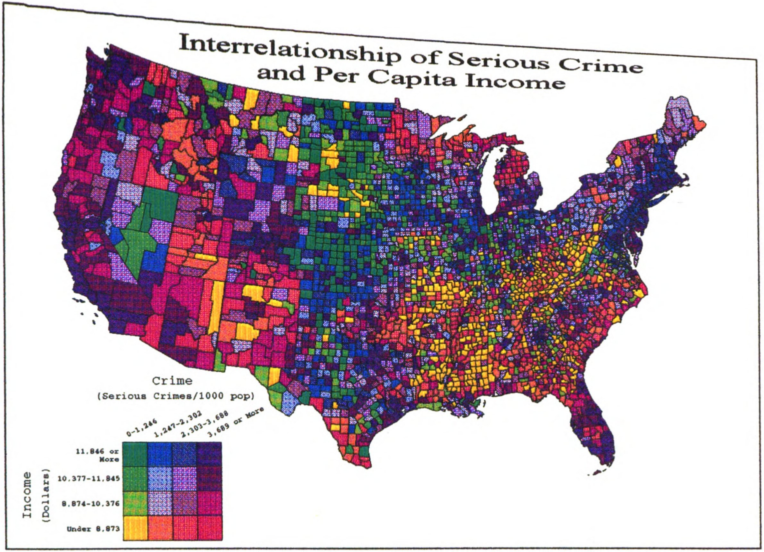


Figure A.1 Map used as an example in the training session

APPENDIX B

EXPERIMENT INSTRUCTIONS

The sets of instructions that were used in the experimental project are contained in this appendix. Each of the three relationships that were mapped had a separate set of instructions. These three instruction sets are identical except for the names of the mapped variables.

Instructions for Serious Crime vs. Unemployment Maps

Part I

Follow the numbered instructions in order. Complete each step before going to the next step.

1. The map represents the relationship between the distributions of two different variables, Serious Crime and Unemployment. Look at and familiarize yourself with the distributions of the two variables and the map legend.

2. A clear transparency has been placed over the map. On this transparency, carefully draw lines around areas that you see as predominantly **high** Serious Crime and **high** Unemployment. Label each of these areas with the letter A.

3. Draw lines around areas that you see as predominantly **low** Serious Crime and **high** Unemployment. Label each of these areas with the letter B.

4. Draw lines around areas that you see as predominantly **low** Serious Crime and **low** Unemployment. Label each of these areas with the letter C.

5. Draw lines around areas that you see as predominantly **high** Serious Crime and **low** Unemployment. Label each of these areas with the letter D.

6. Draw lines around areas that you see as predominantly **average** Serious Crime and **average** Unemployment. Label each of these areas with the letter E.

Part II

When you have finished the first portion, remove the transparency from the map and answer the following questions. Use the letters A, B, C, D, or E for your answers.

1. Which region was easiest to see? ___
2. Which region was the most difficult to see? ___
3. Which region was the largest in total area? ___
4. Which region was most sharply defined? ___

Using the space below, write a short paragraph explaining what this map tells you about these two variables.

Instructions for Educational Attainment vs. Per Capita Income Maps

Part I

Follow the numbered instructions in order. Complete each step before going to the next step.

1. The map represents the relationship between the distributions of two different variables, Educational Attainment and Per Capita Income. Look at and familiarize yourself with the distributions of the two variables and the map legend.

2. A clear transparency has been placed over the map. On this transparency, carefully draw lines around areas that you see as predominantly **high** Educational Attainment and **high** Per Capita Income. Label each of these areas with the letter A.

3. Draw lines around areas that you see as predominantly **low** Educational Attainment and **high** Per Capita Income. Label each of these areas with the letter B.

4. Draw lines around areas that you see as predominantly **low** Educational Attainment and **low** Per Capita Income. Label each of these areas with the letter C.

5. Draw lines around areas that you see as predominantly **high** Educational Attainment and **low** Per Capita Income. Label each of these areas with the letter D.

6. Draw lines around areas that you see as predominantly **average** Educational Attainment and **average** Per Capita Income. Label each of these areas with the letter E.

Part II

When you have finished the first portion, remove the transparency from the map and answer the following questions. Use the letters A, B, C, D, or E for your answers.

1. Which region was easiest to see? ___

2. Which region was the most difficult to see? ___

3. Which region was the largest in total area? ___

4. Which region was most sharply defined? ___

Using the space below, write a short paragraph explaining what this map tells you about these two variables.

