TWO-VARIABLE COLOR MAPPING ON A MICROCOMPUTER

Cyrus B. Dawsey III
Auburn University
Department of Geography
Auburn, AL. 36849

BIOGRAPHICAL SKETCH

Cyrus B. Dawsey received a PhD from the University of Florida and has spent his professional career at Auburn University where he is an Associate Professor of Geography. His research interests have included economic spatial patterns, migration, allocation models, cartography, and computer applications. Dr. Dawsey has maintained a regional focus in Latin America and Brazil where he spent much of his early life.

ABSTRACT

Though two-variable choropleth maps are useful for analyzing the direction, intensity and spatial pattern of a correlation, they are not frequently used because costs associated with conventional production are made high by color layout requirements and by the limited demand for maps of any particular combination of variables. Microcomputers with color output capability make such maps available to a potentially wider audience. A properly programmed system can create an almost unlimited number of unique two-variable maps at little expense beyond the initial investments. This report describes the development and illustrates the results of two-variable choropleth mapping on the Amiga microcomputer.

INTRODUCTION

A two-variable choropleth map can be a useful device for representing the direction, intensity, and spatial patterning of a correlation between two sets of data. If enumeration area observations for variables 1 and 2 are grouped into n and m classes, a rectangular legend with nXm cells displays all category combinations for the variables arrayed along the x and y axes. Most two-variable maps use color to distinguish patterns, so an increase in the perception of a selected hue usually indicates increasing values on the legend axes. Two diagonally opposite corners of the legend, each corresponding to the highest values for one variable and lowest for the other, present cells that are completely saturated in one of the colors. The other two corners are neutrally toned. Low values for both variables are commonly shown in black and high values for both variables in white.

Direction and intensity of a correlation is given by the position of the prevalent map hues relative to the legend diagonals. If most of the map appears in neutral shades from white to black the variables are positively associated, but the relationship is negative if the majority of the areas are vividly saturated in the two colors. The strength of the correlation is roughly indicated by the percentage of the total area mapped in tones along one of the diagonals.

An advantage of two-variable maps is that they provide a distributional representation of the correlation information. Spatial variations in the direction, intensity, and residuals of a relationship between paired variables are more readable on a map than in a table of coefficients. For example, areas of low and high value for each separate variable as well as areas of low and high joint value are immediately identifiable on a single map.

Though often discussed as a useful technique, two-variable mapping is not widely used (Meyer,1975;Olson,1981). Cost is an important factor. Scholarly publishers usually operate on budgets that preclude extensive use of color, while the better heeled cartography companies print maps for a general readership that is seldom interested in multivariable representation. This limited interest may in turn be a function of the scarcity of two-variable maps. A certain familiarity and history of exposure may be necessary before an average person is able to fully comprehend the patterns and relationships depicted on such maps.

Another factor which may limit the availability of two-variable maps is the inherently vast potential for thematic selection. By general consensus, single-variable maps should depict standard items such as landforms, population, or transportation, so heavy production investment is justified by the guaranteed mass demand. No such agreement exists regarding the selection of combinations of variables to be shown on two-variable maps. The relatively small audience for any given map and the high cost of a full color layout have limited most two-variable cartography to demonstration issues. Full use of this type of map for data analysis is relatively rare.

COMPUTER APPLICATIONS

Computer mapping can overcome some of the conditions that limit the use of two-variable maps. Today, relatively inexpensive systems are capable of displaying multi-hued graphics, and they can be programmed to produce choropleth maps. Once the hardware and software have been purchased, a virtually infinite number of multi-colored maps representing varied themes and areas can be generated at no extra cost. High expense per unique map, the drawback of conventional production, can be avoided. Though physical output (hardcopy) may involve money or time outlays, direct analyses of two-variable maps is available to anyone with access to a properly programmed computer connected to a color CRT.

The remainder of this paper describes some aspects of the development of a two-variable choropleth mapping program for the Amiga computer. The introduction of the Amiga 1000 in 1985 generated much excitement in the microcomputer world (Williams,1985;Anderson,1985). Reviewers described it as a landmark machine with capabilities unavailable in any other comparably priced ($1000 - $2000) system. Its sound chip produced music in stereo and talked in English and other languages; its Motorola 68000 processor provided true multitasking by running applications independently of each other; and its architecture was open for easy expansion. The most attractive feature of the Amiga, however, was its ability to produce colorful graphics. Dealers fondly displayed bouncing balls and multi-hued robots to demonstrate the selection palette of over 4000 colors on a screen resolved at up to 640 x 400 pixels. Furthermore, because the hardware was designed to minimize the time required for generating graphic output, convincing animation effects were possible. While written line commands provide complete access to DOS, a mouse controlled environment of icons and windows was the primary user interface.

An effective choropleth mapping system must bring together 3 basic elements; basemap information, the data values to be mapped, and instructions on how to produce the map. Three program modules were designed to accomplish this on the Amiga. The software allowed for the following. Basemap files were created with the mouse which is standard equipment on the Amiga, and the digitizing source consisted of a map outlined on a transparent sheet affixed over the monitor. The cursor was positioned on the screen at successive map boundary intersections or border direction change points, and each was identified as an x-y coordinate of a line segment endpoint by clicking the mouse button. Keyboard editing controls were incorporated, and, once completed, the basemap file was stored to disk.

Development of the second and perhaps most important element, a data base of the variables to be mapped, was straightforward. The number of enumeration units, their names, and the associated data values were stored in a sequential file. Data input to the file was facilitated by a routine which displayed an outline map successively highlighting each area while requesting input of the associated values. The variables to be shown on a two-variable map were stored as separate files.

