Graphical Perception of Continuous Quantitative Maps: the Effects of Spatial Frequency and Colormap Design

Khairi Reda
Indiana University–Purdue University Indianapolis
Indianapolis, IN, USA
redak@iu.edu

Pratik Nalawade
Indiana University–Purdue University Indianapolis
Indianapolis, IN, USA
pnalawad@iupui.edu

Kate Ansah-Koi
Indiana University–Purdue University Indianapolis
Indianapolis, IN, USA
kayansah@iupui.edu

ABSTRACT
Continuous 'pseudocolor' maps visualize how a quantitative attribute varies smoothly over space. These maps are widely used by experts and lay citizens alike for communicating scientific and geographical data. A critical challenge for designers of these maps is selecting a color scheme that is both effective and aesthetically pleasing. Although there exist empirically grounded guidelines for color choice in segmented maps (e.g., choropleths), continuous maps are significantly understudies, and their color-coding guidelines are largely based on expert opinion and design heuristics—many of these guidelines have yet to be verified experimentally. We conducted a series of crowdsourced experiments to investigate how the perception of continuous maps is affected by colormap characteristics and spatial frequency (a measure of data complexity). We find that spatial frequency significantly impacts the effectiveness of color encodes, but the precise effect is task-dependent. While rainbow schemes afforded the highest accuracy in quantity estimation irrespective of spatial complexity, divergent colormaps significantly outperformed other schemes in tasks requiring the perception of high-frequency patterns. We interpret these results in relation to current practices, and devise new and more granular guidelines for color mapping in continuous maps.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Scalar field visualization; continuous colormaps; perception

INTRODUCTION
Continuous, 'pseudocolor' maps visualize how a quantitative attribute varies smoothly over space by mapping data intervals to color gradients. These maps support a range of graphical tasks, from quantity estimation (e.g., estimating air temperature at a specific location), to the comprehension of patterns and structures throughout the image. Continuous maps are common in scientific publications, especially in the physical and climate sciences. However, they are also widely used to disseminate weather information to lay citizens, particularly during inclement conditions. Naturally, the choice of colormap affects the visual appearance of the image and potentially impacts data perception. While this choice is occasionally dictated by convention, often there is no clear agreement on what colormap to use. For example, designers of weather maps employ one of several color schemes to illustrate the geographic distribution of temperatures; some choose the popular *rainbow* scheme while others might employ a diverging blue-to-red scale.

A large body of research has been devoted to understanding how color encoding affects people's perception of information in discrete maps [9]. Cartographers have analyzed the effectiveness of various color schemes for choropleths [29, 7], and contributed robust guidelines and tools for designing segmented colormaps [5, 6, 15].

By contrast, the graphical perception of *continuous maps* for smooth spatial data remains significantly understudied. The few extant studies have produced inconclusive evidence and inconsistent colormap recommendations [8]. For example, while some studies found rainbow colormaps to be accurate for surface interpretation [23, 18], others indicate rainbow to be ineffective, especially as compared to diverging color schemes [3]. Because of these inconsistencies, color encoding advice is largely based on expert opinion and designer intuition, rather than being grounded in empirical evidence. Existing guidelines –such as those discouraging the use of rainbow [4]– are often at odds with the visualization practices of the scientific community, and sometimes even contradict established results [41]. Further research is thus needed to validate current practices [22], and to establish evidence-based guidelines for color-coding in continuous spatial data.

We study the graphical perception of continuous maps, investigating the impact of *Colormap design* and *Spatial frequency*–a measure of spatial variance. We conducted three crowdsourced experiments to test the effectiveness of commonly prescribed colormaps, comparing their accuracy under different kinds of tasks and at increasing levels of spatial frequency. Results indicate that spatial frequency impacts the effectiveness of color encoding, but the precise effect is dependent on the task. We find that rainbow colormaps afford the highest accuracy

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ISBN 978-1-4503-5620-6/18/04...$15.00

DOI: https://doi.org/10.1145/3173574.3173846

in a quantity estimation task irrespective of spatial frequency. However, for pattern-matching tasks, we find that divergent colormaps significantly outperform other schemes when the underlying data exhibits high-spatial variance. These results have significant design implications and suggest complementary perceptual roles for hue-varying and divergent schemes. We distill these findings into new color mapping guidelines for continuous maps, accounting for task and spatial complexity of the data.

RELATED WORK

Color mapping involves the transformation of quantitative or categorical attributes into color by means of a *colormap*. To be of practical use, color mapping must enable the viewer to deduce quantities, distributions, and patterns present in the original data [43]. Researchers outline a number of properties believed to contribute to effective colormap design [39]. A good colormap sequence should be naturally orderable (e.g., from cool blue to warm red), so as to perceptually reflect the order of the originally mapped quantities. Additionally, a colormap should only reflect actual differences in the data without creating artificial boundaries in color [34]. Using these broad principles, researchers developed design tools to provide colormap recommendations for designers. For example, ColorBrewer suggests a set of carefully crafted and validated palettes [15]. Similarly, Colorgorical enables users to generate categorical colormaps on-demand using perceptual optimizations, while allowing for user-provided constraints [13]. The majority these tools, however, are intended for crafting segmented colormaps, and are primarily aimed at discrete map representations (e.g., choropleths). One exception, the PRAVDAColor tool [2], provides color mapping advice for continuous maps based on the data's spatial frequency and the intended task. However, unlike ColorBrewer, this advice is based on design heuristics that have not been verified.

Design Strategies for Continuous Colormaps

Researchers have proposed a number of handpicked and procedurally generated colormaps for continuous data. For instance, Herman and Levkowitz devised a greyscale that maximizes CIELAB differences within the gradient, finding that it reduces estimation errors by 20% compared to a linearly interpolated scale [17]. Greyscale ramps are thought to be effective at revealing shapes and forms, However, they are susceptible to large simultaneous contrast shifts, making them less useful for quantity estimation [41]. One alternative to greyscales involves varying hues instead of lightness, typically via a gradation based on the electromagnetic spectrum. The result is a vivid, fully saturated colormap that looks like a rainbow. Although popular in scientific visualizations, *rainbow* has been the subject of much critique in the visualization community. Experts argue that the order of hues in rainbow is not readily apparent, making it unsuitable for encoding interval data [33]. Moreover, rainbow introduces sharp visual boundaries around its yellow regions, which can be misinterpreted by viewers who might infer nonexistent features in the data [4].

Given the above limitations, researchers proposed many alternatives to rainbow. For instance, 'Spiral' colormaps comprise a limited hue rotation combined with monotonically-increasing luminance. The result is a colormap that spirals up in the hue cone while simultaneously gaining luminance [42]. Similarly, *cubehelix* incorporates sinusoidal RGB variations accompanied by a monotonic buildup in luminance [14]. Kindlmann et al. proposes an isoluminant version of rainbow [21], while Moreland advocates for diverging colormaps, which incorporate two opposing hues at the endpoints while passing through an unsaturated tone (typically white) [24]. These colormaps are thought to provide 'perceptually uniform' alternatives to rainbow, by exerting control over luminance while providing a level of hue variation. Although strongly favored by visualization experts, evidence of their effectiveness remains inconclusive (e.g., see [23, 18] vs. [3]). Our study compares these different design strategies by testing a representative sample of colormaps, including rainbow, Spiral, and diverging schemes.

