

Evaluating Gradient Perception in Color-Coded Scalar Fields

Khairi Reda*

Indiana University–Purdue University Indianapolis

Michael E. Papka†

Argonne National Laboratory & Northern Illinois University

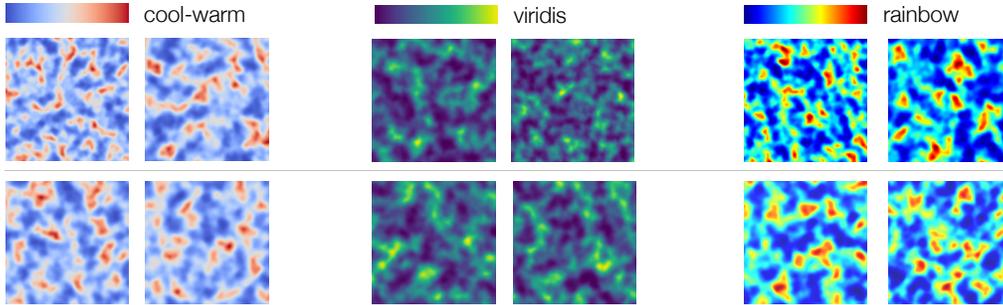


Figure 1: Six example stimuli from our experiment (actual trials showed one pair of maps at a time, separated by a larger horizontal spacing of 150 pixels). The scalar fields (200×200 pixels each) represent synthetic terrain, with color encoding elevation. For each pair of maps (on the same row and under a particular colormap), which image (left or right) shows steeper terrain on average? Steepness is defined by how quickly the elevation changes between adjacent map points. The answers are in §3.1.

ABSTRACT

Color mapping is a commonly used technique for visualizing scalar fields. While there exists advice for choosing effective colormaps, it is unclear if current guidelines apply equally across task types. We study the perception of gradients and evaluate the effectiveness of three colormaps at depicting gradient magnitudes. In a crowd-sourced experiment, we determine the just-noticeable differences (JNDs) at which participants can reliably compare and judge variations in gradient between two scalar fields. We find that participants exhibited lower JNDs with a diverging (*cool-warm*) or a spectral (*rainbow*) scheme, as compared with a monotonic-luminance colormap (*viridis*). The results support a hypothesis that apparent discontinuities in the color ramp may help viewers discern subtle structural differences in gradient. We discuss these findings and highlight future research directions for colormap evaluation.

Index Terms: Human-centered computing—Visualization—Empirical studies in visualization

1 INTRODUCTION

Color mapping is a foundational technique for visualizing how an attribute changes across geography. Continuous colormaps, in particular, map data intervals to color gradations, thereby revealing subtle spatial variations in the data. Such maps are widely used in science and engineering to analyze sensor or simulation data (e.g., air turbulence caused by jet engines). However, they also used by lay citizens, for instance, to keep watch on local weather conditions especially during emergencies. A challenge for map designers is to choose color schemes that are both aesthetically pleasing and accurate. Good colormaps must accurately convey features and summary statistics in the data through apparent color variations.

There is a rich body of guidelines for choosing color sequences [31, 33, 40]. Recently, a number of empirical studies have also been reported in the literature, providing assessment of user performance under an assortment of colormaps and analytic tasks. For

example, Ware et al. tested the resolving power of various color sequences in a task involving the discrimination of small features [38], while others studied how colormap performance is affected by spatial frequency [13, 26]. Although studies have produced evidence consistent with guidelines [8, 19], surprising results have also emerged in some experiments, where the effectiveness of colormaps was strongly dependent on the task [26]. Overall, studies hint at gaps in our understanding, and suggest that existing guidelines may not always dictate the best colormap design strategy for a given task.

