Reading small scalar data fields

Manteau, Constant; Nacenta, Miguel; Mauderer, Michael

Published in:
Proceedings of the 2017 Graphics Interface Conference

Publication date:
2017

Citation for published version (APA):
Manteau, C., Nacenta, M., & Mauderer, M. (2017). Reading small scalar data fields: color scales vs. Detail on Demand vs. FatFonts . In Proceedings of the 2017 Graphics Interface Conference (pp. 50-58). Canada: Canadian Human-Computer Communications Society.

To appear in Proceedings of Graphics Interface 2017 Conference.

Reading Small Scalar Data Fields: Color Scales vs. Detail on Demand vs. FatFonts

Constant Manteau*
University of St Andrews

Miguel Nacenta†
University of St Andrews

Michael Mauderer‡
University of Dundee

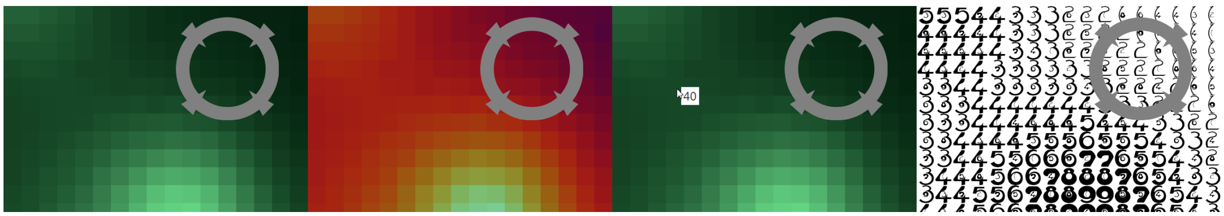

Figure 1: Scalar-field representations: Brightness and HSB scales, Interactive (tooltip) and FatFonts. The examples are excerpts from larger fields, e.g., as seen in Fig. 3.

Abstract

We empirically investigate the advantages and disadvantages of color- and digit-based methods to represent small scalar fields. We compare two types of color scales (one brightness-based and one that varies in hue, saturation and brightness) with an interactive tooltip that shows the scalar value on demand, and with a symbolic glyph-based approach (FatFonts). Three experiments tested three tasks: reading values, comparing values, and finding extrema. The results provide the first empirical comparisons of color scales with symbol-based techniques. The interactive tooltip enabled higher accuracy and shorter times than the color scales for reading values but showed slow completion times and low accuracy for value comparison and extrema finding tasks. The FatFonts technique showed better speed and accuracy for reading and value comparison, and high accuracy for the extrema finding task at the cost of being the slowest for this task.

Index Terms: Human-centered Computing [Visualization]: Empirical studies in visualization—;

1 Introduction

Visualization of scalar fields is common for scientific and non-scientific visualisation. One usual approach to represent these data fields consists of coding the scalar value with brightness or hue; for example, temperature is mapped to colors on a scale from blue to red on weather maps. The exact mapping chosen between magnitudes and colors is arbitrary and can have a large effect on the accuracy of how the visualizations are read [8,45]. It is therefore not surprising that a significant effort in the scientific literature has been devoted to the specification and testing of better scales (also called *color maps* or *heatmaps*); errors and inaccuracies in reading can affect many fields and applications.

Within the area of scalar fields, those that are of low spatial resolution (e.g., in the thousands of samples or less) still play an important role in multiple areas of application. For example, small scalar fields are used to represent confusion and correlation matrices (e.g., [29]), perceptual kernels (e.g., [13]), or other data that is inherently low in resolution [14]. Although color-based representations are dominant, it is possible to represent these data in different ways: by using symbols (digits) and interactivity.

In this paper we compare empirically the traditional approach of representing 2D scalar data through color and monochrome scales [40,45] with two alternatives that have not been compared previously. Our goal is to evaluate if colour scales can be outperformed by alternative representations. The first alternative is an interactive tooltip that complements one of the colour scales and offers an accurate reading of the data for the location of the cursor; although it might seem straightforward that the *detail-on-demand* that this enables could improve current representations in computer-based systems, we have found no literature comparing this with static representations. The second alternative is FatFonts, a recent technique that uses numeral glyphs that also vary in amount of ink (and therefore in brightness) to compose an image [28]. FatFonts is a representative of an emerging group of techniques that use numerals (in addition to visual variables) to represent data.

The empirical evaluation measured accuracy, time, and preference in three tasks: reading values, comparing values, and finding maxima and minima. We found that the detail-on-demand approach can be fast and accurate for reading values, but its accuracy and speed are poor when the task requires to look at multiple values. Finally, we found that using digits to represent the whole field has significant advantages in terms of accuracy for all tasks, and in terms of time for value reading and value comparison. These results, along with a qualitative analysis of technique applicability, are an initial step to a better characterization of techniques for visualizing scalar fields and might help motivate the use of digit-based techniques for small scalar field display.

2 Related Work

A large body of existing work analyses the use of color scales for scalar field visualization. We cannot be comprehensive about color scales but refer the reader to books and surveys [38,39,41,46]. Here we focus on the issues of the traditional forms of scalar field representations and then discuss work on visualization alternatives.

2.1 Color Scales

The basic idea behind color scales is the design of a visual mapping of numeric values to color that allows the user to read the information in a visual form. There is a large

*e-mail: constant@mailoo.org
†e-mail:mans@st-andrews.ac.uk
‡e-mail:michael@michaelmauderer.com

variety of possible mappings, most of which map the scalar value to one or more of the dimensions of a color space (e.g., hue, saturation, brightness, or red, green, blue). Some common examples include (a) rainbow/spectral scales that cycle through all the available hues, (b) single hue with varying saturation, and (c) scales based on brightness.

The perceptual properties of color and color scales have been well studied (e.g., [7, 17, 20, 45]). An often desired property of color scales derived from existing knowledge on color is perceptual linearity: the mapping is chosen so that the *perceived* distance between colors or their differentiability by the human visual system corresponds to the differences in numerical values. To achieve this, color maps have to be carefully designed taking into account aspects of human color perception, e.g., through the use of perceptual color spaces and models [10, 34], and through an understanding of the characteristics of the visual sub-channels (e.g., luminance can convey three times more spatial detail than hue [26]).

One common problem is that the color perception of an element is influenced by the color of its direct surroundings through simultaneous contrast effects [8,11,45]. The effectiveness of the mapping therefore depends on the surrounding color: a dark surround will make the color appear lighter, while a light surround will make it appear darker, resulting in value reading errors. If only hue is used (such as in the common spectral/rainbow/jet scheme) problems of ordering and perceptual artifacts appear. This type of scale has been subject to intense controversy and is generally considered to be problematic at best [5, 22, 35]. Additionally, these scales are highly vulnerable to individual differences in color perception, such as color-blindness [24]. A partial solution for some of these problems is combining hue variation with brightness variation to form a color sequence that traverses a perceptual color space in a spiral fashion. These scales might provide more accurate reading and ameliorate the simultaneous contrast problem [9, 19, 20, 40, 45].

