Weaving versus Blending: a quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color

Haleh Hagh-Shenas, University of Minnesota, Victoria Interrante, University of Minnesota, Christopher Healey, North Carolina State University, Sunghee Kim, Gettysburg College

Abstract—In many applications, situations arise in which researchers would like to be able to understand the individual values of, and relationships between, multiple related scalar variables defined across a common domain. Several strategies have been proposed for representing data in these situations. In this paper we focus on strategies for the visualization of multivariate data that rely on color mixing. In particular, we seek, through a series of controlled observer experiments, to establish a fundamental understanding of the information-carrying capacities of two alternative methods for encoding multivariate information using color: color blending and color weaving. We begin with a baseline experiment, in which we assess participants’ abilities to accurately read numerical data encoded in six different basic color scales defined in the Lab color space. We then assess participants’ abilities to read combinations of 2, 3, 4 and 6 different data values represented in a common region of the domain, encoded using either color blending, in which a single mixed color is formed via linear combination of the individual values in Lab space, or color weaving, in which the original individual colors are displayed side-by-side in a high frequency texture that fills the region. Although our present experiments employ census data, in which a single value of a particular variable applies to a moderately sized region of the domain, our ultimate objective is to facilitate the effective representation of multiple co-located continuous distributions, in which each variable typically assumes a slightly different value at every point across the domain; our use of very high frequency textures in the present study is intended to lay the foundation for these subsequent investigations. A third experiment was conducted to clarify some of the trends regarding the color contrast and its effect on the magnitude of the error that we observed in our earlier experiment. Current results indicate that most participants are reasonably well able to infer the values of individual components in a blended color when the number of components is 2, but that, for most people, this ability degrades rapidly as the number of components increases beyond 2. In contrast, when the component colors are represented side-by-side in a high frequency texture, instead of being blended, most participants’ ability to infer the values of individual components is significantly improved, relative to when the colors are blended, particularly when more than 2 colors are used, and even when the individual colors subtend only 3 minutes of visual angle in the texture. However the information-carrying capacity of the color weaving approach also has its limits. We found that participants’ ability to accurately interpret each of the individual components in a very high frequency color texture typically falls off as the number of components increases from 4 to 6.

We found no significant advantages, in either color blending or color weaving, to using color scales based on component hues that are more widely separated in L a b color space. On the contrary, we found some indications that extra difficulties may arise when opponent hues are employed.

Index Terms—color perception, multi-variate visualization

I. INTRODUCTION

A successful design retain the observers focus on the data and does not distract or confuse them by the design. It avoids confusion or clutter while presenting the detail and complexity of the data. The effective layering of the data is a difficult task in multivariate Visualization. Unfortunately the combined presence of various layering results in an interaction among layers and wrongly implies the colors and patterns which do not exist in the original information. It is thus imperative that the layers are kept as distinct from each other as possible and if they are combined, they do so in a perceptually comprehensible manner. The layering of information must also provide coherence and precision by remaining faithful to the proper relationship among layers of information. Thus different pieces of information can be visually compared to each other.

Small particles are a set of thumbnail graphics which are presented in one page and represent dimensions of one phenomenon. This characteristic makes them a valuable candidate for multivariate visualization because they facilitates comparison and contrast among these aspects and enhance dimensionality. They could also be used to facilitates multiple aspects of the data through ghosting of several displays where each display depicts one dimension of the information at hand.

In the section dedicated to previous work we explain the most related researches to ours especially in cartography as one of the oldest fields which has devoted a considerable amount of investigation to develop and evaluate two dimensional mediums for representing segmented and continuous data. We specifically explain in more details those from which we drew our design strategies. We also reference research studies related to color mixing and multivariate visualization involving color and we will refer to the most effective methods that are commonly used in cartography and visualization for portraying two or three segmented variables on a map.

In cartographic design two general categories of map reading tasks have been identified. The first group pertains to Direct acquisition. These are questions of value discrimination and value estimation or questions involved with conceptual and
deep structure information. The second category is recall and recognition. This group of questions is concerned with tasks that are performed without the map in view. Throughout our experiments we deal with segmented data presented in several maps and in order to start with basic questions we employs data acquisition questions.

