EuroRV3: EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (2017)
K. Lawonn, N. Smit, and D. Cunningham (Editors)

Evaluating the Perceptual Uniformity of Color Sequences for Feature Discrimination

Colin Ware[1], Terece L. Turton[2], Francesca Samsel[2], Roxana Bujack[3], David H. Rogers[3]

[1] Center for Coastal and Ocean Mapping, University of New Hampshire
[2] Center for Agile Technology, University of Texas at Austin, TX, USA
[3]Data Science at Scale Team, Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract

Probably the most common method for visualizing univariate data maps is through pseudocoloring and one of the most commonly cited requirements of a good colormap is that it be perceptually uniform. This means that differences between adjacent colors in the sequence be equally distinct. The practical value of uniformity is for features in the data to be equally distinctive no matter where they lie in the colormap, but there are reasons for thinking that uniformity in terms of feature detection may not be achieved by current methods which are based on the use of uniform color spaces. In this paper we provide a new method for directly evaluating colormaps in terms of their capacity for feature resolution. We apply the method in a study using Amazon Mechanical Turk to evaluate seven colormaps. Among other findings the results show that two new double ended sequences have the highest discriminative power and good uniformity. Ways in which the technique can be applied include the design of colormaps for uniformity, and a method for evaluating colormaps through feature discrimination curves for differently sized features.

Categories and Subject Descriptors (according to ACM CCS): H.1.2 [Models and Principles]: User/Machine Systems—Human Factors H.5.2 [Information Systems]: User Interfaces—Evaluation/methodology H.m [User/Machine Systems]: Miscellaneous—Colormapping

1. Introduction

Pseudocoloring (or color-mapping) is a technique widely used in the sciences to reveal patterns in scalar fields [SMS07, SSM11, ZH16]. Two common principles suggested for effective colormaps are that they should be perceptually uniform [Tru81, PZJ82a, Taj83, Lev96, Mor09, RO86, RKPC99, ZM06, Gre08] and that they should have good overall discriminative power [PZJ82a, Taj83, Mor09, Rhe99, Tru81]. Other principles are that they should be perceptually ordered [Bre04, SB79, PZJ82a, Rhe00, War88] and they should be have no sharp bends in a color space [BRT95, Mor09, SB79]. In the present work we are only concerned with uniformity of colormaps for feature discrimination and their overall discriminative power.

The method that is commonly employed to achieve uniformity is to use one of the many uniform colorspaces that have been developed over the years to support the color industry [RO86, Pha90, ZM06, WVVWVDL08, Che12, MBS*14]. Examples are CIElab and CIEluv. Another method is simply for the designer to adjust the colors so that the rate of change of colors appears perceptually uniform in the sense that the distance between adjacent pairs of colors in the sequence seems constant [PZJ82b, Gre08, RO86].

The uniformity of a colormap is defined based on the perceived differences between color pairs corresponding to the distance of the underlying scalar values they represent [LH92]. A uniform colormap defined in this way should have equally distinct differences through the entire sequence of colors. However, the purpose of colormaps is to resolve *features* in data and there are a number of reasons for thinking that uniformity as defined above may not equate to uniformity in the perception of data features. The uniform color spaces used to achieve uniform colormaps are based on measurements of the just noticeable differences between two adjacent quite large patches of color (1-3 deg fields) [WS82]. Uniform color spaces were created with the paint industry in mind and for relatively large patches of color and so they are unlikely to be effective for small features [ZW97]. It is known that we are relatively less sensitive to chromatic differences (as opposed to luminance differences) for smaller features, unlike the large features on which uniform color spaces are based [Mul85]. Consequently it has been suggested that sequences which embody larger chromatic differences are appropriate for low spatial frequency patterns [RKPC99].

