DATA QUALITY AND CHOROPLETH MAPS: AN EXPERIMENT WITH THE USE OF COLOR

Diane M. Schweizer and Michael F. Goodchild
University of California Santa Barbara
Santa Barbara,CA 93106-4060
diane@geog.ucsb.edu
good@geog.ucsb.edu

ABSTRACT

Thematic attribute uncertainty is an inherent feature of maps produced from geographic databases since all spatial data are of limited accuracy. To minimize the incorrect interpretation of GIS output, accuracy information needs to be readily available. The purpose of this research is to determine if the connotative implicatons of gray are effective in displaying data uncertainty. The Hue Saturation Value (HSV) color model is tested for its ability to simultaneously map data attributes and corresponding data uncertainty. Saturation is used to symbolize data attributes. Value is used to symbolize uncertainty rankings as the value scale corresponds to the gray scale. An experiment designed using this color theory was administered to 101 students. Results of the experimentation indicate that value, when correlated to grayness, is not an effective means of displaying data uncertainty. The authors suggest the inability of value to represent data uncertainty may be attributable to the relationship between grayness and the value scale defined for this experiment. As a result, the authors further suggest maximum grayness occurs towards the center of the gray scale rather than at an endpoint.

DATA UNCERTAINTY

All spatial data are of limited accuracy as spatial databases are abstractions (Goodchild and Gopal 1989) of actual spatial, locational (thematic) and temporal variation. It follows that thematic attribute uncertainty is an inherent feature of maps produced from geographic databases. Uncertainty arises according to the particular data collection and statistical methods employed in producing a map. Within a GIS, the accuracy of spatial data is changed through such manipulations as scale change and data generalization and such operations as buffering and overlay, among others.

While the degree of uncertainty can be minimized through careful attention to data collection and data manipulation practices, data uncertainty cannot be eliminated. The effective and proper use of GIS output is a function of decision-makers' awareness of the error component of GIS output and the decision-makers' ability to interpret such information. To minimize incorrect interpretation of GIS output, accuracy information needs to be readily available. Possibilities for

the communication of data quality information include the use of metadata and attribute tables. An alternative method is to communicate data quality information visually. Visual techniques include the use of color as a means of displaying data quality information. (Buttenfield 1991, Goodchild 1991, MacEachran 1991, McGranaghan 1991).

According to Robinson (1967), when perceiving color on a map, map readers do not clearly distinguish between psychological reactions to color and the color representation as described by the legend. It follows that an effective method of displaying data uncertainty is to use an idea that triggers the notion of uncertainty in the map users' mind. The concept of uncertainty invokes images of fog and haze, as "it is not clear to me, it is foggy". As such, fog is a metaphor for uncertainty. The color which best represents fog is gray. The purpose of this research is to determine if the connotative implications of gray are effective in displaying data uncertainty.

COLOR

Color, or visible light, can be measured according to three physical properties: wavelength, intensity and purity (Thorell and Smith 1989). Wavelength represents the segment of the visible spectrum being reflected, intensity is the strength of the light and purity is the existence of one or more wavelengths. Humans perceive color according to psychological rather than physical properties. (Hilbert 1987, Travis 1991). Three psychological dimensions of color are hue, the perception of wavelength; saturation, the perception of purity; and brightness, the perception of intensity (Evans 1974, Thorell and Smith 1989). A one-to-one correlation between the physical properties of color to the psychological properties of color does not exist (Thorell and Smith 1989). For example, changing the wavelength of visible light may or may not result in a change in hue. The lack of a one to one relationship between the physical and perceptual properties of light makes it difficult to alter color according to color perception on a computer screen. The RGB (Red-Green-Blue) color model is often used to represent color on computer screens, however, it is not perceptually intuitive.

The Hue-Saturation-Value (HSV) color model is a transformation of RGB space and has been developed according to the three psychological dimensions of color. The HSV color model is represented as a hexacone. Hue is measured as an angle between 0 and 360 degrees about the vertical or value axis. Saturation is measured from 0 at the center of the hexacone to 1 along the perimeter of the hexacone. At a saturation of 1, a color is fully saturated and at a saturation of 0, a color is said to be desaturated and the gray scale exists along the value axis. Value ranges from 0 to 1. A value of 0 is black and a value of 1 is white. The HSV model is considered to be a perceptually intuitive color model (Travis 1991, Thorell and Smith 1989) as humans are able to perceive differences along the three individual components.

