Measuring Effective Highlight Colors in Color-Coded Scatterplots

by

Supriya Manikonda Keshavaiah Naidu

B.E., BMS College of Engineering, 2015

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Master of Science

Department of Computer Science

2019

This thesis entitled:
Measuring Effective Highlight Colors in Color-Coded Scatterplots
written by Supriya Manikonda Keshavaiah Naidu
has been approved for the Department of Computer Science

Dr. Danielle Albers Szafir

Dr. Sriram Sankaranarayanan

Dr. R. Benjamin Shapiro

Date ____________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocol #15-0683

Manikonda Keshavaiah Naidu, Supriya (M.S., Computer Science)

Measuring Effective Highlight Colors in Color-Coded Scatterplots

Thesis directed by Dr. Danielle Albers Szafir

Visualization designers commonly use multiple visual channels to draw a user's attention to important information in a visualization. Color is a frequently used channel to highlight important data points. However, recent work shows guidelines for selecting colors to highlight data are based on heuristic design intuitions and offer little insight into effective highlight colors for color-coded visualizations. I present a crowdsourced study measuring the effects of highlight color in color-coded scatterplots. My findings indicate shortcomings in conventional approaches to selecting highlight colors. Empirical measures derived from the study results provide data-driven recommendations for effective highlight colors through preference scores and predictive models.

Dedication

To my family.

Acknowledgements

I am extremely grateful to my committee chair, Dr. Danielle Albers Szafir for her unwavering support and guidance throughout the duration of this work. As my advisor and mentor, she has provided invaluable insight into academic research. I would like to thank my thesis committee members, Dr. Sriram Sankaranarayanan and Dr. R. Benjamin Shapiro for their constructive feedback and advice on this work. I would also like to thank my colleagues in the CUVisuaLab, Steve Smart, Keke Wu, Matt Whitlock and Tetsumichi Umada for their helpful advice and encouragement. Finally, I want to thank Haritha, Sreesha, Abhijit and Ashwin who proofread, offered suggestions and provided moral support.

Contents

Tables

Table

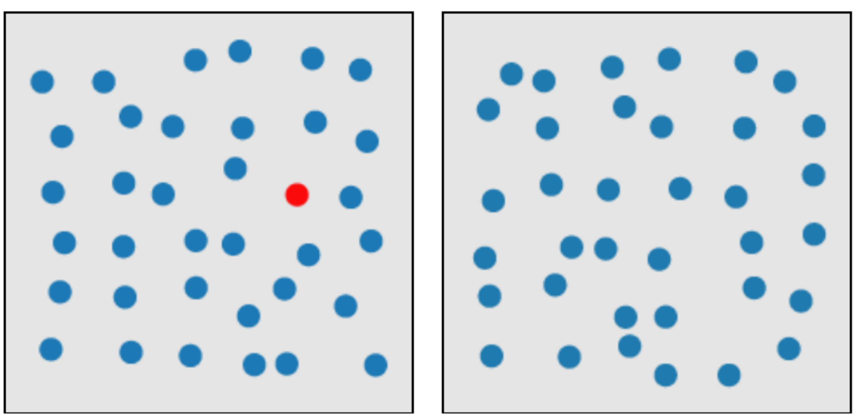
Figures

Figure

Chapter 1

Introduction

An initial look at a visualization should help summarize the information being looked at and immediately locate salient data points in the visualization. This preliminary glance allows certain visual properties to pop-out of a visualization. This phenomenon is more commonly known as "pop-out effect". Visualization designers leverage the use of pop-out effect to reduce the time and cognitive effort required to find relevant data. Color, shape, size, orientation, and motion are a few prevalent features used to create a pop-out effect 1.1. Among these features, color is frequently used to **highlight** important data points because we can readily see well-designed highlight colors as categorically distinct from other encoding colors [64].



(a) (b)

Figure 1.1: An example of searching for the target highlighted in red agaisnt a field of blue context marks, (a) target present; (b) target absent; [33]

Heuristics for selecting effective highlight colors are grounded in design theory. The most common approaches use bright salient colors (reds and yellows) [76, 9, 26] and colors with low lightness and chroma (blacks and grays) [53]. Visualization designers often rely on more sophisticated theories that suggest the use of complementary colors (colors on the opposing side of the conventional color wheel from encoding hues [25, 40, 91]). The goal of this thesis is to measure the effects of highlight color in visualizations and generate preliminary guidelines for choosing effective colors to highlight information in visualizations.

Single homogenous color or a color ramp (sets of colors) is generally used to map data in a visualization. The question that arises here is how do we choose effective colors to highlight a single mark or a set of marks from the context marks in a visualization. In this study, I quantify how color affects a user's ability to distinguish data points in a visualization. I characterized effective highlight colors using color ramps and highlight colors as independent variables for color-coded scatterplots in a crowdsourced study with 302 participants using response time, response accuracy and perceived aesthetics as dependent variables. I use the results from this experiment to evaluate existing methods, measure functional and aesthetic effects of highlight color selections to identify effective highlight color choices for popular color ramps and provide empirically-grounded metrics for utilizing the color channel to highlight information in a visualization.