The contribution of this paper is a methodology for evaluating colormaps in terms of their ability specifically to support *feature*

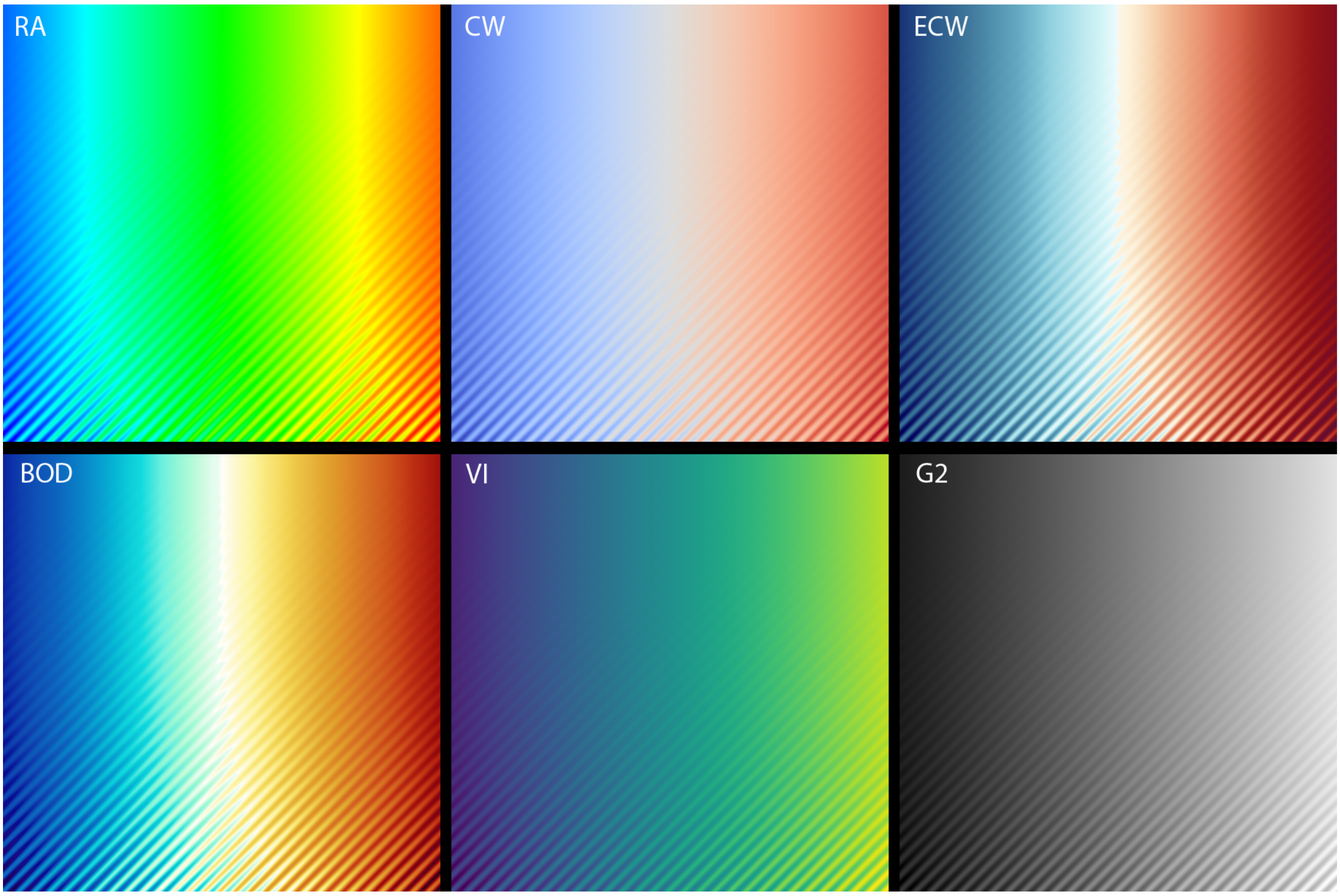

Figure 1: *The same test pattern is shown using six different color sequences. By visually tracing the top boundary of the diagonally striped pattern, it is possible to judge the perceptual uniformity of the sequence for feature discrimination.The seventh pattern, GL, is not shown here to save space. Both colormaps have the same local but different global discriminative powers.*

resolution. It incorporates a standard way of measuring the sensitivity of the human visual system to features, namely sinusoidal grating patterns. Here we provide results from seven sequences evaluated using an Amazon Mechanical Turk (AMT) study.