On the whole, the visualization designers should keep in mind that as MacEachran[1] pointed out no one type of design facilitates all types of map uses equally well and the designers must decide which message they wish to emphasize.

The following sections present our effort in understanding the error which happens in reading the blended colors when overlaying multiple variables versus visualizing the separate layers side by side. Sections III explains our baseline experiment which was designed to assess participants ability to accurately read numerical data encoded via the intensity of a single displayed color. Section IV presents our Experiment which was designed to quantify the differences among the three conditions for visualizing multivariate data when variables were overlapping in a region: blending the colors, representing the overlapping variables by letting the colors coexist in that area with a small size high frequency pattern and again with a bigger size high frequency pattern. Section V follows experiment IV and seeks to investigate the effect of color contrast on the introduced error under two of the above conditions. Finally, section VI explains our conclusion and future works.

II. PREVIOUS WORK

The following studies have employed one or both types of map reading questions in designing their tasks.

Shelton and Glimartin [1], investigated how quickly and accurately map readers viewing choropleth maps on a high resolution monitors were able to identify to which class an areal unit on the map belonged. They provided cartographers with empirical guidelines regarding what level of map reading accuracy might be expected for choropleth maps. The maps in question had between 4 and 8 classes and were produced in shades of gray, green or Magenta.

As expected, increasing the number of classes on the map led to a decrease in the accuracy rate and an increase in the reaction times of participants. Accuracy rate ranged from 91.9 percent for four class maps to 68.2 percent for eight-class maps (averaged for all 3 colors). Hue also affected the accuracy rates and the reaction times. The best accuracy results , 84.5 percent averaged over all number of classes,was obtained with achromatic maps. On the other hand maps shaded with magenta had the least accuracy rate, a rate of 72.8 percent.

In [] Muller asked observers to visually categorize the areas of high, medium, and low densities shown on a continuously shaded choropleth map. The perceived patterns followed closely the three-class categorization derived from an statistical generalization. These results lead to a belief that map readers could see and reorganize elements of a no-class map in a consistent and logical fashion, however these findings proved to be controversial.

In an extensive study[], Mersey investigated the effectiveness of color in symbolizing information on thematic maps.

A large number of subjects evaluated the effectiveness of various graded color schemes. These test maps employed six distinct color schemes and four different number of data categories). 10 tasks were included in the questionnaire, which attempted to duplicate realistic map use tasks dealing with both specific and general map information. Some tasks were performed by conferring to the map directly where others were completed from memory. The research emphasized the complex interactions among several mapping variables: the choice of color scheme, the number of data classes presented on the map and the nature of the task to be performed with maps, more comprehensive than other similar studies before. In particular Mersey investigated the validity of three hypothesis: that map effectiveness decreases if the number of categories on map goes up, that map effectiveness increases as color scheme on the map becomes increasingly ordered, and the nature of the relationship between map effectiveness and either map complexity or color scheme varies depending on the type of tasks.

This research emphasized the importance of asking participants of map reading studies a set of realistic looking questions and showing them realistic looking data. Petchneik [] recognized the critical effect of the meaningfulness of the map content on user’s motivation. A central interest point in cognitive research is decoding the mental processing of the thematic information encoded on a map. Apparently, such cognitive information processing occurs when the maps are meaningful to the user.

In one of the earliest work in multivariate visualization, Eilersten and Groop [] developed the multi pattern sparse colored dot map to portray multiple distributions in the same map. A perception test was employed comparing the relative effectiveness of the multi pattern dot map to single distribution dot maps by asking participants to do various map reading tasks. Analysis of the participants responses was based on the consistency with which region boundaries within a map were found and the internal accuracy of regions identified. Results suggested that three variable color dot maps ware at least as effective as several single black and white dot maps in portraying the regions.

In scientific visualization, several interesting methods have been employed for showing multivariate data with color. Examples of these methods are using various types of glyph for segmented data, and combining color and texture for multivariate visualizations which has been extensively studied in several articles including [Assisted Navigation of Large Information Spaces, Large Datasets at a Glance].