Empirical Evaluations of Continuous Colormaps

It is generally recognized that colormap designs should be adaptive to the intended graphical task [31]. Ware argues that colormaps should monotonically increase their luminance when the goal is to comprehend shapes and spatial features [41]. By contrast, when the goal is to estimate quantities, a colormap should be designed to reduce simultaneous contrast effects, by registering non-monotonic variation in at least one of the three opponent-process channels. Ware confirms this latter hypothesis, finding spectral ramps (i.e., rainbow) to be the most accurate in quantity estimation, but finds little support for the shape perception theory. He then suggests that Spiral colormaps would be ideal for both quantity estimation and form perception [41]. A study by Borkin et al. finds diverging ramps to be significantly more accurate than rainbow when diagnosing heart disease from arterial scans [3]. However, Borkin et al's results contradict two earlier studies, which found spectral schemes to be more accurate in both quantity and surface interpretation [23, 18]. Such inconsistencies highlight a limitation in current literature; prior studies employed widely varying test conditions and tasks, thus complicating their comparison. Our work directly addresses this limitation. We evaluate colormaps under comparable experimental conditions, and in a range of tasks: from quantity estimation to form and pattern comprehension.

Effects of Spatial Frequency

Spatial frequency is a measure of the level of variance (or the amount of information) that is present in a degree of visual angle. Maps with sharp edges and small features will generally convey more information, and thus exhibit higher spatial frequency components. Conversely, maps with broad, smooth surfaces contain less spatial variation, and thus exhibit lower spatial frequency. Spatial frequency is thought to have a critical role in visual perception; some vision theories suggest that the visual cortex operates on a code of spatial frequency, as opposed to a code of straight lines and edges [11, 12]. Moreover, spatial frequency is inversely proportional the average size of visual features in the scene, and size is known to impact color perception [35, 36]. Given these factors, spatial frequency is likely to affect the perception of continuous maps, and possibly modulate the effectiveness of color

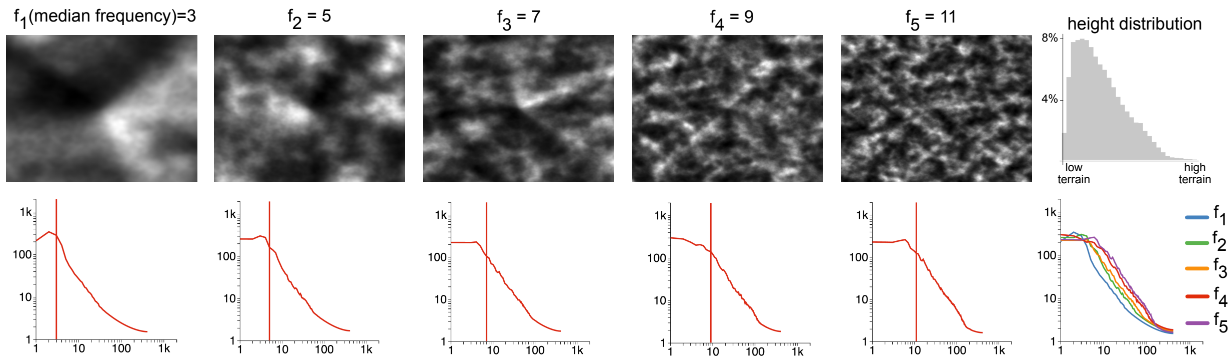

Figure 1. Five example scalar fields used as stimuli in this study. The fields represent the height of procedurally generated terrain (brighter is higher), and are ordered according to their *median spatial frequency* (cycle per 8° of visual angle[‡]). Log-log plots depict the power spectra of each scalar field (combined in the rightmost plot to aid comparison). The position of the median frequency, which splits the power distribution into two approximately equal halves, is illustrated with a vertical line. Although different with respect to spatial frequency characteristics, the maps are very similar in their distribution of height amplitudes (shown in the top-right plot).

schemes to varying degrees. Rogowitz et al. argue for two colormap design strategies depending on spatial frequency; they recommend ramps with monotonically increasing luminance for datasets containing high spatial frequency, and hue- or saturation-varying for low-frequency data [32]. Put differently, we would expect sequential and spiral colormaps (e.g., *cubehelix*) to perform better in scalar fields that have rough surfaces and narrow features. Conversely, we can expect hue-varying ramps (e.g., *rainbow*) to work well with maps that have broad surfaces and low variance. This guideline is consistent with color difference experiments that tested viewers' sensitivity to frequency-modulated Gabor patches [19]. However, it is doubtful whether such experimental results (and the ensuing guideline) generalize to visual analysis tasks on actual scalar fields.

Summary

Color-coding guidelines for continuous maps often come in the form of advice that discourages the use of rainbow [4, 25] and suggests perceptually uniform alternatives [24]. This clinical, omnibus advice is largely based on expert intuition and design heuristics. However, the literature paints a more complex picture, and suggests the choice of color encoding should be based on both task [31, 37] and data characteristics, including spatial frequency [32, 2]. This paper provides a first experimental account of the impact of spatial frequency on people's ability to estimate quantities and perceive patterns in quantitative maps. By studying how spatial complexity modulates the effectiveness of colormap designs, we can devise more nuanced guidelines that are responsive to both data characteristics and viewers' information needs (i.e., tasks).

HYPOTHESES

Building on prior research, we developed three hypotheses:

H1—We expect colormaps comprising large hue variations to be perceived more accurately in scalar fields containing low spatial frequency. Conversely, we expect ramps with monotonically increasing luminance to yield higher accuracy in high-frequency data. These predictions are based on the contrast-sensitivity of our visual perceptual system, which responds more robustly to chromatic and hue variation when assessing broad smooth surfaces, and to lightness differences when resolving small features [32].

H2—In quantity estimation tasks (experiment 1), where the goal is to estimate quantities at specific locations, we expect colormaps having substantial hue variation to perform better. This conjecture assumes that hue-varying ramp will register sinusoidal variations along the chromatic opponent-process channels. Such non-monotonic variations reduce simultaneous contrast shifts because they are less likely to systematically weigh chromatic processing in a particular direction [41].

H3—In tasks requiring the comprehension of forms and structures (experiments 2 and 3), we expect colormaps having monotonically increasing luminance to perform best. Our visual system infers surface information largely from shading cues and luminance variation [27]. Therefore, colormaps that exert linear control over their luminance can be expected to portray forms and structures more effectively.

Although there is existing evidence to back H2 (see an earlier study by Ware [41]), our work aims to replicate and extend these results to account for spatial frequency. To that end, H1 provides a broader (yet untested) prediction of how spatial frequency might impact the performance of continuous colormaps.

METHODOLOGY

In the following sections, we present the results of three crowdsourced experiments to test the above hypotheses. Specifically, we investigate whether the effectiveness commonly prescribed colormaps is modulated by the spatial frequency of the data, and the degree to which this relationship is influenced by variations in luminance, hue, and saturation within the color ramp. Each experiment tests one specific task against nine colormaps and at increasing levels of spatial frequency. The first experiment measures participants' ability to estimate values at specific locations in the map. The second experiment tests participants' accuracy in comparing gradients in larger map swaths. The third and final experiment is aimed at evaluating participants' ability to perceive and match longitudinal patterns in the map. Before delving into the details, we describe our experimental design and stimulus generation procedure.

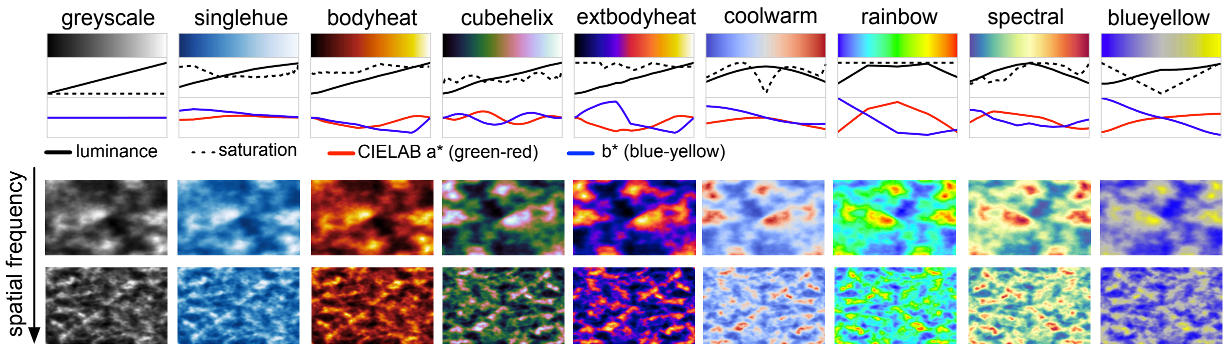

Figure 2. We tested nine colormaps selected to encompass a variety of design characteristics (illustrated by variation in luminance, saturation, and/or hue). Each colormap was tested with multiple scalar fields corresponding to increasing levels of spatial frequency.

