THE ORGANIZATION OF COLOR ON TWO-VARIABLE MAPS

Judy M. Olson
Boston University

There are many types of maps on which color may serve a useful function. Color may be used to enhance or clarify information, it may be used to increase the visual appeal of a map, and in some cases it may be an indispensable ingredient in the system of encoding complex information. An example of the use of color to encode complex information is its use on two-variable "cross-maps" such as those currently produced by the U.S. Bureau of the Census. (See Figure 1, page 251.) 1/ This technique presents two quantitative variables on the same map, and the symbol system, a fairly complex one when considered in detail, is a set of color combinations. The construction of a color scheme for two-variable maps must be based on several logical considerations, or restraints, such that the scheme forms a coherent system. It is the development of this set of color combinations and its logic relative to the map message that is discussed in this paper. Such a color system represents a specific employment of color and is presented as a sort of "case study."

In attempting to construct an outline of the ideas involved in the color scheme, it is necessary to keep clearly in mind the objective, or the intended message, of the map. It is assumed that a map of this type is multipurpose in its intent in that it should convey information about the relationship between the variables (particularly the strength of relationship and the spatial distribution of the evidence) and at the same time maintain the identity of the two separate variables. We can then list several specific messages that should therefore be obtainable from the map. What such a list of messages demands is a color scheme with a very extensive, built-in figure-ground system such that the reader can think, for example, "category two on X" and see it, "anomolies" and see them, and so on. While it cannot be hoped that everything about the distributions can be clearly encoded, this type of map must at least show a larger number of specific messages than either (a) two separate maps of the distributions or (b) the map of residuals which deals with the relationship but does not keep the two variables differentiated. If the cross-map fails to communicate its multiple messages, there is not much point in using the method, except perhaps for its eye catching novelty value.

Given the general purposes of these maps, then, there is one more assumption on which this paper rests. It is that the color scheme will determine the look of the map rather than vice versa. We might develop the color scheme on the basis of the spatial distribution of category combinations but for mechanical as well as theoretical reasons, the color scheme will be developed more or less independently of the spatial distribution here and it is assumed that the map's spatial message will follow.

The considerations in the development of the scheme, then, include:

(1) All colors must be distinguishable. If they are not, there is no reason to divide the data into so large a number of intervals. The colors must be distinguishable not only in the legend but in the map context as well; they must look different from one another on the map and any patch of color on the map must be matchable with its corresponding box in the legend. The alternative to this constraint is a map whose interpretation is based on quite different concepts of pattern recognition than those of traditional map reading, a subject that deserves attention but is too complex to include here.

(2) At the same time that the boxes must be distinguishable from one another, the transitions of colors should progress smoothly and in a visually coherent way. An arbitrary arrangement of sixteen distinguishable colors would make the map impossible to comprehend, and inappropriate "blocking" of colors, i.e., inappropriately similar groups of neighboring colors, may also be misleading. (See Figure 13, page 264.) Ideally, the set of colors should be such that all neighboring pairs look "equally different" from one another.

(3) The individual categories of each distribution should be visually distinguishable or coherent and the two distributions as a whole should be separable from one another. This implies that there is a coherence along each row and along each column and that rows have some element of unity, columns some different element of unity. (See Figure 14, page 264.)

(4) The arrangement of the colors as presented to the map reader in the legend should probably correspond to the arrangement of a scatterdiagram of the pairs of distributions. (See Figure 15, page 264.) In other words, the color representing low values on both distributions should be in the lower left, the color representing high values on both maps should be in the upper right. Perhaps the cells of the appropriate diagonal (the positive is shown in Figure 15) should be emphasized to aid the reader in recognizing the relationship to a scatterdiagram.

(5) Since readers generally associate darker values of colors with higher numerical values in the distribution, tones should progress from lighter to darker on both variables. The combinations should result in approximately equally-dark values along diagonals, progressing from lightest in the lower left to darkest in the upper right. The numbers in Figure 16 (page 264) represent relative degrees of darkness.

(6) Because two-variable maps are intended to show the relationship between the variables, it is important that the reader be able to detect easily the positive and negative evidence on the map. Since the correspondence of extreme values weighs most heavily on the relationship, the extreme categories (legend corners) should stand out the most. Colors which are either visually saturated (i.e., pure in appearance) or are dark in tone might therefore be used for corner cells while relatively desaturated tones might occupy the center cells. (See Figure 17, page 265.)

