

Color Nameability Predicts Inference Accuracy in Spatial Visualizations

Khairi Reda*, Amey A. Salvi*, Jack Gray*, Michael E. Papka†

* Indiana University–Purdue University Indianapolis

† Argonne National Laboratory and Northern Illinois University

Abstract

Color encoding is foundational to visualizing quantitative data. Guidelines for colormap design have traditionally emphasized perceptual principles, such as order and uniformity. However, colors also evoke cognitive and linguistic associations whose role in data interpretation remains underexplored. We study how two linguistic factors, name salience and name variation, affect people's ability to draw inferences from spatial visualizations. In two experiments, we found that participants are better at interpreting visualizations when viewing colors with more salient names (e.g., prototypical 'blue', 'yellow', and 'red' over 'teal', 'beige', and 'maroon'). The effect was robust across four visualization types, but was more pronounced in continuous (e.g., smooth geographical maps) than in similar discrete representations (e.g., choropleths). Participants' accuracy also improved as the number of nameable colors increased, although the latter had a less robust effect. Our findings suggest that color nameability is an important design consideration for quantitative colormaps, and may even outweigh traditional perceptual metrics. In particular, we found that the linguistic associations of color are a better predictor of performance than the perceptual properties of those colors. We discuss the implications and outline research opportunities. The data and materials for this study are available at <https://osf.io/asb7n>

CCS Concepts

• **Human-centered computing** → *Empirical studies in visualization*;

1. Introduction

Color plays a central role in the visual analysis and communication of data. It is rare to find a visualization that does not make use of color in one form or the other. In some representations, such as choropleths and scalar fields, color serves as the principal channel for communicating quantities and spatial patterns.

Designing effective color encodings has been a recurring theme in visualization research. Over the years, the community has developed a repertoire of guidelines for color mapping [ZH16; BTS*17]. For example, researchers recommend colors with distinct hues for categorical data [Mun14]. By contrast, for quantitative data, guidelines suggest perceptually uniform scales that vary predominantly in luminance [RT94]. The assumption here is that gradations of luminance help to convey a sense of order [Mor09]. Moreover, by limiting changes in hue, an encoding emphasizes differences in magnitude, while avoiding the appearance of false categories.

The guidelines above are primarily based on how the visual system perceives colors. However, colors also evoke rich associations, including conceptual [TSG*19], linguistic [BK91], and affective responses [BPS17]. The processing of color stimuli, including visualizations, is thus likely to involve a number of cognitive subsystems that operate on top of the perceptual system. Visualization in-

terpretation may depend on these higher-level cognitive processes more than it depends on the perception of color appearance. Yet, it is still unclear how these multifaceted associations impact one's understanding of visualized data.

Emerging research has challenged the purely perceptual approach to color design. For example, recent work shows that 'colorful' visualizations containing a variety of nameable colors are easier to interpret [RS21]. In particular, the study demonstrates an unusual advantage for rainbow color scales in a visual inference task involving scalar fields. Yet, comparing color nameability with perceptual properties created a confound in that study: as the perceptual difference between colors increased, their names became more distinct. Therefore, it was inconclusive which of the two factors was responsible for the improved performance. Given competing constraints, should a designer optimize for linguistic properties (e.g., by choosing colors with varied and unique names) or is it enough to consider perceptual factors alone (e.g., the perceptual distance and uniformity between colors)? Understanding how the perceptual and linguistic associations of color affect performance is vital for effective visualization design. Furthermore, given the widespread use of color in data displays, it is important to evaluate those effects in a variety of representations.

In this paper, we address two research goals. First, we aim to untangle the linguistic effects of colors from their perceptual appearance. Specifically, we study two nameability metrics for colormaps: *name variation*, a measure of the number of nameable colors, and *name salience*, the degree to which these colors have unique and reliable names. Our central hypothesis is that nameability significantly improves the interpretability of visualizations—even those containing purely quantitative data. Our second goal is to test if the linguistic associations of color are similarly beneficial (or meritless) in a variety of visualization types. We specifically evaluate a mix of continuous and discrete representations: 2D scalar field, heatmaps, choropleth maps, and continuous geographical maps.

We test our hypotheses in two crowdsourced experiments. Participants observed color-coded visualizations and made inferences about the models underlying these visualizations. We find that, across four visualization types, participants were more accurate when viewing encodings with uniquely nameable colors. In particular, incorporating colors that can be described using salient terms, such as ‘red’, ‘yellow’, or ‘blue’ led to more accurate inferences, as compared with similar colors of lesser name salience (e.g., ‘maroon’, ‘teal’, or ‘beige’). These performance variations could not be explained by an alternative perceptual model, suggesting that the benefits of color nameability are cognitive in nature. Our findings provide a deeper understanding of the role of color names in data interpretation. The results also suggest that linguistic properties of color can be a better predictor of usefulness than traditional perceptual measures (e.g., perceptual discriminability). To that end, we contribute concrete metrics that help in gauging colormap utility. Our findings call for new guidelines and color design tools that prioritize nameability as a central metric, even, counterintuitively, for data that lacks inherent linguistic associations.

2. Related Work

Color encoding guidelines for quantitative data have historically emphasized perceptual uniformity [Mor09; vdWS15] and/or luminance monotonicity [RT94; War12]. Diverging color scales have also been recommended, particularly for choropleths [Bre96]. The latter includes carefully crafted ‘spectral’ schemes [Bre97]. Other researchers, however, strongly advise against ‘rainbow’ colormaps for quantitative data [BI07; Mor16; RT98], given their lack of perceptual order and uniformity. On the other hand, hue variation is recommended for categorical data [Mun14]. Selecting ‘semantically resonant’ colors (e.g., yellow for banana and blue for water) is thought to further aid chart interpretation [LFK*13; SLL20].

Empirical studies of color have covered a variety of perceptual tasks, including detection of localized features [KRPC00; WTB*18], quantity estimation [War88; RNA18], and comparison of color values [LH18]. For instance, Ware tested people’s accuracy in estimating quantities in color-coded scalar fields, and found spectral (i.e., rainbow) schemes to be more accurate [War88]. These results were attributed to a reduction in simultaneous contrast effects. For tasks requiring form perception, however, a monotonic luminance ramp is thought to be more effective [RK01; War88].

Studies with domain experts have appeared in the literature [DPR*18; BGP*11]. However, the tasks reported by those

users (e.g., diagnosing heart disease) were often reducible to elementary perceptual operations, such as finding low-lying regions in an arterial scan [BGP*11] or estimating average values [DPR*18]. Few studies have investigated whether color choice impacts performance in inferential tasks, wherein an analyst must infer something about the model or phenomenon that underlies a visualization. Such tasks are likely to be interpretive in nature, and may necessitate color designs that are optimized for cognitive processing as opposed to perceptual precision. Recent work has shown that name distance can be a good metric of color utility in graphical inference [RS21]. However, that work could not uniquely attribute the results to linguistic aspects. We adopt a similar task (i.e., graphical inference), but employ more nuanced nameability metrics. We also study the effects of nameability in a variety of visualizations. Our experiments are specifically designed to separate linguistic and perceptual effects, thus allowing for more specific guidance.

2.1. Color Names

Color names refer to the language terms we use to describe different colors [BK91]. But more than simply serving as linguistic labels, color names are thought to categorize our perception of color [Har87]. Evidence for categorical color perception comes from the fact that people are better at discriminating between colors that have distinctive names. For instance, Winawer et al. found Russian speakers to be faster at discriminating between shades of blue that are otherwise perceptually similar [WWF*07]. The effect was attributed to two words in Russian: *goluboy* for light blue and *siniy* for dark blue. To Russian speakers, these colors are as distinct as brown and orange [Par05]. Similarly, Healey found that people require a minimum name difference to discriminate colors effectively in visualizations [Hea96]. Liu and Heer also suggest that colormap interpretation involves naming and perceptual factors [LH18].