Instructions for Housing Value vs. Cropland Maps

Part I

Follow the numbered instructions in order. Complete each step before going to the next step.

1. The map represents the relationship between the distributions of two different variables, Housing Value and Cropland. Look at and familiarize yourself with the distributions of the two variables and the map legend.

2. A clear transparency has been placed over the map. On this transparency, carefully draw lines around areas that you see as predominantly <u>**high** Housing Value and **high** Cropland</u>. Label each of these areas with the letter A.

3. Draw lines around areas that you see as predominantly <u>**low** Housing Value and **high** Cropland</u>. Label each of these areas with the letter B.

4. Draw lines around areas that you see as predominantly <u>**low** Housing Value and **low** Cropland</u>. Label each of these areas with the letter C.

5. Draw lines around areas that you see as predominantly <u>**high** Housing Value and **low** Cropland</u>. Label each of these areas with the letter D.

6. Draw lines around areas that you see as predominantly <u>**average** Housing Value and **average** Cropland</u>. Label each of these areas with the letter E.

Part II

When you have finished the first portion, remove the transparency from the map and answer the following questions. Use the letters A, B, C, D, or E for your answers.

1. Which region was easiest to see? ___

2. Which region was the most difficult to see? ___

3. Which region was the largest in total area? ___

4. Which region was most sharply defined? ___

Using the space below, write a short paragraph explaining what this map tells you about these two variables.

APPENDIX C

FREQUENCY COUNTS

This appendix contains the frequency counts that were compiled for the five region types subjects were asked to identify on the six test maps. The frequency counts are organized so that the results for the five regions on each test map appear together.