My findings show that the optimal highlight color depends directly on the characteristics of a ramp and, in some cases, contradicts popular heuristic choices. For example, people more easily discriminate a highlighted mark colored in red if the context/distractor marks are not mapped to hues of red. The results also indicate that colors with high luminance have low search time and high response accuracy.

1.1 Problem Description

Interesting and important data points in a visualization are often highlighted using different visual channels, immediately attracting analysts' attention and helping them to find important data without having to actively search for it [34]. Among them, color is a frequently used visual channel because it's often really easy to find brightly colored targets and conventional design states that color is largely separable from other visual channels like size and position [80, 4]. While many studies in psychology have investigated when colors pop-out [14, 54], we lack quantified actionable models of what constitutes a good highlight color.

Studies of highlighting and pop-out in visualization focus on fixed hues in monochromatic visualizations: highlight colors simply need to stand out from constant color distractor marks [26, 76]. However, visualizations also often use color to encode value. The diversity of colors introduced by color encodings complicates highlight color selection, forcing designers to find highlight colors that sufficiently contrast with encoding colors and introducing aesthetic constraints. For instance, a neon green mark contrasts with colors in a red-orange-yellow ramp but results in a visually displeasing graph. We can use alternative channels like size to instead highlight data; however, these channels are often already used to encode data (e.g., mapping size to value) or are key to a visualization's structure (e.g., size and shape on a map). Color remains among the most popular methods for highlighting data: well-designed highlight colors appear categorically distinct from other encoding colors [64].

Existing guidelines have limited basis in empirical work and remain loosely defined, making them difficult to use in practice. Most systems still rely on designer intuitions to select highlight colors, often falling back on **de facto** reds and yellows without conscious consideration of the visualization's design. However, these approaches may work for monochromatic visualizations but may prompt confusing and visually displeasing graphs for color-coded visualizations. In this work, I measure response accuracy, response time, subjective preference, perceived discriminability, and harmony to identify optimal highlight colors for these ramps and to inform a preliminary model for predicting highlight colors.

1.2 Contribution

The primary contribution of this work is a set of empirical measures that inform effective color choices for highlighting data in visualizations. These findings challenge the default uses of reds and yellows as well as hypotheses about reducing saturation to highlight information and reveal some support for the use of complementary hues. The results of this study provide empirical measures that can be used to select highlight colors for popular ramps and ground a preliminary model for automated highlight color selection.

1.3 Thesis Outline

Chapter 1 defines the problem and discusses the existing approaches in highlighting information in data visualizations. Chapter 2 thoroughly discusses prior work accomplished with color perception and aesthetics for visualizations. Chapter 3 details the importance of the problem and explains the conditions to evaluate the hypotheses for effective highlight color choices. Chapter 4 delves into the methodology utilized for the study explaining the different variables, stimuli, and procedure of the study. This chapter also describes the sampling of color ramps and highlight colors. Chapter 5 details the results of the experiment. Chapter 6 presents a final discussion summarizing the contribution. Finally, conclusions, limitations and potential future work are detailed in Chapter 7.

Chapter 2

Related Work

Selecting effective highlight colors for visualizations requires considering both perception (e.g., points that are significantly different from other colors in a visualization) and aesthetics (e.g., colors that are harmonious with any existing encodings in the visualization). I draw on research from visual search, visualization design, and color perception and aesthetics to ground an empirical investigation of effective highlight colors.

2.0.1 Visual Search & Pop-Out

The primary goal of highlighting in visualizations is to facilitate rapid **visual search**—to help people find target datapoints as quickly as possible. Vision science has developed an in-depth understanding of how people find targets in a scene (see Wolfe [96] for a survey). In some instances, an item or set of items can be represented so that it **pops-out** (e.g., a red circle amongst blue circles), immediately attracting one's attention in a form of preattentive processing [34, 86]. The visual features (e.g., size, color, orientation) of an object significantly contribute to visual search [97], with simple features like color processed in ways that can facilitate pop-out [86].

A large body of literature in vision science has explored when and how different visual features might facilitate pop-out and other forms of rapid visual search. For example, Egeth et. al. [18] found that using colors that caused letters to pop-out could help people more quickly find a target within those letters

(e.g., a red 'A' amongst a set of red and black letters). Theeuwes [83] found that rapid changes in lightness pop-out, while changes in hue do not. DeVries et al. [14] found that background color also influences visual search, with search times generally increasing for objects on dark backgrounds. However, preattentive features alone may not be sufficient to define how quickly people can find an object [98]. Factors beyond features such as familiarity with a scene [92] and scene structure [39] also influence visual search.