An alternative approach is to allow users to create their own scales for a specific dataset with a tool [3, 6, 15, 25, 32, 33, 42–44, 47]. Although promising, this approach has obvious drawbacks such as inconsistency across representations [23], and that it requires expertise from users.

2.2 Digit-based Representations

Some of the accuracy problems derived from color scales can be avoided by using symbolic representations of magnitude (i.e., digit glyphs). However, digits by themselves (e.g., arranged on grids) are not ideal for scalar fields because they do not take advantage of the integrative characteristics of the human visual system. Although more accurate for reading values, using digits makes other tasks much harder; for example, finding the largest number in a grid requires much sequential reading, whereas with color scales the eye is almost automatically attracted to the best candidate areas.

A possibility for dynamic displays is to make use of the interactive capability of computers, and provide the symbolic representation only on demand, as suggested by Schneiderman's visualization mantra [37]. This can be done, for example, by providing a tooltip with the value of the point for the location of a cursor or touch. Having the symbolic information only when it is needed might reduce visual overload compared to information-dense digit fields.

Finally, it is possible to show digit based fields that contain both digits and graphical encodings of the information in dynamic displays. One such hybrid technique is FatFonts, which combines visual and symbolic elements to try to simultaneously reap the accuracy of digits and the visual overview of images [28]. Each digit is designed to have an amount of ink (or number of black pixels) proportional to the represented value. For example, a *5* has five times the ink as a *1*, a *4* four times, and so on. The technique encodes values redundantly: in the shape of the glyph (symbolic), and in the amount of ink. This means that, for any given data point, a viewer can read the value from the glyphs, but a grid of values can still be interpreted as an image: darker areas represent higher values (see Fig. 1). To span several orders of magnitude, a smaller digit (1/10 in area) fits within the larger digit (see Fig. 2). This technique is also related to visualizations in spreadsheets that use color on the back of cells for the same purpose (*Conditional Formatting* in Microsoft Excel). These techniques have the drawback that they require multiple pixels per data point, and can pack fewer points in the same space. The original description of the technique contains a detailed discussion of this issue and how FatFonts relates to other techniques [28]. We know of no previous empirical evidence comparing interactive or digit-only approaches. We intend to start filling this gap.

Figure 2: FatFonts numbers 19, 28, 37, 46, 55, 64, 73, 82 and 91. Reproduced with permission from [28].

3 Empirical Investigation

This section describes elements common to the three experiments, starting with the four compared techniques.

3.1 Techniques

We selected four representation techniques for small scalar fields: a *Brightness* scale, an *HSB* scale, a Detail-on-demand technique that we refer to as *Interactive*, and *FatFonts*.

3.1.1 Brightness and HSB

We searched the literature to find static scales (adaptive scales introduce per-participant variability and are not applicable to static media) which were reproducible from the paper description. We settled on the two best performing scales from Spence et al.'s [40]. Although they tested scales that change in fewer discrete steps, these were designed using perceptual principles and based on previous work.

Both scales are based on Munsell's color system. The Brightness scale provides equal intervals in the Munsell value dimension (values in [3,9]), while keeping $H = 2.5G$ and $C = 6$ (a green hue). The HSB (Hue, Saturation, Brightness) scale describes a spiral over the perceptual color space: the lower end of the scale is low in value (dark) and purple in hue, while the high end of the scale is bright and greenish in hue (see Fig. 1, 3). We interpolated values linearly to obtain 100 different levels. Note that color appearance in this paper may vary depending on display, printer or rendering software.

3.1.2 Interactive

The Interactive technique is like the Brighness scale, but with a tooltip (40×40 px) to the right of the cursor that indicates the underlying value (see Fig. 4).

3.1.3 FatFonts

We used two-level FatFonts (values 0 to 99) in the glyphs from Fig. 2. The zero was an empty digit (zero ink). Each number had a size of 40×40 px.

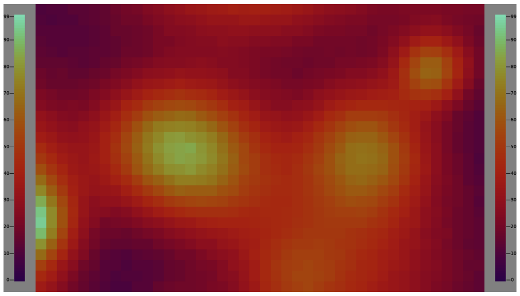

Figure 3: A data field displayed with the HSB technique.

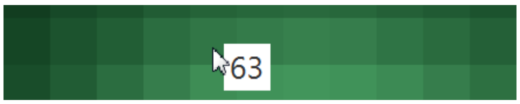

Figure 4: The tooltip of the Interactive technique.

3.2 Apparatus and Stimuli

The experiment took place in a light- and sound-controlled environment. Visual stimuli were presented on a 24 inch EDGE10 EF240a display with 1920×1080px resolution. The white point was set to 6500 K, and the system was calibrated with a Spyder 3 Express colorimeter. The calibration was manually confirmed using a Minolta LS-110 luminance meter. The participant sat in front of the screen with a keyboard on his left and a Logitech M185 mouse on his right (see Fig. 5).

The experiment ran through custom software written with JavaFX on Windows 7, with mouse acceleration turned off. All data fields were 42×27 points, with each data point taking up an area of 40×40 px. This size was chosen to enable reading of digits with the digit-based techniques with the resolution of the display. The side bands housed two legends (one on each side—see Fig. 3) for all techniques except FatFonts. At the beginning of each trial the cursor was reset to the middle of the screen.

We generated 368 different random data fields for use in all experiments. The data generation procedure was based on Sanyal et al.'s [36], but was adapted to represent fields with some noise and a range of fast and slow spatial variations. The fields are generated by adding 40 randomly positioned gaussian curves of random amplitude (20 positive and 20 negative) and random standard deviation between 2 and 12. The overall amplitude of the resulting surface was scaled to be between 50 and 90, and shifted in value randomily up or down with all values staying within the 5 to 95 interval. Finally, white gaussian noise of amplitude 2 was added. This generation procedure results in data fields with plausible dynamic range and data values.

For the comparison task we discarded with replacement data fields without at least three value pairs that differed by 5, 15, and 30 units, and where each of the paired values was on separate horizontal halfs of the field (see Exp. 2).

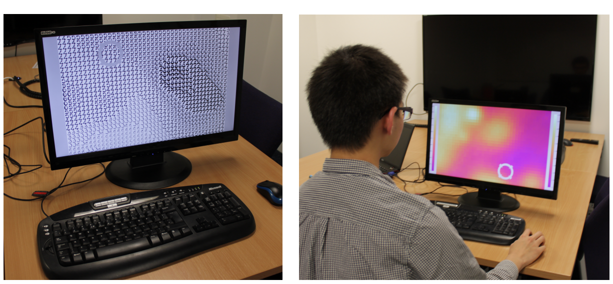

Figure 5: Main apparatus for all experiments.

