

# American Statistical Association

## 1999 Proceedings

*of the*

## Section on Statistical Graphics



Papers presented at the  
Annual Meeting of the American Statistical Association,  
Baltimore, Maryland, August 8–12, 1999,  
under the sponsorship of  
the Section on Statistical Graphics

American Statistical Association ■ 1429 Duke Street, Alexandria, VA 22314

# TABLE OF CONTENTS

## Invited papers by Topic

<b>1. Student Award Winners in Statistical Graphics</b> <b>Organizer/Chair:</b> <i>Michael C. Minnotte</i> , Utah State University	
A Direct Search Method for Projection Pursuit Guided Tour Optimization. <i>Jean-Yves Pip Courbois</i> .....	1
Do Bullets Have Fingerprints? Preliminary Graphical Partitioning by Trace Element Concentration for Bullet Fragment Data. <i>Jennifer Schumi</i> .....	11
Dynamic Graphical Methodology for Multivariate Response Surface. <i>Tingchuan Wang and Javier Cabrera</i> .....	19
<b>2. Network Visualization</b> <b>Organizer:</b> <i>Linton Freeman</i> , University of California <b>Chair:</b> <i>Robert Newcomb</i> , University of California – Irvine	
Representing Two-Mode Binary Data as Galois Lattices. <i>Vincent Duquenne, Annick Le Pape, Mortada Moad, and Karim Teman</i> .....	27
Visualizing Networks with Spring Embedders: Two-mode and Valued Data. <i>Lothar Krempel</i> .....	37
Visualizing Social Groups. <i>Linton C. Freeman</i> .....	47
<b>3. Using Results from Perception and Cognition to Design Statistical Graphics</b> <b>Organizer/Chair:</b> <i>Daniel B. Carr</i> , George Mason University	
Color Use Guidelines for Data Representation. <i>Cynthia A. Brewer</i> .....	55
<b>4. Interactive Statistical Graphics-Mathematical Concepts &amp; Implementation</b> <b>Organizer:</b> <i>Adalbert Wilhelm</i> , University of Augsburg <b>Chair:</b> <i>Matthias Schonlau</i> , NISS/AT&T Laboratories	
A Data Model for Interactive Statistical Graphics. <i>Adalbert F.X. Wilhelm</i> .....	61

## Contributed Papers by Topic

<b>1. At the Border of GIS and Statistical Graphics</b> <b>Organizer/ Chair:</b> <i>Jean-Yves Courbois</i> , Oregon State University	
Using GIS, Spatial Statistics and Visualization to Investigate Environmental Justice. <i>Jeremy L. Mennis</i> .....	71

## COLOR USE GUIDELINES FOR DATA REPRESENTATION

Cynthia A. Brewer

302 Walker Building, The Pennsylvania State University, University Park, PA 16802

**Key Words:** Color, Maps, Visualization, Perception

### INTRODUCTION

Cartographers have been pressed into making use of color and other graphic variables to present our statistical data because we essentially lose location, which is so useful in graphing, to geographic location. A map is a multivariate graph with latitude and longitude already claiming the x and y axes. So, we rely on symbol shape, size, orientation, arrangement, texture, focus, and color to symbolize data values at locations on maps. As color monitors and color printing become permanent fixtures in our offices, researchers from all domains are finding uses for color symbolization that go beyond decoration, and objective guidelines for color use are helpful in many analytical endeavors.

The bare bones of cartographic guidance on color are to use hue to show categorical differences and use lightness to show ordered differences. This is a straightforward approach, but it does require that you be able to systematically separate these two aspects of color. When you decide to use a pure yellow symbol, for example, you have selected a *yellow hue* that is *light* and *fully saturated*. All three of these aspects of the color affect how it will function in your graphic. Color is a three dimensional phenomenon and you will get more benefit from it in your graphic work if you use all three dimensions with care. Additional considerations are to accommodate colorblind readers and anticipate contrast effects. My goals with this paper are to familiarize you with a series of inter-linked research results on color use and also to encourage you to leave behind some assumptions about color that you may have acquired from as long ago as your childhood.