Evans (1974) identifies grayness as darkness and suggests grayness increases along the gray scale with black as the limit for increasing grayness. The goal of this paper is to determine if value, as represented by the HSV color model, is an effective tool for displaying data quality information. The structure of the HSV model appears to lend itself to the representation of data quality because of the ease of manipulation along the value axis. A cross-section of the HSV model, with the following parameters: hue equal to 265 (blue), saturation ranging from 0 to 1 and value ranging from 0.3 to 1 (see Figure 1) is examined in this research.

Cross-section of HSV Hexacone

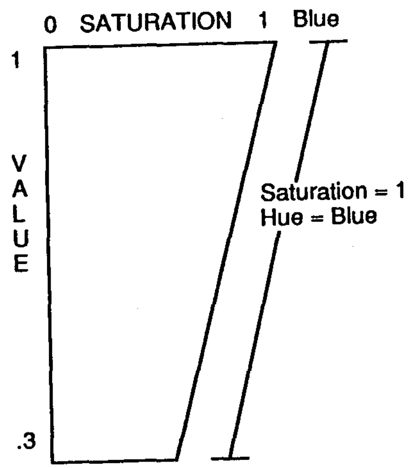

Figure 1

BIVARIATE CHOROPLETH MAPS

The problem cartographers are being faced with when displaying data values and data quality is a bivariate mapping problem. Two distinct sets of variables, as well as their relationship to one another need to be represented in a perceivable way (Carstensen, 1984). (For the purpose of this research, data values shall be referred to as data quantity to prevent confusion between data values and the value component of the HSV model).

Guidelines for color use suggest saturation (Cuff 1972) and value (Cuff 1972, Antes and Chang 1990) are effective in displaying quantitative differences. Attribute data quality and data quantity can both be defined in quantitative terms. Thus, saturation and value are employed to map these two variables.

In this research, value, as described by the HSV color model, is being tested for its ability to map data certainty and saturation is used to

map data quantity. If the use of value is effective in displaying data certainty, the shading schemes for bivariate maps displaying data quality and data quantity can be considered to belong to a separate group from other bivariate choropleth shading techniques. If however, the use of value is ineffective in communicating data uncertainty, data quality and data quantity choropleth maps can benefit from bivariate choropleth shading research.

EXPERIMENT

An experiment was designed to determine if users can effectively interpret data quality when encoded as value and data quantity information when encoded as saturation and displayed simultaneously on a bivariate choropleth map. One hundred and one students verbally screened for color blindness participated in the experiment. Subjects were categorized as experienced or inexperienced map users. Experienced map users were represented by graduate geography students as well as undergraduate students who completed at least one advanced cartography course. Inexperienced map users were represented by undergraduate students from a variety of disciplines taken from introductory level geography classes.

The same bivariate choropleth map of the United States was used for the entire experiment. Participants were informed that all data was simulated and decisions were to be based on the color scheme of the map. To avoid any biases associated with the United States, the quantity variable was identified as an imaginary "tribble" population density. Customary statistical measures of accuracy are described in terms of standard deviation and standard error. For the purpose of this experiment, data quality was presented as a percentage of data quantity to eliminate heteroskedasticity and described as how well the mapped data conforms to actual data.

Although hue, saturation and value are based on the three psychological dimensions of color, increments along the individual axes are not visually equidistant. As such, saturation varied from 0 to 1 and value varied from 0.3 to 1 according to Stevens' equal value gray scale as suggested by Kimerling (1985). Hue was held constant at 265 degrees (blue).

To give the appearance of continuous shading, 15 saturation and 15 value levels were used . The use of 15 value levels and 15 saturation levels requires 225 distinct colors to be displayed on the screen simultaneously. The maximum number of allowable color on the particular hardware device used for this experiment is 256. 225 colors were selected to allow additional colors to be used for other purposes such as text. Continuous mapping rather than discrete mapping was chosen for this experiment as the authors feel continuous mapping provides a more complete understanding of the data being displayed. The concept of choropleth maps without class intervals was introduced by Tobler (1973) as a means of enhancing map readability. Additional research provided by Muller (1979) suggests map readers are able to perceive geographic patterns on continuously shaded maps and can consistently and logically interpret the information being displayed.