(7) To show positive and negative residuals, that is, areas where the values on the second variable are either considerably greater or considerably less than would be expected from the value on the other variable, there should be a coherence in the triangles of cells above and below the main

diagonal. (See Figure 18, page 265.) (Again the figure refers to the case of a positive overall relationship). Residuals are an inherent characteristic of a relationship between variables and the prominence of residuals on the map should reflect the degree of relationship.

(8) Also related to the intention of conveying relationship, positive diagonals (lower left to upper right) and negative diagonals (upper left to lower right) should have visual coherence, particularly the two main, or longest, diagonals but also the secondary ones. (See Figure 19, page 265.) This coherence allows the reader (theoretically, at least) to see strong and weak evidence of a relationship. To alleviate the difficulty of achieving coherence in both the positive and negative diagonals, an alternative is to establish strong coherences for one set of diagonals and then orient the legend according to whether the general relationship is positive or negative. Such an alternative leads to contradictions of several other criteria, however, such as "the darker the more" (See #5).

(9) Because of the difficulty of mentally sorting large numbers of colors in the legend, a legend such as the four-by-four might be visually subdivided into a smaller number of categories, say two-by-two. We might call this a nested categorization. In the example in Figure 20, (page 265) yellow and purple tones would provide evidence of a positive relationship, green and red of a negative relationship, and an equal mixture of all four would indicate no relationship between the two variables. The advantage of this type of four-by-four scheme over a simple two-by-two (yellow, red, purple, green) is that there remains considerable detail about the individual distributions at the same time that the simplicity of a two-by-two scheme is maintained for purposes of depicting the relationship. Whether readers can indeed utilize a matrix large enough to be so nested remains an open question, however.

(10) The color scheme should relate to the data in such a way that the map relationship reflects as closely as possible the statistical relationship between the distributions. To achieve this, it is necessary that units whose values fall close to the line of average relationship on the scatter-diagram also fall in the cells of the main diagonal of the color legend. If we cannot assume that a dominance on the map of the colors in the diagonal indicates a strong relationship, then we must accept these maps as inefficient not only in conveying but even in containing information about the relationship between the variables.

We might be tempted to base the diagonal cells on a regression line, but because there are two regression lines (X on Y and Y on X) and often no reason to choose one of these as opposed to the other, it is necessary to choose some more neutral line of relationship. Equal standard deviation units provide comparability between the values on the two variables and the scatter about the line of equal standard deviations is indeed a function of the closeness of relationship. Hence, the diagonal cells should fall along the line of equal standard deviations and class breaks should be at equivalent standardized values on the two variables. (See Figure 21, page 265.) 2/

(11) To simplify the mechanics of producing the colors, each separate category on each individual distribution should be assigned a specific combination of screens and/or hues. This minimizes the number of printing plate burns (i.e., exposures) to produce the scheme. In the example in Figure 22A category one on the map of the first variable is solid yellow, category two is medium yellow plus a light red, etc., and for the other variable, specific shades of blue (including zero blue) have been assigned. This scheme is similar to the scheme used on the Fort Worth map shown earlier except that on the Fort Worth map yellow was assigned to the category which is blank here. Assuming one color separation sheet for each category on each variable, this scheme (See Figure 22A, page 266) requires three burns onto each of the three color plates -- yellow, red, and blue -- a total of nine. If we choose simply some combination of screen values for each of the sixteen boxes such that as many other principles are observed as possible, we may have as many as 48 burns (3 colors times 16 cells) unless we adopt a different system of color separations than separation by class interval. In addition, we would not be able to demonstrate so easily the basic idea of the two-variable map. With the simpler system, such as the nine-burn or that used in the Fort Worth map, we can demonstrate as follows (Figure 22B): first the map (or legend) of variable one is presented in the red and yellow tones (note the vertical orientation); then the second map (or legend) in tones of blue (note the horizontal orientation). Then the crossed version is presented as a direct combination of the two separate sets of colors, noting that the yellow remains yellow where it does not overlap with other colors, while the yellow plus blue in the upper left results in green, red plus blue in the upper right results in purple, etc.

(12) While the mechanics are certainly of practical importance, the structure of the scheme (i.e., its development as a combination of colors on two individual maps) must also be comprehensible to the map user and the resulting combinations should <u>look</u> like combinations of the specific colors involved. To illustrate, note the cell second from the left and second from the bottom in the final scheme in Figure 22B. While it is indeed a combination of the particular medium yellow and light red of variable one with the light blue of variable two, it <u>looks</u> out of place. [Note: Since problems such as this are highly dependent upon the specific inks and paper used, Figure 22B (page 266) may or may not effectively illustrate the idea.] Either some adjustments must be made in the specific screen value choices for the legends for each of the two individual variables or that specific cell must be adjusted to fit the scheme visually. (See Figure 23, page 266.) Mechanically, it is simpler to adjust the screen choices for the second category of each distribution but there is nothing inherently wrong with adjusting a single cell; the map is, above all, a visual communication device and must be designed as such.