### HUE AND COLOR MIXING

*Hue* is the perceptual dimension of color that we associate with color names, like red, green, blue, and yellow. Dominant wavelength is a similar concept from physics. The rainbow or spectrum arrays saturated hues in wavelength order from long-wavelength reds to short-wavelength blues: red, orange, yellow, green, blue. I have never been quite sure what the hues "indigo and violet" in Newton's spectrum are (many of us learned the spectral mnemonic ROY G BIV as children), and there is some evidence that he selected these additional color names to parallel the seven-step musical scale in addition to categorizing his color perceptions. Purples and magentas are not colors of the

rainbow. They result from mixing red and blue from opposite ends of the spectrum. White sunlight has a full range of visible wavelengths in it, but our TVs and computer screens get by with a reduced set of light primaries for mixing all the other hues: Red, Green, and Blue (RGB).

Magenta, mentioned above, is not one of the basic color names you learned as a child, but you have probably encountered this pinkish red (a red with no yellow) as you work with desktop printer inks or as you communicate with publishers. Cyan (a greenish blue) is another hue making its way into our lexicon from the realm of printing. The primary colors for printing are Cyan, Magenta, and Yellow (CMY). All other hues can be mixed with these ink (or pigment) primaries. Printers often use a black ink with CMY to simplify printing black text and lines and to print dark colors more accurately (rather than mixing black using only CMY).

RGB and CMY? But didn't you learn that red, yellow, and blue were *the* primary colors when you were mixing paints as a child? Let's agree to forget that little piece of information. It will just confuse you in this modern era of computer monitors and color laser printing. This set of colors is, however, akin to the perceptual concept of *unique hues* that do not look like mixtures of other hues: Red, Green, Yellow, and Blue. You may find yourself gravitating toward these unique hues as you develop color codings because they are clearly different from each other. RG and YB are also the *opponent hues* on which our eye-brain color vision system is based. In addition to red, green, yellow, and blue, the other basic colors named in all fully-developed languages are pink, purple, orange, brown, gray, white, and black.

### LIGHTNESS AND SATURATION

As I mentioned above, Hue is only one of three perceptual dimensions of color. If you are using color in analytical work, you also need a good understanding of *lightness* and *saturation*, the other two perceptual dimensions. Lightness is the most important of these dimensions for data representation. Lightness is a relative measure, describing how much light appears to reflect from an object compared to what looks white in the scene. Its relative character makes it a different measure from related terms like Brightness, luminance, and Intensity. Another word for lightness is Value, but that term becomes confusing in quantitative work if you are also describing data values. Saturation is a measure of the vividness of a color, and there are a host of

related terms with slightly varied definitions: Chroma, colorfulness, purity, and intensity (again). Some of these terms are from perception and the scientific realm of psychophysics, some are from the realm of the physics of light, and still others (like shades, tints, and tones) are from art.

Why do I bring up all of these terms? Because, regardless of the context in which you work with it, color is basically a three-dimensional phenomenon. RGB and CMY both have three dimensions but mix color in different media (light vs. pigments). Your software may present you with a color palette based on any number of additional color systems: HLS, HSV, HSB, HVC, L\*J, LAB, IHS... Notice that they always have three dimensions. Which systems will help you in making statistical graphics? The ones that organize the perceptual dimensions of color most accurately. I will return to discussion of color systems after I describe color schemes for graphics.

#### TYPES OF COLOR SCHEMES

When people read your color graphs and other data graphics, they are working with perceptual dimensions of color, even though you may have specified your colors using a mixture system like RGB. Your readers are seeing and thinking about color as 'light desaturated blues,' 'dark saturated oranges,' 'dark grays,' etc. Thus, you can make the most of your graphics by using these perceptual dimensions in ways that parallel the logical structures in your data to allow its organization to be readily perceived. Four scheme types are the basic building blocks for color representation: sequential, qualitative, binary, and diverging. These basic types are described below, and they are illustrated with four example keys (map legends) highlighted in black in Figure 1.