3.3 Participants

Twenty-two participants, aged 19-39, 12 females, volunteered to participate in exchange for a gift card. We asked participants to perform as fast and accurately as possible, and told them that the fastest participant that was not innaccurate would receive an extra award. Participants were screened for color vision deviciencies with a Farnsworth-Munsell test[1]. One participant could not do the Farnsworth-Munsell test for personal reasons but was tested with the standard Ishihara test instead. Three participants were removed before analysis and are not counted above: one because they misunderstood the first experimental task, two more because of moderate color vision deficiency (F-M-CVD above 100).

3.4 General Procedure

The experimenter explained each technique and task before the start of that task's trials, and participants tried the task without a time limit. All participants performed the three experiments in the same order (reading values, finding extrema, value comparison). The order in which the techniques were presented was balanced between participants and each participant saw the techniques in the same order within each of the three tasks. In the first experiment the techniques themselves were also explained in detail.

Participants saw all 368 fields once, 92 with each technique. Stimuli appeared a balanced number of times across techniques. After each experiment, participants filled in a questionnaire on perceived speed and accuracy (7-point Likert scales), and ranked techniques in order of preference for that task. At the end they ranked the techniques overall.

3.5 Measures and Analysis

The software recorded time, accuracy, and mouse movements. Measures of time and accuracy are different depending on the task and are therefore described in the corresponding sections. For each experiment and task, the first half of trials is considered training and is not included in the analysis. Erroneous trials detected by the experimenter (e.g., due to system error or interface mistake) were marked and later removed (a total of 14 out of 8832 data samples).

The main omnibus analyses are RM-ANOVAs of accuracy and log-transformed trial completion time $log_{10}(TCT)$. Logarithmic transformation is to comply with normality assumptions. When averages of times are presented, these are the log-untransformed values in seconds of averaged log-transformed times. When the data was found not to be spherical (i.e., failed Mauchly's test), we applied a Greenhouse-Geisser correction that can be identified by the non-integer values in the reporting of the test's degrees of freedom. All pairwise tests were corrected for multiple comparisons using Holm's method. The subjective data was analysed with Friedman non-parametric tests. Error bars in figures are 95% CIs.

4 Experiment 1: Reading Values

The first experiment tests the viewer's ability to retrieve a value from the scalar field. This task is considered fundamental to information visualization [1,27] and has been evaluated in many empirical tests of color scales, e.g., [7,8,31,45].

4.1 Task, Design and Measures

Upon clicking on a start button, participants saw a scalar field presented with one of the techniques. A randomly placed bullseye in 50% grey highlighted a data point (see Fig. 1).

[1] http://www.color-blindness.com/farnsworth-munsell-100-hue-color-vision-test/

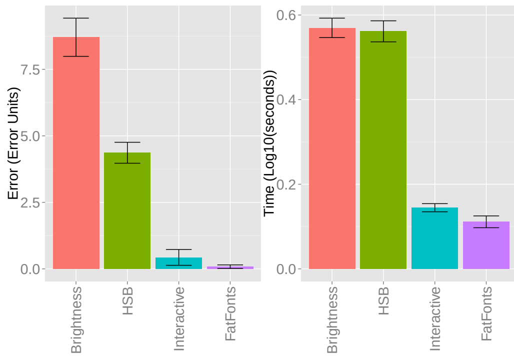

Figure 6: Average absolute error and average trial completion times (Log-tranformed) for the Reading Task.

The participant interpreted the value in the middle of the bullseye and clicked anywhere on the screen, after which a new screen with a textbox allowed to input the value with the keyboard. Bullseyes never appeared closer than two data points away from the edge of the field to avoid display-edge effects. The cursor moved to the center before each trial.

The task is common in the literature (e.g., [6]). Arguably, the digit based techniques have an advantage in this task because the answers are provided in the same response domain as the techniques themselves (digits). However, digit-based representation of numbers is still the best method to measure accuracy (gathering color from participant input is problematic and color is difficult to remember). We consider this a feature of these techniques because digits might be the tool that allows humans to achieve precise quantitative judgments in the first place (cf. [12,16]). The experiment design reflects our belief that the use of digits in the Interactive and FatFont techniques is an intrinsic advantage, not experimental bias.

Participants performed two blocks of 15 trials with each technique. The first block was training (not analyzed). Error in this task means the unsigned subtraction of the value pointed by the bullseye and the participant response (measured in the same units as the data field, which ranges between 3 and 97). Because the data fields are not assigned any semantics, we will call these *error units*.

Measured Time starts when the bullseye appears and ends on the mouse click. It excludes keyboard input time to avoid typing variability, which is extrinsic to our focus.

4.2 Results

Results are split into three sub-sections: accuracy (error, correctness), completion time, and subjective measures.

4.2.1 Accuracy

A one-way RM-ANOVA of mean error magnitude shows a strong effect of technique $F_{2.32,48.72} = 146.80, p < 0.001, \eta^2 = 0.83$. The mean error with FatFonts was smallest ($\mu_{Fat} = 0.08$ units), followed by Interactive ($\mu_{Int} = 0.43$). HSB and Brightness were more than one order of magnitude less accurate, with HSB's average error ($\mu_{HSB} = 4.36$) almost half the size as with the Brightness scale ($\mu_B = 8.71$ units—Fig. 6).

The post-hoc pair analyses show reliable differences between all techniques except between FatFonts and Interactive (all $p < 0.001$ except $p = 0.397$). The difference in proportion of trials with errors (correctness) between Interactive and FatFonts was large (3% vs. 5%), but the numbers are too small to rule out chance. Importantly, 95% of the HSB scale trials and 97% of the Brightness scale answers were not exact.

	Accuracy			Speed			Pref. Rank		
	M	μ	σ^2	M	μ	σ^2	M	μ	σ^2
B	3	2.5	1.4	3	3	1.8	4	3.8	0.4
HSB	3.5	3.1	1.6	3.5	3.6	1.3	3	3	0.2
Int	7	6.7	0.8	6.5	5.9	2.7	2	1.9	0.7
Fat	7	6.7	0.5	7	6.4	0.6	1	1.4	0.3

Table 1: Subjective responses (median, mean, variance) for the Reading Experiment.

4.2.2 Completion Time

A one-way RM-ANOVA of $log_{10}(TCT)$ reveals differences in completion time $F_{1.73,36.26} = 120.61, p < 0.001, \eta^2 = 0.70$. The fastest technique was FatFonts ($\mu_{Fat} = 1.29$ seconds), followed by Interactive ($\mu_{Int} = 1.39$ s), with HSB and Brightness taking longer ($\mu_{HSB} = 3.64$, $\mu_B = 3.71$—see Fig. 6).

The post-hoc tests statistically separate two groups of techniques. On one side the color scales (HSV and Brightness) and on the other FatFonts and Interactive. All comparisons across groups are significant ($p < 0.001$), whereas all comparison within groups were not (Interactive vs. Fatfonts—$p = 0.063$; and HSB vs. Brightness—($p = 0.754$).