Stimuli

We employ digital elevation models (DEM) as stimuli for the three experiments. A DEM represents land elevation, with cells in the 2D scalar field representing terrain height at the corresponding locations. To maintain precise control over spatial frequency and task difficulty, we synthetically generate DEMs using Perlin noise [28], mixing five octaves of the noise function to produce seemingly realistic terrain. The resulting DEMs are then normalized so that their heights span the entire colormap range. All generated maps were 820×630 pixels in size (approximately $16° \times 13°$ of visual angle[‡]).

By varying the scale of the noise function, we obtain scalar fields with different spatial frequency characteristics. The latter is measured by first computing a Fast Fourier Transform (FFT) over the scalar field, and calculating the relative contribution of each carrier frequency from the magnitude of its FFT vector. The result can be illustrated with a power spectra plot for each individual scalar field, with spatial frequency on the x-axis and the contribution of the frequency component on the y-axis. Low frequency fields exhibit a more pronounced right skew in their power spectra. We compute the position of the *median spatial frequency*, which splits the power spectra into approximately equal halves, and use it to order the fields. Fields with larger median frequencies indicate more varied and complex terrain structure. This procedure enabled us to synthesize scalar fields with very similar height distributions while providing precise control over spatial frequency. Figure 1 illustrate examples generated using this method.

Colormaps

We chose nine commonly-used colormaps (listed in Table 1 and illustrated in Figure 2). In addition to a *greyscale* baseline, the colormaps selected reflect five design strategies:

– Sequential: monotonically increasing luminance over a limited number of hues (*singlehue*, *bodyheat*)

– Spiral: monotonically increasingly luminance with multiple hues (*cubehelix*, *extbodyheat*)

– Diverging with uniformly-stepped luminance (*coolwarm*, *spectral*)

– Diverging with uniformly-stepped saturation (*blueyellow*)

– Fully saturated hues (*rainbow*).

All colormaps were interpolated in the CIELAB color space, with the exception of *coolwarm*, *blueyellow*, and *cubehelix* — these were interpolated (as originally intended) in a polar form of the LAB space [24], in the HSL space, and using a tapered RGB helix [14], respectively.

Colormap	Luminance control	Hues	Design strategy
Greyscale: Linear black to white ramp, interpolated in the CIELAB color space.	monotonic increase	-	**luminance ramp**
Singlehue: Monotonically increasing luminance over a single blue hue (from Color Brewer [15]).	monotonic increase	blue	**sequential**
Bodyheat: Monotonically increasing luminance with a limited hue profile similar to a heated metal filament.	monotonic increase	red, yellow	**sequential**
Cubehelix: Monotonically increasing luminance with sinusoidal RGB rotation [14].	monotonic increase	sinusoida RGB	**spiral**
Ext-bodyheat: Monotonically increasing luminance based on bodyheat, but augmented with additional blue and purple hues in the low regions [41].	monotonic increase	blue, red, yellow	**spiral**
Cool-warm: Diverging with blue and red at the endpoints and soft white at the middle [24]. Uniform luminance steps with darker ends and a bright midpoint.	uniform, mid peak	blue, red	**diverging**
Spectral: Diverging, multi-hue encompassing a subset of the rainbow with a yellow middle [15]. Uniform luminance steps with darker ends and a bright midpoint.	uniform, mid peak	limited RGB	**diverging**
Blue-yellow: Diverging, uniformly-stepped saturation with blue and yellow ends and 75% grey in the middle.	-	blue, yellow	**diverging**
Rainbow: Fully saturated hue gradation (blue, green, yellow, red), interpolated in CIELAB	-	saturated RGB	**hue rotation**

Table 1. The nine colormaps evaluated in this study.

Experimental Design

We investigate two independent variables: Colormap and Spatial Frequency. Colormap comprised nine distinct categories (see above), whereas Spatial Frequency is a continuous variable representing the number of cycles in 410 pixels (i.e., half the width of our map stimulus, or approximately 8° of visual angle[‡]). We sampled spatial frequency at five intervals: 3, 5, 7, 9, 11. To systematically study the effect of this variable on the

[‡]Following [36], we derive expected visual angle measurements from pixel dimensions by assuming standard web viewing conditions. W3C-compliant browsers render HTML images at 96 DPI [40], and automatically remap this to compensate for actual display resolution. We assume a viewing distance of 30 inches. Thus, the estimated visual angle for an object of size S *pixels* is $\theta = 2tan^{-1}(\frac{(S/2)/96}{30})$

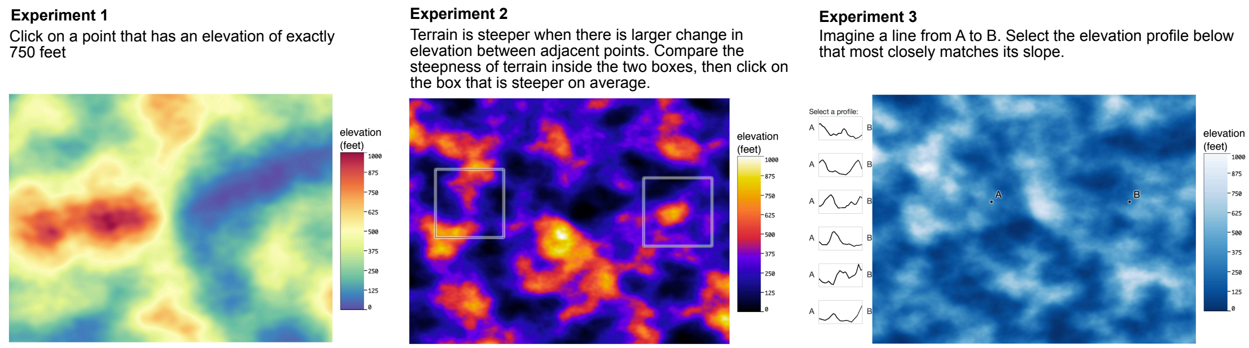

Figure 3. Example stimuli from the three experiments. In experiment 1, participants indicated their response by clicking a point on the map matching a specified elevation. In experiment 2, participants were prompted to select the steeper of the two boxes. In experiment 3, participants were asked to identify the pattern corresponding to the terrain profile between two horizontally displaced markers.

different colormaps, we opted for a factorial design, testing all possible 9×5 Colormap and Spatial Frequency combinations. Given the sheer number of combinations, we opted for a mixed design to make the study feasible. Participants were randomly assigned to one of three experimental conditions (illustrated in Table 2). Each condition comprised 3 of the 9 colormaps (i.e., between-subject) and all 5 frequency levels (within-subject). In effect, every participant saw 3 colormap $\times$ 5 spatial frequency combinations. Participants completed multiple trials with each combination.

To equalize task difficulty, stimuli for a given spatial frequency trial were derived from the same base scalar field with different colormaps applied. This arrangement enables us to make direct comparison between the colormaps for a given frequency. However, it also meant that participants will see the same map three times, albeit with different colormaps. To prevent learning, scalar fields were flipped either horizontally or vertically, resulting in three unique map reflections. The order of colormap presentation was fully counterbalanced across participants to minimize residual learning or fatigue effects.