#### Sequential Schemes

The most basic guidance is to use lightness to represent ordered data. Ordered data can be ordinal (ranked) data or the more sophisticated interval or ratio scales of numerical data. Generally, darker colors are used to represent higher data values; light-to-dark for low-to-high. For example, a graphic showing pollution levels would usually show high levels in a dark color and progress through to lighter colors as pollution levels decrease. I have termed this type of progression a *sequential* color scheme. You need not strictly follow the rule of higher-darker, especially when making graphics to compare variables. For example, it is easier to see the relationships between education and poverty if sequential schemes are organized so that both high education and low poverty are represented with dark colors. The crucial part of the guidance is to pair a monotonic sequence in the data with a lightness sequence.

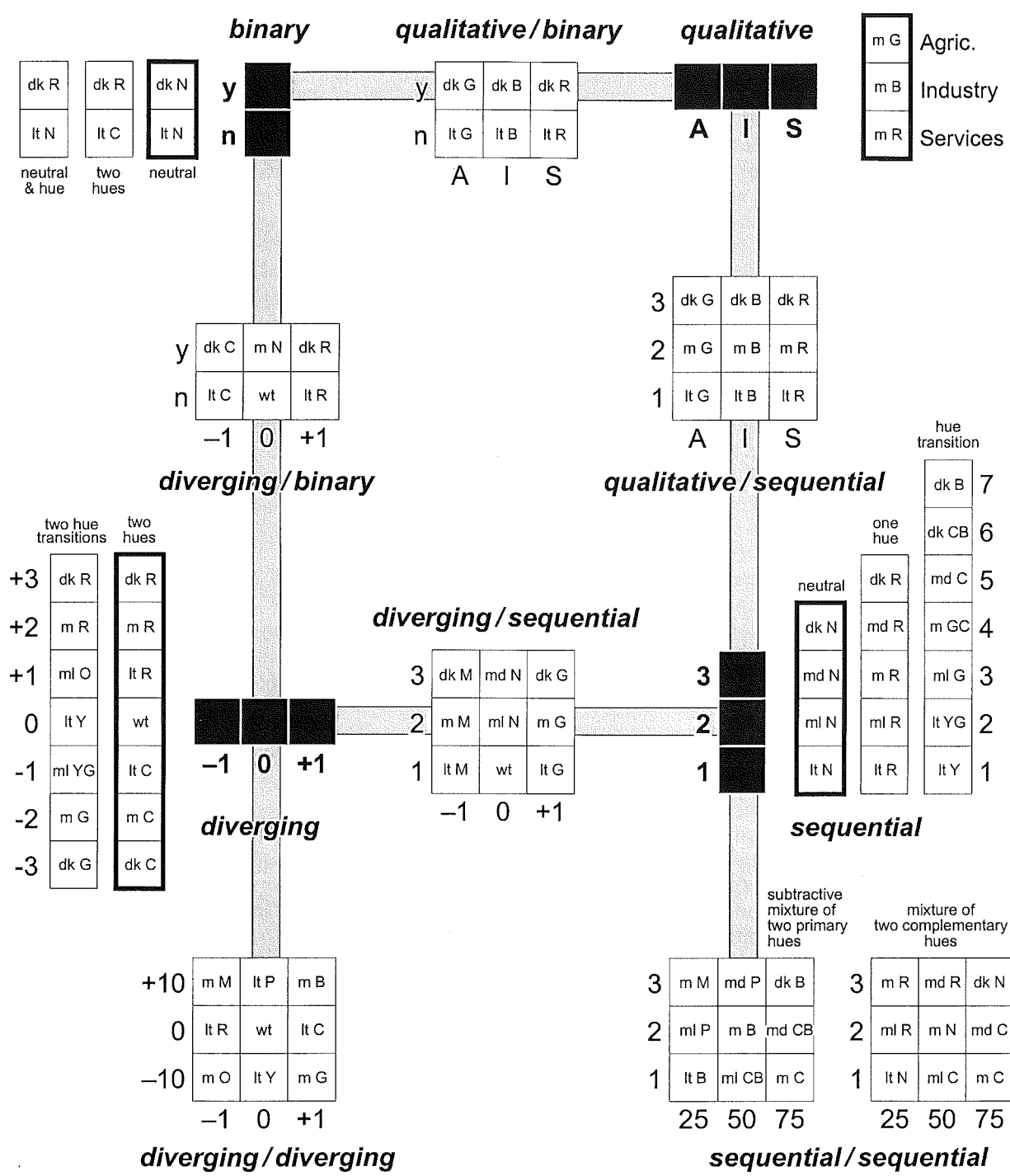
#### Qualitative and Binary Schemes

Categorical differences in data are usually represented with differences in hue. For example, types of government spending, such as military, education, and healthcare, are categories that could be shown with hues like red, green, and blue. Cartographers have termed this a *qualitative* color scheme. Pay attention to the lightness and saturation of the hues you choose for a qualitative scheme. If each category is equally important, you will want each to be fairly similar in its contrast with the background and with each other. For example, a pure yellow will not be as visible on a white background as red, green, or blue. *Binary* schemes are a special case of qualitative schemes for which either a lightness or hue difference (or both) is appropriate for the two categories.

#### Diverging Schemes

Variable comparisons can be accomplished with multiple graphics or can be facilitated by mapping *differences* within a single graphic. Again, careful use of hue and lightness can make this type of analytical graphic quite successful. For example, industrial pollution could be examined by symbolizing increases and decreases in acid rain over a given time period. These data are double-ended and they also have a crucial midpoint in the data range where, in this case, there has been no change in acid-rain levels. A *diverging* color scheme that emphasizes the meaningful midpoint in the data with a light color and then diverges to two different hues works well for this type of graphic. An example scheme would present high increases in dark red, moderate increases in medium red, low increases in light red, negligible change in white, low decreases in light blue, moderate decreases in medium blue, high decreases in dark blue. Emphasizing the two extremes with darker colors and emphasizing the midpoint with the lightest color clearly parallels the structure of the data. In addition to representing 'no change or difference,' a suitable midpoint that might be emphasized in the data range is the mean, median, zero, or other threshold value.