Urness and his colleagues [] introduced the concept of color weaving in the context of flow visualization. In this study the authors presented a new method for simultaneously representing multiple co-located colors through coloring the strokes of a LIC texture for the purpose of flow visualization. Such textures were then used to illustrate the relevant flow data. The perceptual effectiveness of this method and the complication in reading the colors introduced by the gray scale changes in the LIC is still unstudied.

P.E. File (1980) found that when two colors were successively presented the time taken to determine whether the
two colors differed increased as the similarity of the colors increased.

When people read the colored graphics, they are thinking in terms of perceptual dimensions such as saturated green or dark grays. Thus, a successful color design can utilize these perceptual dimensions and map them to the logical structures of the data to allow its organization to be readily perceived [Cynthia].

In [1] Cynthia Brewer describes a set of color use guidelines for visualization in the field of cartography. She specifically presents an efficient method for visualizing two overlapping variables for maps. These guidelines can be applied to choropleth thematic maps, filled isoline maps, and qualitative areal-extent maps.

Depending on the nature of the data she suggests four methods of binary, qualitative, sequential and diverging color schemes to represent two co-located variables in a two dimensional area. Categorical differences, for example, can be shown by a diverging color schemes that emphasizes the meaningful midpoint in the data.

In order to present three variables in cartography, some researchers including Dorling [1] utilizes a triaxial graph to visualize the combined representation of the election census data for three consecutive years. Each side of a triangle was colored according to 36 categories of election results, representing three variables ranging from conservative to liberal, liberal to labor and labor to conservative. Red, blue, yellow was used around triangle.

In [1], Byron proposed extensions to triaxial-graph method for the visual portrayal of soil texture in maps and geographic information systems. He displayed the trivariate nature of soil texture by generating a color legend in shape of a triangular graph known as the soil textural triangle. The three sides of this equilateral triangle characterized the composition of the soil in terms of its percentage of sand, silt, or clay. Soil texture was thus spectrally encoded by assigning one of the additive primaries (red, green, blue) to each axis of the triangle. The relative intensity of each color varied with the amounts of sand, silt, or clay. The color value for a specific location was found by adding the amount of the three primaries. In general, he concludes that this method is more useful for small scale maps that cover large areas. It would be difficult to distinguish variation among the closely related colors if the soils in an area had limited variability.

This sections represented some of the cartographical studies which employed color as the visualization medium. In the latter sections we explain the design and results of our user studies which aims at evaluating and comparing two methods of employing colors in co-located areas: blending v.s. side by side. In the design of our studies we try to closely conform to the guidelines presented in the cartographic field in designing meaningful and realistic looking stimuli.

The focus of this article is multivariate visualization and although a few effective method for color based representing two and three variables with binned data, were discussed in the previous section in case of continuous data the common method is to blend the colors or combine the colors with textures. However the effectiveness of blending in conveying the correct information is questionable. The lack of a qualitative understanding of the error in conveying information by blending the colors is the base of our motivation for designing these experiments.

We hypothesized that despite the complications that is likely to arise when showing the colors side by side, this method has undeniable advantage over blending if the colors are chosen to be perceptually as distant from each other as possible.

In each of the following three experiments participants were asked a direct acquisition map reading task and the accuracy of the answers under each condition was statistically analyzed and compared to each other.

Specifically these three experiments were designed to give visualization researchers a quantitative measure of the accuracy they can expect in representing their data using color with any of the two methods.

III. EXPERIMENT 0

Fig. 1. Snapshot of the map participants saw in Experiment 0

Experiment 0 was designed to assess participants baseline ability to accurately read numerical data encoded via the intensity of a single displayed color. Specifically the focus of this experiment was to reveal any differences in the ability of the observers in achieving map reading tasks with any of the chosen colors. The experiment was also used to provide a comparing baseline for the map reading error for experiment one. The main premise of this experiment is the differences among the capability of each colors in delivering the data was negligible.