Color can, but does not always, pop-out [87]. However, the behavior of this pop-out is not well understood. For example, our abilities to select different objects depends on the configuration of colors in color space [51]. D'Zmura [17] found that visual search partially depends on the full set of colors in a scene: search can be inefficient when a target color is collinear with the remaining colors. Stroud et. al. [77] showed that we can search for collections of discrete colors even when those colors come from a small region in color space. However, we lack systematic insight into when specific colors might pop-out from a collection of colors. As a result, visualization designers tend to use heuristic approaches to select colors that emphasize particular points in a display. Hall et al. surveys methods for emphasizing aspects of a visualization [27]. One of the most common techniques for doing this is **highlighting**: changing the visual features of a target mark such that it pops out from the rest of the data [64, 76].

2.0.2 Visual Search in Visualization

Visualization relies heavily on results from vision science to inform design, drawing on concepts like visual attention [28, 32], ensemble coding [21, 81], and memory [6, 7]. For example, Healey & Enns explored how preattentive visual features could support visual search in multivariate visualizations [35]. However, more recent work seeks to explore not only how vision science concepts can drive design, but how we can apply these concepts to understand how visualizations work through a better understanding of graphical perception.

Several such studies explore aspects of visual search. For example, people can better locate data of interest when the color of that data is semantically relevant [48] or when the visualization's layout organizes the data by relevant features [28], especially as the number of items increases [23].

Saket et. al. [66] noted performance differences across common visualizations for several statistical search tasks, while Kim & Heer [42] categorize the effects of specific visual channels on similar tasks. However, these studies have primarily focused on measuring search performance in general classes of visualizations (e.g., scatterplots versus bar charts). I instead focus on how visualization might support efficient search in specific visualization designs.

Highlighting is a design technique used to direct a viewer's attention to important data, facilitating search and navigation (see Liang & Huang [47] and Robinson [63] for surveys). Visualizations can use any number of visual features to highlight data, typically emphasizing those that can pop-out [34]. For example, Becker & Cleveland use "transient painting" to emphasize information by mapping important data to a special color [3]. Villarroel et al. [89] offer designs for creating multiple categories of highlights on plain text using foreground and background color manipulations. Robinson [64] introduces a set of techniques for highlighting geospatial data.

Studies of pop-out in visualization indicate that focus [64, 90], transparency [64, 61], depth-of-field [45], motion [94], and dichoptic presentation [46] can all effectively highlight data in visualizations. Gutwin et al. [26] found that motion pop-out is more robust in the periphery than color and size.

Waldner et al. [90] model effects of luminance, blur, and other highlighting techniques for multichannel highlighting. Strobelt et al. [76] conducted a crowdsourced study exploring highlighting in text visualizations, including using red fonts and yellow backgrounds, and offer guidelines for selecting highlight techniques based on their results. However, these studies focus on fixed hues (e.g., reds and yellows) and monochromatic visualizations (e.g., black text or grey distractor marks), offering little systematic insight into how to highlight data in color-coded visualizations.

2.1 Color and Aesthetics

While other channels offer effective highlighting cues, channels beyond color often already encode data (e.g., mapping size to value) or are key to the visualization's structure (e.g., size or shape on maps). Color remains among the most popular methods for highlighting important data:

we can readily see well-designed highlight colors as categorically distinct from other encoding colors [64]. Research systems [44, 73], toolkits [37], and commercial tools like ArcGIS [1] and TensorFlow [1] highlight data using color.

While most visualizations choose a reserved color for highlighting data (most often a red or yellow), color behaves in complex ways in data visualization that may preclude fixed-color highlighting from being effective. For example, colors may carry semantic meanings that make it hard to find relevant data [48], background colors may distort meaning or impair search [14, 24], or highlight colors may simply fall too close to other colors in a ramp [50, 79].

Color design tools for visualization largely rely on either designer handcrafting [29, 67] or algorithmic approaches [22, 72]. Other techniques aim to support specific tasks like comparison [84] or optimizing cluster separation [93]. However, these efforts focus on general encoding design, not on selecting colors to emphasize particular components of a display. Other methods highlight given ranges in data. For example, PRAVDAColor uses rule-based approaches to decide how and when to use particular portions of a color ramp to highlight relevant ranges in data [5]. Samsel et al. [67] combine ramps to highlight specific ranges of interest by manipulating luminance contrasts. Elmqvist et al. [19] introduce lenses that refine color encodings on regions of interest. These efforts offer approaches for enhancing and directing attention in visualizations to different data ranges; however, they do not explore generalizable methods for selecting specific highlight colors.