4.2.3 Subjective Measures

Friedman non-parametric ANOVAs show differences between techniques on subjective evaluations of accuracy ($\chi^2(3) = 53.42, p < 0.001$), speed ($\chi^2(3) = 44.67, p < 0.001$), and preference ($\chi^2(3) = 45.16, p < 0.001$). Averages and medians of the three questions are best for FatFonts in speed and equal to best (tied with Interactive) in accuracy, with Brightness and HSB trailing far behind. FatFonts were also preferred overall to any of the other techniques (see Tbl. 1).

4.3 Summary and Discussion

In this reading task Fatfonts and Interactive data fields result in more accurate readings of values; traditional scales showed error sizes at least one order of magnitude larger. This is not surprising, since the digit techniques are designed for this purpose, but the differences are large: almost all values were read incorrectly with the scales, but accurately with the digit techniques. More interesting are the time results which show that reading values takes almost three times longer when using Brightness and HSB scales than when retrieving the value with FatFonts and the Interactive scale. This is likely due to the additional step of checking the legend. Subjective results generally follow the quantitative analysis–participants recognized the advantages of the digit-based techniques. These results also offer a useful comparison between the time costs of moving the cursor to a data point and reading a legend. For small scalar fields on interactive media (e.g., weather temperature maps on tablets), interactive retrieval of values could be better than a legend.

5 Experiment 2: Comparing Values

This experiment tests participants' ability to compare values from separate locations of a data field. The task is common in existing evaluations of color scales, e.g., [7,11].

5.1 Task, Design and Measures

Clicking a start button revealed a scalar field shown with one of the techniques. Two bullseyes appeared on screen in separate positions, one on each of the left-right halfs of the screen. The participant compared the two values and pressed the left or the right arrow key to indicate the largest.

The bullseyes were positioned so that their value differences were 5, 15 or 30 units. Participants completed a total of 30

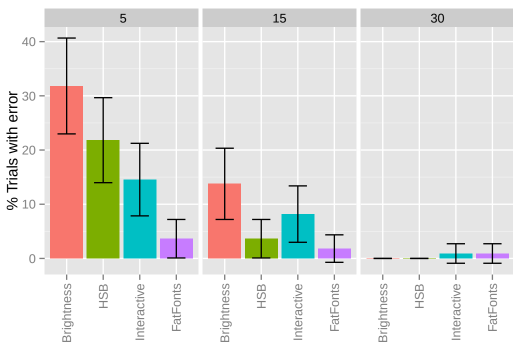

Figure 7: Erroneous trials for the comparison task, by value difference.

trials (10 for each value difference, in random order) in two blocks of 15 (block 1 is training). The software measured correctness and time from stimulus onset until the keypress.

5.2 Results

5.2.1 Accuracy

A two-way RM-ANOVA of the proportion of errors (correctness), with technique and value difference as main factors found main effects of technique ($F_{3,63} = 10.87, p < 0.001, \eta^2 = 0.14$) and value difference ($F_{1.51,31.61} = 55.16, p < 0.001, \eta^2 = 0.28$), as well as an interaction ($F_{3.87,81.29} = 6.75, p < 0.001, \eta^2 = 0.13$). Fig. 7 shows how error proportions diminish with the increase in value differences, to reach almost zero with all techniques.

When considering all value differences, the most accurate technique was FatFonts ($\mu_{FatFonts} = 2\%$ errors), ahead of Interactive and HSB, which have accuracy ($\mu_{Int} = 8\%$, $\mu_{HSB} = 8\%$). The Brightness scale was the least accurate ($\mu_B = 15\%$). Pairwise post-hocs show differences between all techniques (all $p < 0.05$), except for the HSB-Interactive pair ($p = 0.822$).

5.2.2 Completion Time

A two-way RM-ANOVA of $log_{10}(TCT)$ with technique and value difference as main factors found main effects of technique ($F_{2.08,43.77} = 30.86, p < 0.001, \eta^2 = 0.20$) and value difference ($F_{1.39,29.21} = 83.78, p < 0.001, \eta^2 = 0.20$), and an interaction ($F_{6,126} = 2.92, p < 0.05, \eta^2 = 0.02$). Fig. 8 shows that TCT diminishes with larger value differences (i.e., more different values are faster to judge). The pattern holds across differences: the Interactive technique takes longer than the second slowest (FatFonts) for all value difference levels ($\mu_{int(5)} = 2.25s$, 37% longer than FatFonts for 5, $\mu_{int(15)} = 1.98s$, 44% for 15, and $\mu_{int(30)} = 1.48s$, 16% for 30). Pairwise comparisons collapsed across value differences show differences between Interactive and the others (all $p < 0.001$). The interaction reflects that participants used color, not tooltip, for larger differences with the Interactive technique.

5.2.3 Subjective Measures

The non-parametric tests of the subjective responses indicate perceived subjective differences in accuracy ($\chi^2(3) = 25.12, p < 0.001$), speed ($\chi^2(3) = 8.37, p < 0.05$), and preference ($\chi^2(3) = 18.05, p < 0.001$). The medians and averages suggest a FatFonts advantage in accuracy, an advantage of HSB in speed (with FatFonts as a close second), and a preference for FatFonts (see Tbl. 2).

5.3 Summary and Discussion

The results show that FatFonts provides an advantage in terms of the viewer's ability to differentiate values. However, the advantage diminishes as the differences increase in size. For the large value difference conditions (roughly 30% of the range) the errors were very small for all techniques. The mean errors for the color scales confirm previous results where the best scales can only help reliably differentiate sufficiently distant values, and where the color is superior to just brightness. Interestingly, the interactive technique showed a surprisingly large number of errors for a technique that also provides a symbolic component. We speculate that participants are sometimes overconfident and rely on the color scale without querying the value with the tooltip.

In terms of speed, color scales and FatFonts are roughly equivalent, and the Interactive technique takes longer, probably due to mouse movement. However, we also suspect that deciding whether to trust color or further move the cursor might have played a part in these delays. Interactive is a hybrid technique, and there are other HCI examples where hybrids impose cognitive overheads [30].

The subjective results follow roughly the same pattern: participants perceive accuracy and speed differences between techniques and generally prefer Fatfonts.

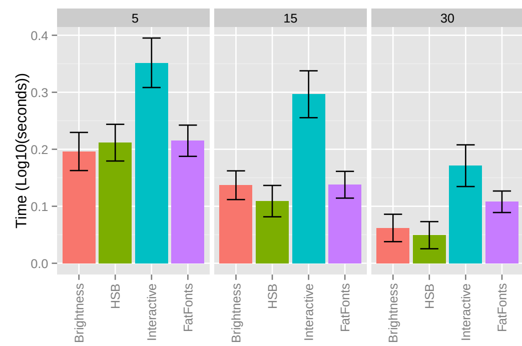

Figure 8: TCTs for the comparison task, by value difference.

	Accuracy			Speed			Pref. Rank		
	M	μ	σ^2	M	μ	σ^2	M	μ	σ^2
B	4	4.1	2	5	5	3	4	3.2	1.2
HSB	5	4.9	1.6	6	5.7	1.9	2	2	0.7
Int	6	6	1	5	4.9	1.6	3	2.9	0.8
Fat	7	6.2	2	6	5.7	1.9	1.5	1.8	1

Table 2: Subjective responses for the Comparison Experiment.