Spectral schemes, or rainbow schemes, are popular in scientific visualizations and news media graphics such as a daily weather page. Unfortunately, this scheme is often misused as a sequential scheme with poorly used lightness differences. The most informative use of a spectral scheme is as a diverging scheme. Let dark red and dark blue mark the extremes in the data and position light yellow to emphasize a meaningful midrange (such as no change in acid rain levels). The full sequence is familiar: dark red, red-orange, orange-yellow, yellow, yellow-green, green-blue, and dark blue, for example. People like the multi-hue character of the scheme, and the variety of hues also helps distinguish symbol categories in the graphic.



**Figure 1. Example color sets for typology of map color schemes**

*Adapted from Brewer 1994 (Proceedings of the International Society for Optical Engineering, Vol. 2171).*

**Relative Lightness**  
 wt: white  
 lt: light  
 ml: medium-light  
 m: medium  
 md: medium-dark  
 dk: dark

**Hue**  
 R: Red  
 O: Orange  
 Y: Yellow  
 YG: Yellow-Green  
 G: Green  
 GC: Green-Cyan  
 C: Cyan  
 CB: Cyan-Blue  
 B: Blue  
 P: Purple  
 M: Magenta  
 N: Neutral (no hue)

Structuring the lightness sequences in the scheme to parallel the characteristics of the data produces an enlightening visualization of the data.

### Two-Variable Schemes

These basic types of schemes—sequential, qualitative, binary, diverging—may be combined to show two related variables on maps or to add two additional variables to a graph. There are six useful combinations of the one-variable schemes: sequential-sequential, sequential-qualitative, binary-qualitative, binary-diverging, sequential-diverging, and diverging-diverging. For example, stage of cancer at diagnosis (sequential) could be represented with healthcare provider (qualitative) on a graph with axes for age and income of a sample of patients. Darker colors of one hue in the lower-right portion of the graph would show that older and poorer patients have later-stage diagnoses with a particular provider. Examples of each scheme are shown with European economic data maps at <http://www.essc.psu.edu/~cbrewer/ColorSch/SchHome.html>. A black-and-white representation of example map keys and their interrelationships is shown in Figure 1 (a color version of this figure is at the above website).

### Additional Issues

Saturation can be systematically varied to represent ordered data, but people are not accustomed to accurately comparing saturation levels, especially between hues. There are also usually few perceivable steps available in saturation differences. Therefore, I don't recommend using saturation as the primary symbolization for a data set. However, ignoring saturation can produce some mighty confusing graphics, where individual colors stand out strongly from other symbols for no apparent reason. Attempt to use saturation to either bolster a lightness sequence in an ordered manner (such as light-desaturated to dark-saturated) or emphasize categories that tend to have smaller symbols in the graphic.