Selecting effective highlight colors requires carefully balancing aesthetic and perceptual constraints. Colorgorical balances these elements by combining models of discriminability [58, 74] with models of aesthetics [68, 69]. However, aesthetics are influenced by factors like affect [2] and audience demographics [60] that are difficult to capture **a priori**. Further, different applications may weight objective performance (e.g., finding data quickly) and subjective performance (e.g., positive aesthetics) differently depending on the application. While approaches like Color Measures [10] provide quantitative assessments across both objective and heuristic components of color encoding design, these approaches provide little concrete insight into how to select useful supplemental colors

[1] `https://www.arcgis.com/index.html`

to support tasks like highlighting. I instead seek to holistically evaluate perceptual and aesthetic factors in highlight color choices. I do so by measuring how well different colors support highlighting data in popular color ramps in an experimental study.

Chapter 3

Conditions and Hypothesis

Color is a frequently used visual channel to highlight data in a visualization. Current understanding of efficient highlight colors stems from prior work where distractor marks mapped to a homogenous color were scattered around a target mark of another color or a subset of marks are focused through contrasting background [49]. Earlier work also quantified that search can be inefficient when a highlight color is collinear in color space with the remaining colors in a visualization [54]. The common-sense approach of using a contrasting color to highlight information (red against a set of blue marks) will work for a set of homogeneous distractors. However, visualizations often have large variance in the colors present in a scene: the diversity of colors reflects the variance in the data. I account for this diversity by measuring how well people can find a given target point in a color-coded scatterplot.

Apart from search time, aesthetic quality is a salient factor in a visualization. Schloss et al. [68] evaluated the role aesthetics play on preference and harmony for color combinations. Visualization designers apply results from former aesthetic studies to create pleasing and effective visualizations. However, designers lacking this knowledge and experience often draw from hunches and personal preferences while creating visualizations. Color combinations have ecological and perceptual factors that affect its aesthetic quality [69, 2, 43]. I tested three factors for aesthetic quality: color combination preference, color harmony and discriminability. In this work, I measure highlight color performance across a broad variety of popular ramps from data gathered with search effectiveness (accuracy and response time) and aesthetic preference (subjective usability, harmony,

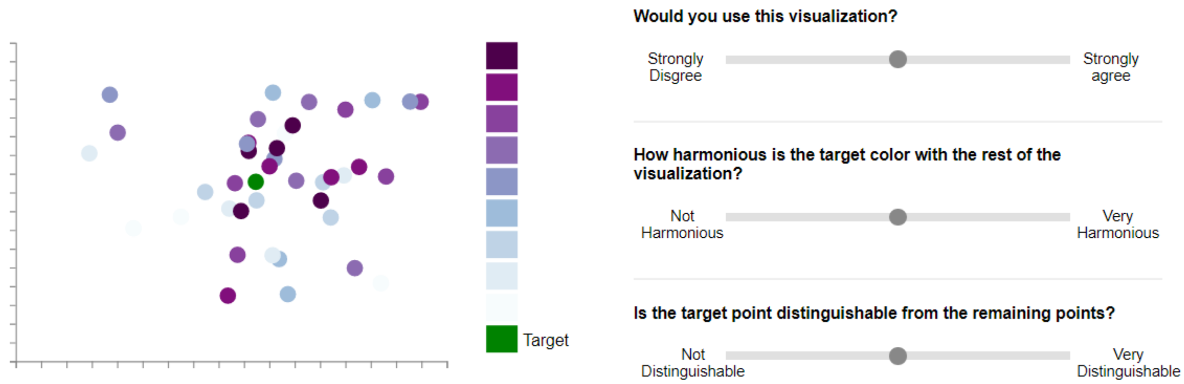


Figure 3.1: Participants clicked on a target mark colored using a given highlight color. Marks were embedded in a scatterplot with 40 distractor marks mapped to random colors in a color ramp. Participants then reported their aesthetic impressions of the scatterplot overall.

and discriminability) as factors in a crowdsourced experiment.

I chose to use Scatterplots with circular marks of identical size for the experiment to avert potential effects of size as prior work quantified that size has an effect on color difference perception [79, 78]. I used a search and click-target method previously used for color comparison experiments [49, 14]. Participants were asked to click on a mark of a given color in a scatterplot and then reported subjective impressions of the scatterplot using a set of sliders (Figure 3.1). To measure functional and aesthetic effects of highlight color choices, I gathered existing sequential color ramps from common sources (ColorBrewer and Tableau[1]) and ran an edge-to-edge sampling of the CIELAB perceptual color space to create a candidate set of highlight colors (Figure 4.4). A detailed explanation of the perceptual sampling of highlight colors is described in section 4.2.2.

Building on prior knowledge from visualization, vision science, and aesthetics, I hypothesized that:

H1–*Colors with high luminance will have higher accuracy, lower response times, and higher subjective discriminability.*

I expect that light colors will be more discriminable than dark colors. This hypothesis stems from earlier work that presents luminance as the best channel for distinguishing color differences

[1] https://www.tableau.com/

[5, 65, 67, 95] and prior studies that demonstrate effects of luminance in highlighting and pop-out [26, 47, 90].