Condition	Colormaps tested	Spatial frequencies tested
1	greyscale, cubehelix, rainbow	3, 5, 7, 9, 11
2	singlehue, extbodyheat, spectral	3, 5, 7, 9, 11
3	bodyheat, coolwarm, blueyellow	3, 5, 7, 9, 11

Table 2. Three experimental conditions, each included 3 of 9 colormaps (i.e., between-subject variation), and all 5 levels of spatial frequency.

EXPERIMENT 1: QUANTITY ESTIMATION

The first experiment tests participants' ability to identify locations on the map matching specified elevations. Participants were instructed to "Click on a point that has an elevation of exactly *[H]* feet". Five different values for H were tested: 0, 250, 500, 750, and 1000 feet. These values correspond to the three quartiles of the color scale as well as the min and max.

Participants

We recruited 90 participants from Amazon Mechanical Turk (50 females, 40 males) with a mean age of 34.64 ($STD = 9.53$ years). Participants were first screened for color-vision deficiency using a 14-panel Ishihara test, and had to correctly guess the number in 12 of the 14 panels to qualify. We restricted the study to participants with a screen resolution of at least 1280×800 to ensure the experimental interface would fit their display. Participants received a base reward of \$0.50 and a maximum bonus of \$3.00 based on the percentage of correctly solved tasks (for a possible total of \$3.50).

Procedure

After signing up for the study, participants were directed to an external link that displayed the experiment within a web interface. Participants entered their MTurk ID, and were presented with an information sheet about the study. They were then presented with the color-vision qualification test. Those who successfully passed the test were given a set of 6 training trials, and provided with feedback on their accuracy. Participants had to identify a location that is within a 5% margin from the specified height before proceeding to the next training trial.

The main portion of the experiment consisted of 3 rounds, one with each of the 3 colormaps. In each round, participants saw the five spatial frequency levels in ascending order, providing a progression from simple to more complex maps. The order of colormap presentation was fully counterbalanced across participants using a Latin square design. Participants completed 5 trials with each colormap and spatial frequency combination, corresponding to the 5 tested quantities (0, 250, 500, 750, and 1000) presented in random order. A color scale was displayed to the right of the map, and the range of the scale was fixed at [0–1000] feet (see Figure 3). In each trial, participants first saw the question and clicked on 'Show Map' to reveal the stimulus. They then indicated their response by clicking on the map to mark their selected location, and clicked 'Next'. To aid participants in accurately selecting locations, the mouse cursor was changed to a crosshair with a hollowed-out center (so as not to obscure the focal pixel).

Results

We computed an 'error' measurement for each response by taking the absolute difference between the requested elevation and the elevation at the point clicked by the participant. We then applied the following log transform [10, 16]:

$$log_2(error) = log_2(|judged\ percent - true\ percent| + 1/8)$$

We removed three participants from the analysis (amounting to 3.3% of subjects) because their overall accuracy was two standard deviations below the mean accuracy for all participants ($M = 85.36\%, STD = 9.48\%$). We analyze the results

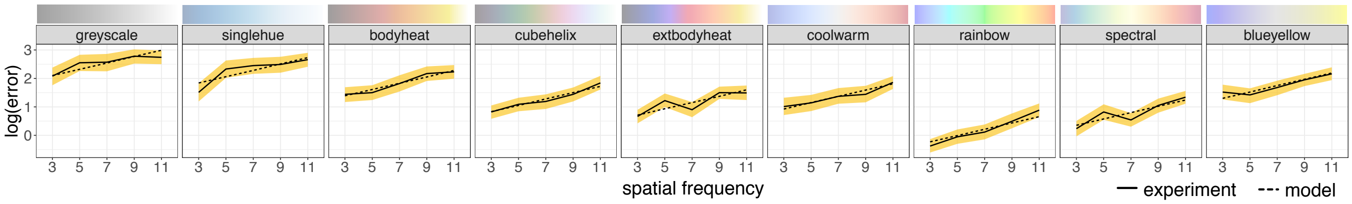

Figure 4. Mean log of error in quantity estimation (experiment 1 vs. model). Ribbons represent 95% CIs of the experimental results.

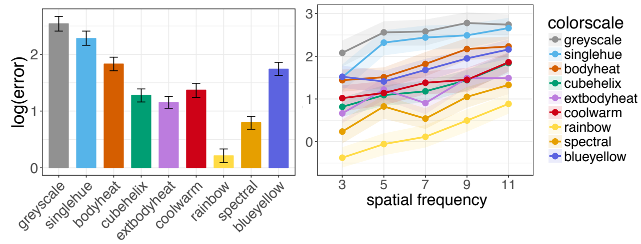

Figure 5. Mean log error in quantity estimation by colormap (left) and by colormap × spatial frequency. Intervals are 95% CIs.

Coefficient	Estimate	$\|t\ value\|$	p
(Intercept)	1.77	6.628	**
singlehue	-0.26	2.253	*
bodyheat	-0.71	6.118	***
cubehelix	-1.27	16.467	***
extbodyheat	-1.39	12.073	***
coolwarm	-1.18	10.122	***
rainbow	-2.33	30.264	***
spectral	-1.75	15.172	***
blueyellow	-0.80	6.871	***
Spatial Frequency	0.11	16.967	***

Table 3. Effects of colormap and spatial frequency on the log of error in quantity estimation. The intercept represents *greyscale* as colormap ($*** = p < 0.001, ** = p < 0.01, * = p < 0.05$)

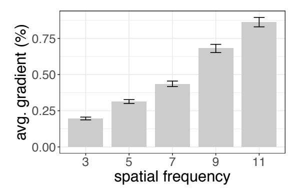

Figure 6. Average local gradient (i.e., terrain slope) at locations selected by participants. Error bars are 95% CIs.

by fitting the log of error to a linear, mixed-effects model comprising two fixed effects (colormap, frequency) and two random effects. The first random effect accounts for individual variations among participants and the second for intra-trial variations (recall that trials comprised different test elevations).

Figure 4 illustrates the mean log error by colormap and spatial frequency (the dashed trendline represents the model). The experimental results are combined in Figure 5 to ease comparison. A likelihood ratio test indicates the overall model is significant ($\chi^2(18) = 1481.7, p < 0.001$). To test for interaction between spatial frequency and colormap, we fit a reduced model that accounts for both frequency and colormap, but not their interaction. There was no significant difference between the full and the reduced model ($\chi^2(8) = 10.695, p = 0.219$), thus ruling out an interaction between colormap and spatial frequency. We will therefore interpret the reduced model which accounts for both factors independently. Table 3 illustrates the model coefficients.

The model predicts that a step-increase in spatial frequency yields a 0.11 increase in the log of estimation error. The difference in estimation error between the highest (*f=11*) and lowest (*f=3*) frequency levels is approximately 0.9 orders of magnitude. The effect of color encoding was equally evident; all the colormaps were significantly better than *greyscale*. However, the gain in accuracy was markedly different between the colormaps. *Rainbow* had the largest impact on estimation accuracy, reducing error by approximately 2.3 orders of magnitude compared to *greyscale*. The runner-up was *spectral*, which also contains substantial hue variation. However, *spectral* reduced error by 1.75 orders of magnitude only. On the other hand, Spiral colormaps (*extbodyheat*, *cubehelix*), which comprise multiple hues over a monotonically increasing luminance, decreased estimation errors by approximately 1.3-1.4 orders of magnitude, compared to 0.8-1.2 for Diverging ramps (*blueyellow*, *coolwarm*). Sequential schemes (*singlehue* and *bodyheat*) had the least impact on error, with a mere improvement of 0.26-0.71 orders of magnitude relative to *greyscale*.