1.1. Test Pattern

Our test pattern is based on a simple linear ramp from left to right, starting at 0.1 and ending at 0.9. Superimposed on this ramp is a diagonal sinusoidal pattern which has a decreasing amplitude from bottom to top. The amplitude of the sinusoidal signal increases according to a power law from top to bottom:

$$a = c/2 * 2^{(1+(d-y)/p)} \quad (1)$$

where c is the starting amplitude, d is the distance from the top, y is the position at which the pattern starts (from the top), and p is the amplitude doubling interval. The combined patterns have a 0-1 range at the lower edge.

In all of the examples shown here, $c = 0.001$, $s = 40$ pixels, $p = 80$ pixels. This results in feature amplitude that is 128 times larger at the bottom edge. Since the sine pattern varies between ± 1, this results in a maximum peak-to-peak difference of 0.256. The pattern wavelength in the examples is 15 pixels. Note that the feature amplitude here is defined as a proportion of the data range displayed using a colormap.

Figure 1 shows the test pattern pseudocolored by six of the following seven sequences. Colormaps were chosen because of common use, theoretical interest or performance.

RA Rainbow. This particular rainbow is the one found in ParaView [AGL05]. Rainbow sequences are frequently criticized because they are not perceptually ordered [BI07].
CW Cool-warm. A widely used divergent sequence [Mor09].
ECW An extended cool-warm developed by artist F. Samsel and the Data Science at Scale team at Los Alamos National Laboratory [STB*].
BOD Blue-Orange. Another extended cool-warm that goes into the oranges, also developed by F. Samsel [STB*].
VI Viridis. A color sequence thought to have good uniformity and designed to be more colorblind-safe [vS15].
G2 A perceptually uniform grey scale using CIElab $L*$.

GL A linear (gamma corrected) grey.

When the patterns are viewed on a screen they should each subtend 600 pixels. The spatial frequency of the sinusoid will depend on the pixel size and the viewing distance. For example, if these patterns are viewed on a 45 pixel/cm screen from a distance of 56 cm, the spatial frequency will be 3 cycles per degree of visual angle. Because the printing process inevitably distorts the colors that are seen, images should be viewed on a monitor. Ideally, to guarantee correct viewing, the versions provided in the supplementary material should be viewed on a PC or other display with a gamma of 2.2.

By simple inspection of the upper boundary where the pattern becomes invisible it is possible to immediately see that these patterns differ greatly in terms of resolving power. In particular, the ParaView rainbow is extremely non-uniform, with a very low resolving power in the central green region and a high resolving power at the ends.

2. Experiment to Compare Color Sequences

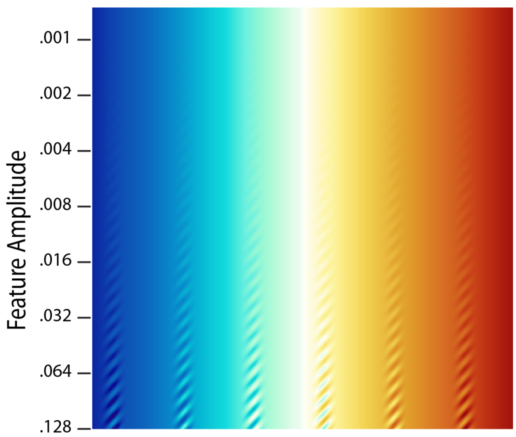

Figure 2: *One of a set of five test patterns rendered with the BOD colormap. The participant task was to click on the highest feature visible in each of the six patterned columns. The scale is provided here to show how feature amplitude doubles every 80 pixels. It was not shown to the participants.*