Not all colors are equal. Berlin and Kay argued that certain colors are psychologically ‘salient’ given their association with basic names in many languages [BK91]. Black, white, red, green, yellow, and blue are examples of monolexic terms thought to represent an especially salient part of the color space [HMNP72]. Such colors tend to exhibit ‘stability of reference’ [Bol78], with most people agreeing on their name associations. Still, the role of color names in visualization interpretation remains underexplored, except perhaps in categorical assignments [SS15; SLWF18; GLS17; LFC*20]. We investigate how color nameability impacts quantitative data interpretation, a domain with no obvious linguistic associations.



The computer vision and visualization communities have taken interest in building color name models, given their potential applications [Moj05; BTBV02; CSH08]. One model by Heer and Stone [HS12] builds on a very large survey [Mun10]. The model is a discretization of the CIELAB color space, accompanied by metrics for color name distance and salience. We employ this model for our nameability measures, given its empirical foundation and the useful operations it provides.



3. Research Questions & Methods

We conducted two crowdsourced experiments to understand how the linguistic associations of color affect people’s ability to make

inferences from quantitative, spatial data. The experiments were designed to answer three research questions:

RQ1 — How does color nameability affect the interpretation of visualizations? We conceptualize nameability as the ability to categorize and refer to colors by their names. We consider two metrics to quantify this construct: **color name variation**, a measure of the number of nameable colors in a given color scale. For example, the sequence  incorporates approximately three distinct color names: purple, pink, and white. By contrast, a rainbow sequence of  incorporates a wider variety of names: blue, cyan, yellow, orange, and red. Rainbow thus has higher name variation. We expect this metric to correlate with interpretability.

The second nameability factor we consider is **color name salience**, the degree to which a color is associated with a unique term. To illustrate, consider , which most people associate with the prototypical term ‘red’. When a large majority of observers refer to a color uniquely using the same term, we can say that the color is endowed with a highly salient name. On the other hand, take , which is likely to evoke more varied responses, such as ‘maroon’, ‘dark red’, ‘crimson’, and possibly even plain ‘red’. Because of its multiple name associations, this color exhibits relatively lower name saliency. We expect people to perform better with colors of more salient names.

RQ2 — To what extent does color nameability affect performance, beyond that which is sufficient for perceptual discriminability? In other words, is it enough to select colors that are perceived as different, or is it important that those colors be also associated with distinct and unique names? For example, the earlier sequence of  incorporates purple, pink, and white, which are empirically salient colors. By comparison  incorporates teal and beige—names that are less salient on average. This is despite the fact that these two sequences have virtually identical perceptual properties. We hypothesize that it is easier to interpret visualizations encoded with the first color scale than with the second.

RQ3 — Assuming there are advantages to encodings that optimize color nameability, do these benefits hold true across different visualizations? We are specifically interested in the effects of nameability on discrete visualizations, such as choropleths and heatmaps, versus similar continuous representations, like smooth geographical maps and scalar fields.

We provide a deeper discussion of our hypotheses in §4, and formally define our two nameability metrics in §5. But first, we introduce the graphical inference task employed in our experiments.

3.1. Modeling Interpretive and Inferential Tasks

Empirical research on colormap characteristics have emphasized elementary perceptual tasks, such as value estimation [War88; RNA18] and feature detection [WTB*18]. Though useful, these studies do not adequately reflect the more interpretive analyses people undertake with visualizations. Consider, as an example, an epidemiologist who is modeling the spread of an infectious disease by analyzing simulation results from a handful of competing models. These simulations may be visualized as choropleths that depict infection rates at the county level (e.g., similar to Figure 1-left).

Here, the researcher may be interested in comparing the simulations on whole to understand where the models converge or differ. The researcher could conclude that the models are in agreement, even though the simulations might exhibit slight variations. Alternatively, the researcher may declare one of the models to be unique in that it exhibits a markedly distinct outcome. In conducting this kind of visual analysis, the researcher may perform elementary perceptual queries, such as estimating the color value at landmark locations in the map. However, it is unlikely that the analysis can be completed alone by a sequence of value retrievals. Instead, the researcher will likely resort to assessing the broader spatial patterns; patterns of infection from several maps need to be compared, differences attended to, and observations integrated [RTB08], before one can begin to characterize the relationships between the models.

Inferential tasks, where the analyst must infer model characteristics from an ensemble of visualizations, abound in both the computational and natural sciences. The focus here is not on the recovery of quantities but rather on patterns and distributions, which enable one to infer *qualitative* differences between visualizations. These operations may necessitate color encodings that facilitate *attention* to patterns, perhaps at the expense of quantitative precision.

We operationalize notions of ‘interpretation’ and ‘inference’ in our experiments by appropriating a graphical inference task [BCH*09; WCHB10]. This task was conceived as a visual equivalent to statistical hypothesis testing. It is meant to test if an observer can tell apart the ‘real’ data from other distributions that could have arisen by chance. The setup involves concealing a plot of the real data in a lineup that also contains several other ‘decoy’ plots sampled from a ‘null’ distribution. An observer is then asked to identify the plot that “doesn’t belong”, with the correct answer being the real plot. Successful discrimination amounts to making a (correct) inference that the two models (i.e., the real and the null) represent distinct phenomena. The analogy to statistical inference makes the test useful for controlling the rate of false discovery. However, it can also be used to compare the statistical power of different visualization techniques [HFMC12] or, for that matter, different colormap designs. If a particular colormap makes it easier to discriminate between models, then that colormap can be said to have higher power. We adopt this task, and use it to test if colormaps with better nameability can improve graphical inference.

3.2. Synthesizing Lineup Visualizations

We start by synthesizing a *target* model. We use a mixture of 2D Gaussian kernels that placed randomly within a 2D domain. We use anywhere between 3 and 6 kernels, randomly varying the parameters of those kernels (centroid, standard deviation, X-Y correlation) within a fixed range (determined through piloting). The individual kernels are integrated to form a joint probability density function. We then obtain a corresponding *decoy* model by perturbing the parameters of the kernels in the target model. The degree of perturbation dictates the ensuing target-decoy dissimilarity; larger perturbation amounts to higher target-decoy divergence, which would be easier to discriminate visually. We quantify the latter using KL divergence, the entropy-distance between the two models.

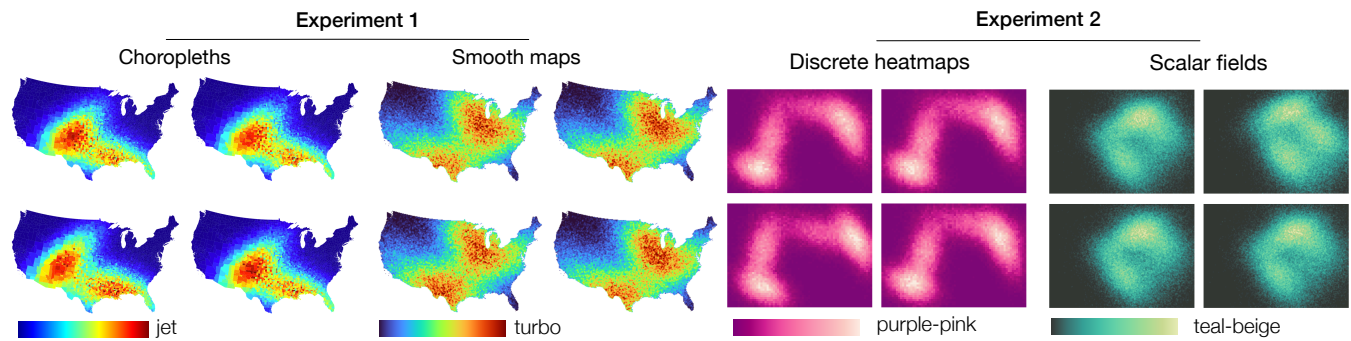


Figure 1: Example lineup stimuli from the two experiments. The observer is tasked with identifying the plot that doesn't belong. The answers can be found on the last page of the paper.