The experiment consisted of having students respond to a series of 31 questions administered from IBM PS/2 computers. For each question, subjects were asked to make a selection of either a higher quality value or a higher quantity value from a selection of two states. More specifically, along the bottom of the screen either the phrase "Choose the more certain tribble population density" or "Choose the higher tribble population density" was displayed. The two possible states to chose from were highlighted in white to eliminate requiring prior knowledge of state names. Selections were made using a mouse. The experiment consists of two parts.

The purpose of Part I was to determine if there is an intuitive association between value and data quality. Additionally, results of this portion of the experiment will determine if users were differentiating between saturation and value and if they were consistently associating saturation with one variable and value with the other.

In Part I of the experiment, legend boxes did not appear. Subjects were presented with a series of 16 questions. For each question, subjects were asked to select the state with either higher data certainty or higher population density from a choice of two states. Before the maps appeared, subjects were required to read an information screen explaining the nature of the experiment. The information screen familiarized the subjects with the concept of mapping data quality as it is a relatively unfamiliar concept. Of 90 subjects participating, 40 were experienced map users and 50 were inexperienced.

Part II of the experiment was designed to determine if subjects associated the amount of value with data uncertainty despite the symbolism presented by the legend. This portion of the experiment consisted of a control legend, Legend A and a test legend, Legend B. See Figure 2 for a description of each legend. In Legend A, saturation represented population density and value represented the level of data uncertainty. In Legend B, the axes are reversed. That is, value represented population density and saturation represented data uncertainty. Forty-nine subjects, 20 experienced and 29 inexperienced, were administered control legend A against 52 subjects, 20 experienced and 32 inexperienced, administered test legend B.

Subjects were presented with a series of 15 questions. The legend was constant throughout this part of the experiment. As in Part I, subjects were asked to select the state of either higher data quality or higher quantity from two possible choices.

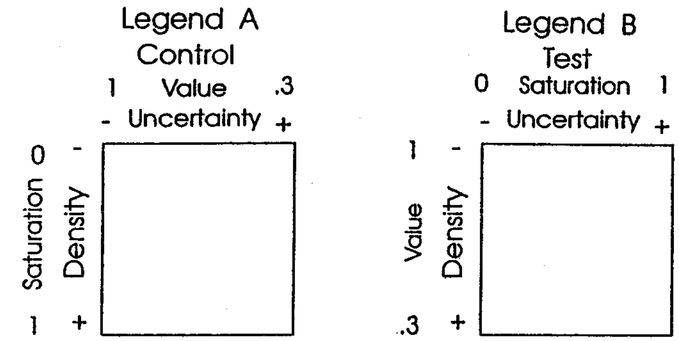

Figure 2

RESULTS

Part I : No Legend

The mean of percent correctly scored for experienced users is 54% with a standard deviation of 21%. For inexperienced users, the mean is 48% with a standard deviation of 20%. The similarity between mean scores and standard deviations for the two populations raises the question of whether or not choices were made randomly. If the selections made were random, a probability of .5 can be assigned to obtaining a correct choice and a probability of .5 can be assigned to obtaining an incorrect choice. Random selection by users would show a proportion of correct choices for each of the 16 questions clustered about .5.

Plots of the proportion of correct choices (see Figures 3 a,b) do not show a distribution clustered about .5 indicating the choices made were not random. The pattern of choices for the two populations resemble each other indicating certain criteria existed when making selections. Questions 10, 11 and 14 were examined to determine if users were attempting to differentiate between saturation and value. For each of these questions, one state had a value of 1 and the other state had a saturation of 0 providing a clear distinction between saturation and value. In other words, one choice appeared blue, with no indication of gray and the other choice appeared gray with no indication of blue. If users were making a consistent association between saturation and value and the two variables, either consistently high or consistently low scores are expected for each of these questions. However, this is not the case, questions 10, 11 and 14 received scores of 68%, 52% and 50%, respectively for experienced users and 56%, 32% and 43%, respectively for inexperienced users.