2.1. Stimulus Patterns

In order to use the pattern in an empirical evaluation, we created a modified version, shown in Figure 2 for the Blue-Orange sequence. Sets of patterns such as this were used to compare seven color sequences in an experiment using Amazon Mechanical Turk. In each test pattern six discrete vertical stripes of the sine pattern were generated by modulating the sine pattern with a set of horizontally oriented Gaussian windows. Each of the stripes is horizontally separated by 100 pixels. A set of five such patterns was generated with starting offsets of 10, 30, 50, 70, and 90 pixels. This resulted in a set providing 30 equally spaced sampling columns across the 600 pixel pattern supporting discrete selections which are easy to interpret.

2.2. Task

On each trial the participants' task was to click on each of the six points where the vertical pattern became invisible. Each participant saw the entire set of five patterns or 30 columns in this way.

2.3. Participants

The experiment gathered 147 Mturk participants. Doing studies online carries the risk of possible contamination by individuals who have some type of color vision deficiency (CVD). In order to minimize these effects, a version of the Farnsworth D-15 color cap arrangement test [CJJ93] was coded in a Qualtrics study. This study is launched at intervals to develop an ongoing list of colorblind Mturkers. Any Mturker self-identifying with any type of CVD from the Farnsworth D-15 studies is excluded from color-based studies, including this one. As of this writing, this list has 343 Mturkers. Additionally, as part of the study, people were asked to state their CVD status and any potential participant who self-identified as CVD was prevented from continuing with this study. An analysis of gender distribution shows 58% of our participants are male. CVD rates are approximately 9% and 0.5% for males and females, respectively. Given the gender distribution and CVD rates, we estimate potentially eight CVD individuals in a sample size of 147 people. However, this should be considered an upper limit given the use of the exclusion list of CVD Mturkers, the self-exclusion request within the study and the data-scrubbing which may have caught people unaware of their CVD status who were unable to properly see the columns of features. We therefore do not consider CVD contamination to impact the results of this study.

2.4. Procedure

The experiment was coded using the *Heat Map* question type within the Qualtrics survey software. A Qualtrics device check was used to prohibit the use of mobile devices.

Participants were given an explanation of the task and shown a before/after example demonstrating how they needed to click at the top on each of the six vertical columns of features. They were then given a training question that was used to validate their understanding of the task. The example/training image used a desaturated rainbow not seen within the actual study. Any participant whose click pattern indicated that they did not understand the task was rejected. Additionally, during data analysis, any participant who did not click once along each of the six columns or who consistently clicked only along the top or along the bottom was rejected. These requirements resulted in 121 valid participants with between 27 and 32 participants for each colormap. Each participant saw two of the possible seven sequences but a participant was prevented from seeing two very similar sequences. For example, a participant who saw the standard cool/warm would not also see the extended cool/warm. The five stimuli images for each sequence were shown in a random order and the two colorscales were seen sequentially in a randomized order.

2.5. Results

The feature discrimination curves for the seven test colormaps are shown in Figure 3. The upper part of Figure 3 shows results for