The last step in generating a lineup is to sample both models. Three plots are sampled from the decoy, and one is sampled from the target. The plots are then displayed to participants in a 2×2 layout (see Figure 1). The participant is tasked with identifying the plot that does not belong. Note that in addition to the principal target-decoy variation, the decoy plots will also vary among themselves due to the random sampling. The observer thus needs to consider which plot features reflect true differences between the underlying models, as opposed to fluctuations arising from the sampling process. These added complexities make for a more interpretive task, as compared with tasks that focus on value estimation [RNA18; RP19] or image similarity [DPR*18]. Furthermore, the ontological parallel to statistical inference provides a useful approach to evaluating color utility for data interpretation purposes. Note also that the synthetic procedure allows for lineups with a variety of characteristics and patterns to emerge, due to interactions between multiple Gaussians. Model-differentiating features also occur at different color levels (e.g., as hotspots, middle, or lower values).

4. Hypothesis Development

Visualization guidelines suggest that colors should “faithfully represent the magnitude of continuous data” [RK01]. However, it is unclear if quantitative precision is always important to the interpretation of visualizations [BCF20]. Many analyses can be completed by operating on visual proxies [JOEF19]—task-relevant visual features that allow the viewer to entirely skip quantitative interpretation. Consider the visual inference task introduced in §3.1 (illustrated in Figure 1). Here, an observer may resort to comparing the color layout in different visualizations in order to identify the plot with a seemingly ‘odd’ distribution. This comparison on whole can be fairly complex, so the observer could attempt to break up the task into several sub-comparisons [Gle17]. One potential strategy is to selectively attend to a narrow range of quantities (i.e., colors) at a time. For example, the observer may start by comparing the hotspots, and then switch to comparing valleys. At a more granular level, the observer could selectively attend to individual quartiles in the color scale (e.g., high, mid-high, mid-low, or low values), comparing the shapes occupied by these quartiles in different plots.

This kind of shape- or pattern-based matching, we conjecture, is easier when colors appear categorized. Evidence from experimental

psychology suggests that people can easily attend to a shape or pattern defined by color, but only if that pattern spans a single color category [HP07]. Color encodings that create a categorized look may present an advantage here by making it easier to selectively attend to specific distributional features (e.g., hotspots, valleys, or middle values). Prototypical color categories (e.g., red, yellow, and blue) tend to come more readily to attention than others [Ros73]. While visual salience is modulated by low-level vision, task-driven attention is likely controlled by cognitive and linguistic features, including color names.

In short, we conjecture that categorizable colors aid selective attention in visualizations, making it easier to access task-relevant features for detailed inspection. “Linguistically codable” colors [Hei72], we speculate, serve as salient anchors, allowing viewers to compare and reason over complex patterns. These attentional benefits cannot be explained alone by perceptual discriminability, but are rather (at least partially) attributable to color nameability. The more linguistically salient the colors, we argue, the more useful they are as attentional anchors. Accordingly, we posit the following hypotheses:

H1 — Color categorization is a useful property when making graphical inferences about visualizations. Categorization tendency can be increased by incorporating colors with a diverse set of names. We quantify the latter with a metric referred to as *color name variation* (see §5 for definition). We expect people to be more accurate when viewing color encodings of higher name variation.

H2 — Colors with salient linguistic terms (e.g., ‘red’ over ‘crimson’) will further improve accuracy in graphical inference. An example of a highly name-salient colormap is *jet* (see Figure 2). It consists of saturated colors associated with basic terms. Compare to *turbo*, which has a similar rainbow appearance but blends colors that are less prototypical, giving it lower *name salience*. We expect people to be more accurate when viewing colors associated with more salient names.

H3 — The linguistic associations of colors are a better predictor of performance than their perceptual properties. Specifically, we expect name salience and name variation to provide a better correlation to accuracy than perceptual distance in the LAB space.

H4 — We hypothesize that name salience and name variation will have similar positive effects across different visualizations. More

specifically, we expect the effects to hold in smooth (e.g., scalar fields) and discrete representations (e.g., choropleths).

H1 was proposed by Reda and Szafir [RS21], who found support for it. However, they were unable to uniquely attribute the observed benefits to linguistic associations or to perceptual factors (color name variation was highly correlated with perceptual discriminability in their setup). Our second hypothesis (H2) provides further nuance by considering not only the distinctiveness of color names but also their linguistic prominence. Name saliency, in particular, is not tightly correlated with perceptual discriminability (see Figure 3). We take advantage of this fact by testing colormaps with similar perceptual discriminability, but that which exhibit different name-saliency levels. Thus, in addition to independently testing for effects of name saliency and name variation, our study allows us to test an overarching hypothesis: the linguistic properties of colors are a better predictor of performance than perceptual discriminability, in a graphical inference task (H3). The counterargument (\neg H3) is that linguistic properties do not explain participants' performance any more than what can be readily explained by perceptual differences. Lastly, H4 generalizes these effects across several visualization types.

5. Linguistic and Perceptual Metrics

We formalize two metrics of color nameability: *name saliency* and *name variation*, both of which are specified in terms of operations over Heer and Stone's color name model [HS12]. We then define a measure of *perceptual discriminability* using the CIELAB ΔE^{76} distance function. Each of these three metrics is a function that takes a continuous colormap C as input and outputs a real number. A colormap is a mapping $C : [0, 1] \rightarrow \text{CIELAB}$, with $c = C_i$ representing the color value at the i th normalized coordinate. For example, $C_{0.5}$ is the color at the center of the scale.

5.1. Color Name Saliency

Heer and Stone define name saliency as the "degree to which a color value is uniquely named" [HS12]. They argued that saliency for a given color can be measured by obtaining a distribution of names given to that color by independent respondents, and measuring the focality of that distribution. A narrow distribution, in which virtually all respondents agree on the same unique term, implies almost zero entropy (i.e., no randomness), and thus high name saliency. We generalize this notion to a continuous colormap by computing the entropy at n uniformly spaced points on the scale, and averaging those measurements:

$$\text{Name Saliency}(C) = \frac{1}{n} \sum_{i=0}^{n-1} \mathcal{S}(C_{\frac{i}{n-1}})$$

$$\text{where } \mathcal{S}(c) = -H(p(W|c)) = \sum_{w \in W} p(w|c) \log p(w|c)$$

\mathcal{S} is the name saliency metric from [HS12], defined as the negative entropy of the distribution $p(W|c)$, where $p(w|c)$ is the probability of a name w given the color c , and W is the set of all color names reported by surveyed respondents. Heer and Stone observed

an empirical entropy range of $[-4.5, 0]$, which we map to $[0, 1]$; the closer the value to 1, the higher the name saliency.

5.2. Color Name Variation

This metric is meant to approximate the number of distinctly nameable colors in a scale. It is computed by sampling the colormap at n uniformly spaced points, and summing the probabilistic *name distance* between pairs of adjacent color samples:

$$\text{Name Variation}(C) = \sum_{i=1}^{n-1} \Delta_{\text{cos}}(C_{\frac{i-1}{n-1}}, C_{\frac{i}{n-1}})$$

Δ_{cos} is a cosine name-distance metric [HS12], which takes two LAB colors and outputs the probability they evoke different names. We opted for cosine distance over an alternative Hellinger metric due to its simplicity (the two metrics tend to produce qualitatively similar distance results).

5.3. Perceptual Discriminability

Perceptual discriminability refers to the likelihood an observer can discriminate between different colors [SLL20]. The likelihood should be proportional to the distance between those colors. For a continuous colormap, we define perceptual discriminability as the summation of distances between pairs of adjacent colors. The cumulative distance corresponds to the arc length of a colormap's curve in the CIELAB color space. A log transform reduces skew and improves model fit:

$$\text{Perceptual Discriminability}(C) = \log[\text{LAB Length}(C)]$$

$$= \log \sum_{i=1}^{n-1} \Delta E^{76}(C_{\frac{i-1}{n-1}}, C_{\frac{i}{n-1}})$$

In all the above metrics, we employ $n = 9$, sampling the colormap uniformly at nine locations to compute the above nameability and perceptual discriminability measures. The choice of an odd number of samples allows us to sample at exactly the middle of the scale, which can be an especially salient color in some designs (e.g., diverging ramps). Note that there is generally strong correlation between color name variation and perceptual discriminability (Pearson's $r = 0.946$). Intuitively, the more distant the colors, the more distinct the names. We therefore use one or the other in our model construction. On the other hand, name saliency exhibits a weak relationship with perceptual discriminability (see Figure 3). Nameability thus cannot be explained by perception alone, but rather requires knowledge of linguistic associations [BK91].