6 Experiment 3: Finding Extrema

This experiment tests a maxima and minima finding task. This task is referred to as *Find Extremum* in Amar et al.'s task categorization [1].

6.1 Task, Design and Measures

Clicking on a start button revealed a scalar field without bullseyes. Participants had to find the location of the absolute maximum or the minimum (depending on the trial block), and click on it. The task was repeated a total of 32 times for each technique in four blocks of eight trials each: training max., training min., max., and min.

We measured error in two ways: unsigned differences between the actual extremum and the chosen point (magnitude), and proportion of actual extrema clicked across trials.

6.2 Results

6.2.1 Accuracy

A two-way RM-ANOVA of average error found an effect of technique on accuracy ($F_{2.16,45.35} = 30.08, p < 0.001, \eta^2 = 0.29$), but not of the type of extremum (maximum vs. minimum—$F_{1,21} = 0.19, p = 0.670, \eta^2 = 0.00$). The interaction was significant ($F_{2.16,45.35} = 30.08, p < 0.001, \eta^2 = 0.29$), reflecting how HSB has increased errors for finding minimums (Fig. 9).

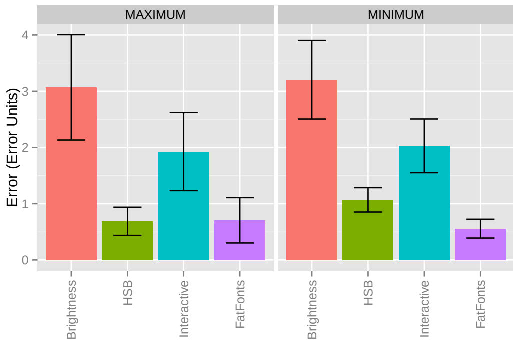

Figure 9: Average error for the extrema task.

Post-hoc tests found differences between all techniques (all $p < 0.02$) except between FatFonts and HSB ($p = 0.195$). This splits the techniques into two groups: FatFonts and HSB are the most accurate ($\mu_{Fat} = 0.63$ error units, $\mu_{HSB} = 0.88$), then Interactive ($\mu_{Int} = 1.98$), then Brightness ($\mu_B = 3.14$).

The RM-ANOVA of the proportion of correct answers (correctness) reveals effects of technique ($F_{3,63} = 23.86, p < 0.001, \eta^2 = 0.28$), as well as type of extremum ($F_{3,63} = 23.86, p < 0.001, \eta^2 = 0.28$), but not of the interaction ($F_{0.76,15.98} = 0.87, p = 0.437, \eta^2 = 0.01$). Pairwise comparisons are significant (all $p < 0.02$), except HSB vs. Interactive ($p = 0.577$). FatFonts is most accurate ($\mu_{Fat} = 23\%$ errors), followed by HSB ($\mu_{HSB} = 43\%$), then Interactive ($\mu_{Int} = 46\%$), and finally the Brightness scale ($\mu_B = 61\%$—Fig. 10).

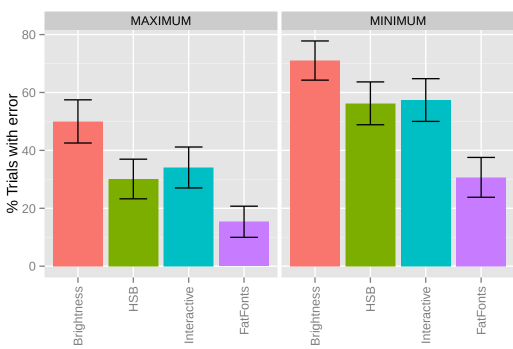

Figure 10: Average trials with error for the extrema task.

6.2.2 Completion Time

A two-way RM-ANOVA of the trial completion time found effects of technique ($F_{2.39,50.17} = 93.03, p < 0.001, \eta^2 = 0.42$) and the type of extrema ($F_{1,21} = 40.05, p < 0.001, \eta^2 = 0.04$), but not an interaction ($F_{3,63} = 1.84, p = 0.150, \eta^2 = 0.00$). Fig. 11 shows that the task took the longest with FatFonts ($\mu_{FatFonts} = 5.31$ s), with Interactive closely behind ($\mu_{Int} = 4.37$ s), and with Brightness and HSB having roughly halved TCTs ($\mu_B = 2.29$ s, $\mu_B = 2.29$ s). The post-hoc tests confirm differences among all techniques (all post-hocs $p < 0.03$), except between HSB and Brightness ($p = 0.250$).

6.2.3 Subjective Measures

The subjective measures analysis indicates that technique had an effect on perceived accuracy ($\chi^2(3) = 26.25, p < 0.001$)

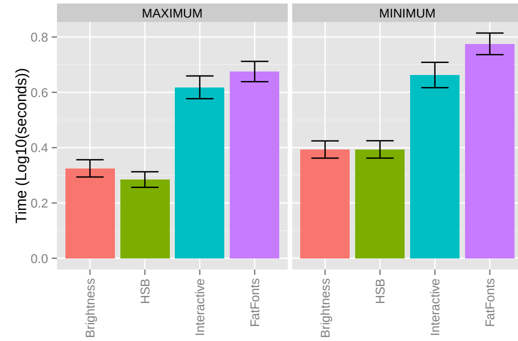

Figure 11: Average TCT for the extrema task for min. and max.

	Accuracy			Speed			Pref. Rank		
	M	μ	σ^2	M	μ	σ^2	M	μ	σ^2
B	3.5	3.9	2.1	5	5	2.1	3.5	3.1	1.1
HSB	4	4.3	1.6	5	5	2.8	2	2	1
Int	6	5.6	1	4.5	4.5	2.4	3	2.6	0.8
Fat	6	5.4	1.7	5	4.6	1.5	2	2.3	1.6

Table 3: Subjective responses (median, mean, variance) for the Extrema Experiment.

and preference ($\chi^2(3) = 9.22, p < 0.05$), but not on perceived speed ($\chi^2(3) = 1.02, p = 0.796$). Table 3 shows that HSB was the preferred technique.

6.3 Summary and Discussion

Finding extrema is a more demanding task, as evidenced by the large proportion of trials with errors and the longer completion times. Scales have a speed advantage here because they do not require the legend as in the first task. We also speculate that the availability of exact data with the Interactive and FatFonts techniques induces participants to be more systematic and therefore slower.

Surprisingly, besides being fastest, the HSB scale also achieves very low magnitude errors (comparable with FatFonts). The choice of technique here should depend on the importance of the different aspects of accuracy. If error magnitude is important HSB is the best; if finding the actual coordinate of the maximum (not one close enough in value) then using FatFonts is better (approx. 47% fewer errors).