H2–*Low chroma colors will result in low objective performance and low preference scores*

Visualization designers frequently use gray hues for less important data elements or contextual elements such as labels and axes while reserving bright, high chroma colors to highlight important data elements [53]. Grey colors also have lower contrast with the scatterplot's background, likely reducing search performance [4, 41]. However, prior studies suggest that grey tones in colorful visualizations may help emphasize data values [64]. Further, high chroma colors can create neons or other overly bright tones that may reduce aesthetics.

H3–*Reds and yellows will generate low search time and high response accuracy.*

Many visualization systems employ red or yellow as default highlight colors [44, 73, 75]. Prior studies that tested color for search and identification tasks show that red and yellow hues have a low mean response time [9] and facilitate highlighting in monochromatic visualizations [26, 76]. As a result, I expect red and yellow hues to generally have low search times and high accuracy except in cases where red or yellow are part of the ramp itself.

H4–*Ramp and highlight colors with complementary relationships will generate high aesthetic scores and high search performance.*

Complementary colors like blue-orange pairs have high contrast and are generally considered aesthetically harmonious [4, 25, 40, 41]. As a result, design heuristics suggest that such colors may make optimal highlight choices [91]. I anticipate that this theory carries over to visualizations with sequential ramps that tend to contain relatively few hues.

Chapter 4

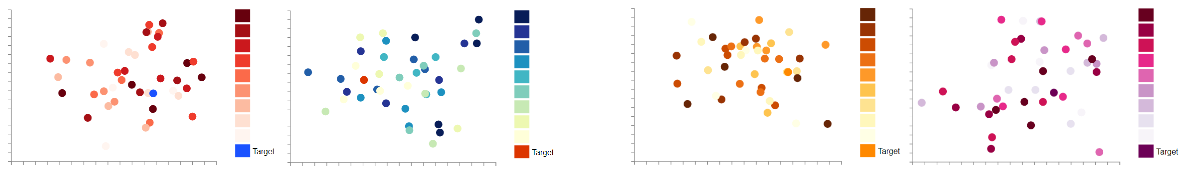
Methods

I conducted a 25 (color ramp, within participants) $\times$ 50 (highlight color, between participants) mixed factors experiment on Amazon's Mechanical Turk to measure the perceptual and aesthetic factors that influence target identification, focusing on one standard visualization type: scatterplots. The study asked participants to find and click on a data point with a particular target color in a series of 78 scatterplots following the search-and-click methodologies from Lindsey et al. [49] and De Vries et al. [14]. I tested two independent variables—color ramp and highlight color—that were counterbalanced across participants and five dependent variables—response accuracy, time to locate the target, subjective preference, subjective discriminability and subjective color harmony—with subjective factors measured using a Likert-type slider ranging from -10 to $+10$.

4.1 Stimuli

I measured perceptual and aesthetic factors using scatterplots rendered using D3 [11]. Scatterplots were rendered on a 375 $\times$ 250 pixel white background using 1-pixel mid-gray axes. A legend mapping the color ramp and highlight color was rendered on a 30 $\times$ 240 pixel white background. The ramp in the legend was ordered from dark to light with the target color at the bottom, following the conventions recommended in Gramazio et al. [24].

Each scatterplot contained 1 target mark and 40 distractor marks. Distractor marks were positioned following a random sampling of a standard normal distribution ($\mu = 0, \sigma = 1$), with x- and y-values between 5 and 160 pixels. Target marks were mapped to a random value between 30



(a) (b)

Figure 4.1: I measure how effectively different highlight colors support visual search and positive aesthetics in color-coded scatterplots. The results provide data-driven recommendations for effective highlight colors with different ramps and scaffold a preliminary model for predicting optimal highlight colors. The four scatterplots above highlight the diverse combinations of color ramp and candidate highlight colors used in the experiment.

and 140 pixels to ensure that targets fell within the data distribution to avoid positional pop-out. Target and distractor marks were mapped to a constant size of 14-pixel diameter circular marks. Figure 4.1 shows examples of the scatterplots used in the study. Conditions and limitations of prior color perception experiments were studied to consider a holistic set of design factors that are generally seen in real-world visualizations. Marks were randomly positioned to avoid confounds from distance comparisons [8] and could overlap one another. This data distribution reflects the common scenario where target points are not situated in an isolated region of a visualization. To avoid occlusion, target marks were always rendered on top of any distractor marks.

Above each scatterplot I provided a brief statement about the task "Locate and click on the target mark in the scatterplot." I also had a statement encouraging the participants to complete the task and answer the questions as quickly and accurately as possible and a counter showing the progress through the study. The scatterplot was rendered on the left side of the screen. When the target mark is located and clicked on, three slider questions to the right of the scatterplot. To minimize bias from slider design, sliders were rendered with the selector at 0 (midscale) and without tick marks [52]. The participants were required to answer three questions regarding the task and the aesthetics of the colors in the scatterplot for each stimulus. Question framing was determined through piloting:

(1) Would you use this visualization?