6. Colormap Selection

We selected a set of 6 colormaps for our experiments representing various designs (see Figure 2). We first started with a corpus of 235 color scales that were collected from a variety of sources by Smart et al. [SWS19]. We then selected three colormap pairs with increasing degrees of name variation and, by extension, increasing levels

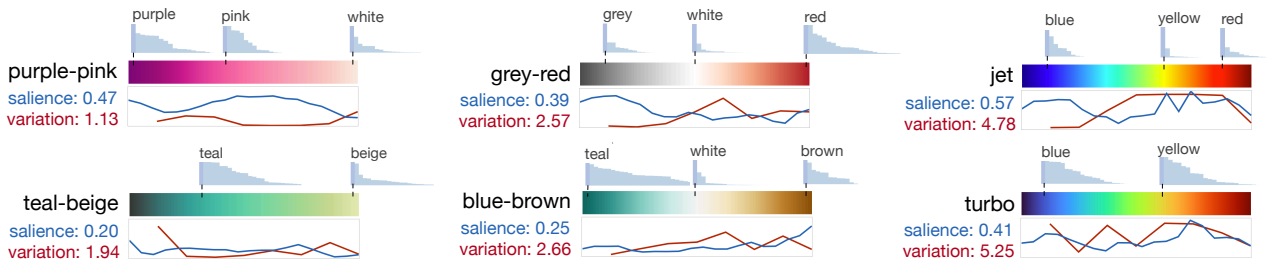


Figure 2: We selected six colormaps for Experiments 1 and 2. The top row represents scales with relatively high color name saliency. Curves show the running name saliency (blue) and name distance (red) measurements across the scale. The summary measures are in text. To illustrate the concept of name saliency, we show the distribution of names given to a color by survey respondents, along with the most common name. A narrower distribution implies a color is more strongly associated with a specific term, contributing to higher name saliency.

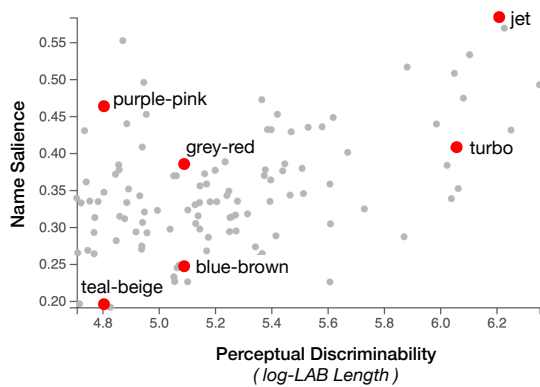


Figure 3: Perceptual discriminability vs. color name saliency for a 235-colormap corpus. The six selected colormaps are highlighted.

of perceptual discriminability: *<teal-beige and purple-pink>*—low name variation, *<blue-brown and grey-red>*—medium name variation, *<turbo and jet>*—high name variation. The three pairs also cover three distinct designs: multi-hue with monotonically increasing luminance, diverging scheme, and a rainbow design, respectively. For each pair, we purposely selected one colormap with low name saliency and another with high saliency. Specifically, *teal-beige*, *blue-brown*, and *turbo* have a relatively low name saliency (.2, .25, and .41, respectively), whereas *purple-pink*, *grey-red*, *jet* enjoy higher saliency (.47, .39, and .57). Figure 3 plots the relationship between name saliency and perceptual discriminability, highlighting the six selected colormaps.

This colormap selection enables us to independently estimate the effects of name variation (H1) and name saliency (H2). If the latter indeed improves interpretation, we would expect to see differences in performance within the pairs, which are otherwise virtually identical in perceptual discriminability. This near equivalence further allows us to attribute within-pair differences (e.g., between *teal-beige* and *purple-pink* or between *turbo* and *jet*) to linguistic rather than perceptual factors (H3).

7. Experiment I

We evaluate the six colormaps above in a graphical inference task with two visualization types: 1) a choropleth map of the US with counties as geographical units and 2) a smooth map spanning the same geography, but with the data varying continuously. Figure 1 illustrates the two conditions. We employ the synthetic procedure described in §3.2, sampling three plots from the decoy model and one plot from the target. We aggregate sample data by county in the choropleth, and normalize by county area. In the smooth map condition, the samples are aggregated pixel-by-pixel.

7.1. Pre-study

Participants' ability to infer the correct target may differ across the two visualizations, given differences in how model samples are aggregated. We therefore conducted a pre-study in order to determine appropriate levels of target-decoy divergence (i.e., difficulty). We recruited 50 participants from Amazon Mechanical Turk who were screened for color-vision deficiency. Participants completed 19 trials with each of the two visualizations, for a total of 38 trials. Each trial consisted of a freshly generated lineup. In both visualizations, we modulated trial difficulty by synthesizing target-decoy pairs with varying levels of KL divergence. Specifically, we uniformly sampled trials with divergence between 6% and 30%, providing a range of barely noticeable to highly salient lineups. Lower divergence implies less noticeable differences between the target and the decoys and, thus, more difficult judgement. We used the *viridis* colormap (a default in many systems) to visualize all lineups in the pre-study. We fitted the results to a logistic regression model to predict the expected inference accuracy for the two visualizations as a function of KL divergence.

We did not find significant difference between the smooth map and the choropleth (Wald's $Z = 0.833$, $p = 0.8389$). This suggests that people can make inferences with similar accuracy in both representations. We used the logistic model to predict a divergence level that would lead to an expected accuracy of 62.5%—halfway between chance ($P=0.25$) and perfect accuracy. Pooling results from the two visualizations, the model predicts a KL divergence level of 14.86%, which we adopt for all stimuli in the main experiment.

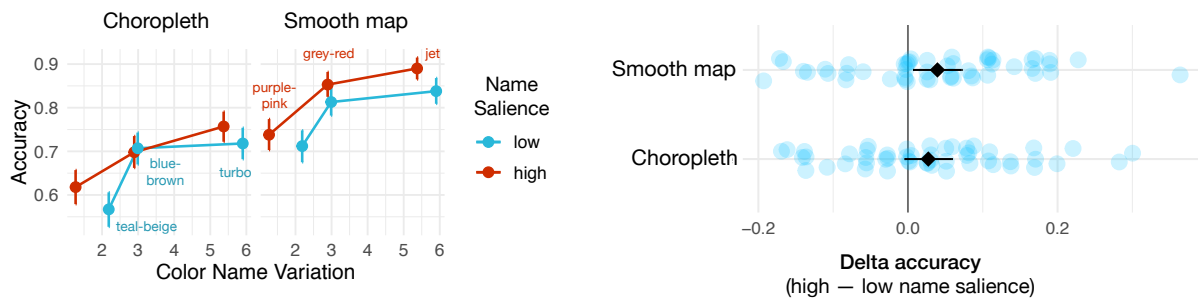


Figure 4: Participants' inference performance in Experiment 1. Left: Average accuracy in the six colormaps as a function of name variation, name salience, and visualization. Right: Performance difference between highly name-salient and perceptually equivalent, low-salience colormaps. Dots depict differences in accuracy for individual subjects. Diamonds depict group averages. Error bars are 95% CI.

7.2. Participants

For the main experiment, we recruited 100 participants (59 males, 41 females) from Amazon Mechanical Turk. We recruited workers who are US residents and who have a minimum task-approval rate of 97%, excluding participants in the pre-study. Participants completed a color vision test consisting of 14 Ishihara panels prior to enrolling in the experiment. They received a \$3 compensation.

7.3. Experiment Design

We employ a mixed design: visualization type (choropleth or smooth map) was varied between-subjects, whereas colormap (6 designs) was a within-subject factor. Half of the participants (50 individuals) saw only choropleth maps, and half saw smooth maps. All participants were tested with the six colormaps. The experiment was blocked by colormaps (6 blocks in total), with randomized block order. Participants completed 12 trials with each block, for a total of 72 trials. A trial consisted of a newly generated lineup with 4 choropleths or smooth maps. Individual maps were sized to fit within a 350×250 -pixels bounding box. A color scale was displayed to the right of the lineup. Participants were instructed to “click on the plot that doesn't belong”. After making a choice, they confirmed their selection by pressing Enter, and moved to the next trial. All lineups consisted of three decoys and one target plot (the correct answer). All trials had a fixed KL divergence of 14.86% (based on results from the pre-study).