Hue for qualitative data and lightness for quantitative data should also not be taken as strict and exclusive guidelines. Quantitative color schemes may include plenty of hue variation, but they should first and foremost be obviously ordered by lightness. For example, a part spectral scheme of yellow, light-orange, medium-red, dark red is a common sequence for ordered data. It uses different hues, but they are also obviously ordered by lightness which makes the sequential nature of the data obvious to the person reading the graphic. Qualitative schemes benefit from smaller variations in lightness, and saturation, so that colors are readily discerned, but the use of lightness should not erroneously suggest that the data are ordered or that one category is more important than another.

### PROBLEMS IN COLOR APPEARANCE

People who are colorblind can still see lightness differences and can see a fairly wide range of hue differences. Approximately four percent of the population have some degree of color vision impairment (approximately eight percent of men and less than one percent of women). 'Red-green colorblindness' is the most common type of impairment, but this type includes confusion of other hue combinations as well (other pairs of hues will not look different, such as pink and blue). Expected color confusions can be rigorously modeled in color order systems, but these results are difficult to apply without color measurement equipment. Guidelines specified by color names (rather than color measurements) eliminate a wider range of color combinations than necessary, but they are usable in the context of designing statistical graphics. The following pairs of hues are not confused by people with the most common types of color vision impairments: red-blue, red-purple, orange-blue, orange-purple, brown-blue, brown-purple, yellow-blue, yellow-purple, yellow-gray, and blue-gray. These ten color pairs are from a total of 36 pairs of basic color names, so many pairs are confusing. For example, any combinations among red-orange-yellow-green are potentially confusing.

I discussed the popularity and usefulness of spectral schemes in the previous section, but these schemes also contain numerous colors that may be confusing to colorblind readers. Keep in mind that colorblind readers can see lightness differences, so a well designed yellow-orange-red scheme will still be perceived as a lightness sequence (color vision impairments offer another reason to work carefully with lightness steps between hues in sequential schemes). Skipping yellowish-greens and using a lightness sequence of bluish greens, blues, and purples completes a spectral diverging scheme with hues that are not confused with the oranges and reds of similar lightness.

Color appearance is affected by context. Small colored objects will be more difficult to identify than larger colored areas. Thus, you will be able to distinguish fewer colors with small point symbols or thin lines. Different surroundings also change the appearance of a color. Small color areas tend to appear more similar to their surroundings because of a perceptual process called assimilation. Conversely, contrast between a larger patch of color and its surrounding color will enhance the difference between them in a process called simultaneous contrast. For example, a color of medium lightness will look darker on a light background and lighter on a dark background. A gray will look greenish on a red background (its opponent complement) and a green will look yellowish on a blue background. These perceptual interactions between colors can change perceptions of individual

symbols substantially so they may not match corresponding colors in the key (saturation is particularly susceptible to change).

There is little that can be done to prevent contrast effects from occurring because the data distribution being represented controls the positions of the symbols. The best approach is to look carefully at the colors in the data graphic. Do not select colors by comparing them only in the key where they are seen in one order on a uniform background. It is also important to look at your selection of colors in the final graphic format (as a projected slide or overhead transparency, color inkjet print, lithographic proof, color photocopy, etc). Be sure you can identify examples of each color symbol and can tell colors apart throughout the graphic.

### COLOR PREFERENCE

I recommend that you do not become overly concerned about which colors your audience will like. Everyone seems to have an opinion about color aesthetics, and members of your audience will undoubtedly have differing opinions based on their own color preferences. There has been a substantial amount of loosely structured research on color preferences. Regardless of context, it seems that most people like blue and don't like yellow, but that is overly simplistic guidance for use in multi-color contexts. People also like graphics with many colors, so focus your attention on organizing the perceptual dimensions of color so that your data is presented clearly, whether or not you have picked everyone's favorite colors.