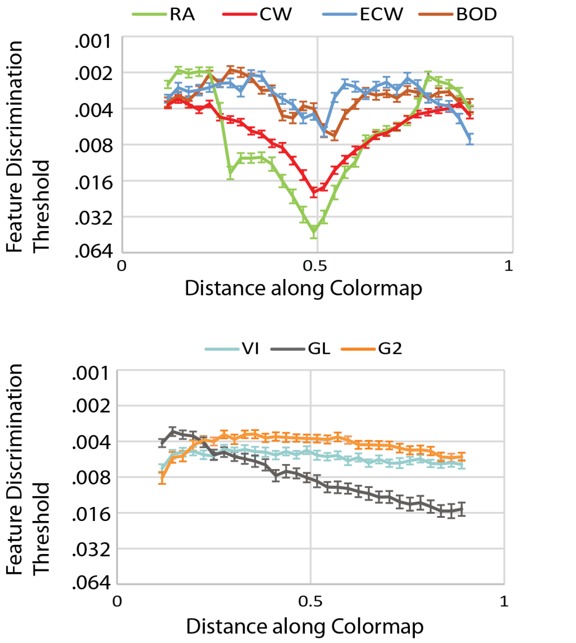

Figure 3: *Results Summary. Above: feature discrimination curves for the double ended and rainbow sequences. Below: feature discrimination curves for the Luminance increasing sequences. Note that the y axis is logarithmically spaced and inverted; a low threshold indicates greater feature resolving power. Vertical bars represent one standard error.*

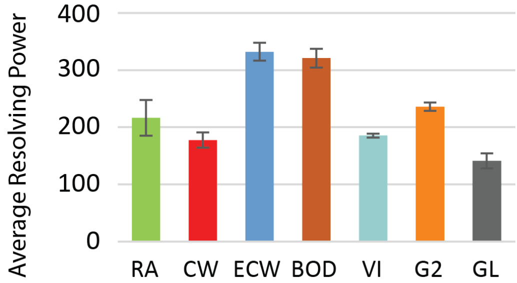

Figure 4: *Average feature resolving power for each of the seven tested color sequences.*

the Rainbow sequence as well as the three divergent colormaps. The lower graph shows results for the three luminance increasing colormaps. These plots use an inverted logarithmically based scale since a smaller threshold indicates greater resolving power. Assuming that participants used typical PCs or laptops, a spatial frequency of 3 cycles per degree is a reasonable estimate for the spatial frequency of the stimulus patterns.

To provide a metric for overall resolving power R_λ for a particular spatial frequency λ, we propose a weighted average of the inverse of the feature discrimination threshold:

$$R_\lambda = \sum_{i=1}^{n} p_i \Big/ \sum_{i=1}^{n} p_i t_{i\lambda} \qquad (2)$$

where n is the number of threshold sample points, p_i is the width of sample i of the sequence, and $t_{i\lambda}$ is the threshold amplitude for that interval. This simplifies to a simple average if the samples are equal as is the case here. The results are summarized in Figure 4.

3. Discussion

Overall the results reveal clear differences between different colormaps with quite small inter-subject variability, suggesting that this method is indeed an effective and easily applied test of the feature resolving power of colormaps.

The divergent BOD and ECW colormaps provide the best resolving power over most of their extents. These curves also provide good uniformity. In contrast the Rainbow colormap is very weak in the central portion and highly non-uniform. The Cool Warm (CW) colormap is also non-uniform, but it varies more smoothly. Comparing the three curves that increase monotonically in Luminance, Viridis is the most perceptually uniform of all of the colormaps, showing only a slight fall off at the dark end. However, the perceptual grey colormap has better resolving power overall, probably because it has maximal luminance variation. As expected, the physically linear sequence has greater resolving power at the dark end but increasingly less as it approaches white.

There are a number of ways in which this new method can be applied. By changing the spatial frequency of the sinusoidal pattern we can investigate the ability of different colormaps to convey information about data features of different sizes. But in order to adequately test lower spatial frequencies, larger test patterns will be required. Additionally, the basic data pattern (underlying all the examples in Figure 1) could be incorporated into an interactive design tool such as the one described in Guitard and Ware [GW90]. This would make it easier to design sequences with uniform resolving power. For this purpose, a set of data patterns with three different spatial frequencies is provided as supplementary material.

There are obvious tradeoffs in using AMT for a study such as this. Disadvantages are uncertainties regarding the exact viewing parameters. Advantages are a large and diverse subject population and ecological validity arising from the variation of viewing monitors and other viewing parameters. This is valuable because it mimics the diversity found in the world of scientific visualization.