A. Choosing the Colors

We defined the base colors by choosing six evenly spaced points around a circle of constant saturation in a plane of constant luminance in a region of the La*b* color space that
fit within our monitor gamut. Two circles were chosen at L=62 and L=97. Figure 2 shows the maximum, minimum and the mid point slice in the portion of the color space we chose. From L=62 to L=97, we created six different perceptually linear single-hued color ramps, by continuously increasing the luminance and saturation values of each of six different base colors. In order to test the observers ability to read the minimum and maximum values accurately, we extended our color ramp on both sides such that the final color ramps included a continuous range of colors between the minimum and maximum values in addition to the right and left tail which consisted of colors which didn’t exist in the maps. Color ramps are presented in Figure 3. The maximum value for each variable was tied to a color at L=62 and the minimum value was mapped to the corresponding color at L=97.

B. Method

1) Apparatus: The stimuli consisted of six maps of the twelve midwestern United States, in which each state was filled with a different constant color, from a single color ramp, representing the value of a particular data attribute for that state. The actual data values were obtained from 2000 census data, but the particular assignment of data values to states was randomized to prevent people from using domain knowledge about the midwestern US to increase the accuracy of their responses. The six data distributions were: median household income, percentage of the population that had graduated from high school, percentage of the population that had graduated from college, percentage of the population living below the poverty line, median cost of a single family dwelling, and home ownership rate. Each color ramp was tied to one specific variable throughout experiment 0 and Experiment 1. The map size was chosen to be equal to the map size in experiment 1 which was the maximum possible map size we could have on our Monitors when showing the legends for all the six variables. The background was chosen to be a neutral gray with the L value halfway between the largest and smallest L values in our color ramps(L=79).

A total number of 216 displays were shown to the observers. For each variable a map reading question was asked for each of the states present. This amounted to 72 (12x6) displays, each of which was repeated 3 times to ensure the accuracy of the results (72*3). These displays were shown randomly to each of the participants.

A chin rest was employed to fix the participant’s eyes’ distance to the monitor at the appropriate level. The display’s specifications can be viewed in Table ??.

2) Procedure: Participants were naive to the purpose of the experiment other than knowing that its goal was investigating color perception. Participants color vision ability was assessed using an online version of a collection of Ishihara plates, the validity of a subset of which had been informally assessed for us by a person with known red-green color perception deficiencies. A question appeared on each display asking the observers to evaluate the color in a particular state by clicking on and dragging the small slider -a black triangle-. On each trial in this experiment, the participants task was to identify the value of a particular data attribute for a particular state by: reading the color from the map, setting a slider to the matching color on the provided color scale, and then clicking on the state to indicate that their selection is final. Figure ??(fig:Exp0snap) shows a screen shot from one trial. All participants were provided with a printed outline map showing the correspondence of names to states, and all trials on which the state was mis-identified were discarded.

The clicking on the state was added as an extra assurance that the observer was in fact looking at the correct state when making the choice of color value.

3) Participants: There were a total number of 9 participants for this experiment. The participants were both graduate, undergraduates and faculty members from University of Minnesota, Gettysburg College and NCSU. Participants were compensated 5 dollars per half an hour for completing the experiment and they took on average an hour and a half to complete the study.

C. Result

Participants were uniformly able to perform the task in experiment 0 with fairly good accuracy. The average relative error, computed over all participants and all colors, was 6.02%, with a standard error of 0.57%. Figure ??(fig:Exp0Meds) shows a scatter plot of the average errors for each trial, computed over the nine participants. The error bars show the extents of the 95% confidence intervals, and each point is color-coded according to the base color of the color scale used for
Fig. 2. Three slices of the LAB color space used to construct the color ramps. From left to right: at L=62, L= 79.5, L=97

Fig. 4. Scatter plot of actual vs. averaged user answers that trial. A single number giving the average of the median relative errors per observer, was added into the overall chart of results from experiment 1. We found that the difference among different colors were negligible. Figure , the accuracy had a logarithmic relationship to the decreasing L value.

IV. EXPERIMENT 1

The goal of this experiment was to quantify the differences among the three conditions for visualizing multivariate data when variables are overlapping in a region: blending the colors representing each overlapping variable, representing the overlapping variables by letting the colors coexist in that area with a small size high frequency pattern and again with a bigger size high frequency pattern.