7.4. Procedure

Participants first completed the color vision test, followed by a brief tutorial. They then completed a training session with 24 practice trials (4 practice trials \times 6 colormaps). During practice, participants were provided with feedback and informed of the correct plot. They then completed the analyzed trials, in which no feedback was provided. In addition to the trials, we randomly inserted 2 engagement checks per block, for a total of 12 checks throughout the experiments. Engagement checks consisted of lineups with high target-decoy divergence (36%), making them very easy to answer (accuracy on the engagement checks was 96.3% on average). Participants who failed the majority of the checks were removed from the analysis. We recruited new participants in replacement of those excluded, until we reached 100 individuals.

7.5. Results

Participants completed the experiment in 25.97 minutes on average ($\sigma = 9.47$). Altogether, they provided 7,200 binary rated judgments (correct for inferring the target plot). On average, participants achieved 74.6% accuracy at the task ($\sigma = 15.3\%$). Figure 4 shows the observed accuracies as a function of colormap. We fit these results to two logistic regression models: A **linguistic effects** model that predicts participants' accuracy using a colormap's name variation and name salience as independent variables. The model also includes a discrete variable representing the visualization type (choropleth or smooth map), plus interaction terms between visualization and the two linguistic factors. A second **perceptual** model predicts performance based on a colormap's perceptual discriminability (i.e., log-LAB Length). The model similarly includes a fixed effect of visualization type and its interaction with perceptual discriminability. Both models also include a fixed effect that accounts for the target-decoy divergence in a given trial. While divergence is set to a fixed level, the stimulus synthesis procedure introduces inevitable but small variations between the trials, which we account for. Lastly, we incorporate a random intercept to account for individual variations among participants.

The linguistic model predicts based on the nameability of a colormap, enabling us to independently test for effects of color name variation (H1) and name salience (H2). By contrast, the perceptual model relies solely on the appearance of colors, thus providing a more parsimonious and purely perceptual explanation of the results. We analyze the two models separately and then compare their likelihoods to infer which of the two is more plausible (H3).

7.5.1. Linguistic Model

The model indicates color name variation to be a significant predictor of inference performance (Wald's $Z = 5.227, p < 0.0001$). Adding one distinctive name to the color scale improves the odds of correct inference by 1.14 (95% CI: 1.09 — 1.20). For example, comparing *jet* and *grey-red* (a difference of 2.21 in the number of names), the odds of inference are 1.34 times higher with *jet*. The model also indicates a significant main effect of color name salience ($Z = 2.409, p < 0.05$); a unit-increment here increases inference odds by a factor of 2.20 (CI: 1.16 — 4.17). As an example, the difference between the high- and low-salience colormaps is

0.19 on average. Accordingly, we would expect people to be 1.16 times more accurate when viewing a highly name-salient colormap.

We found no evidence of interaction between visualization type and color name salience ($Z = 0.728, p = 0.467$), or between visualization and name variation ($Z = 1.204, p = 0.228$). Nameability seems to have a similar positive effect in both visualizations. There was, however, a main effect of visualization type ($Z = 2.011, p < 0.05$); participants who saw the smooth map made more accurate inferences (by a factor of 1.63, CI: 1.01 — 2.64).

7.5.2. Perceptual Model

Perceptual discriminability is a significant predictor of performance ($Z = 6.942, p < 0.001$). A step increase in log-LAB Length improves the odds of inference by a factor of 1.6 (CI: 1.4 — 1.83). There was no evidence of interaction between visualization type and perceptual discriminability ($Z = 1.630, p = 0.1030$).

7.5.3. Model Comparison and Discussion

We use AIC scores to compare the goodness-of-fit for the two models (lower AIC is better). Table 1 provides this metric along with other criteria. The AIC score for the linguistic model is lower than the perceptual model ($\Delta\text{AIC}=3.126$). The difference, being greater than 2, indicates that the former provides a significantly better fit [BA02]. The likelihood ratio indicates that the empirical data is 36.6 times more likely to be observed under the linguistic model than under the perceptual model.

Model	Parameters	AIC	logLik	Deviance
Linguistic	7	7502.6	-3743.3	7486.6
Perceptual	5	7505.7	-3746.9	7493.7

Table 1: Goodness-of-fit for the linguistic and perceptual models.

Overall, we find that color nameability is a good predictor of performance. Both name salience and name variation significantly affected performance. Specifically, incorporating colors with varied and unique names improved inferences about maps. These results support H1 and H2. Importantly, the results cannot be explained by considering perceptual factors alone. Instead, we obtained significantly better explainability with the linguistic model, thus lending support to H3. Lastly, we did not find evidence for an interaction between nameability and visualization type, which lends support to H4. However, a post-hoc analysis of the differences between the high and low name-salient colormaps suggests the potential for a zero effect in choropleths (2.72%, CI: -0.04 — 5.88%). In contrast, the highly salient color scales provided a non-zero advantage in smooth maps (3.94%, CI: 0.68 — 7.19%). Note that these differences are not significant in our models, which account for individual variations among subjects. Nevertheless, the data in Figure 4-right begs a question: could color name salience have a different effect in smooth versus discrete visualizations? We address this question in a followup experiment.

8. Experiment II

Experiment I demonstrates that nameable color encodings confer a reliable advantage. Yet, the results hint that the benefit could

differ across visualization types. In particular, name salience exerts a smaller effect in choropleths than in smooth maps. A potential explanation is the differences in the spatial structure between the two representations. Specifically, the choropleth consisted of irregular geography defined by approximately 3,000 US counties. When sampled from a model, the choropleth shows faster change in color between adjacent counties. The resulting color-layout exhibits more ‘randomness’, requiring increased attentional resources [HW12]. By contrast, the smooth map exhibits more gradual change, with identical-name colors forming larger connected shapes that may be easier to attend to. A structured color pattern may thus confer higher advantage for salient names. To eliminate this confound, Experiment 2 tests two other visualizations: a smooth scalar field and a discrete heatmap (see Figure 1). Unlike the choropleth, however, the heatmap utilizes regular bins (10×10 pixels each), resulting in a more contiguous spatial structure, despite the discrete look.

8.1. Participants, Experiment Design, and Procedures

We recruited 120 participants from Amazon Mechanical Turk (78 males, 41 females, and 1 did not identify). All participants successfully passed a color-vision test. They were compensated with \$3. The design mirrored that of Experiment 1: visualization type (heatmap or scalar field) was varied between-subjects. Half of the participants (60 individuals) saw only heatmaps and the other half saw scalar fields. The size of individual plots was fixed at 350×250 pixels. Colormap was a within-subjects factor, with all participants encountering the 6 colormaps. The procedures were similar to Experiment 1: participants first saw a brief tutorial and completed 24 practice trials. The main experiment consisted of 6 colormap blocks presented in random order. Blocks comprised 12 trials each, for a total of 72 trials (plus 12 engagement checks). Based on a pre-study with 50 participants, we set the target-decoy divergence to 16.2% in all trials for a 62.5% expected accuracy.

8.2. Results

Participants completed the experiment in 25.1 minutes on average ($\sigma = 9.2$). They provided 8,640 judgements with an average accuracy of 65.9% ($\sigma = 16\%$). Figure 5 plots the observed accuracies. As in Experiment 1, we fit the results to two logistic models representing competing explanations: a linguistic model, with color name salience and name variation as predictors, and a perceptual discriminability model that uses log-LAB Length as a predictor.