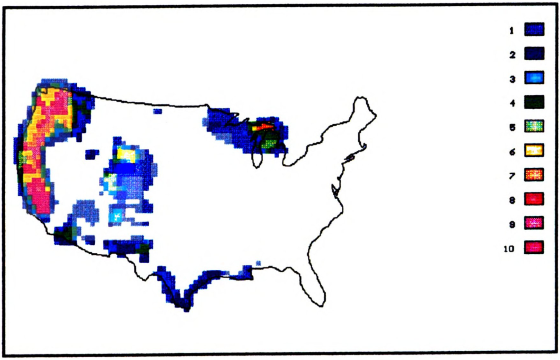


Figure C.1 Frequency counts for Map 1 - Region A

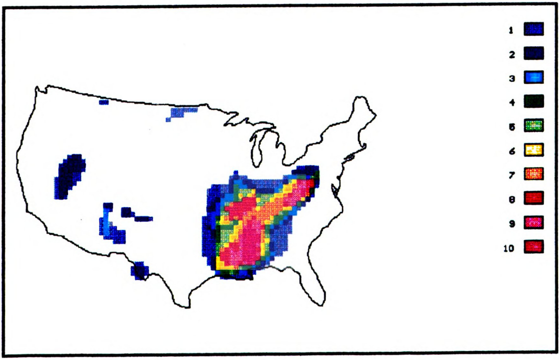


Figure C.1 (continued) - Region B

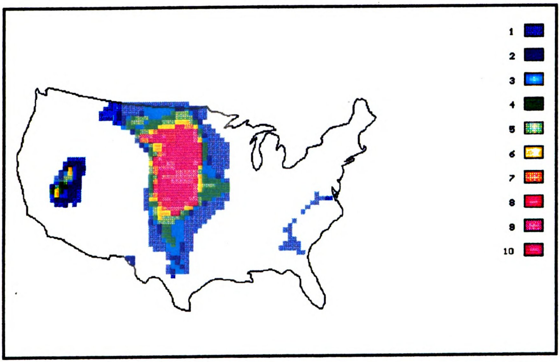


Figure C.1 (continued) - Region C

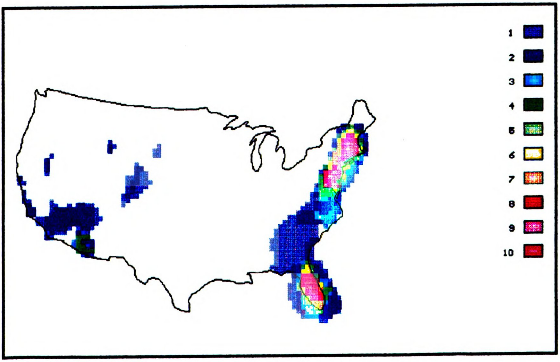


Figure C.1 (continued) - Region D

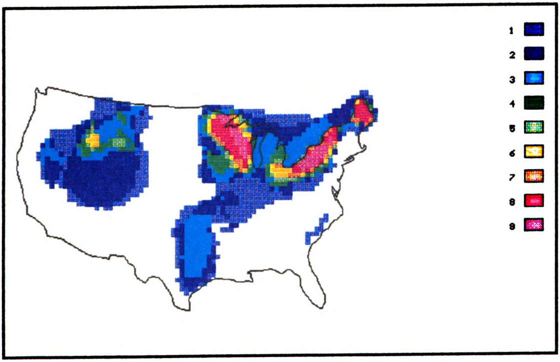


Figure C.1 (continued) - Region E

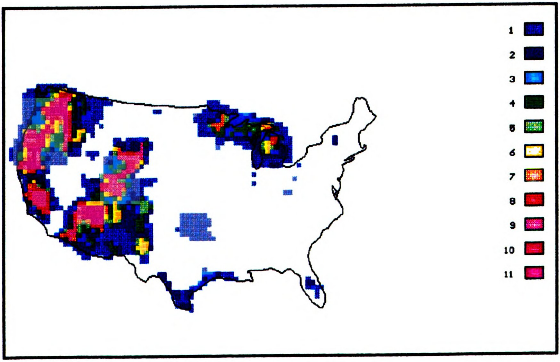


Figure C.2 Frequency counts for Map 2 - Region A

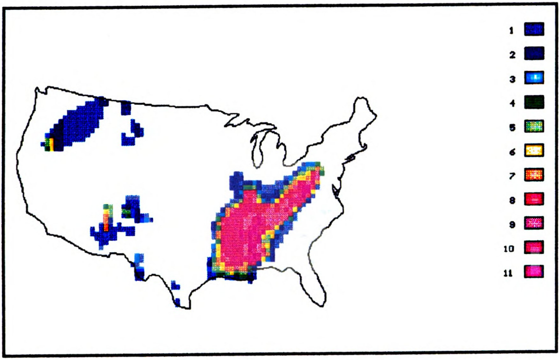


Figure C.2 (continued) - Region B

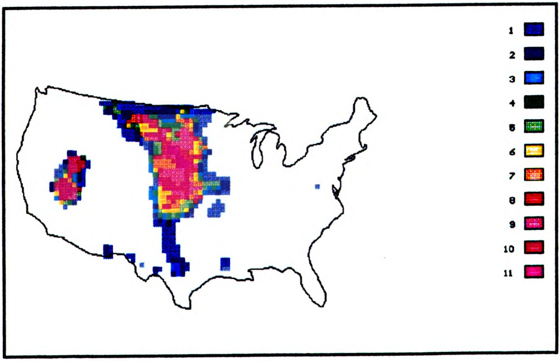


Figure C.2 (continued) - Region C

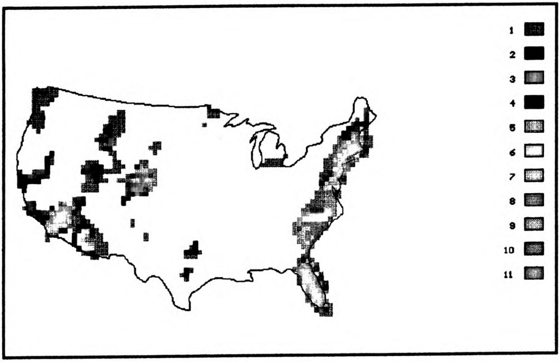


Figure C.2 (continued) - Region D

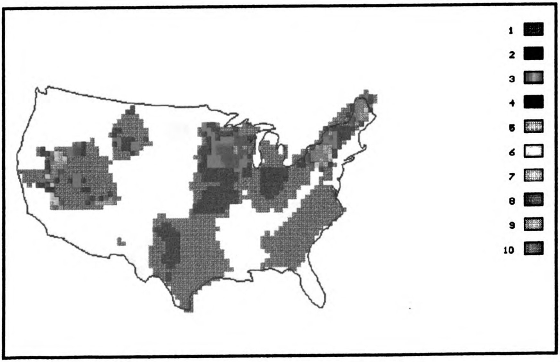


Figure C.2 (continued) - Region E

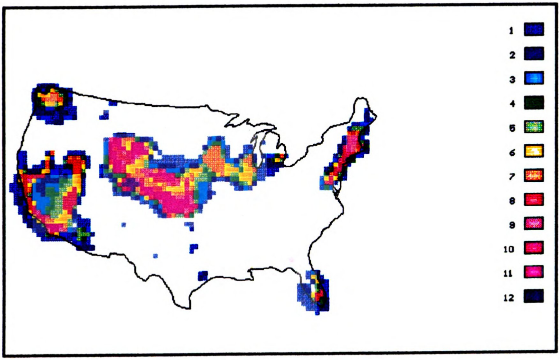


Figure C.3 Frequency counts for Map 3 - Region A

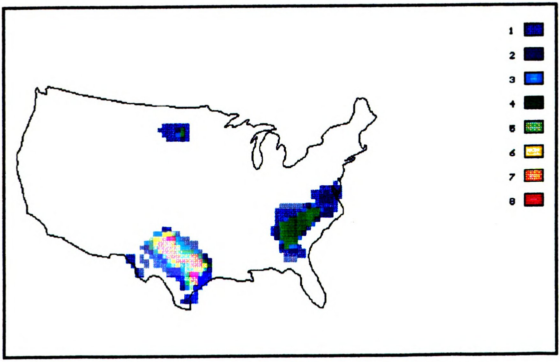