Fig. 5. Median relative error in the map reading task for our nine participants

A. Method

1) Apparatus: The stimuli in experiment 1 consisted of a series of maps of the twelve midwestern United States in which the values of either two, three, four or six different data distributions were simultaneously represented via either color-blending, in which the separate color layers were made semi-transparent and then overlaid to form a single composite representation, or color-weaving, in which the separate color layers were individually sampled at independent pixels defined by a random noise function and them stitched together to form a finely patchworked, unified representation. We tested noise patterns of two different spatial frequencies: small noise, in which each pixel subtended 3 minutes of visual angle, and large noise, in which each pixel subtended 6 minutes of visual angle, and participants viewed all images from a fixed position enforced by a chin rest. Screen shots of the sample stimuli are
This experiment included a total number of 153 displays. The same map size as in experiment 0 was used. However this time, 2 or more variables were overlapping across all the states. All the 2 ways (15 displays), 3 ways (20 display), 4 ways (15 displays) and 6 way(1 display) combination of the 6 variables were considered (51 trials). Three conditions for combining the colors were investigated. Under the first condition, the 51 displays were made by blending the overlapping variables in each state. The second condition included 51 displays in which colors coexisted in an area by following a 2 pixel size noise pattern and finally for the third condition the colors coexisted by covering different levels of a 4 pixel size noise pattern. The patterns were made by filling the area of the map by a noise pattern of the appropriate size in Photoshop, equalizing the histogram to get approximately the same number of pixels for each of the gray level values and finally posterizing the pattern into 2, 3, 4 and 6 different levels. This procedure ensured that we approximately had the same number of pixels belonging to each of the coexisting colors in the states.

In order to remove the effect of any learning, for each of cases of the 2 variable, 3 variable, 4 variable and 6 variable we chose a random swapping of the values belonging to each state while at the same time ensuring that for each of the overlapping conditions, the same maps were used in making the Blend, small size noise and big size noise versions.

In order present a set of cognitively organized tasks to the observer, each observer saw a random order of the following sets: All the blended images, all the small noise images and all the 4 pixel size noise images. Inside each of the blended, small noise and big noise groups, they randomly saw all the 2 variable, 3 variable, 4 and 6 variables respectively.

2) Procedure: Due to the large number of displays we were not able to test all the states nor were we able to include repeated measures. The instruction guided the participants to only look at the state of Iowa and make their color evaluations by adjusting up to 6 sliders each of which corresponding to one of the overlapping or coexisting variables.

3) Participants: 18 people participated in this experiment. 6 of the participants were from university of Minnesota, 6 were recruited from Gettysburg College and 6 from NCSU. All the participants were compensated by getting 5 dollars gift certificates per half an hour it took them to complete the experiment.

B. Data Analysis and General Results

Our results from experiment 1 indicate that error rates were significantly lower when the original color information was available via the high frequency texture than when the colors were blended. In the case of the blended representation, error rates steadily rose as the number of components increased (a trend that we found statistically significant in an ANOVA analysis). We observed weak evidence of a similar effect in the case of the woven textures, but it was not statistically significant.

Additionally, the ANOVA analysis showed that both the type of color mixing and the number of variables had a
significant main effect. Within mixing type, blend was significantly different from, and worse than, noise. There was however no significant difference between the two sizes of noise. Regrading the number of variables, 3 and 4 variables were worse than 2 and 6 variable was worse than 3 or 4; 3 and 4 variables however were not significantly different.

Looking at the blended colors alone, 3 or 4 variables were worse than 2, and performance with 6 was worse than with 3 or 4. But there was no significant difference between 3 and 4.

Looking at the small noise alone, performance was worse with 6 than with 2, 3 or 4 but performance with 2, 3 and 4 wasn’t significantly different.

Looking at the large noise alone, performance was worse with 6 than with 2, 3 or 4 but performance with 2, 3 and 4 wasn’t significantly different.