Proportion of Correct Choices By Question--Experienced Users, Part I

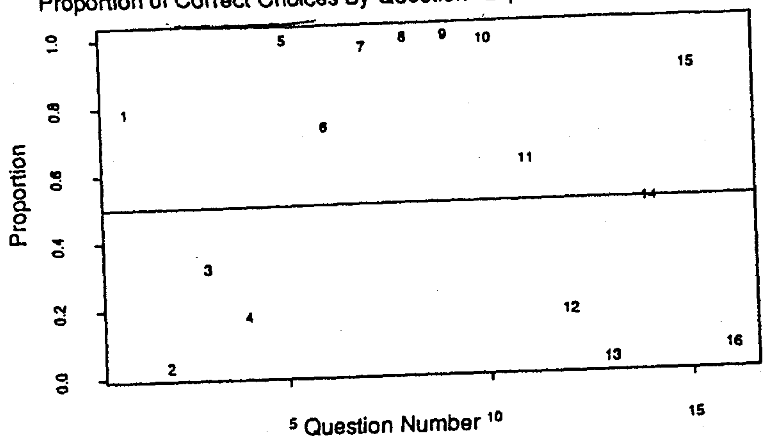

Figure 3a

Proportion of Correct Choices By Question--Inexperienced Users, Part I

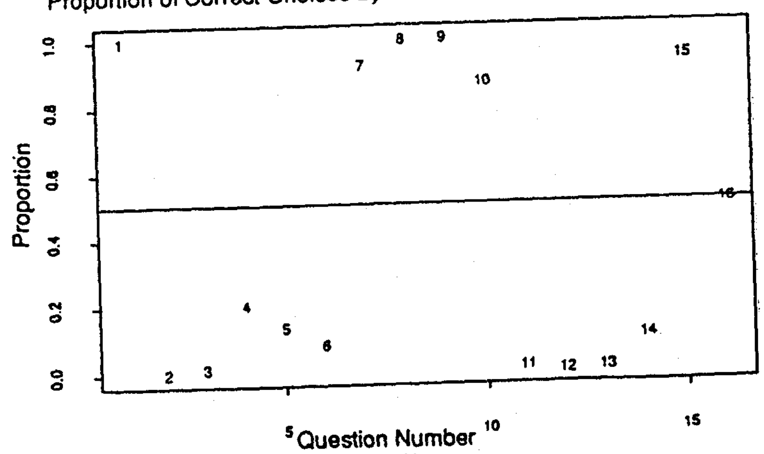

Figure 3b

To determine methods of selection made by both populations, individual questions were examined. The pattern of response between the two populations is evident from the proportions of correct choices of questions 1, 2, 4, 7-10, 12, 13 and 15. Although a legend did not appear in Part I, results were scored according to the legend scheme assigned to the control legend. Questions were categorized as high scoring, low scoring or undetermined. The criteria for the categorization are as follows: if 33% or less of the population answered the question correctly, the question is low scoring; if 67% or more of the population answered the question correctly, it is high scoring; if more than 33% and less than 64% of the population answered the question correctly, it is undetermined.

Of the questions which had similar response patterns between the two populations, questions 8 and 9 are categorized as high scoring

and questions 2 and 13 are categorized as low scoring. The remaining questions, 1, 4, 7, 10, 12 and 15 are undetermined. Examination of the high and low scoring questions suggests users may have been making some selections based on the common cartographic assumption, "dark is more". A general agreement exists within the cartographic community that regions of greater data magnitude should be represented by darker symbols (Cuff 1974, Keates 1973, Robinson et al. 1984, Robinson 1967)

Based on the assumptions that less value is darker than more value (Evans 1974), more saturation is darker than less saturation and value is darker than saturation, all responses could be categorized as one of the following: a) both selections had equal amounts of saturation with differing amounts of value in which case the state with the lower value was termed darker; b) both selections had equal amounts of value with differing amounts of saturation in which case the state with the higher saturation was termed darker; c) the relative amounts of saturation and value were symmetrical about the upper left to lower right diagonal in which case the state with the lower value was termed darker.

The correct choice for question 9 (high scoring) could also be categorized as "darker" . The other high scoring question, question 8, did not follow the pattern of "dark is more". Question 8 asked for the more certain data. Although the incorrect selection is the "darker" state, the correct selection is a bright and vivid blue, with no gray content and appears to be the obvious choice. The questions which received low scores had "lighter" colors representing more.