In this work, we empirically study a task that has not been sufficiently characterized: the perception of gradients in scalar fields. We evaluate colormap effectiveness in helping people compare and judge spatial variations. Specifically, we design a model task to determine the just-noticeable differences (JND) at which participants can reliably contrast the gradients in two scalar fields. Concurrently, we test a hypothesis that apparent color bands in the image (e.g., hue boundaries in spectral and diverging schemes) may, counter-intuitively, facilitate this task. Comparing three colormaps (diverging *cool-warm*, *rainbow*, and *viridis*), we find that *viridis* (monotonic luminance) exhibits higher JND levels, indicating lower sensitivity. The results suggest that color discretization may aid the perception of net gradients, along with possibly other distributional characteristics. We discuss the findings and outline future research directions.

2 RELATED WORK

There exist guidelines for choosing quantitative color sequences [33, 40]. Although the vocabulary can be ambiguous [6], guidelines generally agree in spirit on three principles: 1) *perceptual uniformity*: even distances between adjacent colors; 2) *order*: a viewer should be able to intuitively order colors from the scale; 3) *continuity*: the colormap should produce a smooth image devoid of artificial boundaries. Given these principles, most researchers advocate for colors with monotonically increasing luminance, while discouraging the use of ‘spectral’ schemes (e.g., *rainbow*) [30]. The tendency for *rainbows* to create boundaries between hues (sometimes referred to as a ‘hue banding’ effect) is believed to mislead viewers [23, 33]. In this work, we explore an alternative hypothesis: that apparent color bands may help in tasks that involve reasoning about gradients.

It is worthy to note that ‘hue banding’ is not limited to spectral schemes, as several other colormaps also induce discontinuities. For

*e-mail: redak@iu.edu

†e-mail: papka@anl.gov

instance, in *cool-warm* (a perceptual, diverging scheme), one observes a relatively pronounced boundary in the middle of the scale. This edge seemingly discretizes the scale into distinct blue, grey, and red regions, though to a lesser extent than *rainbow* (see Figure 1). It is possible, however, to minimize discretization. *Viridis*, for instance, is defined as a series of equidistant spline interpolations [36] in the CAM02-UCS space [20]. By making ‘soft’ curves in a perceptual color space, *viridis* mostly avoids inducing color boundaries. The result is a generally smooth ramp that arguably satisfies the continuity principle. Liu and Heer found that viewers can accurately judge color distances in *viridis* [19], suggesting minimal separation.

2.1 Tasks and Empirical Evaluations

Pseudocolor maps support a range of graphical tasks, from quantity estimation (e.g., reading temperatures at specific locations) to the comprehension of forms and patterns [37]. These maps are also used for comparative analysis, where a viewer might compare two datasets, or different parts of the same dataset [14]. The former is a frequently recurring task in computational science. For example, climate scientists often compare scalar fields representing projected climate conditions under different scenarios or model parameters [9]. In looking at these simulations, one could be interested in small-scale features, but often, when drawing comparison, one is concerned with finding structural differences between the maps [21]. And while the discriminability of small features and the accuracy of point estimates has been studied [37, 38], less attention has been devoted to how people extract and compare summary features in multiple scalar fields. In one experiment, Dasgupta et al. tested climate scientists’ accuracy in comparing the mean value of two maps. They found participants to be generally more accurate with Brewer’s blue scheme (monotonic luminance [11]), compared to a standard or Kindlemann’s *rainbow* [8]. However, they reported higher accuracy for rainbow maps in a different task. In a study with doctors, Borkin et al. found diverging colormaps to be more accurate than *rainbow* when diagnosing heart disease from arterial scans [4].

The task used in this work also involves comparing the structure of two scalar fields. Specifically, we measure participants’ accuracy in judging the overall gradient in two images. We employ a two-alternative forced choice (2AFC) design in which participants make a single choice between two options. The 2AFC procedure offers a psychologically grounded method for assessing sensitivity [3], compared to eliciting numerical or Likert estimates [8]. Notably, 2AFC has been applied in a number of visualization studies [12, 39]. Our task is similar to the gradient perception experiment reported by Reda et al. [26]. This study extends that work by testing an explanatory hypothesis, and by modeling the thresholds (JNDs) at which individuals become sensitive to gradient variations. Note that our notion of JND refers to just-noticeable *differences in gradient* when comparing two scalar fields—not to be confused with JND as applied to color distance. As this task may be affected by spatial frequency [2, 29], we study the latter as a second independent factor.