Also surprising is the bad performance of the Interactive technique. Participants took slightly less time to find the extrema with the Interactive technique than with the slowest technique (FatFonts), but the level of errors (of both types) was almost on-par with the worst technique (Brightness). We speculate that the large errors are probably due to the sequential nature of the technique, which forces participants to systematically scan areas for results, a tedious and error-prone process. FatFonts is not that different in this respect (the participant still needs to redirect their sight sequentially to read the candidate numbers), but it might put less strain in short time memory, since the image in FatFonts is unchanging. The subjective results suggest that HSB was recognized by participants as very good for this task.

7 Additional Analyses

This section reports the analyses that affect all three experiments and an exploratory analysis of contrast effects.

7.1 Overall Preference

A non-parametric Friedman test shows technique differences in preference rankings ($\chi^2(3) = 23.79, p < 0.001$). Participants ranked FatFonts first, HSB and Interactive as tied seconds, and Brightness worst. Fatfonts was best for 12 participants, and Brightness was worst for 15.

7.2 Contrast Effects

Color and brightness perception can be influenced by the surround. To explore whether this was the source for the high levels of error found with the HSB and Brightness scale, we performed a post-hoc analysis of the data in the reading task (inspired by [45]). For every trial we computed a surround value s for the observed value using a 3×3 gaussian mask with the center set to 0. Based on s we classified each trial as having an equal, lighter or darker surround relative to the central color patch. We excluded trials where surround was equal to the central color patch and then computed the trial's error e. A 2×2 repeated-measure ANOVA with participant as random factor shows that the surround affects error ($F_{1.39,29.18} = 4.12, p < 0.05, \eta^2 = 0.02$). The interaction ($F_{2.43,51.07} = 6.29, p < 0.01, \eta^2 = 0.09$) tells us that not all techniques are affected in the same way: surround affects error most with the Brightness scale, somewhat with the HSB scale, but not with the digit-based techniques.

8 General Discussion

We discuss results and implications of the study in five subsections: Confirmatory Results, Interactive, FatFonts, qualitative analysis, and limitations and future work.

8.1 Confirmatory Results

Our exploratory analysis of the reading task offers confirmation of a previous analysis performed by Ware [45] of the effect of surrounding colours in value perception. Both color scales are vulnerable to the effects of the surround, helping generalize Ware's results to more scenarios (Ware tested center patches of colour surrounded by a circular gradient). This is a partial explanation of *why* color scales suffer from reading accuracy problems.

In the tested tasks and with our chosen scales we have shown that a colour scale that uses multiple color-perceptual variations (HSB) was superior to one that only uses brightness. The HSB scale was more accurate than the brightness scale in all tasks (50.12%, 55.84% and 27.99% fewer errors in the three tasks). Readers should, however, be careful to generalize these results. Not all color scales are good—rainbow scales are particularly bad according to many [5]—and the small fields that we tested might not generalize to higher resolutions.

8.2 Interactive

We included an interactive tooltip in our study because we thought it would be a valuable addition to color scales. Obviously, an interactive data tooltip cannot be used in static (e.g., print) scenarios, but when a scalar field is in electronic form, it is straightforward to add this functionality; interactive graphics are starting to appear in scientific publications (e.g., [2]). The tooltip might provide the advantages of accessing data points in both forms (symbolic and graphical) without having to clutter and occlude the content, and without having to radically change the visuals.

The interactive tool did deliver advantages in the reading task (required only 38.28% of the time needed with HSB, 37.57% of the time needed with Brightness, and only 9.86% of the error magnitude of HSB). Surprisingly, Interactive is worse for the comparison task (worst in time and equivalent in accuracy to the HSB scale) and the extrema task (where accuracy is similar to the worst scale—Brightness—and completion time is similar to FatFonts—the slowest technique). This indicates that participants took the worst of the color and the digit-based worlds. We believe this is due to: a) the costs of interactivity (i.e., pointing time); b) the overhead of deciding to act upon digits vs. color; and, c) a false sense of confidence in color perception over digit reading.

8.3 FatFonts and Digit-Based Techniques

Our results show that FatFonts is best (or equal to best) in accuracy for all three tasks, and also fastest (or equal to fastest) in reading and comparison. Although these results should be carefully interpreted in the scope of these experiments and limited tasks, they suggest that digit-based static representations offer promise in small matrix applications (e.g., confusion matrices).

Being able to access the symbolic part of the data seems faster and more accurate than trying to read or compare values from a legend. It can also enable operations that correspond to the symbolic domain, such as calculating how much higher a value is than another, or detecting values above a threshold. FatFonts is also limited in several ways, most notably resolution, and its performance is unknowin other tasks such as pattern finding. We discuss these issues below; see also the original paper for a discussion on FatFonts' limitations and applicability [28].

8.4 Qualitative Analysis and Design Implications

The most obvious drawback of FatFonts is resolution. To achieve accuracy and speed with FatFonts, a designer will trade in spatial resolution. The interactive tooltip technique is a good alternative for high spatial resolution data.

Although resolution is becoming less of a problem (displays are becoming larger and denser), FatFonts is not a technique for all situations. Interactive tooltips have the advantage over FatFonts that the digit-based precision of a single value is potentially unlimited, whereas with FatFonts this is usually limited to two or three orders of magnitude depending on the precision and resolution that needs to be achieved. However, the reader should still consider that this is still about an order of magnitude more precision than with color scales. Our results also warn against choosing tooltips for comparisons and finding extrema. Adding a tooltip might result in slower or less accurate readings.

Designers should also consider a range of important technique characteristics. The Interactive technique requires a computer, whereas the other three techniques are fine in print. The color scales are vulnerable to uncalibrated displays [18], to the display circumstances (e.g., lighting conditions) [4], and to variability in perceptual ability of people (e.g., color blindness) [24]. These problems do not affect FatFonts, and may be partially mitigated by adding a tooltip. Table 4 provides a summary of issues to be considered by practitioners, including conclusions derived from the empirical data.

8.5 Limitations and Future Work

In this paper we have investigated three tasks that are important for the use of scalar field representations and are named in task analysis of InfoVis (e.g., [1,27]); however, other low-level perceptual tasks might also be relevant in certain scenarios, such as cluster finding [21].

Although we believe that FatFonts represents techniques that display data visually using number glyphs, further research is required to compare it with other digit-based techniques such as Excel formatted cells. Similarly, emerging color scales should be compared to digit-based approaches.

Importantly, our version of the Interactive technique is based on the Brightness scale, which turned to be the worst of the two scales. We do not know know if an HSB-interactive

	B	HSB	Int.	Fat.
High Resolution	++	++	++	−−
Potential Precision	−−	−	++	+
Works on Static	++	++	−−	++
Context Independent	−−	−−	−	++
Person Independent	−	−−	+	++
Accuracy (Read)	−−	−	++	++
Accuracy (Comp)	−−	−	−	+
Accuracy (MaxMin)	−−	−	−	++
Speed (Read)	−−	−−	+	++
Speed (Comp)	+	+	−−	+
Speed (MaxMin)	++	++	−−	−−
Overall Preference	−	+	+	++

Table 4: Quantitative and qualitative analysis summary.

combination can be better than FatFonts or HSB. This question is worth exploring in subsequent experiments.