The next step in this analysis is to determine why the two populations did not follow a similar pattern for the remaining questions, questions 3, 5, 6, 11, 14 and 16. The particular colors chosen to represent the two states for questions 3, 5, and 6 were located close to each other in the control legend thus making it difficult to perceive a color difference. For questions 11 and 14, one choice was gray with no blue and the other blue with no gray and neither state appeared "darker". Thus, these two questions may have been answered arbitrarily. The choices for question 16 both contained value components and saturation components. As is evident from the previous selections, value has not been associated with data quality and thus this selection may have been made arbitrarily. Figure 4 shows where the color symbols would appear on the control legend for each question along with the correct choice. The two possible choices for each question are the endpoints of each line. The correct choice is denoted by a circle and the users' choice is identified with an 'X'. For the undetermined questions, the users' choice is omitted.

As data quality information is not customarily presented on choropleth maps, question type was compared to score to determine if users had more difficulty answering those questions asking for data certainty. Both populations scored slightly lower on those questions asking for data certainty. However, the differences between the two question

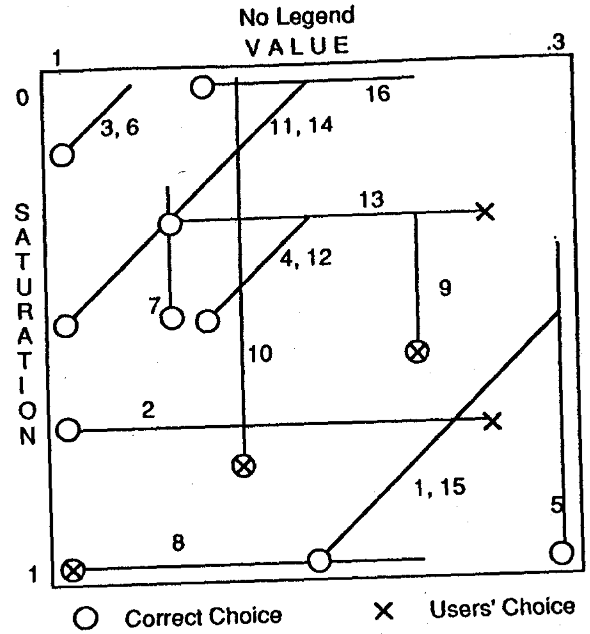

Figure 4

types are not great enough to suggest users are unable to interpret data quality information.

Part II: With a Legend Box

Legend A: Control Legend

A two-sample t-test based on a 95% confidence level indicates there is not a significant difference between the two populations for Legend A. The mean score is 72% with a standard deviation of 23% for experienced users and the mean score is 67% with a standard deviation of 19% for inexperienced users.

As in part I, a similar pattern of responses exists between the two populations. The responses for questions 1-7, and 9-13 identify this pattern. This group of questions can be divided into high scoring questions, questions 2, 3, 6, 9, 10 and 12 and undetermined questions, questions 1, 4, 5, 7, 11 and 13. See Figure 5 for the location of the color symbol on the legend for the two choices for each question.

Of the high-scoring questions, two of the correct selections (2 and 12) can also be categorized as "darker". For the remaining high scoring questions, 3, 6, 9 and 10, the correct choice is "lighter". This indicates that map readers were paying more attention to the legend box than to individual biases. However, since the overall scores indicate some misinterpretation of the legend box, individual biases may have influenced some of the decisions more than the legend.

Legend B: Test Legend

As for Legend A, a two-sample t-test based on the 95% confidence level indicates there is not a significant difference between the two populations. The mean score is 79% with a standard deviation of 19% for experienced users and the mean score is 70% with a standard deviation of 24% for inexperienced users.

The similar pattern of responses between the two populations is not as evident for test Legend B as for the absence of a legend and the presence of control Legend A. The larger difference between the mean scores for Legend B than for Legend A is an indication of the weaker response pattern.

Further examination of the results suggests users were making selections based on the "dark is more" assumption. Based on the test legend, 67% of the correct answers (questions 1, 3, 4, 5, 6, 7, 8, 9, 10, 15) could also be categorized as "darker". Questions 1, 3-10, and 12-14 are high scoring questions and questions 2, 11 and 15 are undetermined questions. Except for questions 12, 13, and 14, all of the high scoring questions have "darker" colors representing more. Of the undetermined questions, questions 2 and 11 have "lighter" colors representing more and question 15 has a "darker" color representing more. See Figure 6 for the location of the color symbol on the legend for the two choices for each question.