-10–Strongly Disagree, +10–Strongly Agree

(2) How harmonious is the target color with the rest of the visualization?

-10–Not Harmonious, +10–Very Harmonious

(3) Is the target point distinguishable from the remaining points?

-10–Not Distinguishable,+10–Very Distinguishable

4.2 Data

The density of circular marks may influence the time required in locating the target. To simulate realistic scenarios, I constructed synthetic data. The x and y coordinates used to plot the data points were generated using a standard normal distribution ($\mu = 0, \sigma = 1$), with x- and y-values between 5 and 160 pixels. The range for the target highlight mark was between 30 and 140. This gap between the target and distractor marks controls the target mark from being rendered near the edges of the gamut because if the marks are rendered out of the scatterplot area, it will not be visible to participants thereby affecting the dependent variables.

4.2.1 Color Ramps

The ramps used as stimuli are sequential color ramps that are often used to encode quantitative data. I chose to use 48 sequential ramps designed by ColorBrewer [29] and Tableau, two popular sources of ramps often used by expert and novice visualization designers. The 48 ramps that form the combined Brewer and Tableau corpus include both single-hue and multi-hue ramps. I focused on sequential ramps as they contain a smooth progression across hues, making it easier to assess the algorithmic relationship between the colors in the ramp and the corresponding highlight colors.

Since two sources were used, there was overlap and the final 25 color ramps used in the study were hand-picked by eliminating multiple ramps with similar hues.

As ramps from both sources contained a number of visually similar ramps, I manually assessed the corpus and removed near-duplicates, resulting in 25 total ramps: 17 ramps from ColorBrewer

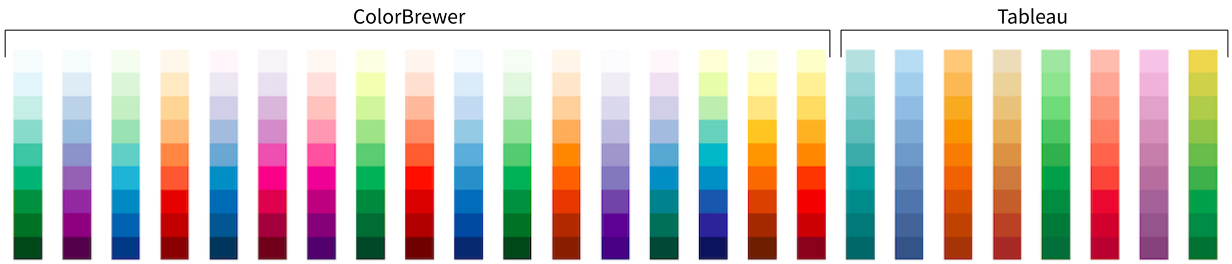


Figure 4.2: The study evaluated highlight colors on 25 nine-color ramps from ColorBrewer and Tableau. I removed 23 near-duplicate ramps from an initial collection of 48 ramps from these tools to generate the final set.

and 8 ramps from Tableau (Figure 4.2). Color ramps range from three to twelve classes of color. Participants mean response time was high in studies for seven and nine color classes [31]. In practice, visualizations are cluttered with data and nine is the upper bound for number of colors in sequential ramps [30]. Therefore, I chose nine color classes for all ramps. Tableau ramps with more or less than nine colors were retargeted by first fitting the colors to a cubic B-Spline and then using arc length interpolation in CIELAB to resample nine equidistant points along the resulting curve. While the colors found in the ramp corpus determine which highlight colors are most effective, the final corpus was assembled to reflect a relatively even distribution of nameable hues to minimize bias from the tested ramps. The final corpus contained 214 unique colors.

4.2.2 Highlight Colors

CIELAB is a color space composed of three key axes: $L*$ (lightness) with a range of 0 to 100, $a*$ (the amount of red or green) with a range of -128 to +128 and $b*$ (the amount of green or yellow) with a range of -128 to +128. In this color space, one unit corresponds to 1 just-noticeable-difference (JND). I used CIELAB space to systematically sample through $L*$, $a*$, and $b*$ axes with 1 step increments. This yielded a pool of 224 unique colors. I employed three approaches to restrict the highlight colors to a reasonable number that can be used in the study.