Figure C.3 (continued) - Region B

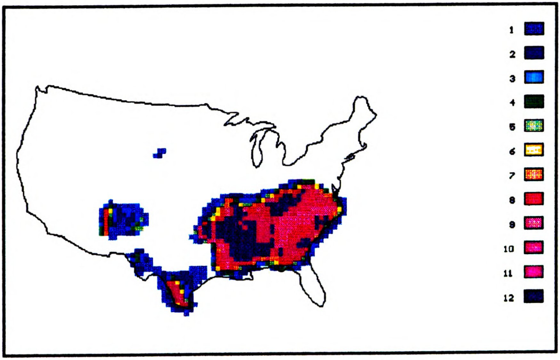


Figure C.3 (continued) - Region C

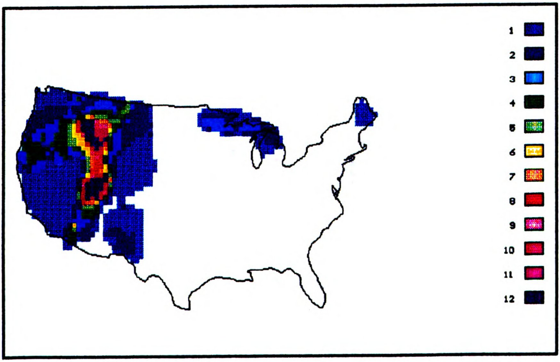


Figure C.3 (continued) - Region D

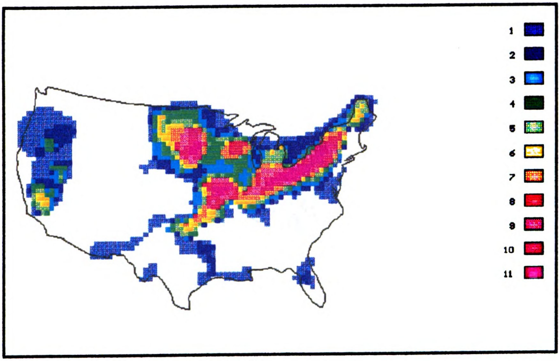


Figure C.3 (continued) - Region E

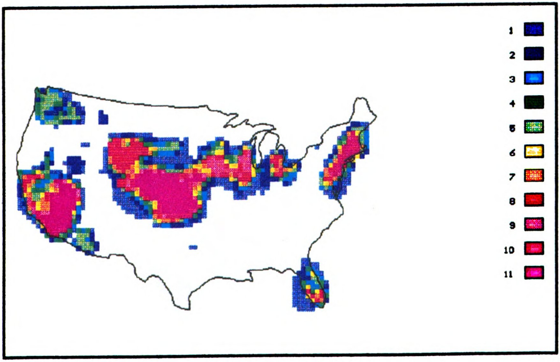


Figure C.4 Frequency counts for Map 4 - Region A

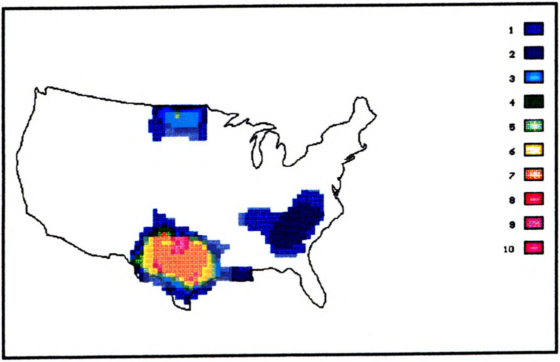


Figure C.4 (continued) - Region B

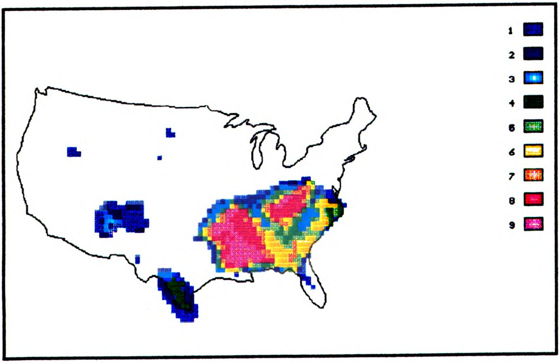


Figure C.4 (continued) - Region C

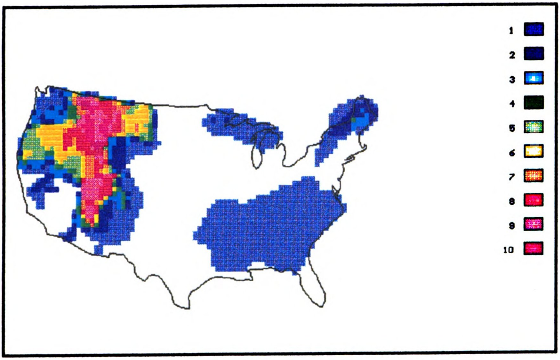


Figure C.4 (continued) - Region D

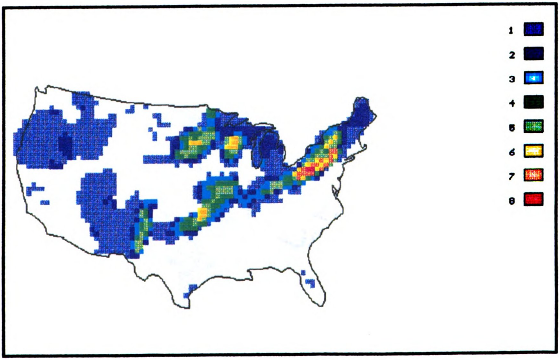


Figure C.4 (continued) - Region E

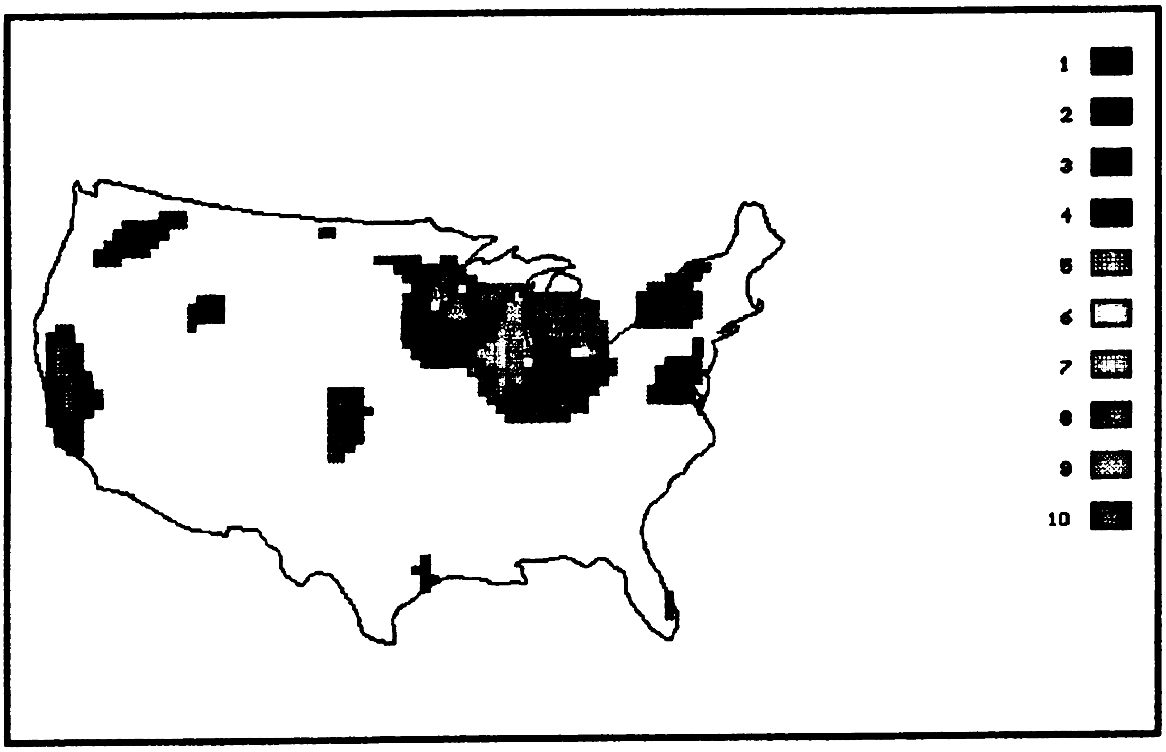


Figure C.5 Frequency counts for Map 5 - Region A

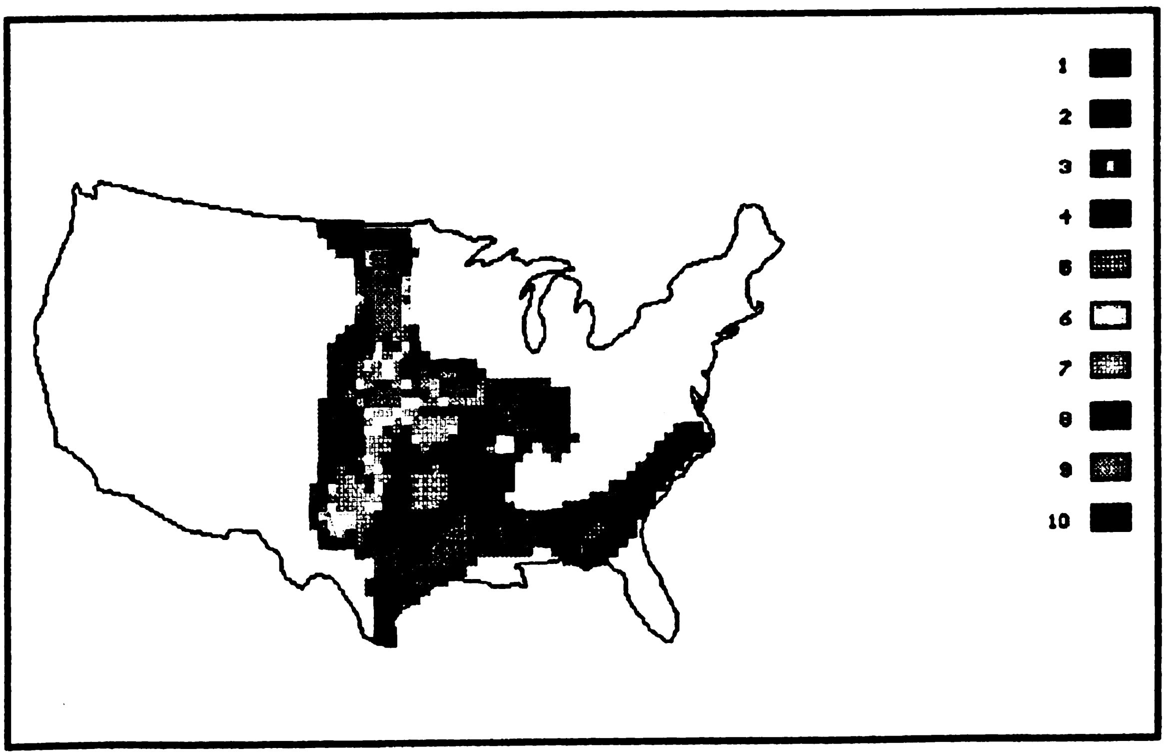


Figure C.5 (continued) - Region B

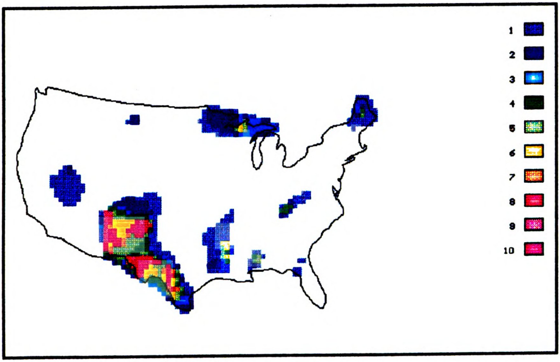


Figure C.5 (continued) - Region C