The fact that we did not find interaction between colormap and spatial frequency implies that the relative effectiveness of the different colormaps is stable across all spatial frequency levels tested. *Rainbow* is thus expected to be the most accurate colormap for quantity estimation, regardless of how spatially complex the data is. However, estimation accuracy will decrease comparably for all colormaps as the data becomes more spatially varied. This could reflect a combination of perceptual and motor difficulty in locating and clicking the intended location, due the larger local gradients encountered in high-frequency maps (see Figure 6).

EXPERIMENT 2: GRADIENT PERCEPTION

The second experiment tests participants' accuracy in comparing and judging the steepness of gradients. The ability to judge how fast the encoded quantities change between adjacent map locations is important in many contexts.

Participants

We recruited 126 participants (50 females, 74 males, 2 others) with a mean age of 35.62 years ($STD = 9.55$ years). Participants had an overall success rate of 67.75% ($STD = 11.41\%$). Ten participants (7.9% of subjects) were dropped from the analysis because their overall accuracy was worse than chance, having correctly answered less than 50% of trials in a two-alternatives forced choice experiment.

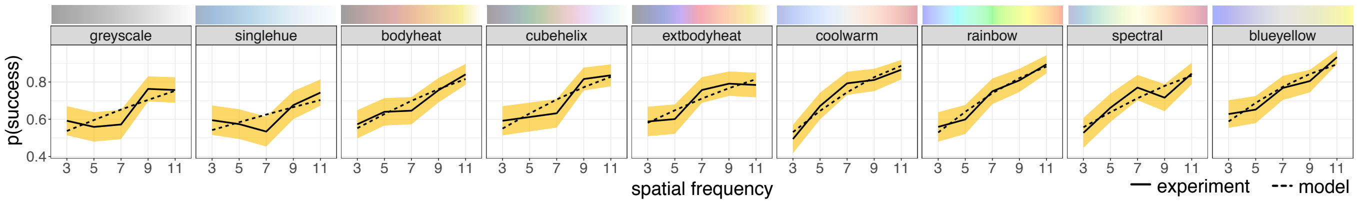

Figure 7. Probability of successful gradient judgment (experiment 2 vs. model). Ribbons represent 95% CIs of the experimental results.

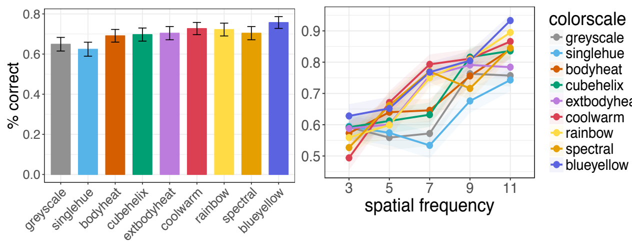

Figure 8. Percentage of correctly answered trials in a gradient perception task (experiment 2). Intervals are 95% CIs.

a. Main effects

Coef.	Est.	$\|z\|$	p
(Intercept)	0.80	0.516	
singlehue	1.14	0.415	
bodyheat	0.94	0.183	
cubehelix	0.90	0.354	
extbodyheat	1.13	0.388	
coolwarm	0.66	1.316	
rainbow	0.65	1.354	
spectral	0.92	0.259	
blueyellow	0.91	0.281	
Frequency	1.15	4.633	***

b. Interaction effects (colormap x frequency)

Coef.	Est.	$\|z\|$	p
singlehue	0.96	0.910	
bodyheat	1.04	1.020	
cubehelix	1.06	1.303	
extbodyheat	1.03	0.596	
coolwarm	1.14	3.008	**
rainbow	1.13	2.876	**
spectral	1.06	1.330	
blueyellow	1.12	2.471	*

Table 4. Main effects of colormap and spatial frequency on success odds in gradient judgment (a) and their interaction. Coefficients shown correspond to the exponented model estimates to reflect odd-ratios. The intercept represents *greyscale* ($*** = p < 0.001$, $** = p < 0.01$, $* = p < 0.05$)

Procedure

Each trail consisted of a map with two squares juxtaposed on top (see Figure 3). Participants were prompted to "compare the steepness of terrain inside the two boxes" and "click on the box that is steeper on average." The two boxes were identical in size (175×175 pixels, or $3.5° \times 3.5°$ of visual angle). However, terrain steepness, which was calculated by taking the average first derivative within each box, was varied systematically. The gradient-ratio between the flatter and the steeper boxes was fixed to one of four levels: $0.8, 0.83, 0.86, 0.9(\pm 0.05)$. A lower ratio implies larger and potentially more perceptible slope difference, making the task easier. However, the two boxes encompassed terrain with identical height ranges to reduce variability in the appearance of their peaks (a potential confound in slope judgment [26]).

Participants first completed a set of 6 training trials that included feedback, before proceeding to the main trials. The order of stimuli was similar to the previous experiment; the study consisted of 3 rounds, one with each of the 3 colormaps the participant was assigned to see. Each round encompassed all 5 spatial frequency levels. Participants completed 4 trials with each colormap and frequency combination, spanning a range of easy to difficult tests (a total of 60 trials). The order of colormap presentation was fully counterbalanced across participants.

Results

Figure 7 illustrates participants' probability of correctly identifying the steeper gradient. The experimental data is shown separately in Figure 8. We fit the results to a logistic regression model comprising two fixed effects (colormap, frequency). The model also included two random effects to account for individual differences among participants and intra-trial variations (recall that trials varied in difficulty). The model essentially predicts the odds of correctly identifying the steeper gradient. A likelihood ratio test indicates the overall model is significant ($\chi^2(17) = 421.65, p < 0.001$). The model correctly predicts 74.67% of outcomes. We find significant interaction between colormap and spatial frequency ($\chi^2(8) = 27.81, p < 0.001$). Table 4 shows model coefficients.

The main-effect coefficients for all colormaps were not significant, indicating that all colormaps perform comparably to *greyscale* at low spatial frequencies. Participants are thus unlikely to benefit from the use of color when judging gradients in low-variance data. The main effect of spatial frequency, however, is significant. The model estimates that a step-increase in spatial frequency improves the odds of correct judgment by 15%. Estimating gradients appear to be easier in maps with more complex spatial structures.

The model indicates several noteworthy interactions. Although the use of color had no significant effect in low-frequency maps, several colormaps significantly outperformed *greyscale* at high frequency. The divergent *coolwarm* improved participants' success odds by 14% for every step-increase in spatial frequency. Similarly, *rainbow* and *blueyellow* increased the odds by approximately 13% and 12%, respectively. Notably, these three colormaps contain substantial variation in saturation (*coolwarm* and *blueyellow*) or hue (*rainbow*). All other colormaps tested were not reliably different from *greyscale*.

EXPERIMENT 3: PATTERN PERCEPTION

Having tested accuracy in quantity estimation and gradient perception, we now evaluate participants' ability to integrate these two skills. Experiment 3 required participants to extract a longitudinal pattern from the map and match it to an external representation, a task originally devised by Hyslop [18].

Participants

We recruited 165 participants (79 females, 84 males, 2 others). The mean participant age was 36.04 years ($STD = 11.71$). Overall, participants had a mean success rate of 78.51% in matching the correct pattern ($STD = 19.31\%$). We dropped

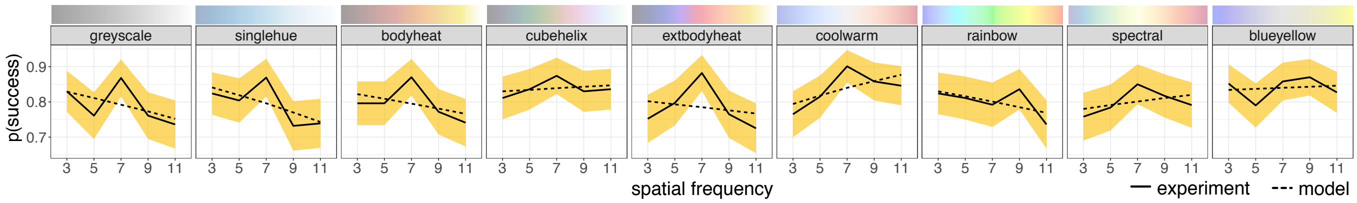

Figure 9. Probability of successful pattern matching (experiment 3 vs. model). Ribbons denote 95% CIs of the experimental data.