4.2.2.1 Normalized color difference

Distance between colors is calculated with CIELAB ΔE. I used a normalized formula for $\Delta E_{(p,s)}$ described in [79] to remove the colors that have a $\Delta E_{(p,s)}$ less than 1.

$$\Delta E_{(p,s)} = \sqrt{(\frac{\Delta L}{ND_L(p,s)})^2 + (\frac{\Delta a}{ND_a(p,s)})^2 + (\frac{\Delta b}{ND_b(p,s)})^2} \tag{4.1}$$

The noticeable difference ($ND_L(p,s)$, $ND_a(p,s)$ and $ND_b(p,s)$) formulas were calculated specifically for circular marks with a diameter of 14-pixels. The p-value was set to 0.5. This reduced the preliminary pool to 198 candidate colors.

$$ND_L(p,s) = \frac{p}{0.0937 - \frac{0.0085}{diameter}} \tag{4.2}$$

$$ND_a(p,s) = \frac{p}{0.0775 - \frac{0.0121}{diameter}} \tag{4.3}$$

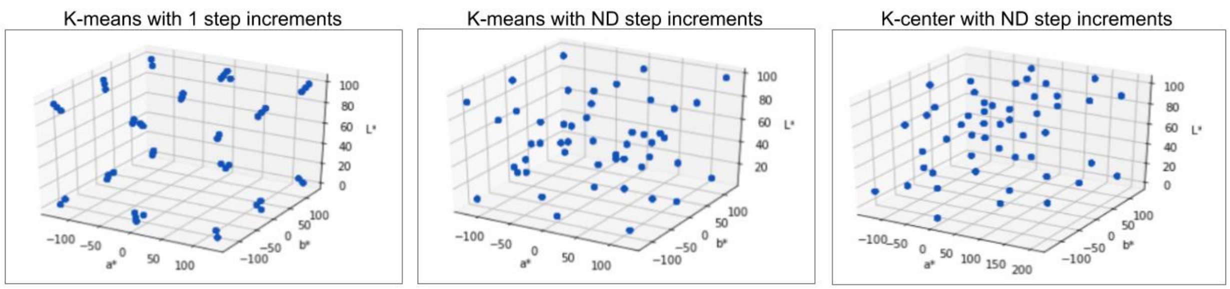
$$ND_b(p,s) = \frac{p}{0.0611 - \frac{0.0096}{diameter}} \tag{4.4}$$

It wasn't feasible to handpick colors from the candidate list of colors because of its size. I chose to shift to an algorithmic approach, specifically to clustering and approximation algorithms, to reduce the number of candidate colors.

4.2.2.2 K-means

K-means is a clustering algorithm that partitions n observations into k clusters. Each observation is put into the cluster with the nearest mean. I used k-means with $k = 50$ to create clusters of similar colors. The means or centers of these clusters create the final set of highlight colors. Applying k-means to the sampled and normalized pool of colors with 1 step increments along the $L*$, $a*$, and $b*$ axes yielded clusters closely packed at the edges of the three axes in the LAB space (Figure 4.3(a)).

The noticeable difference parameters calculated for the normalized color difference have different values for each of the axes. Utilizing the calculated $\mathrm{ND}_L(p,s)$, $\mathrm{ND}_a(p,s)$ and $\mathrm{ND}_b(p,s)$ values for the $L*$, $a*$, and $b*$ axis step increments respectively yielded clusters scattered across the color space. As large spans of CIELAB contain greens and blues while yellows encompass a small volume [38], I examined the algorithmic outputs to inspect for bias in named hues. (Figure 4.3(b)).



(a) (b) (c)

Figure 4.3: Visual representation of the sampled highlight colors the three axes in CIELAB.

4.2.2.3 K-center

K-center is an approximation algorithm. K-center creates a subset of k points to minimize the maximum distance of any point to its closest center [88, 12]. The distance formula I used in creating these clusters is a standard Euclidean distance. This resulted in a more balanced distribution across nameable colors and used the resulting central colors for the preliminary candidate set.

The dense collection of points near the corner indicates varying hues at the color regions with lesser area when compared to the dominant blue and the green hues (Figure 4.3(c)). I further adjusted for oversampling by replacing eight blue and green colors visually similar to others in the sample with additional red and yellow shades. This oversampling of reds and yellows allowed me to adjust the corpus to reflect the heavy reliance of conventional systems on red and yellow to highlight data. Two hues of gray were also added for [**H2**]. Figure 4.4 contains the final set of candidate colors.