We found a significant effect of name salience (Wald’s $Z = 3.186, p < 0.01$): a unit-increase in salience increases the odds of inference by a factor of 2.57 (95% CI: 1.44 — 4.59). There was no evidence of interaction between name salience and visualization type, indicating that the effect is similar in scalar fields and heatmaps. The effect of color name variation was not significant ($Z = 1.827, p = 0.0678$). Adding one distinct color name multiplies the odds by 1.04. Yet, the effect was not reliable (CI: 0.99 — 1.09) with a possible odds ratio of 1 (i.e., no effect). The perceptual model shows a significant effect for perceptual discriminability ($Z = 4.163, p < 0.001$). A step-increase in log-LAB Length increases the odds by a factor of 1.28 (CI: 1.14 — 1.44).

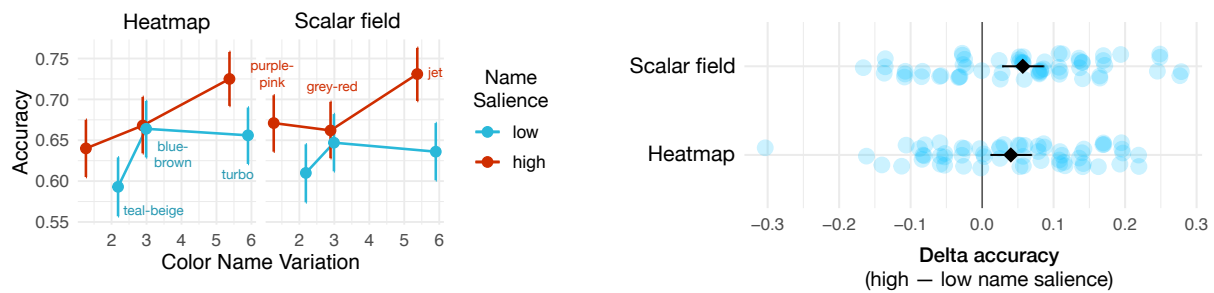


Figure 5: Performance in Experiment 2. Left: Average accuracy in the six colormaps. Right: Difference in performance between high and low name-saliency colormaps. Dots depict performance deltas for individual subjects. Diamonds depict group average. Error bars are 95% CI.

8.2.1. Model Comparison and Discussion

Table 2 gives fit scores for the two models. We similarly use AIC to select between the two models (lower is better). The score for the linguistic model is lower than the perceptual model ($\Delta AIC = 14.56$). The difference is substantially larger than 2, providing strong evidence in favor of the linguistic model.

Model	Parameters	AIC	logLik	Deviance
Linguistic	7	10419.7	-5201.9	10403.7
Perceptual	5	10434.3	-5211.1	10422.3

Table 2: Goodness-of-fit for the models in Experiment 2.

Overall, the effect of name saliency on inference (odds multiplier: 2.57, CI: 1.44 — 4.59) was comparable to that of Experiment 1 (2.20, CI: 1.16 — 4.17). However, the confidence intervals suggest a more robust impact for name saliency in this experiment. The results support H2. On the other hand, the evidence for color name variation (H1) is weak and non-significant ($p = 0.0678$). The estimated effect, while likely positive (1.04), is substantially weaker when compared to Experiment 1 (1.14). Even so, comparing the linguistic and perceptual models indicates that benefits of nameability cannot be attributed to perceptual properties alone, with the AIC criterion showing strong evidence in favor of H3. Lastly, we found no evidence of interaction between either nameability metrics and visualization type. In both visualizations, participants' accuracy improved reliably with high-saliency colormaps (mean delta accuracy: 5.69%, CI: 2.83 — 8.56% in scalar fields vs. 4.02%, CI: 1.19 — 6.86% in heatmaps). These results are consistent with H4. While 5.7% might seem like a small delta, the impact of name saliency amounts to approximately 9% better inference (CI: 4% — 13%), considering the baseline task accuracy (65.9%).

9. General Discussion

Guidelines for color design in visualization have long emphasized perceptual properties. However, colors also evoke semantic and linguistic associations, which are known to influence the interpretation of categorical data [SLL20]. We sought to understand how color names affect people's ability to make inferences about quantitative visualizations. We studied metrics that capture two nameability aspects: the distinctiveness and uniqueness (i.e., saliency) of

color names. We theorized that these linguistic properties provide a better indicator of colormap usefulness for graphical inference.

9.1. Nameability Predicts Accuracy in Graphical Inference

Results from both experiments show that color nameability plays a role in interpreting visualizations. Moreover, nameability appears to have greater influence on task-performance than traditional perceptual properties, such as order and uniformity. As an example, the most accurate colormap in both experiments was *jet*. The advantage for this scale can be explained by the variety and saliency of its color names (prototypical blue, green, yellow, orange, and red). By comparison, *turbo*, an almost equivalent rainbow performed less well. Notably, *turbo* incorporates similar but less saturated hues in an attempt to improve perceptual uniformity [Mik19]. However, these less prototypical colors gave rise to lower name saliency, and ultimately, lower accuracy. Similar differences were found between *purple-pink* (high name-saliency) and *teal-beige* (low saliency), with the former performing better despite the two sharing identical perceptual discriminability. The difference between *grey-red* and *blue-brown*, however, was less clear, possibly due to the smaller saliency gap between this pair.

In modeling the effects of color name variation and saliency, we found support of both. Evidence in favor of name variation (H1), however, was marginally weaker in the second experiment ($p = 0.0678$). On the other hand, both experiments provided positive evidence for name saliency (H2). It is noteworthy that while color name variation is highly correlated with perceptual discriminability ($r = .95$), name saliency does not exhibit the same strong association. This means that one cannot predict the nameability of colors (and by extension their usefulness) from their perceptual properties alone (H3). Indeed, in both experiments, the linguistic model consistently outperformed a more parsimonious LAB-based model. The results suggest cognitive benefits to color names, which appear to transcend low-level color appearance. These benefits were observed in all four visualizations we evaluated (H4), although they may be weaker in choropleths. The latter is possibly due to nonuniform geographies causing a less structured color distribution, which may strain one's attention [HW12; BDM*16].

The above results partially—but not entirely—confirm recent work, which found a positive effect for color name variation [RS21]. However, this work extends those earlier results in several impor-

tant ways. Specifically, we found a robust effect for name salience, which turned out to be a more reliable predictor of performance than name variation alone. In other words, it is not enough to select colors that vary in name, but it is beneficial to incorporate colors with salient and prototypical names (e.g., ‘blue’, ‘yellow’ over ‘teal’ and ‘beige’). We have shown that nameability is a better predictor of colormap utility than perceptual characteristics, a contribution with theoretical and practical implications. Lastly, we demonstrated that nameability effects present across a variety of visualizations. This not only indicates that above nameability-centric guidelines are robust, but also points to language-driven processes involved in interpreting quantitative colormaps. It is important to note, however, that these results may not generalize to other analysis tasks. Prior research suggests that perceptual uniformity and luminance monotonicity are still important factors [LH18; DPR*18; RNA18]. Colormap designers may therefore need to weigh these traditional perceptual properties in addition to nameability.

9.2. Color Categorization Considered Helpful

A central question is: why do color names matter, especially when the data being represented is quantitative, with no obvious name associations? A potential explanation starts by recognizing that color names serve to discretize a continuous data domain into a number of color categories that are easy to reference. Though researchers have long argued against this kind of categorization [BI07; QPCM19], we suggest that it can be helpful in the more interpretive tasks like graphical inference. Here, the observer has to identify a plot containing an unusual data distribution. This is a useful model task because it requires one to distinguish between percepts that truly differentiate models from others that are due to chance (i.e., random sampling). The observer then weighs those percepts against four competing hypotheses (corresponding to each of the four plots being the true target).

We speculate that people are better at performing tasks such as the above when the representation affords visual grouping (or ‘chunking’). A color scale with categorizable colors serves that exact purpose. For example, *jet* affords grouping of quantities into rough structures, such as hotpots (red), mid-high values (yellow), mid-low (cyan), and valleys (blue). An observer can then attend to those structures more easily throughout the task, as in when comparing different simulations. Visual chunking is thought to be a crucial visuospatial strategy in STEM [SWD*20], and color representations that facilitate this strategy can be advantageous. Conversely, encodings that give rise to a smooth, highly uniform representation may complicate some analyses. This observation may explain why practitioners continue to prefer colorful encodings (e.g., rainbows), to the dismay of visualization researchers [Mor16].