The third step in assembling a choropleth mapping package consisted of writing a module through which a user could actually produce a map. The developed program included a menu driven main loop that called several subroutines to accomplish the following:

1. Read in a basemap file.
2. Read the data files.
3. Compute value range interval categories.
4. Set colors to the interval categories.
5. Draw the outline map.
6. Paint each area unit with the proper color.
7. Quit and close all files.

Steps 1 and 2 were simple disk filing operations which read the previously digitized basemap and the data into appropriate arrays. The identification of the value range intervals in step 3 involved inputting the number of categories (2 to 4 for each variable) and then selecting a "manual" or "automatic" option. "Manual" allowed the user to identify the value range break points while "automatic" caused a programmed routine to develop the categories. The "automatic" algorithm compensated for skewed data and attempted to include approximately the same number of observations within each interval while maintaining a standard range width for all but two of the intervals. Top and bottom interval ranges (ie. from the minimum value to the lowest break point and from the maximum value to the highest break point) were allowed to be of variable width and different from that of the standard width of the other interval ranges.

Color selection (step 4) was from a pre-established gradient of light to dark shades in a red hue for the first variable and blue for the second. Once mapped, the hue bases could be exchanged or combined with green. Choropleth mapping was performed by a routine which painted each area unit polygon of the displayed basemap with a color value matched to the proper range interval for the combined variables.

RANGE INTERVALS AND COLOR SETTINGS

Computerized two-variable choropleth mapping is little different from single variable mapping except for two tasks: setting the class intervals and determining the color hues. Range intervals based on a bivariate distribution can be developed in several ways. Eyton described the methods and produced two maps combining income and education levels for counties in the United States (Eyton,1984). The first map included classes based on a simple rectangular 3X3 red and blue frequency count matrix, while the second reduced the matrix dimensions to 2X2 and added a category corresponding to areas computed to be within a 50% equiprobability ellipse about the income-education relationship. Olson described criteria for effective interval development and legend representation, while Tobler and Eyton showed that classification is not necessary (Olson,1972;Tobler,1973;Eyton,1984).

Intervals based on simple frequency counts are used for this study. The class generating algorithm attempts to minimize the variation in the number of observation grouped into each category while maintaining a uniform interval, but it does not compute probability estimates.

Minimizing the inter-class count variance for the separate variables on a two-variable map does not ensure representation in all categories. Perfectly correlated variables, for example, might have equal count representation among the classes for each of the separate variables, yet most of the cells of the rectangular matrix would include no observation at all. Of the 9 cells in a 3X3 legend, only the colors of the 3 diagonal cells would appear on the map.

The Amiga raster display maintains color information in a maximum of 5 bitplanes, and binary coding of the bitplanes allows complete control of up to 32 different colors on the screen at one time. The colors are produced by separately setting hexadecimal red, green, and blue software switches, so the displayed colors can be drawn from a potential 4096 different shades (16X16X16). Several programming languages implemented on the Amiga provide access to the color controls and bitplanes through commands to library functions imbedded in the executive and operation systems. BASIC, for example, includes a Palette command for selecting the colors directly. Added display coloring on the Amiga can be achieved with sprites, a half-bright mode and a unique hold and modify control. These options allow concurrent display of the full 4096 colors, but their access and control is more difficult.

Several color settings were examined in reference to accepted criteria for good legend display (Olson,1972). Manipulation of two color registers (red and blue) proportionally to the two variables while holding the other register (green) constant created colors which provided appropriate gradient change, but the diagonal neutral tones did not range from black to white. If the green setting was at a minimum value, the tones along the diagonal of positive correlation progressed from black (lowest class for both variables) to deep lavender (highest for both variables). The more pleasing black to white diagonal transition was achieved by increasing the setting for the non-active color (green) proportionally to the oblique distance from the corner of lowest combined values (black). The black to white diagonal shift, usually requiring complementary colors, was thus obtained with primary colors.

CONCLUSION

As demonstrated, two-variable mapping is easily performed on a personal computer and at practically no cost once the initial investments have been made. Though the Amiga is not the only nor the best color graphics system, it is very inexpensive when compared to computers with comparable capabilities. What was done on the Amiga can also be accomplished on the Atari ST, Mac II, or the new PS II line from IBM. Two-variable mapping on a computer allows a cartographer to overcome the drawbacks of high cost and fragmented thematic demand which are associated with conventional production processes. The new technology available to geographers makes accessible previously impossible options for analyzing spatial data.

REFERENCES

Anderson, J. (1985) "Amiga," Creative Computing, 11 (9): 32-41.

Eyton, J. R. (1984) Complementary Color, Two-Variable Maps," Annals of the Association of American Geographers, 74: 477-90.

Meyer, M., Broom, F., and Schweitzer, R. (1975) "Color

Statistical Mapping by the US Bureau of the Census," The American Cartographer, 2: 100-17

Olson, J.M. (1972) "Class Interval Systems on Maps of Observed Correlation Distributions," The Canadian Cartographer, 9: 122-31

Olson, J. M. (1981) "Spectrally Encoded Two-Variable Maps," Annals of the Association of American Geographers, 71: 259-76.

Tobler, W. R. (1973) "Choropleth Maps Without Class Intervals ?" Geographical Analysis, 3:262-65

Williams, G., Edwards, J., and Robinson, P. (1985) "The Amiga," Byte, 10 (8): 83-100