3 EXPERIMENT

We conducted a crowdsourced experiment to investigate how colormap choice impacts people’s ability to perceive and compare gradients in 2D scalar fields. The gradient, as the first order derivative, is defined over a neighborhood of points. We therefore test participants’ sensitivity to the net *average gradient magnitude* within a field. Our task thus represents an ensemble or summary task [7, 34].

The **average gradient** is an important summary characteristic to communicate in scalar field visualizations. It enables people to compare and identify structural differences between multiple datasets. Moreover, the same measure can cue the viewer to regions showing faster spatial changes within a single field. Such regions could be interpreted as having higher levels of ‘activity’, and consequently be deemed interesting for further detailed inspection. For example, in

simulating the noise created by jet engines, physicists are often interested in regions of high turbulence within the jet plume [15], which can be spotted by looking for areas with higher average gradient. Likewise, subtle fluctuations in the cosmic microwave background, which are similarly marked by larger gradients, may indicate imprints from the early universe [17]. Therefore, the degree to which colormaps accurately depict net gradients in scalar fields (or sub-regions thereof), and the threshold at which people can perceive subtle gradient variations, carry significant practical importance. From a theoretical point of view, we expect this task to benefit from discontinuities in the color ramp (see §3.6 for elaboration).

3.1 Stimuli and Task

We employ synthetically generated scalar fields to precisely control stimuli characteristics. Fields were procedurally generated using a Simplex noise algorithm [18] with random seeds. We used a blend of multiple noise octaves to produce realistic datasets. The resulting fields were re-normalized to span the extents of the color ramp. All generated fields measured 200×200 pixels, subtending 4° of visual angle when viewed from a distance of 30 inches at 96 DPI. Our stimulus maps are thus larger than a single ‘feature’ yet compact enough for a perceptual judgment task. In generating each stimulus, we vary the Simplex parameters using a stochastic optimization, until yielding the desired gradient levels. This enables us to generate map pairs with almost identical characteristics (e.g., similar distribution of scalar values) while precisely varying gradients as needed.

Participants were told that the maps represented terrain, with elevation depicted in color. In each trial, participants were presented with two randomly generated scalar fields displayed side-by-side (150 pixels apart) and encoded using the same colormap. A color scale was displayed to the right. Participants were asked to click on the map that represented “steeper terrain on average”, with steepness indicated by “faster change in elevation between adjacent map points.” This task wording was inspired by Padilla et al. [24]. Participants made a two-alternative forced choice (2AFC) judgment (i.e., ‘left’ or ‘right’ map) based on the perceived average steepness. Upon selection, participants received a new pair of maps for a new trial.

Figure 1 shows six example stimuli.¹ The ground truth, net gradient is calculated by convolving a 3×3 Sobel filter. The sum of gradients from all pixels was averaged, yielding an *average gradient magnitude* for each field. This ground truth measurement was used to determine whether a participant had judged correctly.

3.2 Study Design

We investigate two independent variables: Colormap and Spatial Frequency. We tested three colormaps representing different design strategies and varying levels of hue discretization:

- **Rainbow**: we used the ‘Jet’ variant, a previously default colormap in Matlab. *Rainbow* is often perceived as highly discretized [5].
- **Cool-warm**: a diverging blue-red scheme designed for scientific visualizations [22]. It exhibits some discretization.
- **Viridis**: a smooth, perceptually uniform map over blue, green, and yellow hues [36]. It is thought to exhibit minimal discretization.

Spatial frequency reflects the level of variation within a degree of visual angle. We control spatial frequency by adjusting the base average gradient in the stimulus fields. We test two spatial frequency levels (see Figure 2): low (base avg. gradient: 25.5%) and high (40.0%). Hereafter, we express gradient magnitudes in percentage points indicating the change in scalar value over a one-pixel step. This value reflects the output of the 3×3 Sobel filter times hundred.