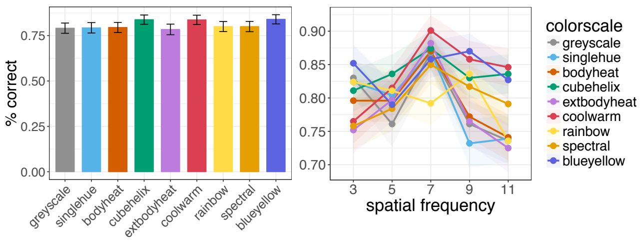

Figure 10. Percentage of correctly answered trials in experiment 3. Intervals are 95% CIs.

a. Main effects

Coef.	**Est.**	$\lvert z \rvert$	p
(Intercept)	9.03	6.502	***
singlehue	1.34	0.649	
bodyheat	0.80	0.503	
cubehelix	0.77	0.681	
extbodyheat	0.81	0.466	
coolwarm	0.44	1.868	.
rainbow	0.98	0.064	
spectral	0.57	1.284	
blueyellow	0.73	0.697	
Frequency	0.93	2.058	*

b. Interaction effects (colormap x frequency)

Coef.	**Est.**	$\lvert z \rvert$	p
singlehue	0.98	0.450	
bodyheat	1.02	0.456	
cubehelix	1.09	1.781	.
extbodyheat	1.04	0.797	
coolwarm	1.17	3.143	**
rainbow	1.01	0.259	
spectral	1.11	2.198	*
blueyellow	1.09	1.683	.

Table 5. Main effects of colormap and spatial frequency on success odds in experiment 3 (a) and their interaction. Coefficients depict exponented model estimates to reflect odd-ratios. The intercept correspond to *greyscale* ($*** = p < 0.001$, $** = p < 0.01$, $* = p < 0.05$, $. = p < 0.1$)

seven participants from the analysis (4.2% of subjects) whose overall accuracy was two standard deviations below the mean.

Procedure

Participants first completed a set of 6 training trials that included feedback, before proceeding to the main experiment. Each trial consisted of a map with two markers labeled A and B (see Figure 3). The markers were horizontally displaced by 350 pixels (7°of visual angle). Participants were given the following prompt: "Imagine a line from A to B. Select the elevation profile below that most closely matches its slope". They then selected a choice among a set of 6 patterns, including the actual elevation profile and 5 other distractors. Distractors were generated from the same map so as to reflect similar spatial frequency characteristics, and had to be 65-70% similar to the actual profile (as measured by dynamic warping [20]). Additionally, profiles and distractors were selected to not have peaks or valleys at the endpoints. These criteria, determined after a pilot, ensure similar task difficulty across the trials.

The order of stimuli was similar to the previous two experiments; the study consisted of 3 rounds, one with each of the 3 colormaps the participant was assigned to see, and encompassing the 5 spatial frequency levels. Thus, every participant saw 3×5 colormap and frequency combinations, and completed 3 pattern matching trials with each combination, for a total of 45 trials. As in the previous experiments, the order of colormap presentation was fully counterbalanced.

Results

We fit the results to a logistic regression model comprising two fixed effects (colormap, frequency) and two random effects to account for individual differences and intra-trial variations. Figure 9 shows the odds of successful profile matching. The experimental results are illustrated separately in Figure 10. A likelihood ratio test indicates the model is significant ($\chi^2(17) = 39.467, p < 0.001$). We find significant interaction between colormap and spatial frequency ($\chi^2(8) = 18.131, p < 0.05$); the relative effectiveness of the colormaps appears to vary with spatial frequency.

Overall, we find a significant, detrimental main effect of spatial frequency on pattern perception, as indicated by a 0.93 *Frequency* coefficient (Table 5.a). This translates to a 7% drop in the odds of correctly matching the profile, for every step-increase in spatial frequency. The main effect coefficients for all colormaps were not significant, indicating that the use of color at low spatial frequency is unlikely to improve pattern perception, as compared to a plain *greyscale* ramp.

Colormap performance begins to diverge at high spatial frequency. Only two colormaps have significant and large enough odds-ratio coefficients (i.e., > 1.07) to overcome the frequency-induced perceptual difficulty: *spectral* and *coolwarm* increased the odds of correct pattern matching by 11–17%, respectively, for a every step-increase in spatial frequency (after adjusting for frequency effects alone). Additionally, *blueyellow* and *cubehelix* were associated with a 9% improvement, but the advantage was not reliable ($p < 0.1$). On the other hand, *extbodyheat*, *bodyheat*, *rainbow* had small (and insignificant) odds-ratio coefficients (0.98–1.04, < 1.07), indicating that, similar to *greyscale*, they are associated with lower success odds in complex maps.

In short, only two of the tested colormaps (*coolwarm* and *spectral*) appear to reliably support pattern perception at high spatial frequency. Both consist of a diverging ramp with uniformly-stepped luminance. All other colormaps (including *greyscale*) suffered as data complexity increased.

DISCUSSION AND GUIDELINES

Our work sheds new light on how spatial complexity impacts the perception of continuous color-coded maps. The experiments also led to some surprising findings that are at odds with current guidelines. We interpret these results and accordingly

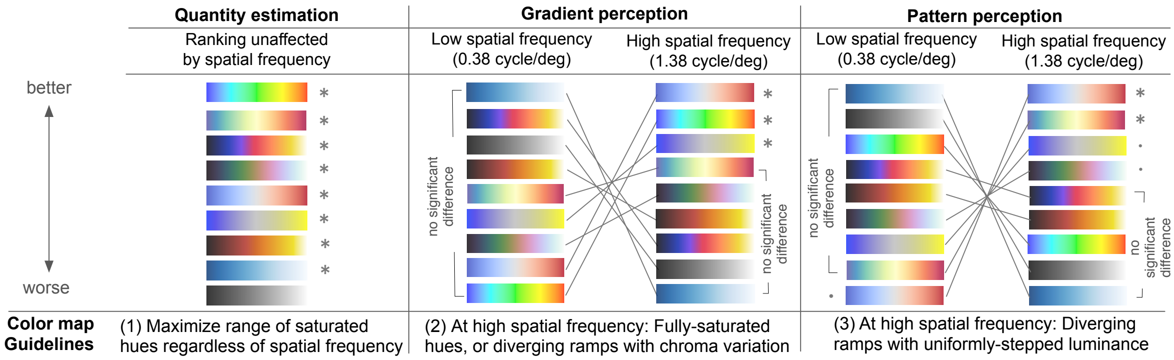

Figure 11. Model-derived colormap ranking and guidelines by task and spatial frequency ($* = p < 0.05, . = p < 0.1$, relative to *greyscale*)

devise new task- and frequency-aware color mapping guidelines (indicated by ★). We also rank the tested colormaps and summarize our guidelines in Figure 11.

Quantity Estimation

Our first hypothesis (H1) predicts hue- and saturation-varying ramps to be more accurate at low spatial frequencies, and ramps with monotonically increasing luminance to be more accurate at high frequencies. As discussed, H1 is based on the relative contrast-sensitivity of our visual system [32]. A quantity estimation task (experiment 1) shows no interaction between colormap and spatial frequency. While increased spatial complexity is associated with higher estimation error, the effect is similar across all colormaps. We thus reject H1.