### MIXING PRIMARIES

We want to control the perceptual dimensions of color, but we are often working with color mixture systems of RGB or CMYK (K for black) to specify colors. Basically, these systems work in opposite ways. Increasing amounts of CMY ink mix to darker and darker colors (these are the subtractive primaries) and increasing amounts of RGB light mix to produce lighter colors, with all three mixing to white (these are the additive primaries). I will provide some basic rules of thumb for mixing CMYK that may assist your color design work;

1. Set hue using *proportions* of the two higher percentages of CMY (Example: 20%C with 40%Y is approximately same hue as 50C-100Y)
2. Set lightness using overall magnitude of CMK percentages, since Y will stay light (Example: 15M-5K is lighter than 30M-10K)
3. Set saturation using the lowest percentage of CMY or use K (Example: 20M-20Y is more saturated than 20M-20Y-5K or 20M-20Y-5C)

4. Create systematic perceptual changes with systematic percentage changes (Example: 5-15-35-65 will look more evenly spaced than 5-10-55-65)
5. Equal percentage steps do not look like equal visual steps; use bigger steps in higher percentages (Example: 5 to 15 looks more different than 80 to 90)
6. Do not use all four inks: desaturate/darken with either K or the least percentage of CMY (Example: 20C-40Y-8K will be easier to work with than 20C-5M-40Y-5K)

### PERCEPTUAL COLOR SYSTEMS

CMYK color mixing, as described above, requires a fair bit of practice, and alternative solutions are to select colors using a perceptual color space or to limit yourself to a palette of colors provided on a color chart offered by the software. Some of the most rigorously designed color systems from color science are Munsell, CIELAB, OSA-UCS, and NCS. All of these systems use spectral order, with end points connected through non-spectral purples, to define the order of hues around a central vertical axis defined by one of the dimensions akin to lightness. Saturation or a similar measure increases as you move away from the central lightness axis. Each system also claims to be at least partly perceptually scaled, so that equal distances in color space produce equal color difference perceptions. Munsell uses hue, value, chroma (HVC) structured in cylindrical coordinates. The CIELAB system, standardized by the Commission Internationale de l'Eclairage, uses rectangular coordinates with RG and YB opponent hues defining perpendicular axes intersecting at the vertical lightness axis. The Optical Society of America Uniform Color Scales (OSA-UCS) has a sophisticated lattice geometry that provides 12 perceptually equidistant neighbors for each color and can be sliced on multiple planes to yield numerous well-structured color sequences. The Swedish Natural Color System (NCS) also has a cylindrical structure that is based on perceptual mixtures of the four unique hues and white/black.

Systematic paths through perceptual color systems produce systematic color schemes with pre-determined perceptual characteristics. A simple example of using Munsell for color scheme design is demonstrated by selecting a set of symbol colors for a diverging scheme. Without going into the details of Munsell color notation, you could choose common hues for all colors in each of two sequences above and below the data midpoint. For example, we used a brown (5YR) and blue-green (5BG) hue pair for the *Atlas of U.S. Mortality*. Each hue was used with the same lightness and saturation steps (value steps of 4/, 6/, 8/ and chroma steps of /8, /6, /4) that converged on a light gray median class (N 9/). This systematic perceptual structure takes

a lot of fiddling in CMYK but is effortless in Munsell. The CMYK percentages to print these colors on a particular printer are much less systematic and were arrived at through rules of thumb and a dose of trial and error: 40M-100Y-47K, 20M-55Y-18K, 6M-17Y-2K, 2K, 22C-6Y, 61C-20Y-15K, 100C-35Y-40K. Tektronix was promoting a Munsell-like HVC interface for a while and LAB is showing up in illustration software now. If you want to make ready use of perceptually ordered color in data graphics, push the software vendors to integrate the perceptual color order systems into your visualization tools.

Computer science offers a few poorer cousins to these perceptual spaces that may also turn up in your software interface, such as HSV and HLS. They are easy mathematical transformations of RGB, and they seem to be perceptual systems because they make use of the hue, lightness/value, saturation terminology. But take a close look; don't be fooled. Perceptual color dimensions are poorly scaled by the color specifications

that are provided in these and some other systems. For example, saturation and lightness are confounded, so a saturation scale may also contain a wide range of lightnesses (for example, it may progress from white to green which is a combination of both lightness and saturation). Likewise, hue and lightness are confounded so, for example, a saturated yellow and saturated blue may be designated as the same 'lightness' but have wide differences in perceived lightness. These flaws make the systems difficult to use to control the look of a color scheme in a systematic manner. If much tweaking is required to achieve the desired effect, the system offers little benefit over grappling with raw specifications in RGB or CMY.