To conclude and summarize, we believe that this study opens up several questions and further areas of inquiry. The most important pieces of future work concern the extrapolation of results to: a) denser scalar fields; b) other tasks; c) other glyph-based techniques; d) other color scales.

9 Replicability and Recomputability

This paper and its analysis are written using LATEX, with its statistics implemented in R and embedded into the LATEXsource with Sweave. Compiling the paper from the source will recalculate all the analyses. This enables any researcher to access our exact computations for most numbers in this text. The auxiliary file also contains the collected data, images of all the stimuli, and the source code of the experimental software.

10 Conclusion

Although research in scalar field representations has shown that color scales are vulnerable to systematic error due to perceptual effects, most efforts have focused on the development of better color scales. In this paper we empirically compared two types of color scales with two alternatives approaches that use digits (FatFonts and an interactive tooltip). We measured participant performance and preference for three tasks (value reading, value comparison and extrema finding) in small scalar fields (42×27 spatial samples). Our findings and analysis provide the following main insights:

- Although interactive tooltips are good for reading values, they result in relatively poor accuracy and slow times in comparisons and extrema finding tasks.
- When the spatial resolution required is small, FatFonts provide the best option for all three tasks in accuracy, and in terms of speed for reading and comparison.
- Color scales show important value-reading errors due to the visual context, whereas digit-based techniques do not appear affected.

Although more studies are needed, our analysis and findings can help designers make better decisions that result in faster and more accurate access to the data.

11 Acknowledgments

This research is supported by the Engineering and Physical Sciences Research Council of the United Kingdom under research award EP/M008932/1.

References

[1] Robert Amar, James Eagan, and John Stasko. 2005. Low-level components of analytic activity in information visualization. In *Information Visualization (INFOVIS '05)*. IEEE Computer Society, Washington, DC, USA, 111–117. DOI:http://dx.doi.org/10.1109/INFVIS.2005.1532136

[2] David G Barnes, Michail Vidiassov, Bernhard Ruthensteiner, Christopher J Fluke, Michelle R Quayle, and Colin R McHenry. 2013. Embedding and Publishing Interactive, 3-Dimensional, Scientific Figures in Portable Document Format (PDF) Files. *PLoS ONE* 8, 9 (2013). DOI:http://dx.doi.org/10.1371/journal.pone.0069446

[3] Lloyd A Bergman, Lawrence D and Rogowitz, Bernice E and Treinish. 1995. A rule-based tool for assisting colormap selection. In *Proceedings of the 6th conference on Visualization*. IEEE Computer Society, 118. DOI:http://dx.doi.org/10.1109/VISUAL.1995.480803

[4] Simone Bianco and Andrea G Citrolo. 2013. High contrast color sets under multiple illuminants. In *Proceedings of the 4th international conference on Computational Color Imaging*. Springer, 133–142. DOI:http://dx.doi.org/10.1007/978-3-642-36700-7_11

[5] David Borland and Russell M. Taylor. 2007. Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications* 27, 2 (2007), 14–17. DOI:http://dx.doi.org/10.1109/MCG.2007.323435

[6] Leonard A Breslow, J Gregory Trafton, J Malcolm McCurry, and Raj M Ratwani. 2010. An algorithm for generating color scales for both categorical and ordinal coding. *Color Research and Application* 35, 1 (2010), 18–28. DOI:http://dx.doi.org/10.1002/col.20559

[7] Leonard a Breslow, J Gregory Trafton, and Raj M Ratwani. 2009. A perceptual process approach to selecting color scales for complex visualizations. *Journal of Experimental Psychology: Applied* 15, 1 (March 2009), 25–34. DOI:http://dx.doi.org/10.1037/a0015085

[8] Cynthia A Brewer. 1996. Prediction of Simultaneous Contrast between Map Colors with HuntâĂŹs Model of Color Appearance. *Color Research and Application* 21, 3 (1996), 221–235. DOI:http://dx.doi.org/10.1002/(SICI)1520-6378(199606)21:3<221::AID-COL4>3.0.CO;2-U

[9] Cynthia A Brewer, Alan M MacEachren, Linda W Pickle, and Douglas Herrmann. 1997. Mapping Mortality: Evaluating Color Schemes for Choropleth Maps. *Annals of the Association of American Geographers* 87, 3 (1997), 411–438. DOI:http://dx.doi.org/10.1111/1467-8306.00061

[10] Robert Carter and Rafael Huertas. 2010. Ultra-large color difference and small subtense. *Color Research and Application* 35, 1 (Feb. 2010), 4–17. DOI:http://dx.doi.org/10.1002/col.20557

[11] William S. Cleveland and R. McGill. 1983. A Color-Caused Optical Illusion on a Statistical Graph. *American Statistician* 37, 2 (1983), 101–105. DOI:http://dx.doi.org/10.1080/00031305.1983.10482720/00031305.1983.10482720#

[12] Stanislas Dehaene. 1999. *The number sense. How the mind creates mathematics.* Number 16. OUP USA. 74–75 pages. DOI:http://dx.doi.org/10.2307/2589308

[13] Çağatay Demiralp, Michael S Bernstein, and Jeffrey Heer. 2014. Learning Perceptual Kernels for Visualization Design. *IEEE Transactions on Visualization & Computer Graphics* 20, 12 (Dec. 2014), 1933–1942. DOI:http://dx.doi.org/10.1109/TVCG.2014.2346978

[14] Pietro Genovese, Giulia Schiroli, Giulia Escobar, Tiziano Di Tomaso, Claudia Firrito, Andrea Calabria, Davide Moi, Roberta Mazzieri, Chiara Bonini, Michael C Holmes, Philip D Gregory, Mirjam van der Burg, Bernhard Gentner, Eugenio Montini, Angelo Lombardo, and Luigi Naldini. 2014. Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* 510, 7504 (2014), 235–40. DOI:http://dx.doi.

org/10.1038/nature13420

[15] Denis Horvat, Borut Zalik, MarjanSlak Slak Rupnik, Domen Mongus, and Borut Žalik. 2013. Visualising the Attributes of Biological Cells, Based on Human Perception. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*. Vol. 7947. Springer, 386–399. DOI:http://dx.doi.org/10.1007/978-3-642-39146-0_36

[16] Georges Ifrah and E F Harding David Bellos Sophie Wood and Ian Monk. 2000. *The Universal History of Numbers: From Prehistory to the Invention of the Computer*. Wiley. xxiv, 633 pages.