On the other hands, results provide support for H2, which predicts that hue-varying ramps will lead to more accurate estimation. Indeed, the top performing colormaps (*rainbow* and *spectral*) contain substantial hue variation. Results from experiment 1 thus replicate earlier findings by Ware [41], but also extend them to show that spatial frequency have no apparent impact on the effectiveness of hue-varying ramps. Our data shows that *rainbow* and *spectral* are the most accurate among the colormaps tested, even at the highest levels of spatial frequency. Altogether, these results lend further support to the theory that lookup errors in color-coded maps are largely caused by systematic simultaneous contrast shifts [41], rather than being affected by contrast sensitivity modulation [32]. These shifts are best counteracted with colormaps that vary *non-monotonically* along one or more perceptual channels.

A corollary result is that mixing monotonic luminance with hue variation would lead to significant accuracy loss. Indeed, data from experiment 1 indicates that *rainbow* is approximately an order of magnitude more accurate than *extbodyheat* and *cubehelix*. These Spiral colormaps are designed to be more accurate *rainbow* alternatives for interval data [4, 24]. Contrary, we find that they reduce accuracy compared to a purely hue-varying ramp. This finding suggests that, when estimating a continuously coded spatial quantity, people benefit most from a large dynamic hue range. Incorporating monotonically increasing lightness within the colormap would necessarily reduce the hue range, thereby diminishing accuracy.

★ **Guideline 1**: We recommend maximizing hue variation to improve quantity estimation irrespective of spatial frequency.

Gradient Perception

Gradient perception allows people to distinguish how quickly the encoded attribute changes between adjacent locations, an essential skill when evaluating the distribution and variance of spatial data. We find that the task is strongly modulated by the data's spatial complexity; increased spatial frequency appears to enhance the perception of gradients. This is unsurprising, as maps with jagged surfaces are likely to exhibit more pronounced —and thus more perceptible— differences in slope. Colormap effectiveness was also impacted by spatial frequency; color encoding did not help participants' distinguish gradients at low frequency levels, as all colormaps showed similar performance to *greyscale*. However, three colormaps demonstrated significant advantage at high frequencies. *Coolwarm*, *rainbow*, and *blueyellow* improved perception odds by 12-14% for every step-increase in spatial frequency. All three employed one of two design strategies: a diverging ramp with varying saturation, or a fully saturated hue rotation.

The above results contradict H1, which predicts hue- and saturation-varying colormaps to perform better at low frequencies. In fact, we see the opposite. The results also do not support H3, which predicts better performance for monotonically-luminant ramps in structure perception tasks. In fact, all three top-performing ramps exhibit non-monotonic luminance.

★ **Guideline 2**: For tasks requiring gradient perception at high spatial frequency, we recommend a range of fully saturated hues (e.g., *rainbow*), or diverging chroma-varying ramps (e.g., *coolwarm* or *blueyellow*).

Pattern Perception

Experiment 3 prompted participants to match the elevation profile along a horizontal path with an external pattern. We expected colormaps with monotonically increasing luminance to be more accurate at this task (H3), but results were not entirely consistent with this prediction. While all tested colormaps had comparable performance at low spatial frequency, only two colormaps, *coolwarm* and *spectral*, gave participants higher odds of successfully matching the pattern at high frequency. Both colormaps comprise a diverging ramp with uniformly-stepped (though not strictly monotonic) luminance. By contrast, sequential and spiral ramps performed just as poorly as *greyscale* in complex maps, and so did *rainbow*.

The above result are consistent with Moreland's argument that diverging ramps provide "maximal perceptual resolution" (through increasing and decreasing luminance intervals) [24], potentially enabling high-frequency patterns to be resolved more easily. Our results may also explain why diverging schemes performed better in medical diagnosis [3]; we suspect such tasks to require the analysis of potentially high-frequency features (e.g., small tissue aberrations).

★ **Guideline 3**: We recommend diverging ramps with equidistant luminance steps (e.g., *coolwarm* and *spectral*) to support the perception of longitudinal patterns at high spatial frequency. *Rainbow*, Sequential, and Spiral schemes should be avoided in complex maps, especially if the task involves the analysis and matching of fine-grained features.

Yet Another Look at the Rainbow

Results of experiments 1 and 2 may shed a light on why *rainbow* remains a popular choice among scientists [25], despite being considered a bad choice by the visualization community [4, 30]. Our data reveals that, counterintuitively, *rainbow* is robust for estimating a smoothly varying quantitative attribute, regardless of spatial complexity. Moreover, rainbow provides good support for gradient estimation at high spatial frequency. These two tasks correspond to elementary visual analytic primitives, including characterizing distributions, determining ranges, and filtering [1]. Moreover, studies show that when experts attempt to form a mental model about a visualization, they first go through a time-consuming process of extracting quantitative data "at a rather detailed level" [38]. For instance, a weather forecaster will lookup pressure and wind changes, estimating current readings at landmark locations in the map before making a forecast. Our data suggests that *rainbow* provides good support for these tasks, making it a potentially reasonable choice for weather forecasters.

Critique of *rainbow* centers on its tendency to create sharp visual boundaries, particularly around its yellow regions [4]. Experts also criticize the use of fully saturated hues [24], which result in non-uniform perceptual steps within the color ramp. The common intuition is that these two factors combined will inevitably distort the perception of quantities. We do not see evidence to support this hypothesis. In fact, to the contrary, attempts to 'linearize' the rainbow, by monotonically increasing the luminance of hues, could reduce estimation accuracy by up to an order of magnitude.

★ **Guideline 4**: Rather than entirely discouraging the use of *rainbow*, we suggest that it can be a reasonable design choice for conveying spatial distributions and variances, and in tasks that require quantitative, as opposed to geometric, precision. However, *rainbow* has a number of limitations. The use of green and red hues is problematic for people with color deficiency. Moreover, rainbow is probably ineffective at revealing high-frequency patterns. Interestingly, these shortcomings are balanced by diverging ramps (e.g., *coolwarm*), which, although quantitatively inaccurate, appear to support pattern perception at high spatial frequency. We thus argue that hue-varying and diverging colormaps support orthogonal tasks in continuous maps, and should therefore be considered as complementary, rather than mutually exclusive choices.

LIMITATIONS AND FUTURE WORK

There are some limitations to our work that should be considered. First, as with other crowdsourced graphical perception studies, we gain access to a larger pool of participants, but sacrifice some experimental control [16]. Particularly relevant to our study is the variations in participants' monitors, including color calibration and display resolution, as well as the illumination conditions in their homes or offices — all of which can impact color perception. We could not control these factors, but attempted to counteract their variation by involving a larger sample (*N=381*). Although we expect crowdsourcing to improve the ecological validity of results and guidelines, uncontrolled variations can potentially reduce our ability to detect small but otherwise significant differences in performance between tested conditions. Future lab studies should therefore be attempted to replicate our findings with added controls.

Second, our study employed a limited set of tasks designed to measure elementary perceptual operators, including quantity estimation, gradient perception, and pattern matching. There is an opportunity to test higher-level tasks that mimic scientific analyses more closely, including the identification and comparison of larger map features (e.g., fronts, ridges). Additionally, some of the tasks we tested could be re-evaluated in more authentic formulations. For instance, a metric task could require participants to estimate the quantity at a specific location on the map. This formulation is arguably more realistic than the task we tested, which simply asked participants to click *any* location thought to match a specified quantity.

Third, our analysis was focused exclusively on spatial frequency, and there are good reasons to consider this factor [11, 32]. However, there are also additional data characteristics to consider, including, for instance, the distribution of amplitudes within the map. Such factors will influence the distribution of colors in the image and may thus impact perception.

Lastly, we limited our study to synthetically generated scalar fields to precisely vary spatial frequency while controlling for other confounds. However, synthetic stimuli may also introduce (unknown) perceptual or cognitive biases. Therefore, additional studies are needed to replicate our findings with datasets from real-world domains (e.g., meteorology, geophysics, or oceanography), and with domain experts. We also restricted this study to participants with normal color vision. Therefore, our results may not generalize to approximately 5% of the population who have some form of color deficiency.