We employ a full-factorial (3×2) within-subject design: all participants were tested with three colormaps at the two spatial frequency

¹The correct answers for the stimuli in Figure 1: top row (starting with the leftmost pair): *left map, right, left*; bottom row: *left, right, right*.

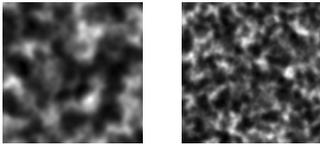


Figure 2: Examples of low (left) and high spatial frequency stimuli.

levels. The experiment was blocked by colormap. For each colormap, participants completed 2 sets corresponding to low and high spatial frequency, prior to moving to the next colormap. Each set consisted of 40 judgment trials. At the beginning of a set, the gradient difference between the two fields was initialized to 3%, and subsequently adapted throughout the set according to a staircase procedure (more on this in §3.3.1). Participants completed 6 sets (3 colormaps \times 2 spatial frequency levels), for a total of 240 (6×40) trials. The order of colormap and spatial frequency presentation was fully counterbalanced across participants. In addition to the actual trials, we included 18 engagement checks that were randomly distributed. The checks comprised easy judgment (13% gradient difference—roughly $4 \times$ easier than the starting difficulty).

3.3 Procedure

Participants were first screened for color vision deficiency using 14 Ishihara panels. After a brief tutorial, they completed 45 practice trials through which they were exposed to the three colormaps at increasing levels of spatial frequency. During practice, they were provided with feedback on whether they had judged correctly. After training, participants completed the 240 actual trials. At the end of the experiment, we asked participants to provide a brief description of the strategy they followed in the task, including any “visual features or characteristics” they used to decide between the two maps. Lastly, participants completed a demographic survey.

3.3.1 Staircase Procedure

The difference in average gradient between the two maps was adapted by a 1-up, 3-down staircase procedure: if the participant answers a trial correctly, the difference in gradient for the next stimulus is decreased by one step. On the other hand, if the participant makes an incorrect judgment, the difference is increased by three steps. We define one step as 0.285% gradient (selected after a pilot). The procedure thus dynamically adjusts task difficulty based on a participant’s performance (increasing gradient difference makes the judgment easier). Ultimately, the procedure converges to the individual’s JND: the difference at which the participant can discriminate correctly 75% of the time (halfway between chance and perfect reliability). For a visual reference, the top row in Figure 1 shows map pairs where the gradient difference is $2 \times$ the average JND observed (i.e., easy judgments), whereas the bottom row is at the mean JND.

3.4 Participants

We recruited 105 participants (60 males, 45 females) from Amazon Mechanical Turk. Participants had a mean age of 37.2 years (STD : 11.1). They were compensated with \$3 upon completing the experiment. We excluded participants who failed the color vision test, or those who had low accuracy ($<70\%$) in the engagement checks (50% accuracy is chance). New participants were recruited in replacement of those excluded, until we reached our sample size.

3.5 Analysis

We fit the responses of each participant to a psychometric function (PF) [12, 16]. The PF in our study models the probability of correctly choosing the steeper of two scalar fields based on the difference in their ground-truth, average gradient magnitude. We use a Gaussian cumulative distribution function as our PF. The model enables us to

estimate two parameters: 1) the **JND**, which represents the difference in average gradient between the maps at which a participant can make correct judgment 75% of the time, and 2) the **spread** (equivalent to the standard deviation of the Gaussian), which estimates the degree to which the judgment is susceptible to noise. For each individual participant, we fit 6 PFs, one for every combination of colormap and spatial frequency level. We then assess how the *JND* and *spread* are impacted by colormap and spatial frequency.

The PF provides an interpretable model for our task: A colormap with lower JND allows participants to discriminate smaller differences in gradient. Similarly, a small spread implies that, once the JND is met, the judgment is most likely to be accurate. Conversely a larger spread implies a flatter response curve with more ambiguous judgments. The intuition behind the PF is illustrated in Figure 3.