(13) 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. Five categories, on the other hand, can be smooth and aesthetically appealing but lead to both mechanical problems and severe problems of distinguishability, especially in the map context. (See Figure 24, page 266 .)

(14) Finally, there is the problem of the rectangular nature of the legend cells and the effect on the map message. This results in a two-fold problem, one of which is that the reader expects to see only the side-sharing legend colors next to one another in a map representing a highly-organized spatial arrangement of two variables. Ideally, however, if the two individual distributions are highly related, one should expect the colors of the diagonal to dominate the map and these cells share only a point in the legend rather than a side. This is a rather subtle problem but an important one so far as the reader's ability to grasp the map as a whole is concerned. Secondly, the message of relationship is affected and areas of equal residual value (i.e., of equal distance from the line of equal standard deviations) may not look "equally residual" on the map. For example, a map unit may fall into an off-diagonal cell even though its X and Y values are very close to the line of equal standard deviations as illustrated by the observation indicated with a large dot in Figure 25A (page 266). To alleviate such a problem, perhaps we should consider a color scheme arranged parallel to the line of equal standard deviations while at the same time visually maintaining the categories for the two individual distributions. (Figure 25B, page 266). Such a scheme might include two distinctly different patterns such as stripes and coarse dots to represent the two distributions but for simplicity, in Figure 25B they are simply labelled with appropriate category numbers. A combination of hue and value change indicates closeness to the line of equal standard deviations: yellow indicates very close; light red and light blue, slight negative and positive residuals, respectively; dark red and dark blue, the larger residuals. This represents quite a change in the technique itself and hence, shall be pursued no further here.

CONCLUDING COMMENTS

These numerous considerations do not complete the list of things that should be kept in mind in constructing a color scheme for a two-variable map. They simply illustrate the complexity of the problem. The real test of utility of such maps comes not in simply developing a scheme of colors which theoretically obeys the maximum number of constraints; the development of the color scheme is basically a preliminary step. The real test is whether people can deal with the maps themselves and can accurately extract both spatial and statistical information. No amount of examination of separate components of the color scheme will tell us whether a two-dimensional array of colors can be grasped as a coherent whole to say nothing of whether the structure of that scheme coincides with the structure of information that is intended to be grasped from the map, thus enabling the reader to assimilate its information.

The use of color to encode relatively complex sets of data requires considerable forethought. The list of constraints to be kept in mind is long and some will undoubtedly be sacrificed in final decisions, but while there is little hope of an ideal system, we can at least strive for the best possible.

Two final notes I would like to include are first, that while some schemes are certainly better than others, it takes at least as much effort to consciously develop the "worst" scheme as to develop the "best." That is encouraging; even when our schemes have gross faults, if we have exerted any positive effort at all we find that they do contain a lot of information. Whether our readers can extract that information is yet to be studied. Secondly, these two-variable color-coded maps probably require a reasonable amount of mental effort and perhaps more than other representations of the same information. But the compelling fascination of the coding scheme as well as the initial attraction of attention induced by their colorfulness probably means that we at least have a captivated reader. While our concerns with good mapping should indeed take us beyond such a consideration, the captivated reader is certainly a good start in map communication.

ACKNOWLEDGEMENTS

This paper has resulted from a summer of employment at the U.S. Bureau of the Census. Many of the ideas presented here have developed through discussions with colleagues at the Bureau and in academia; the lack of references reflects only the difficulty of crediting the appropriate persons and not a lack of appreciation on the part of the author.

REFERENCES

1. Barabba, Vincent P. "The Graphic Presentation of Statistics," pp. 89-103 in The Census Bureau: A Numerator and Denominator for Measuring Change, Technical Paper 37, U.S. Government Printing Office, Washington, D.C., 1975. (Describes the rationale for such maps in decision-making and describes the overlapping color code.)

2. Olson, Judy. "Class Interval Systems on Maps of Observed Correlated Distributions," The Canadian Cartographer, Vol. 9, No. 2 (Dec. 1972), pp. 122-131.

3. Broome, Frederick R. "Micrographics: A New Approach to Cartography at the Census Bureau," Proceedings of the Fall Convention, ACSM, Orlando, Florida, 1973, pp. 265-283. (Describes the techniques used to produce color separations using the computer and the computer-output-on-microfilm unit.)