[17] Alan D Kalvin, Bernice E Rogowitz, Adar Pelah, and Aron Cohen. 2000. Building perceptual color maps for visualizing interval data. *Electronic Imaging* 3959, 1 (2000), 323–335. DOI:http://dx.doi.org/doi:10.1117/12.387169

[18] Alan D Kalvin and Thomas J Watson. 2002. Graduated profiling: enumerating and generating perceptual colormaps for uncalibrated computer displays. In *Electronic Imaging*, Vol. 4662. 218–225. DOI:http://dx.doi.org/10.1117/12.469517

[19] Sarah Kindlmann, Gordon and Reinhard, Erik and Creem. 2002. Face-based luminance matching for perceptual colormap generation. In *Proceedings of the conference on Visualization*. IEEE, IEEE, 299–306. DOI:http://dx.doi.org/10.1109/VISUAL.2002.1183788

[20] Haim Levkowitz and Gabor T Herman. 1992. Color scales for image data. *IEEE Computer Graphics and Applications* 12, 1 (1992), 72–80. DOI:http://dx.doi.org/10.1109/38.135886

[21] Stephan Lewandowsky, Douglas J Herrmann, John T Behrens, Shu-Chen Li, Linda Pickle, and Jared B Jobe. 1993. Perception of clusters in statistical maps. *Applied Cognitive Psychology* 7, 6 (1993), 533–551. DOI:http://dx.doi.org/10.1002/acp.2350070606

[22] Adam Light and Patrick J Bartlein. 2004. The end of the rainbow? Color schemes for improved data graphics. (2004). DOI:http://dx.doi.org/10.1029/2004EO400002

[23] James McNames. 2006. An effective color scale for simultaneous color and gray-scale publications. *IEEE Signal Processing Magazine* 23, 1 (2006), 82–87. DOI:http://dx.doi.org/10.1109/MSP.2006.1593340

[24] Gary W Meyer and Donald P Greenberg. 1988. Color-defective vision and computer graphics displays. *IEEE Computer Graphics and Applications* 8, 5 (1988), 28–40. DOI:http://dx.doi.org/10.1109/38.7759

[25] Kenneth Moreland. 2009. Diverging color maps for scientific visualization. In *Advances in Visual Computing*. Springer, 92–103. DOI:http://dx.doi.org/10.1007/978-3-642-10520-3_9

[26] Kathy T Mullen. 1985. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of physiology* 359, 1 (1985), 381–400. DOI:http://dx.doi.org/10.1113/jphysiol.1985.sp015591

[27] Tamara Munzner. 2014. *Visualization Analysis and Design*. CRC Press.

[28] Miguel Nacenta, Uta Hinrichs, and Sheelagh Carpendale. 2012. FatFonts : Combining the Symbolic and Visual Aspects of Numbers. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 407–414. DOI:http://dx.doi.org/10.1145/2254556.2254636

[29] Andrew J Peters, Simon X Chen, and Takaki Komiyama. 2014. Emergence of reproducible spatiotemporal activity during motor learning. *Nature* 510, 7504 (2014), 263–7. DOI:http://dx.doi.org/10.1038/nature13235

[30] Umar Rashid, Miguel A Nacenta, and Aaron Quigley. 2012. The cost of display switching: a comparison of mobile, large display and hybrid UI configurations. In *International Working Conference on Advanced Visual Interfaces*. ACM, 99–106. DOI:http://dx.doi.org/10.1145/2254556.2254577

[31] Penny L Rheingans. 1992. Color, change, and control of quantitative data display. *Proceedings of the 3rd conference on Visualization* (1992), 252–259. DOI:http://dx.doi.org/10.1109/VISUAL.1992.235201

[32] Philip K Robertson. 1988. Visualizing color gamuts: A user interface for the effective use of perceptual color spaces in data displays. *Computer Graphics and Applications* 8, 5 (1988), 50–64. DOI:http://dx.doi.org/10.1109/38.7761

[33] Philip K Robertson and John F O'Callaghan. 1986. The generation of color sequences for univariate and bivariate mapping. *Computer Graphics and Applications* 6, 2 (1986), 24–32. DOI:http://dx.doi.org/10.1109/MCG.1986.276688

[34] Philip K Robertson and John F O'Callaghan. 1988. The application of perceptual color spaces to the display of remotely sensed imagery. *Transactions on Geoscience and Remote Sensing* 26, 1 (1988), 49–59. DOI:http://dx.doi.org/10.1109/36.2999

[35] Bernice E. Rogowitz and Lloyd A Treinish. 1998. Data visualization: the end of the rainbow. *IEEE Spectrum* 35, 12 (1998), 52–59. DOI:http://dx.doi.org/10.1109/6.736450

[36] Jibonananda Sanyal, Song Zhang, Gargi Bhattacharya, Phil Amburn, and Robert J Moorhead. 2009. A user study to compare four uncertainty visualization methods for 1d and 2d datasets. *Transactions on Visualization and Computer Graphics* 15, 6 (2009). DOI:http://dx.doi.org/10.1109/TVCG.2009.114

[37] Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Symposium on Visual Languages*. IEEE, 336–343. DOI:http://dx.doi.org/10.1109/VL.1996.545307

[38] Samuel Silva, Joaquim Madeira, and Beatriz Sousa Santos. 2007. There is more to color scales than meets the eye: a review on the use of color in visualization. *Information Visualization* (2007), 943–950. DOI:http://dx.doi.org/10.1109/IV.2007.113

[39] Samuel Silva, Beatriz Sousa Santos, and Joaquim Madeira. 2011. Using color in visualization: A survey. *Computers and Graphics* 35, 2 (April 2011), 320–333. DOI:http://dx.doi.org/10.1016/j.cag.2010.11.015

[40] Ian Spence, Natasha Kutlesa, and David L Rose. 1999. Using Color to Code Quantity in Spatial Displays. *Journal of Experimental Psychology: Applied* 5, 4 (1999), 393–412. DOI:http://dx.doi.org/10.1037/1076-898X.5.4.393

[41] Maureen Stone. 2002. *A Field Guide to Digital Color*. Vol. 3. CRC Press.

[42] Christian Tominski, Georg Fuchs, and Heidrun Schumann. 2008. Task-driven color coding. In *Proceedings of the International Conference on Information Visualisation*. IEEE, 373–380. DOI:http://dx.doi.org/10.1109/IV.2008.24

[43] Lujin Wang, Joachim Giesen, Kevin T McDonnell, Peter Zolliker, and Klaus Mueller. 2008. Color design for illustrative visualization. In *Transactions on Visualization and Computer Graphics*, Vol. 14. IEEE, 1739–1746. DOI:http://dx.doi.org/10.1109/TVCG.2008.118

[44] Lujin Wang and Klaus Mueller. 2008. Harmonic colormaps for volume visualization. In *Proceedings of the Fifth Eurographics / IEEE VGTC Conference on Point-Based Graphics*, Vol. Vi. Eurographics Association, 33–39.

[45] Colin Ware. 1988. Color sequences for univariate maps: Theory, experiments and principles. *Computer Graphics and Applications* (1988). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7760

[46] Colin Ware. 2004. *Information visualization: perception for design* (3rd editio ed.). Elsevier. 486 pages. DOI:http://dx.doi.org/10.1016/B978-0-12-381464-7.00018-1

[47] Martijn Wijffelaars, Roel Vliegen, Jarke J Van Wijk, and Erik J. Van Der Linden. 2008. Generating color palettes using intuitive parameters. In *Computer Graphics Forum*, Vol. 27. Wiley Online Library, 743–750. DOI:http://dx.doi.org/10.1111/j.1467-8659.2008.01203.x