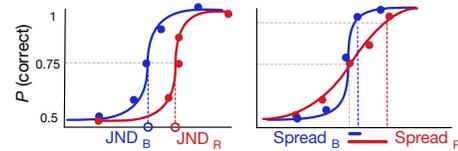


Figure 3: We model the probability of comparing two maps correctly (Y axis) as a psychometric function of the gradient difference between them (X axis). The red curve on the left shows a hypothetical colormap with higher JND, requiring larger difference for discrimination. The curves on the right have similar JNDs but red has wider spread.

3.6 Hypotheses

We developed two primary hypotheses²:

H1—We hypothesize the discontinuities in the color ramp will enhance the perception of gradients in scalar fields. We conjecture that viewers take advantage of the emergent, discrete color features (e.g., red and blue bands in *rainbow* or *cool-warm*), and estimate their apparent *size* and *frequency* as proxies for gradient. Heuristically, smaller-sized and more-frequent color patches can be indicative of larger net gradients. It is known that people can extract such estimates from a scene in parallel [1, 32, 34]. Since both *cool-warm* and *rainbow* have a tendency to create discrete features in the image, we expect them to result in higher sensitivity to gradients and, thus, exhibit lower JNDs compared to *viridis*.

H2—We expect the above heuristic to become less effective at high spatial frequency. This is because the latter will make any discrete features progressively smaller, making it harder to obtain reliable distributional estimates.

3.7 Results³

Participants completed the experiment in 23.2 minutes on average. In total, they provided 25,200 gradient judgments. Their average accuracy was 74.3%, indicating that the staircase procedure was effective at converging to the 75% JND thresholds we sought. Average accuracy in the engagement checks (easy judgments) was 96.9%.

3.7.1 Just-noticeable differences (JND)

Figure 4 shows the individual JNDs—the gradient difference levels at which subjects achieve 75% accuracy (lower JND is better). Means are illustrated in Figure 5. We employ a linear mixed-effect model to estimate the impact of colormap and spatial frequency on the JND. The model also incorporates a random intercept to account for variations among individual participants. We find a significant main effect of colormap ($F(2, 520) = 14.771, p < 0.001$). Pairwise

²We preregistered our experimental design and hypotheses prior to data collection: <https://osf.io/ng9qw>

³The experiment data and materials are at: <https://osf.io/ew638/>

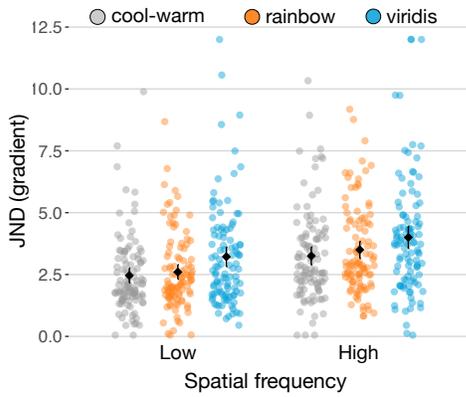


Figure 4: Fitted JND thresholds (lower is better sensitivity). Each colored dot represents the JND of one subject, showing the gradient threshold at which the subject can discriminate between two scalar fields. Diamonds depict group means ($\pm 95\%$ confidence intervals).

tests with Tukey’s adjustment show significant differences between *cool-warm* and *viridis* ($p < 0.001$) and between *rainbow* and *viridis* ($p < 0.001$). No significant difference was noted between *cool-warm* and *rainbow*. It appears that *cool-warm* (JND gradient: 2.86%, CI: 2.62–3.09%) and *rainbow* (JND: 3.05%, CI: 2.82–3.28%) afford higher sensitivity to gradients as compared to *viridis* (JND: 3.61%, CI: 3.31–3.91%). The estimated effect size (Cohen’s d) is 0.376 for *cool-warm* over *viridis* and 0.264 for *rainbow* over *viridis*.

We found a significant main effect of spatial frequency ($F(1, 520) = 49.272, p < 0.001$). The JNDs at high spatial frequency were 29.7% greater compared with low the frequency condition. We found no interaction between colormap and spatial frequency; the latter reduces sensitivity equally in all tested colormaps.