CONCLUSIONS

We conducted three experiments to investigate the effects of spatial frequency and colormap characteristics on the perception of continuous, pseudocolor maps. Our results indicate that spatial frequency impacts judgment of the encoded quantities and structures. While viewers' quantity estimation accuracy exhibited a predictable response, increased data complexity had a more nuanced effect on gradient and pattern comprehension, the impact of which was dependent on the colormap used. Designers should therefore consider both the type of task and the spatial complexity of the underlying data. We re-examined current guidelines and devised new recommendations for color-coding of continuous spatial data.

REFERENCES

1. Robert Amar, James Eagan, and John Stasko. 2005. Low-level components of analytic activity in information visualization. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*. IEEE, 111–117.
2. Lawrence D Bergman, Bernice E Rogowitz, and Lloyd A Treinish. 1995. A rule-based tool for assisting colormap selection. In *Proceedings of the 6th conference on Visualization'95*. IEEE Computer Society, 118.
3. Michelle Borkin, Krzysztof Gajos, Amanda Peters, Dimitrios Mitsouras, Simone Melchionna, Frank Rybicki, Charles Feldman, and Hanspeter Pfister. 2011. Evaluation of artery visualizations for heart disease diagnosis. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2479–2488.
4. David Borland and Russell M Taylor Ii. 2007. Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications* 27, 2 (2007).
5. Cynthia A Brewer. 1994. *Visualization in Modern Cartography*. Elsevier Science, Chapter Color use guidelines for mapping and visualization.
6. Cynthia A Brewer. 1996. Guidelines for selecting colors for diverging schemes on maps. *The Cartographic Journal* 33, 2 (1996), 79–86.
7. 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.
8. Roxana Bujack, Terece L Turton, Francesca Samsel, Colin Ware, David H Rogers, and James Ahrens. 2017. The Good, the Bad, and the Ugly: A Theoretical Framework for the Assessment of Continuous Colormaps. *IEEE Transactions on Visualization and Computer Graphics* (2017).
9. William S Cleveland and William S Cleveland. 1983. A color-caused optical illusion on a statistical graph. *The American Statistician* 37, 2 (1983), 101–105.
10. William S Cleveland and Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554.
11. Russel De Valois and Karen De Valouis. 1990. *Spatial Vision*. Oxford University Press.
12. Russell L De Valois, Duane G Albrecht, and Lisa G Thorell. 1978. Cortical cells: bar and edge detectors, or spatial frequency filters? In *Frontiers in visual science*. Springer, 544–556.
13. Connor C Gramazio, David H Laidlaw, and Karen B Schloss. 2017. Colorgorical: Creating discriminable and preferable color palettes for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 521–530.
14. D. A. Green. 2011. A colour scheme for the display of astronomical intensity images. *Bulletin of the Astronomical Society of India* 39 (June 2011), 289–295.
15. Mark Harrower and Cynthia A Brewer. 2003. ColorBrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal* 40, 1 (2003), 27–37.
16. Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 203–212.
17. GT Herman and H Levkowitz. 1992. Color scales for image data. *IEEE Computer Graphics and Applications* 12, 1 (1992), 72–80.
18. Michael D Hyslop. 2006. *A comparison of spectral color and greyscale continuous-tone map perception*. Master's thesis. Michigan State University.
19. Alan David Kalvin, Bernice E Rogowitz, Adar Pelah, and Aron Cohen. 2000. Building perceptual color maps for visualizing interval data. In *Human Vision and Electronic Imaging V*, Vol. 3959. International Society for Optics and Photonics, 323–336.
20. Eamonn Keogh and Chotirat Ann Ratanamahatana. 2005. Exact indexing of dynamic time warping. *Knowledge and Information Systems* 7, 3 (2005), 358–386.
21. Gordon Kindlmann, Erik Reinhard, and Sarah Creem. 2002. Face-based luminance matching for perceptual colormap generation. In *Proceedings of the IEEE Conference on Visualization '02*. IEEE Computer Society, 299–306.
22. Robert Kosara. 2016. An Empire Built On Sand: Reexamining What We Think We Know About Visualization. In *Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization*. ACM, 162–168.
23. Mark P Kumler and Richard E Groop. 1990. Continuous-tone mapping of smooth surfaces. *Cartography and Geographic Information Systems* 17, 4 (1990), 279–289.
24. Kenneth Moreland. 2009. Diverging color maps for scientific visualization. In *International Symposium on Visual Computing*. Springer, 92–103.
25. Kenneth Moreland. 2016. Why We Use Bad Color Maps and What You Can Do About It. *Electronic Imaging* 2016, 16 (2016), 1–6.
26. Lace Padilla, P Samuel Quinan, Miriah Meyer, and Sarah H Creem-Regehr. 2017. Evaluating the Impact of Binning 2D Scalar Fields. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 431–440.
27. Stephen E Palmer. 1999. *Vision science: Photons to phenomenology*. MIT press.

28. Ken Perlin. 1985. An image synthesizer. *ACM SIGGRAPH Computer Graphics* 19, 3 (1985), 287–296.
29. Linda Williams Pickle. 2003. Usability testing of map designs. In *Proceedings of Symposium on the Interface of Computing Science and Statistics*. 42–56.
30. P Samuel Quinan and Miriah Meyer. 2016. Visually comparing weather features in forecasts. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 389–398.
31. Penny L Rheingans. 2000. Task-based color scale design. In *28th AIPR Workshop: 3D Visualization for Data Exploration and Decision Making*. International Society for Optics and Photonics, 35–43.
32. Bernice E Rogowitz and Lloyd A Treinish. 1994. Using perceptual rules in interactive visualization. In *IST/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics, 287–295.
33. Bernice E Rogowitz, Lloyd A Treinish, Steve Bryson, and others. 1996. How not to lie with visualization. *Computers in Physics* 10, 3 (1996), 268–273.
34. Samuel Silva, Beatriz Sousa Santos, and Joaquim Madeira. 2011. Using color in visualization: A survey. *Computers Graphics* 35, 2 (2011), 320–333.
35. Maureen Stone, Danielle Albers Szafir, and Vidya Setlur. 2014. An engineering model for color difference as a function of size. In *Color and Imaging Conference*, Vol. 2014. Society for Imaging Science and Technology, 253–258.
36. Danielle Albers Szafir. 2017. Modeling Color Difference for Visualization Design. *IEEE Transactions on Visualization and Computer Graphics* (2017).
37. Christian Tominski, Georg Fuchs, and Heidrun Schumann. 2008. Task-driven color coding. In *Information Visualisation, 2008. IV'08. 12th International Conference*. IEEE, 373–380.
38. J Gregory Trafton, Susan S Kirschenbaum, Ted L Tsui, Robert T Miyamoto, James A Ballas, and Paula D Raymond. 2000. Turning pictures into numbers: extracting and generating information from complex visualizations. *International Journal of Human-Computer Studies* 53, 5 (2000), 827–850.
39. Bruce E Trumbo. 1981. A theory for coloring bivariate statistical maps. *The American Statistician* 35, 4 (1981), 220–226.
40. W3C. 2016. CSS Values and Units Module Level 3. (2016). `http://www.w3.org/TR/css3-values/#absolute-lengths`
41. Colin Ware. 1988. Color sequences for univariate maps: Theory, experiments and principles. *IEEE Computer Graphics and Applications* 8, 5 (1988), 41–49.
42. Colin Ware. 2012. *Information visualization: perception for design*. Elsevier.
43. Liang Zhou and Charles D Hansen. 2016. A survey of colormaps in visualization. *IEEE Transactions on Visualization and Computer Graphics* 22, 8 (2016), 2051–2069.