Colors with salient names are possibly easier to attend to, not just because nameability improves discriminability (as the experiment with Russian speakers shows [WWF*07]), but also because salient colors are “linguistically codable” [Hei72]. For instance, it may be easier to think about ‘red’ vs. ‘blue’ patterns, than it is to describe ‘maroon’ and ‘teal’ features. Salient linguistic associations are not only useful for communication, but could also aid one’s own problem solving loop by increasing the opportunity to use verbal working memory.

10. Limitations and Future Research

Our results suggest that color design tools should help designers explicitly think about color names. Examples from the community already exist. For instance, Heer and Stone provide a set of sample applications that demonstrate new image editing capabilities based on color names [HS12]. Colorgorical generates categorical palettes based partly on name distance and uniqueness [GLS17]. However, quantitative color mapping tools (e.g., CCC [NCS*19] and Color Crafter [SWS19]) still only consider perceptual factors. An opportunity exists to augment these tools with nameability metrics and, in the process, extend the design space of quantitative colormaps.

Although we show advantages to colorful encodings (e.g., rainbows), it is important to note other shortcomings: rainbows may exclude people with color-vision deficiency. The green area in an RGB rainbow also suffers from low perceptual discriminability, which could impede the perception of small features in that region [WTB*18]. Designers thus need to balance nameability with other considerations, including accessibility. Future design tools could aid in optimizing for these potentially competing constraints.

While emphasizing color names appears to be useful in graphical inference, the results may not generalize to other tasks. It is crucial to understand circumstances under which nameability aids or hinders performance. In particular, future research may consider a potential trade off between perceptual precision (i.e., the ability to accurately retrieve and compare color-coded values in a visualization) versus enabling the observer to think categorically about color, which may lead to lower quantitative precision. A related question is how well-suited these different colormap affordances are for various data distributions, beyond the Gaussian-mixture models employed in this study.

11. Conclusion

Color is a principal channel for encoding quantitative data. Visualization research has so far focused on the perceptual characteristics of colors, painting an incomplete picture that minimizes other considerations. We investigated how color nameability affects people’s ability to interpret and make inferences about visualizations. Results from two studies suggest that people are more accurate when viewing colors associated with salient names. Colorful visualizations with a variety of nameable colors also seemed more interpretable, though that latter effect was not as robust. Together, however, these two linguistic properties provided a better explanation of participants’ performance than did the perceptual color properties. Our findings have theoretical and practical implications for visualization design. They also suggest a need for design tools to incorporate color nameability metrics, even for encodings that are purely quantitative.

Answers for the lineups in Figure 1: From left to right: bottom-left quadrant; bottom-left; bottom-left; top-right.

Acknowledgements

This research was supported by NSF award 1942429, and by a grant from the Argonne Leadership Computing Facility, a US Department of Energy Science User Facility operated under contract DE-AC02-06CH11357.

References

- [BA02] BURNHAM, KENNETH P and ANDERSON, DAVID R. "A practical information-theoretic approach". *Model selection and multimodel inference, 2nd ed.* Springer, New York 2 (2002) 8.
- [BCF20] BERTINI, ENRICO, CORRELL, MICHAEL, and FRANCONERI, STEVEN. "Why Shouldn't All Charts Be Scatter Plots? Beyond Precision-Driven Visualizations". *arXiv preprint arXiv:2008.11310* (2020) 4.
- [BCH*09] BUJA, ANDREAS, COOK, DIANNE, HOFMANN, HEIKE, et al. "Statistical inference for exploratory data analysis and model diagnostics". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367.1906 (2009), 4361–4383 3.
- [BDM*16] BEECHAM, ROGER, DYKES, JASON, MEULEMANS, WOUTER, et al. "Map lineups: effects of spatial structure on graphical inference". *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2016), 391–400 9.
- [BGP*11] BORKIN, MICHELLE, GAJOS, KRZYSZTOF, PETERS, AMANDA, et al. "Evaluation of artery visualizations for heart disease diagnosis". *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), 2479–2488 2.
- [BI07] BORLAND, DAVID and II, RUSSELL M TAYLOR. "Rainbow color map (still) considered harmful". *IEEE Computer Graphics and Applications* 27.2 (2007) 2, 10.
- [BK91] BERLIN, BRENT and KAY, PAUL. *Basic color terms: Their universality and evolution*. Univ of California Press, 1991 1, 2, 5.
- [Bol78] BOLTON, RALPH. "Black, white, and red all over: the riddle of color term salience". *Ethnology* 17.3 (1978), 287–311 2.
- [BPS17] BARTRAM, LYN, PATRA, ABHISEKH, and STONE, MAUREEN. "Affective color in visualization". *Proceedings of the 2017 CHI conference on human factors in computing systems*. 2017, 1364–1374 1.
- [Bre96] BREWER, CYNTHIA A. "Guidelines for selecting colors for diverging schemes on maps". *The Cartographic Journal* 33.2 (1996), 79–86 2.
- [Bre97] BREWER, CYNTHIA A. "Spectral schemes: Controversial color use on maps". *Cartography and Geographic Information Systems* 24.4 (1997), 203–220 2.
- [BTBV02] BENAVENTE, ROBERT, TOUS, FRANCESC, BALDRICH, RAMON, and VANRELL, MARIA. "Statistical modelling of a colour naming space". *Conference on Colour in Graphics, Imaging, and Vision*. Vol. 2002. 1. Society for Imaging Science and Technology. 2002, 406–411 2.
- [BTS*17] BUJACK, ROXANA, TURTON, TERECE L, SAMSEL, FRANCESCA, et al. "The Good, the Bad, and the Ugly: A Theoretical Framework for the Assessment of Continuous Colormaps". *IEEE Transactions on Visualization and Computer Graphics* (2017) 1.
- [CSH08] CHUANG, JASON, STONE, MAUREEN, and HANRAHAN, PAT. "A probabilistic model of the categorical association between colors". *Color and Imaging Conference*. Vol. 2008. 1. Society for Imaging Science and Technology. 2008, 6–11 2.
- [DPR*18] DASGUPTA, ARITRA, POCO, JORGE, ROGOWITZ, BERNICE, et al. "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) 2, 4, 10.
- [Gle17] GLEICHER, MICHAEL. "Considerations for visualizing comparison". *IEEE transactions on visualization and computer graphics* 24.1 (2017), 413–423 4.
- [GLS17] GRAMAZIO, CONNOR C, LAIDLAW, DAVID H, and SCHLOSS, KAREN B. "Colorgical: Creating discriminable and preferable color palettes for information visualization". *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), 521–530 2, 10.
- [Har87] HARNAD, STEVAN. "Psychophysical and cognitive aspects of categorical perception: A critical overview". *Categorical perception: The groundwork of cognition*. Cambridge University Press, 1987, 1–52 2.
- [Hea96] HEALEY, CHRISTOPHER G. "Choosing effective colours for data visualization". *Proceedings of IEEE Visualization '96*. IEEE. 1996, 263–270 2.
- [Hei72] HEIDER, ELEANOR R. "Universals in color naming and memory." *Journal of experimental psychology* 93.1 (1972), 10 4, 10.
- [HFMC12] HOFMANN, HEIKE, FOLLETT, LENDIE, MAJUMDER, MAHBUBUL, and COOK, DIANNE. "Graphical tests for power comparison of competing designs". *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), 2441–2448 3.
- [HMNP72] HAYS, DAVID G, MARGOLIS, ENID, NAROLL, RAOUL, and PERKINS, DALE REVERE. "Color term salience". *American Anthropologist* 74.5 (1972), 1107–1121 2.
- [HP07] HUANG, LIQIANG and PASHLER, HAROLD. "A Boolean map theory of visual attention." *Psychological review* 114.3 (2007), 599 4.
- [HS12] HEER, JEFFREY and STONE, MAUREEN. "Color naming models for color selection, image editing and palette design". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012, 1007–1016 2, 5, 10.
- [HW12] HAROZ, STEVE and WHITNEY, DAVID. "How capacity limits of attention influence information visualization effectiveness". *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), 2402–2410 8, 9.
- [JOEF19] JARDINE, NICOLE, ONDOV, BRIAN D, ELMQVIST, NIKLAS, and FRANCONERI, STEVEN. "The Perceptual Proxies of Visual Comparison". *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2019), 1012–1021 4.
- [KRPC00] KALVIN, ALAN DAVID, ROGOWITZ, BERNICE E, PELAH, ADAR, and COHEN, ARON. "Building perceptual color maps for visualizing interval data". *Human Vision and Electronic Imaging V*. Vol. 3959. International Society for Optics and Photonics. 2000, 323–336 2.
- [LFC*20] LU, KECHENG, FENG, MI, CHEN, XIN, et al. "Palettailor: Discriminable Colorization for Categorical Data". *IEEE Transactions on Visualization and Computer Graphics* (2020) 2.
- [LFK*13] LIN, SHARON, FORTUNA, JULIE, KULKARNI, CHINMAY, et al. "Selecting semantically-resonant colors for data visualization". *Computer Graphics Forum*. Vol. 32. 3pt4. Wiley Online Library. 2013, 401–410 2.
- [LH18] LIU, YANG and HEER, JEFFREY. "Somewhere over the rainbow: An empirical assessment of quantitative colormaps". *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. 2018, 598 2, 10.
- [Mik19] MIKHAILOV, ANTON. *Turbo, An Improved Rainbow Colormap for Visualization*. [Online; accessed 4-December-2020]. 2019. URL: <https://ai.googleblog.com/2019/08/turbo-improved-rainbow-colormap-for.html> 9.
- [Moj05] MOJSILOVIC, ALEKSANDRA. "A computational model for color naming and describing color composition of images". *IEEE Transactions on Image processing* 14.5 (2005), 690–699 2.
- [Mor09] MORELAND, KENNETH. "Diverging color maps for scientific visualization". *International Symposium on Visual Computing*. Springer. 2009, 92–103 1, 2.
- [Mor16] MORELAND, KENNETH. "Why We Use Bad Color Maps and What You Can Do About It". *Electronic Imaging* 2016.16 (2016), 1–6 2, 10.
- [Mun10] MUNROE, RANDALL. *XKCD Color Names Survey*. [Online; accessed 4-December-2020]. 2010. URL: <https://blog.xkcd.com/2010/05/03/color-survey-results/> 2.
- [Mun14] MUNZNER, TAMARA. *Visualization analysis and design*. CRC Press, 2014 1, 2.