3.7.2 Spread

The spread reflects the effect of uncertainty when comparing two maps, and is measured in the same unit as the JND (gradient magnitude). A linear mixed-effect model shows no significant main effects of colormap ($F(2, 520) = 0.111$) or spatial frequency ($F(1, 520) = 0.055$). No interaction was noted ($F(2, 520) = 1.389$). Subjects experienced similar ambiguity when judging gradients with any of the three colormaps (Mean spread=1.23%, CI: 1.08–1.37%).

4 DISCUSSION

The JND levels for *rainbow* and *cool-warm* were quite similar. This suggests that the two colormaps perform similarly in tasks involving the comparison of scalar datasets, and where gradient magnitude is a main object of comparison. Participants exhibited lower JNDs with *cool-warm* and *rainbow* as compared with *viridis*, indicating that the latter requires larger gradient differences for effective discrimination. The observed effect size (0.264–0.376) is small to moderate. The results are inline with our prediction and provide evidence in support of H1. Namely, that discontinuities in color ramps may aid the perception of 2D gradients. This is because, although discrete color bands are essentially artifacts [5], viewers may rely on such features as proxies for gradient. For instance, a viewer could estimate the apparent size of the blue and red patches in Figure 1, or how frequently they seem to occur in the image. People can quickly and accurately make such distributional estimates [1, 32, 34]. Estimates about the size and frequency of peaks and valleys, in particular, can be reliable proxies for spatial variance. Such features are made salient in *rainbow* and *cool-warm* thanks to hue discretization. This kind of segmentation may also enhance ensemble perception [10]. By contrast, to the extent that *viridis* is perceived uniformly, viewers may find it difficult to apply the above heuristic.

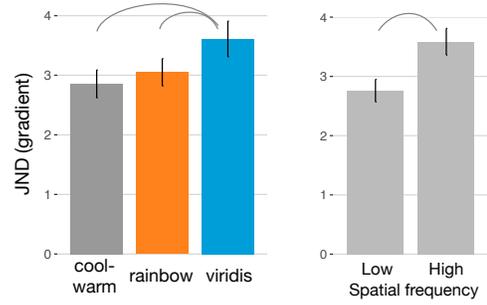


Figure 5: Mean JNDs by colormap and spatial frequency ($\pm 95\%$ confidence intervals). Arcs indicate significant post hoc differences.

Unsurprisingly, discriminating gradients is more difficult at high spatial frequency as evidenced by the larger JNDs. This is consistent with Weber’s law, which may govern the perception of statistical properties in visualizations [27, 28]. That said, our second hypothesis (H2) failed to materialize; we observe that *cool-warm* and *rainbow* continue to outperform *viridis* even at high spatial frequency. It may be that viewers can still infer distributional characteristics about fields, even as the discrete color features become smaller.

Overall, the experiment suggests that hue-based discretization may not always be a bad thing. Color bands may aid some tasks where the goal is to compare structural differences, including net gradients. Such tasks arise frequently in scientific visualization. Weather forecasters, for instance, typically start by making quantitative estimates (e.g., of air pressure) at landmark locations, and look for variations in those values to form an initial mental model [35]. Earlier studies show spectral schemes to be effective in quantity estimation [26, 37]. This study contributes evidence that spectral and diverging sequences may help people in reasoning about certain spatial characteristics, more so than monotonic-luminance ramps. Although our study does not rule out an alternative explanation, the results advance the hypothesis (H1). This, in turn, could partially explain why spectral schemes remain a popular choice among practitioners [25]. It is important to note, however, that *rainbow* colormaps can be problematic in other tasks [4, 5], and can exclude people with color vision deficiency. Therefore, they should still not be recommended as standalone visualizations.

Our findings suggest a new avenue for evaluating colormaps, by testing their ability to convey distributional and structural characteristics in scalar data. Such tasks, we argue, are important for many users, but are not sufficiently addressed in formal studies. The model task introduced here could be adopted to measure the JND thresholds for other spatial statistics, such as autocorrelation and amplitude skewness. Results can then be used to derive task-specific color encoding guidelines.