In conclusion, I encourage you to be a critical consumer of color information, color systems, and color statistical graphics. Organize the perceptual dimensions of color in your graphic to parallel the logical orderings in your data, and choose color tools that make this task easy.

#### BIBLIOGRAPHY OF RELATED RESEARCH PAPERS BY CYNTHIA A. BREWER

*I prepared this paper for an invited presentation in the Theme Session titled "Using results from perception and cognition to design statistical graphics" organized by Dan Carr. At Dan's request, I have synthesized my previous research papers. These papers are listed below, and each includes a review of literature by other researchers working in these areas.*

- MacEachren, A.M., C.A. Brewer, and L.W. Pickle, "Visualizing Georeferenced Data: Representing Reliability of Health Statistics," *Environment & Planning A*, Vol. 30, No. 9 (September 1998), pp. 1547-1561.
- Brewer, C.A., "Spectral Schemes: Controversial Color Use on Maps," *Cartography and Geographic Information Systems*, Vol. 24, No. 4. (October 1997), pp. 203-220.
- Brewer, C.A., A.M. MacEachren, L.W. Pickle, and D.J. Herrmann, "Mapping Mortality: Evaluating Color Schemes for Choropleth Maps," *Annals of the Association of American Geographers*, Vol. 87, No. 3 (September 1997), pp. 411-438.
- Brewer, C.A., "Evaluation of a Model for Predicting Simultaneous Contrast on Color Maps," *The Professional Geographer*, Vol 49, No. 3. (August 1997), pp. 280-294.
- Olson, J.M., and C.A. Brewer, "An Evaluation of Color Selections to Accommodate Map Users with Color Vision Impairments," *Annals of the Association of American Geographers*, Vol. 87, No. 1 (March 1997), pp. 103-134.
- Brewer, C.A., "Guidelines for Selecting Colors for Diverging Schemes on Maps," *The Cartographic Journal* Vol. 33, No. 2 (December 1996), pp. 79-86.
- Brewer, C.A., "Prediction of Simultaneous Contrast between Map Colors with Hunt's Model of Color Appearance," *Color Research and Application*, Vol. 21, No. 3 (June 1996), pp. 221-235.
- Brewer, C.A., and K.A. Marlow, "Color Representation of Aspect and Slope Simultaneously," *Proceedings, Eleventh International Symposium on Computer-Assisted Cartography (Auto-Carto-11)*, Minneapolis, October/November 1993, pp. 328-337.
- Brewer, C.A., "Color Use Guidelines for Mapping and Visualization," Chapter 7 (pp. 123-147) in *Visualization in Modern Cartography*, edited by A.M. MacEachren and D.R.F. Taylor, 1994, Elsevier Science, Tarrytown, NY.
- Brewer, C.A., "Guidelines for Use of the Perceptual Dimensions of Color for Mapping and Visualization," *Color Hard Copy and Graphic Arts III*, edited by J. Bares, Proceedings of the International Society for Optical Engineering (SPIE), San José, February 1994, Vol. 2171, pp. 54-63.
- Brewer, C.A., "Review of Colour Terms and Simultaneous Contrast Research for Cartography," *Cartographica*, Vol. 29, No. 3&4 (Autumn/Winter 1992), pp. 20-30.
- Brewer, C.A., "The Effect of Color on the Perception of Map Scale," *Student Honors Competition Winning Papers*, proceedings edited by E. Wingert, Cartography Specialty Group Publication #3, AAG, Toronto, April 1990, pp. 1-12.
- Brewer, C.A., "Color Chart Use in Map Design," *Cartographic Perspectives*, No. 4 (Winter 1989-90), pp. 3-10.
- Brewer, C.A., "The Development of Process-Printed Munsell Charts for Selecting Map Colors," *The American Cartographer*, Vol. 16, No. 4 (October 1989), pp. 269-278.