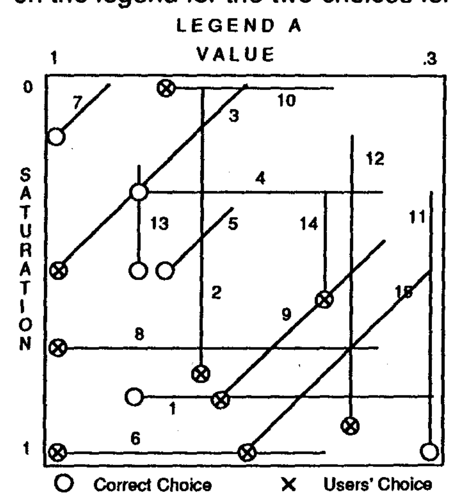

Figure 5

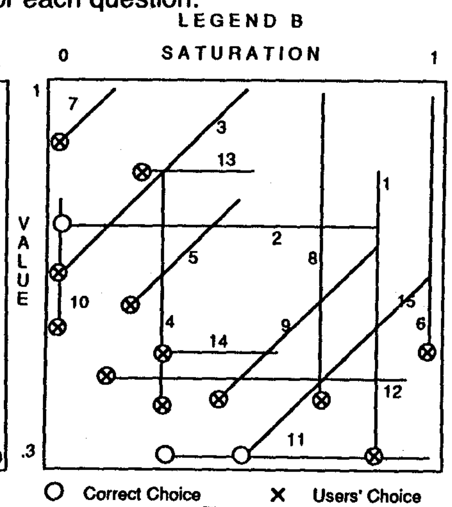

Figure 6

COMPARISON OF LEGEND A AND LEGEND B

An important factor in this work is the comparison of the two legends. If users score high on the control legend and low on the test legend then the assertion can be made that value is effective in displaying data quality. Previous analysis of the individual legend schemes does not support this hypothesis. For both legends, neither consistent high scores nor consistent low scores were recorded for individual questions providing a clear distinction between saturation and value. This indicates that the legends were not easy to interpret and that users were not able to differentiate between saturation and value when making selections.

As there is no difference between the experienced and inexperienced users for the two legends, a two-sample t-test was taken to determine if there is a difference between the scores for Legend A and Legend B. The computed test statistic indicates there is not a significant difference between the two legend types based on a 95% confidence interval.

DISCUSSION

The inconsistency among users' selections in the absence of a legend is not unreasonable. Maps do not customarily display quality information and the introduction of such a variable can be confusing to map readers. Verbal communication with participants indicated the mapping of data quality is an unfamiliar concept and selections were made arbitrarily or according to individual biases. This suggests education of users may improve the effectiveness of displaying data quality.

The addition of a legend significantly improved the results with minor differences according to which legend was used. Thus, there is not an intuitive association between value and uncertainty. The increase in mean scores indicates data uncertainty can be effectively portrayed on bivariate choropleth maps. However, data uncertainty should be treated as would any other variable. Improvements upon current bivariate mapping techniques may provide for a higher level of understanding among users.

The lower scores for the control legend than for the test legend were contrary to what was expected. Examination of the results indicates users may have been making selections based on the common cartographic assumption "dark is more" rather than differentiating between saturation and value. The variable being tested did not affect users' selections.

For the test legend, users were not differentiating between saturation and value and associating saturation with one variable and value with the other. Rather, they were combining saturation and value when making selections for both variables.

Researchers of color vision have determined that the human perception of color is psychological and physiological, not physical. (Travis 1991, Hilbert 1987). Although humans can distinguish between variations in hue, saturation and value, people do not equate color with amounts of these three variables. Thus, individuals perceive color according to individual ideas and concepts. In order for

value to successfully represent data quality, map readers would need to become familiar with this association through repeated use. The learned association between value and data uncertainty would not be successful or practical for cartography. By requiring value to be representative of data certainty, an unnecessary restriction is placed on the cartographer.

CONCLUSION

Although the use of value as described by the HSV model cannot be accepted as an effective measure of mapping data quality, the following conclusions can be drawn from this experimentation.

1. There is not an intuitive association between data quality and value as displayed by the HSV color model.
2. There is not a significant difference in scores according to the particular legend displayed or the level of experience of map users.
3. Although data uncertainty is an unfamiliar concept, there is not a significant difference in how users responded to data quality questions as compared to data quantity questions.