5 CONCLUSION

We studied the perception of gradients in color-coded scalar fields. In a crowdsourced experiment, we sought to determine the just-noticeable differences when comparing 2D scalar data. We find that participants can discriminate smaller gradient variations when looking at a diverging or a spectral scheme, as compared with a monotonic-luminance colormap. The results support a hypothesis that discontinuities in color ramps may help in certain spatial tasks. Our findings suggest a new avenue for colormap evaluation, with a focus on assessing sensitivity to distributional characteristics.

ACKNOWLEDGEMENTS

This work was supported in part by the Argonne Leadership Computing Facility, which is a U.S. Department of Energy Office of Science User Facility operated under contract DE-AC02-06CH11357.

REFERENCES

- [1] D. Ariely. Seeing sets: Representation by statistical properties. *Psychological Science*, 12(2):157–162, 2001.
- [2] L. D. Bergman, B. E. Rogowitz, and L. A. Treinish. A rule-based tool for assisting colormap selection. In *Proceedings of the 6th conference on Visualization '95*, p. 118. IEEE Computer Society, 1995.
- [3] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, 113(4):700, 2006.
- [4] M. Borkin, K. Gajos, A. Peters, D. Mitsouras, S. Melchionna, F. Rybicki, C. Feldman, and H. Pfister. Evaluation of artery visualizations for heart disease diagnosis. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2479–2488, 2011.
- [5] D. Borland and R. M. T. Ii. Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications*, 27(2), 2007.
- [6] R. Bujack, T. L. Turton, F. Samsel, C. Ware, D. H. Rogers, and J. Ahrens. The good, the bad, and the ugly: A theoretical framework for the assessment of continuous colormaps. *IEEE Transactions on Visualization and Computer Graphics*, 2017.
- [7] J. Chen, G. Zhang, W. Chiou, D. H. Laidlaw, and A. P. Auchus. Measuring the effects of scalar and spherical colormaps on ensembles of dmri tubes. *IEEE Transactions on Visualization and Computer Graphics*, 2019.
- [8] A. Dasgupta, J. Poco, B. Rogowitz, K. Han, E. Bertini, and C. T. Silva. The effect of color scales on climate scientists' objective and subjective performance in spatial data analysis tasks. *IEEE Transactions on Visualization and Computer Graphics*, 2018.
- [9] A. Dasgupta, J. Poco, Y. Wei, R. Cook, E. Bertini, and C. T. Silva. Bridging theory with practice: An exploratory study of visualization use and design for climate model comparison. *IEEE Transactions on Visualization and Computer Graphics*, 21(9):996–1014, 2015.
- [10] S. L. Franconeri, D. K. Bemis, and G. A. Alvarez. Number estimation relies on a set of segmented objects. *Cognition*, 113(1):1–13, 2009.
- [11] M. Harrower and C. A. Brewer. Colorbrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.
- [12] A. Kale, F. Nguyen, M. Kay, and J. Hullman. Hypothetical outcome plots help untrained observers judge trends in ambiguous data. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):892–902, 2019.
- [13] A. D. Kalvin, B. E. Rogowitz, A. Pelah, and A. Cohen. Building perceptual color maps for visualizing interval data. In *Human Vision and Electronic Imaging V*, vol. 3959, pp. 323–336. International Society for Optics and Photonics, 2000.
- [14] J. Kehler and H. Hauser. Visualization and visual analysis of multi-faceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):495–513, 2012.
- [15] J. Kim, D. J. Bodony, and J. B. Freund. Adjoint-based control of loud events in a turbulent jet. *Journal of Fluid Mechanics*, 741:28–59, 2014.
- [16] K. Knoblauch and L. T. Maloney. *Modeling Psychophysical Data in R*, vol. 32. Springer Science & Business Media, 2012.
- [17] E. Komatsu, K. Smith, J. Dunkley, C. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Nolta, L. Page, et al. Seven-year wilkinson microwave anisotropy probe (wmap*) observations: cosmological interpretation. *The Astrophysical Journal Supplement Series*, 192(2):18, 2011.