In future work we plan to compare the results from this test with predictions based on color theory. In particular, we predict that chromatic variation in a colormap will be relatively less important for feature resolution with high spatial frequency patterns and more important with low spatial frequency patterns. We also plan to use the method to evaluate many of the other interesting colormaps in common use.

Acknowledgments

This material is based upon work supported by Dr. Lucy Nowell of the U.S. Department of Energy Office of Science, Advanced Scientific Computing Research under Award Numbers DE-AS52-06NA25396, DE-SC-0012438, and DE-SC-0012516. The authors would like to thank Dr. James Ahrens.

References

[AGL05] AHRENS J., GEVECI B., LAW C.: ParaView: An end-user tool for large-data visualization. In *Visualization Handbook*, Hansen C. D., Johnson C. R., (Eds.). Butterworth-Heinemann, Burlington, 2005, pp. 717–731. URL: http://www.sciencedirect.com/science/article/pii/B9780123875822500381, doi:http://dx.doi.org/10.1016/B978-012387582-2/50038-1. 2

[BI07] BORLAND D., II R. M. T.: Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications 27*, 2 (2007), 14–17. doi:http://doi.ieeecomputersociety.org/10.1109/MCG.2007.46. 2

[Bre04] BREWER C.: *Designing better Maps: A Guide for GIS users*. Environmental Systems Research, 2004. 1

[BRT95] BERGMAN L. D., ROGOWITZ B. E., TREINISH L. A.: A rule-based tool for assisting colormap selection. In *Proceedings of the 6th conference on Visualization'95* (1995), IEEE Computer Society, p. 118. 1

[Che12] CHEUNG V.: *Uniform Color Spaces*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 161–169. URL: http://dx.doi.org/10.1007/978-3-540-79567-4_14, doi:10.1007/978-3-540-79567-4_14. 1

[CJJ93] CJ B., JC G., J H.: Comparison of the Farnsworth-Munsell 100-hue, the Farnsworth D-15, and the Anthony D-15 desaturated color tests. *Archives of Ophthalmology 111*, 5 (1993), 639–641. doi:10.1001/archopht.1993.01090050073032. 3

[Gre08] GRESH D. L.: *Self-Corrected Perceptual Colormaps*. Tech. rep., IBM, 2008. RC24542 (W0804-104). 1

[GW90] GUITARD R., WARE C.: A color sequence editor. *ACM Transactions on Graphics 9*, 3 (1990), 338–341. URL: http://scholars.unh.edu/ccom/913. 4

[Lev96] LEVKOWITZ H.: Perceptual steps along color scales. *International Journal of Imaging Systems and Technology 7*, 2 (1996), 97–101. 1

[LH92] LEVKOWITZ H., HERMAN G. T.: The design and evaluation of color scales for image data. *IEEE Computer Graphics and Applications 12*, 1 (1992), 72–80. 1

[MBS*14] MITTELSTÄDT S., BERNARD J., SCHRECK T., STEIGER M., KOHLHAMMER J., KEIM D. A.: Revisiting Perceptually Optimized Color Mapping for High-Dimensional Data Analysis. In *EuroVis - Short Papers* (2014), Elmqvist N., Hlawitschka M., Kennedy J., (Eds.), The Eurographics Association. doi:10.2312/eurovisshort.20141163. 1

[Mor09] MORELAND K.: Diverging color maps for scientific visualization. In *International Symposium on Visual Computing* (2009), Springer, pp. 92–103. 1, 2

[Mul85] MULLEN K. T.: The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology 359* (1985), 381–400. 1

[Pha90] PHAM B.: Spline-based color sequences for univariate, bivariate and trivariate mapping. In *Proceedings of the 1st conference on Visualization'90* (1990), IEEE Computer Society Press, pp. 202–208. 1

[PZJ82a] PIZER S. H., ZIMMERMAN J. B., JOHNSTON R. E.: Concepts of the display of medical images. *IEEE Transactions on Nuclear Science 29*, 4 (1982), 1322–1330. 1