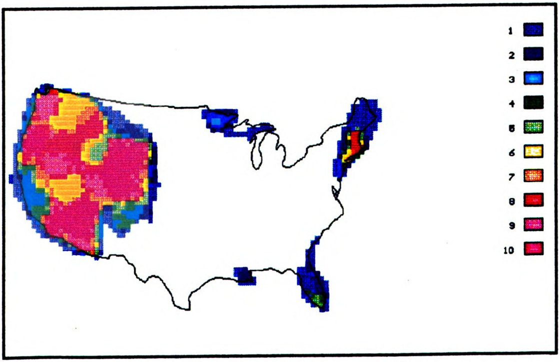


Figure C.5 (continued) - Region D

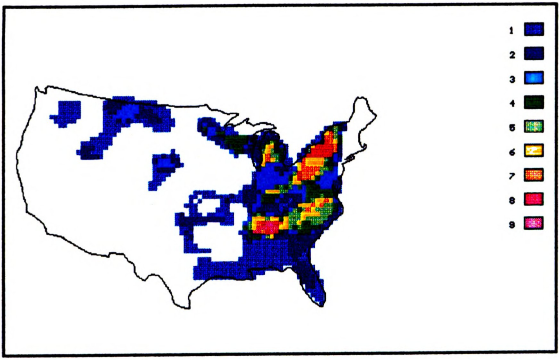


Figure C.5 (continued) - Region E

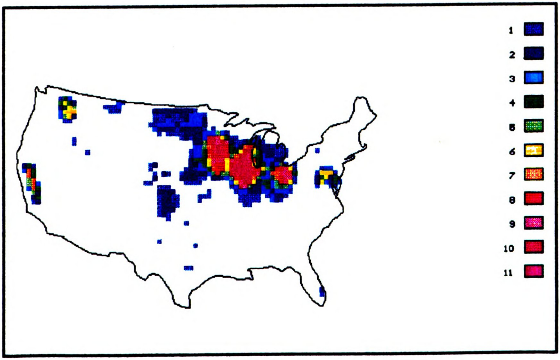


Figure C.6 Frequency counts for Map 6 - Region A

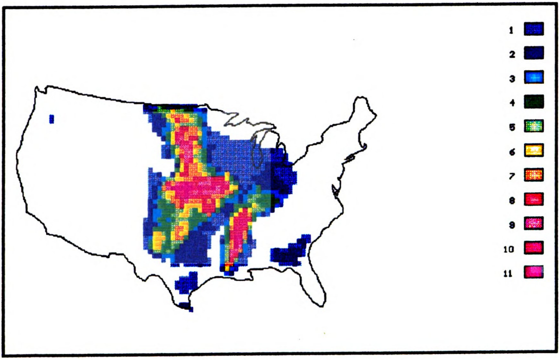


Figure C.6 (continued) - Region B

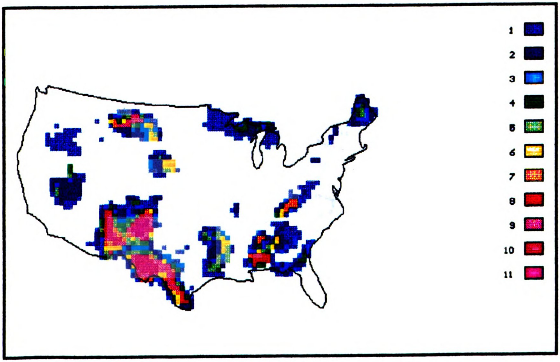


Figure C.6 (continued) - Region C

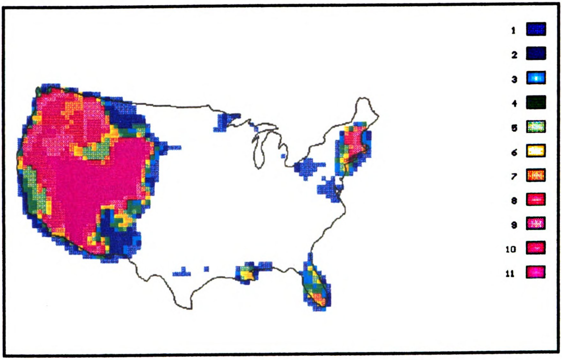


Figure C.6 (continued) - Region D

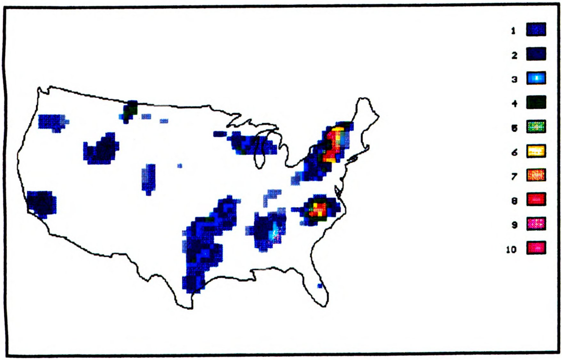


Figure C.6 (continued) - Region E

BIBLIOGRAPHY

BIBLIOGRAPHY

Carstensen, Laurence W., Jr. "Bivariate Choropleth Mapping: the Effects of Axis Scaling." The American Cartographer 13, no. 1 (1986): 27-42.

_____. "A Continuous Shading Scheme for Two-Variable Mapping." Cartographica 19, nos. 3,4 (1982): 53-70.

_____. "Perceptions of Variable Similarity on Bivariate Choroplethic Maps." The Cartographic Journal 21 (June 1984): 23-29.

Eilertsen, Jill P. "The Effectiveness of Color Dot Maps in Region Portrayal." M.A. thesis, Michigan State University, 1980.