- [18] A. Lagae, S. Lefebvre, R. L. Cook, T. DeRose, G. Drettakis, D. S. Ebert, J. P. Lewis, K. Perlin, and M. Zwicker. State of the art in procedural noise functions. In *Eurographics (STARs)*, pp. 1–19, 2010.
- [19] Y. Liu and J. Heer. Somewhere over the rainbow: An empirical assessment of quantitative colormaps. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 598. ACM, 2018.
- [20] M. R. Luo and C. Li. Ciecmap02 and its recent developments. In *Advanced Color Image Processing and Analysis*, pp. 19–58. Springer, 2013.
- [21] A. M. MacEachren. *How maps work: representation, visualization, and design*. Guilford Press, 2004.
- [22] K. Moreland. Diverging color maps for scientific visualization. In *International Symposium on Visual Computing*, pp. 92–103. Springer, 2009.
- [23] K. Moreland. Why we use bad color maps and what you can do about it. *Electronic Imaging*, 2016(16):1–6, 2016.
- [24] L. Padilla, P. S. Quinan, M. Meyer, and S. H. Creem-Regehr. Evaluating the impact of binning 2d scalar fields. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):431–440, 2017.
- [25] P. S. Quinan and M. Meyer. Visually comparing weather features in forecasts. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):389–398, 2016.
- [26] K. Reda, P. Nalawade, and K. Ansah-Koi. Graphical perception of continuous quantitative maps: the effects of spatial frequency and colormap design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 272. ACM, 2018.
- [27] R. A. Rensink. On the prospects for a science of visualization. In *Handbook of Human Centric Visualization*, pp. 147–175. Springer, 2014.
- [28] R. A. Rensink and G. Baldrige. The perception of correlation in scatterplots. In *Computer Graphics Forum*, vol. 29, pp. 1203–1210. Wiley Online Library, 2010.
- [29] B. E. Rogowitz and L. A. Treinish. Using perceptual rules in interactive visualization. In *IST/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*, pp. 287–295. International Society for Optics and Photonics, 1994.
- [30] B. E. Rogowitz and L. A. Treinish. Data visualization: the end of the rainbow. *IEEE spectrum*, 35(12):52–59, 1998.
- [31] B. E. Rogowitz, L. A. Treinish, S. Bryson, et al. How not to lie with visualization. *Computers in Physics*, 10(3):268–273, 1996.
- [32] J. Ross and D. C. Burr. Vision senses number directly. *Journal of vision*, 10(2):10–10, 2010.
- [33] S. Silva, B. S. Santos, and J. Madeira. Using color in visualization: A survey. *Computers Graphics*, 35(2):320–333, 2011.
- [34] D. A. Szafir, S. Haroz, M. Gleicher, and S. Franconeri. Four types of ensemble coding in data visualizations. *Journal of vision*, 16(5):11–11, 2016.
- [35] J. G. Trafton, S. S. Kirschenbaum, T. L. Tsui, R. T. Miyamoto, J. A. Ballas, and P. D. Raymond. Turning pictures into numbers: extracting and generating information from complex visualizations. *International Journal of Human-Computer Studies*, 53(5):827–850, 2000.
- [36] S. van der Walt and N. Smith. Matplotlib colormaps. <https://bids.github.io/colormap/>, 2015. [Online; accessed 7-June-2019].
- [37] C. Ware. Color sequences for univariate maps: Theory, experiments and principles. *IEEE Computer Graphics and Applications*, 8(5):41–49, 1988.
- [38] C. Ware, T. L. Turton, R. Bujack, F. Samsel, P. Shrivastava, and D. H. Rogers. Measuring and modeling the feature detection threshold functions of colormaps. *IEEE Transactions on Visualization and Computer Graphics*, 2018.
- [39] F. Yang, L. T. Harrison, R. A. Rensink, S. L. Franconeri, and R. Chang. Correlation judgment and visualization features: A comparative study. *IEEE Transactions on Visualization and Computer Graphics*, 25(3):1474–1488, 2018.
- [40] L. Zhou and C. D. Hansen. A survey of colormaps in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(8):2051–2069, 2016.