These results indicate users are able to interpret data quality information. As there is not a significant difference of scores according to the particular legend scheme nor the particular variable, bivariate choropleth maps which display data quality and data quantity should not be treated differently than bivariate choropleth maps in general. Thus, maps displaying data quantity and data quality should benefit from bivariate mapping research. Additionally, as the scores increased significantly between the absence and the presence of a legend, the particular legend employed is an important factor to the degree of understanding of map interpretation.

A reason for the lack of an association between data quality and value is the relationship between value and gray. As suggested by Evans (1974), grayness increases along the value scale with black as the limit. From the results of this study, the authors suggest an alternate interpretation of grayness. That is, maximum grayness is achieved toward the center of the gray scale with grayness decreasing from this point towards white and from this point towards black. The point of maximum grayness is not yet defined. In this regard, the representation used in this experiment associates data quality with darkness rather than grayness. An interesting addition to this experiment would be to see how map users' interpretations of data quality information change when mapped according to the definition of maximum grayness occuring towards the center of the gray scale as well as when mapped according to previously defined bivariate mapping techniques and different color models. Results of such experimentation would determine if the outcome from this experiment could be attributed to the particular definition of grayness, the color model or the shading technique.

REFERENCES

Antes, James R. and K. Chang, 1990, "An empirical analysis of the design principles for quantitative and qualitative area symbols", *Cartography and Geographic Information Systems,* Vol. 17, No. 4 pp. 271-277

Burger, Peter and Duncan Gillies, 1989, Interactive Computer Graphics, Addison Wesley Publishing Company, Inc.

Buttenfield, Barbara P., 1991, "Visualizing Cartographic Metadata", NCGIA Technical Paper 91-26.

Carstensen, Laurence W., Jr., 1984, "Perceptions of Variable Similarity on Bivariate Choroplethic Maps", *The Cartographic Journal,* Vol. 21, pp. 23-29

Cuff, David J., 1972, "Value versus chroma in color schemes on quantitative maps", *The Canadian Cartographer,* Vol. 9, No. 2, December 1972, pp. 143-140

Cuff, David J., 1974, "Impending conflict in color guidelines for maps of statistical surfaces", *The Canadian Cartographer,* Vol. 11, No. 1, June 1974, pp. 54-58

Evans, Ralph M., 1974, Perception of Color, John Wiley & Sons, Inc., New York, NY.

Goodchild, Michael, F. and Sucharita Gopal, Editors, 1989, Accuracy of Spatial Databases, Taylor and Francis

Goodchild, Michael, F., 1991, "Position Paper for I7 Specialist Meeting",NCGIA Technical Paper 91-26.

Hilbert, David R., 1987, Color and Color Perception, Center for the Study of Language and Information, Lecture Notes Number 9

Keates, J.S., 1973, Cartographic Design and Production, Longman Press, New York.

Kimerling, A. Jon, 1985, "The comparison of equal-value gray scales", *The American Cartographer,* Vol. 12, No. 2, 1985, pp. 132-142.

McGranaghan, Matthew, 1991, "A view on visualizing spatial data quality", NCGIA Technical Paper 91-26.

MacEachran, Alan M., 1991, "Visualization of Data Uncertainty: Representational Issues", NCGIA Technical Paper 91-26.

Muller, Jean-Claude, 1979, "Perception of Continuously Shaded Maps", *Annals of the Association of American Geographers,* Vol. 69, No. 2, pp. 240-249.

Robinson, Arthur H., 1967, "Psychological Aspects of Color in Cartography", *International Yearbook of Cartography,* 7, pp. 50-61.

Robinson, Arthur H., Randall D. Sale, Joel L. Morrison and Philip C. Muehrcke, 1984, Elements of Cartography, fifth edition. John Wiley & Sons, New York.

Thorell, L.G. and W.J. Smith, 1989, Using Computer Color Effectively, Prentice Hall, Englewood Cliffs, New Jersey

Tobler, Waldo R., 1973, "Choropleth Maps Without Class Intervals?", *Geographical Analysis,* Vol. 5, pp. 262-265.

Travis, David, 1991, Effective Color Displays, Academic Press Inc., San Diego, CA