Eyton, J. Ronald. "Complementary-Color, Two-Variable Maps." Annals of the Association of American Geographers 74, no. 3 (1984): 477-490.

Groop, Richard E. "Color Dot Density Choropleth Maps, 1992" Unpublished manuscript, Department of Geography, Michigan State University.

Haggett, P. Geography: A Modern Synthesis. Third ed. New York: Harper and Row, Publishers, 1983.

IDRISI Version 4.0, 1992. Clark University, Worcester, Massachusetts.

Jenks, George F. "Contemporary Statistical Maps--Evidence of Spatial and Graphic Ignorance." The American Cartographer 3, no. 1 (1976): 11-19.

_____. "'Pointillism' as a Cartographic Technique." The Professional Geographer 5 (1953): 4-6.

Lavin, Stephen. "Region Perception Variability on Choropleth Maps: Pattern Complexity Effects." Ph.D diss., University of Kansas, 1979.

Lavin, Stephen, and J. Clark Archer. "Computer-produced Unclassed Bivariate Choropleth Maps." The American Cartographer 11, no. 1 (1984): 49-57.

MapViewer Version 1.1, 1992. Golden Software, Golden, Colorado.

McCarty, H., and N. Salisbury. "Visual Comparison of Isopleth Maps as a Means of Determining Correlations Between Spatially Distributed Phenomena." Paper No. 3. Department of Geography, University of Iowa (1961).

Meyer, Morton A., Frederick R. Broome, and Richard H. Schweitzer, Jr. "Color Statistical Mapping by the U.S. Bureau of the Census." The American Cartographer 2, no. 2 (1975): 100-117.

Olson, Judy M. "The Organization of Color on Two-Variable Maps." In Proceedings of the Second International Symposium on Computer-Assisted Cartography, 289-294. Falls Church, Virginia: American Society of Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, 1975.

_____. "Spectrally Encoded Two-Variable Maps." Annals of the Association of American Geographers 17, no. 2 (June 1981): 259-276.

Rader, Charles Peirce. "Representing Regions: Geography, Cartography, and Spatial Understanding." Ph.D. diss., Michigan State University, 1995.

Robinson, Arthur H. "Mapping the Correspondence of Isarithmic Maps." Annals of the Association of American Geographers 52 (1962): 414-425.

Rogers, Jill Eilertsen, and Richard E. Groop. "Regional Portrayal with Multi-Pattern Color Dot Maps." Cartographica 18, no. 4 (1981): 51-64.

Smith, Richard M. "The Development of Black and White Two-Variable Maps: Ongoing Cartographic Research." In Applications of Geographic Research, ed. H. Winters and M. Winters, 145-54. East Lansing, Michigan: Department of Geography, Michigan State University, 1977.

Wainer, Howard, and Carl M. Francolini. "An Empirical Inquiry Concerning Human Understanding of Two-Variable Color Maps." The American Statistician 34, no. 2 (May 1980): 81-93.