Looking at each mixture of 2, 3, 4, 6 variables alone, performance was worse with blend than with noise, but performance with the different sizes of noise wasn’t significantly different.

Of an interesting note was that despite the convictions of participants that the colors could not be differentiated from each other in the case of blending more than 2 variables, all participants were still able to perform the task with an accuracy rate of % 70 or better.

It is well known that the different surroundings can change the appearance of a color. A perceptual process called assimilation causes the small color areas to appear more similar to their surroundings. The process of simultaneous contrast enhances the differences among colors between a larger patch of color and its surrounding color. These perceptual interaction between colors can change the perception of individual colors and hence add to the error in the map reading tasks. Unfortunately there has not been an effective method to minimize these issues and the designers need to keep the presence of these problems in mind.

Because the color values that we used in experiment 1 were defined according to actual data values, the distribution of measurements that we collected - while highly representative of what would be encountered in a typical visualization application - was not sufficient to answer all of the questions that came up about peoples ability (or inability) to accurately decompose a composite color into its constituent components. Our main goals in experiment 2 were to specifically explore the relationships between hue separation, luminance difference, and the error rates in peoples judgments of the values of individual components in a bi-variate representation. Specifically, we sought to answer the two questions: 1) will error rates be greater, smaller, or the same for the reading of color combinations in which the hues are separated by 60, 120, or 180 in La*b* space?; and 2) will error rates be greater, smaller, or the same for the reading of color combinations in which the luminance values of the individual components are nearly equal, moderately close, or relatively widely separated?

V. EXPERIMENT 2

Because the color values that we used in experiment 1 were defined according to actual data values, the distribution of
first set except that for each display the colors L values were swapped. Three examples of the displays can be seen in Figures 11, 12 and 13.

2) Procedure: The task in experiment 2 was similar to the task in experiment one. Participants were asked to read the two variables present in the map in the state of Iowa. In order to remove any possible influence from the surrounding colors all other states were colored with white.

3) Participants: Four graduate students who had passed the Ishihara test participated in this study. It took the participants approximately an hour and a half to two hours to finish the study.

B. Data Analysis and General Results

In a 3-way ANOVA analysis of the effects of mixture type, hue difference and luminance difference, we found significant main effects of both mixture type and luminance difference, with errors being higher when colors were combined by blending than when they were interwoven, and with the errors being smallest when the luminance values of the component colors were most similar. Looking at separately at individual subsets of the data, defined by blend type and hue angle, we found a significant main effect of luminance difference in all cases (marked in red where statistically significant) except where the hues were directly complementary. We did not find hue angle to have a significant effect on error rates in any subset of the data.

VI. DISCUSSION

The results of our three experiments indicate that color weaving is consistently more effective than color blending for conveying the values of individual data distributions in a multivariate visualization. Error rates remain low for woven combinations of 2, 3 and 4 different colors and only begin to rise to a statistically significant extent when the number of component colors increases to six. This advantage exists despite the potential of complications due to simultaneous contrast effects, and persists even when the area subtended by each patch of continuous color is very small.

Although the problem of inferring the values of the component colors in a blended mixture is inherently ill-posed, observers are able to perform this task fairly accurately, within a moderately constrained domain, when presented with pairs of component colors that have nearly equal luminance values, although errors rise as the luminance values of the component colors begin to differ.

We found no significant advantages, in either color blending or color weaving, to using color scales based on component hues that are more widely separated in La*b* color space. On the contrary, we found some indications that extra difficulties may arise when opponent hues are employed.

VII. CONCLUSION AND FUTURE WORK

ACKNOWLEDGMENT

This research was supported by the National Science Foundation through grants ACI-9875368 and CNS-0323471, and by the University of Minnesota through the Digital Technology Center. We are grateful to Dr Gary Meyer, Seth Barrier and Clement Shimizu for helping us to select colors and verify the validity of the representation using their Lab color space exploration tools. We are also very grateful to all of the participants in our experiments for their dedicated and conscientious efforts throughout a long and tedious task.

REFERENCES
Fig. 11. Experiment 2

Fig. 12. Experiment 2

Sunghee Kim Biography text here.


Fig. 13. Experiment 2