[PZJ82b] PIZER S. M., ZIMMERMAN J. B., JOHNSTON R. E.: Contrast transmission in medical image display. In *1st International Symposium on Medical Imaging and Image Interpretation* (1982), International Society for Optics and Photonics, pp. 2–9. 1

[Rhe99] RHEINGANS P.: Task-based color scale design. In *In Proceedings Applied Image and Pattern Recognition. SPIE* (1999), International Society for Optics and Photonics, pp. 35–43. 1

[Rhe00] RHEINGANS P. L.: Task-based color scale design. In *28th AIPR Workshop: 3D Visualization for Data Exploration and Decision Making* (2000), International Society for Optics and Photonics, pp. 35–43. 1

[RKPC99] ROGOWITZ B., KALVIN A., PELAH A., COHEN A.: Which trajectories through which perceptually uniform color spaces produce appropriate color scales for interval data? In *7th Color and Imaging Conference* (1999), Society for Imaging Science and Technology, pp. 321–326. 1

[RO86] ROBERTSON P. K., O'CALLAGHAN J. F.: The generation of color sequences for univariate and bivariate mapping. *IEEE Computer Graphics and Applications 6*, 2 (1986), 24–32. 1

[SB79] SLOAN K. R., BROWN C. M.: Color map techniques. *Computer Graphics and Image Processing 10*, 4 (1979), 297–317. 1

[SMS07] SILVA S., MADEIRA J., SANTOS B. S.: There is more to color scales than meets the eye: a review on the use of color in visualization. In *Information Visualization, 2007. IV'07. 11th International Conference* (2007), IEEE, pp. 943–950. 1

[SSM11] SILVA S., SANTOS B. S., MADEIRA J.: Using color in visualization: A survey. *Computers & Graphics 35*, 2 (2011), 320–333. 1

[STB*] SAMSEL F., TURTON T., BUJACK R., ROGERS D., AHRENS J., ABRAM G., WARE C.: Data science at scale sci-vis color website. URL: sciviscolor.org. 2

[Taj83] TAJIMA J.: Uniform color scale applications to computer graphics. *Computer Vision, Graphics, and Image Processing 21*, 3 (1983), 305–325. 1

[Tru81] TRUMBO B. E.: A theory for coloring bivariate statistical maps. *The American Statistician 35*, 4 (1981), 220–226. 1

[vS15] VAN DER WALT S., SMITH N.: Matplotlib documentation update. bids.github.io/colormap/, 2015. 2

[War88] WARE C.: Color sequences for univariate maps: Theory, experiments and principles. *IEEE Computer Graphics and Applications 8*, 5 (1988), 41–49. 1

[WS82] WYSZECKI G., STILES W. S.: *Color science*, vol. 8. Wiley New York, 1982. 1

[WVVWVDL08] WIJFFELAARS M., VLIEGEN R., VAN WIJK J. J., VAN DER LINDEN E.-J.: Generating color palettes using intuitive parameters. *Computer Graphics Forum 27*, 3 (2008), 743–750. URL: http://dx.doi.org/10.1111/j.1467-8659.2008.01203.x, doi:10.1111/j.1467-8659.2008.01203.x. 1

[ZH16] ZHOU L., HANSEN C.: A survey of colormaps in visualization. *IEEE Transactions on Visualization and Computer Graphics 22*, 8 (2016), 2051–2069. 1

[ZM06] ZHANG H., MONTAG E.: Perceptual color scales for univariate and bivariate data display. *The Society for Imaging Science and Technology (IS&T)* (2006). 1

[ZW97] ZHANG X., WANDELL B. A.: A spatial extension of CIELAB for digital color-image reproduction. *Journal of the Society for Information Display 5*, 1 (1997), 61–63. URL: http://dx.doi.org/10.1889/1.1985127, doi:10.1889/1.1985127. 1