- [NCS*19] NARDINI, P., CHEN, M., SAMSEL, F., et al. "The Making of Continuous Colormaps". *IEEE Transactions on Visualization and Computer Graphics* (2019), 1–1. DOI: [10.1109/TVCG.2019.2961674](https://doi.org/10.1109/TVCG.2019.2961674) 10.
- [Par05] PARAMEI, GALINA V. "Singing the Russian blues: An argument for culturally basic color terms". *Cross-cultural research* 39.1 (2005), 10–38 2.
- [QPCM19] QUINAN, P SAMUEL, PADILLA, LM, CREEM-REGEHR, SARAH H, and MEYER, MIRIAH. "Examining Implicit Discretization in Spectral Schemes". *Computer Graphics Forum*. Vol. 38. 3. Wiley Online Library. 2019, 363–374 10.
- [RK01] ROGOWITZ, BERNICE E and KALVIN, ALAN D. "The "Which Blair Project": a quick visual method for evaluating perceptual color maps". *Visualization, 2001. VIS'01. Proceedings.* IEEE. 2001, 183–556 2, 4.
- [RNA18] REDA, KHAIRI, NALAWADE, PRATIK, and ANSAH-KOI, KATE. "Graphical perception of continuous quantitative maps: the effects of spatial frequency and colormap design". *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. 2018, 272 2–4, 10.
- [Ros73] ROSCH, ELEANOR H. "Natural categories". *Cognitive psychology* 4.3 (1973), 328–350 4.
- [RP19] REDA, KHAIRI and PAPKA, MICHAEL E. "Evaluating Gradient Perception in Color-Coded Scalar Fields". *2019 IEEE Visualization Conference (VIS)*. IEEE. 2019, 271–275 4.
- [RS21] REDA, K. and SZAFIR, D. A. "Rainbows Revisited: Modeling Effective Colormap Design for Graphical Inference". *IEEE Transactions on Visualization and Computer Graphics* 27.2 (2021), 1032–1042. DOI: [10.1109/TVCG.2020.3030439](https://doi.org/10.1109/TVCG.2020.3030439) 1, 2, 5, 9.
- [RT94] ROGOWITZ, BERNICE E and TREINISH, LLOYD A. "Using perceptual rules in interactive visualization". *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics. 1994, 287–295 1, 2.
- [RT98] ROGOWITZ, BERNICE E and TREINISH, LLOYD A. "Data visualization: the end of the rainbow". *IEEE spectrum* 35.12 (1998), 52–59 2.
- [RTB08] RATWANI, RAJ M, TRAFTON, J GREGORY, and BOEHM-DAVIS, DEBORAH A. "Thinking graphically: Connecting vision and cognition during graph comprehension." *Journal of Experimental Psychology: Applied* 14.1 (2008), 36 3.
- [SLL20] SCHLOSS, KAREN B, LEGGON, ZACHARY, and LESSARD, LAURENT. "Semantic Discriminability for Visual Communication". *IEEE Transactions on Visualization and Computer Graphics* (2020) 2, 5, 9.
- [SLWF18] SCHLOSS, KAREN B, LESSARD, LAURENT, WALMSLEY, CHARLOTTE S, and FOLEY, KATHLEEN. "Color inference in visual communication: the meaning of colors in recycling". *Cognitive research: principles and implications* 3.1 (2018), 1–17 2.
- [SS15] SETLUR, VIDYA and STONE, MAUREEN C. "A linguistic approach to categorical color assignment for data visualization". *IEEE transactions on visualization and computer graphics* 22.1 (2015), 698–707 2.
- [SWD*20] STIEFF, MIKE, WERNER, STEPHANIE, DESUTTER, DANE, et al. "Visual chunking as a strategy for spatial thinking in STEM". *Cognitive Research: Principles and Implications* 5 (2020), 1–15 10.
- [SWS19] SMART, STEPHEN, WU, KEKE, and SZAFIR, DANIELLE ALBERS. "Color Crafting: Automating the Construction of Designer Quality Color Ramps". *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2019), 1215–1225 5, 10.
- [TSG*19] THAM, DIANA SU YUN, SOWDEN, PAUL T, GRANDISON, ALEXANDRA, et al. "A systematic investigation of conceptual color associations." *Journal of Experimental Psychology: General* (2019) 1.
- [vdWS15] Van der WALT, STÉFAN and SMITH, NATHANIEL. *Matplotlib colormaps*. <https://bids.github.io/colormap/>. [Online; accessed 20-April-2020]. 2015. URL: <https://bids.github.io/colormap/2>.
- [War12] WARE, COLIN. *Information Visualization: Perception for Design*. Elsevier, 2012 2.
- [War88] WARE, COLIN. "Color sequences for univariate maps: Theory, experiments and principles". *IEEE Computer Graphics and Applications* 8.5 (1988), 41–49 2, 3.
- [WCHB10] WICKHAM, HADLEY, COOK, DIANNE, HOFMANN, HEIKE, and BUJA, ANDREAS. "Graphical inference for infovis". *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), 973–979 3.
- [WTB*18] WARE, COLIN, TURTON, TERECE LOUISE, BUJACK, ROXANA, et al. "Measuring and Modeling the Feature Detection Threshold Functions of Colormaps". *IEEE Transactions on Visualization and Computer Graphics* (2018) 2, 3, 10.
- [WWF*07] WINAWER, JONATHAN, WITTHOFT, NATHAN, FRANK, MICHAEL C, et al. "Russian blues reveal effects of language on color discrimination". *Proceedings of the national academy of sciences* 104.19 (2007), 7780–7785 2, 10.
- [ZH16] ZHOU, LIANG and HANSEN, CHARLES D. "A survey of colormaps in visualization". *IEEE Transactions on Visualization and Computer Graphics* 22.8 (2016), 2051–2069 1.