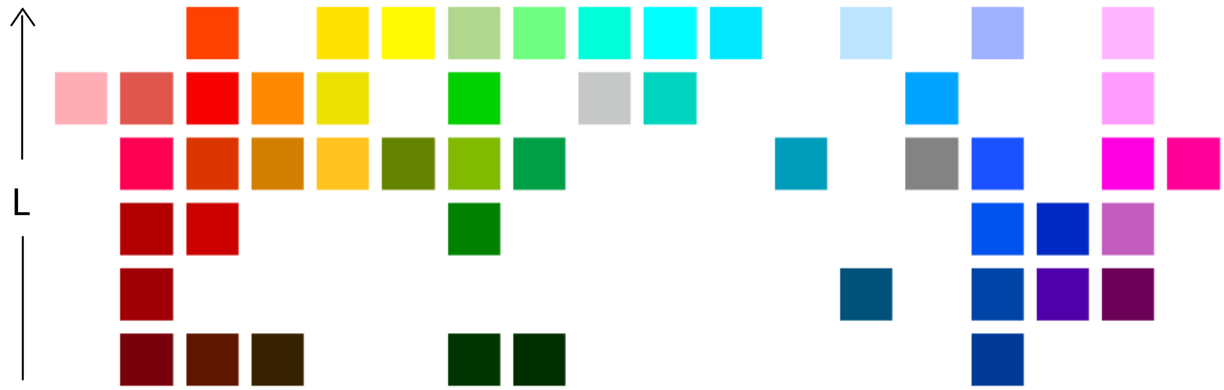


Figure 4.4: I selected 50 candidate highlight colors using k-centers applied to a uniform sampling of CIELAB, manually adjusting the colors to oversample reds and yellows given their frequent use as highlight colors in visualization systems.

4.3 Procedure

The study consisted of six phases: (1) Consent, (2) Color Vision Screening, (3) Tutorial, (4) Aesthetic Reference, (5) Formal Study, (6) Demographic Questionnaire

First, each participant provided informed consent in accordance with the IRB protocol to participate in the study. Participants were then screened for color vision deficiencies using four Ishihara plates. Piloting successfully caught color vision deficiencies although variations in color representation across displays limit the effectiveness of an Ishihara test. To account for issues the screening may have missed, I additionally asked participants to self-report any CVD as part of the demographic questionnaire. Data from participants failing the screening or self-reporting CVD were excluded from our analysis. A tutorial session followed the successful completion of the vision screening where participants received instructions about the task they will be performing in the study. The tutorial had five scatterplots presented serially, each using a black-grey sequential ramp and brightly colored targets. The aesthetic questions were not included in the tutorial. Participants had to correctly locate and click on the target mark in each scatterplot before moving onto the next phase of the study.

A reference page preceded the formal study that presented all of the color combinations the participant would see during the study. Following best practices for aesthetic evaluation [60], this page provided participants with an aesthetic reference to anchor judgments for the subjective Likert-type questions. After viewing the reference, participants advanced to the formal study. The formal study contained 78 trials presented serially (75 test stimuli and 3 engagement checks). The stimuli were rendered in a randomized order to mitigate learning or fatigue effects. Each participant saw three combinations of one color ramp and a highlight color. Each combination occurred at most once per participant. Participants clicked a button to begin the formal trials, allowing them to read the instructions on the study page. Each stimulus was rendered then revealed for up to 15 seconds to ensure focused participation and to control for variation in response time due to multitasking [70]. After participants clicked on a mark or time ran out, the scatterplot remained visible but no further click input was allowed. Subjective questions then appeared to the right of the scatterplot with accompanying sliders (Figure 4.1). Once participants completed these questions, they clicked a button to move to the next trial. After each trial, a gray box covered the scatterplot and legend to reduce potential contrast effects between trials. To ensure honest participation three engagement checks were constructed with the greys ramp used in the tutorial. As these trials had high contrast between target and distractor marks and relatively easy to locate targets, I excluded the data from participants who failed to answer the engagement check trials correctly.

A demographic questionnaire followed the formal study. This included an opportunity for participants to include optional comments and upon completion of the questionnaire, they were compensated for their participation.

4.4 Analysis

Equal numbers of samples were collected for all combinations of the independent variables. The independent variables were counterbalanced between participants and each participant saw each of the 25 tested color ramps three times, each time using a different highlight color. I analyzed the effects of a color ramp and highlight color (decomposed by color axis in CIELCh) across five dependent measures (accuracy, response time, subjective preference, subjective discriminability, and subjective color harmony) using four-way ANCOVAs with interparticipant variation and question order as random covariates

4.5 Participants

I recruited participants using Amazon's Mechanical Turk, a web-based crowdsourcing platform, as it provides a large pool of participants for large-scale studies and allowed me to measure visualization performance in realistic contexts. Prior studies have shown that models of graphical and color perception in visualization can be reliably constructed using crowdsourcing [13, 36, 50, 57, 71, 79]. In validation studies, crowdsourced models demonstrate a high degree of accuracy in predicting color discrimination compared to traditional laboratory metrics applied to web-based designs [82]. I included all ramps within-subjects to mitigate per-ramp biases and randomized highlight color selection across participants to randomly distribute viewing variance across the dataset. These conditions increase the ecological validity of the results as visualizations are rarely viewed in perfect conditions.

Chapter 5

Experiment

Scatterplots of varying shapes have been used in past studies to test the effects of highlight color against distractor marks. I chose to use circular marks in the experiment as color science models have used diagonally symmetric marks to encode data. This also mitigates possible effects of size on color which have proven to happen with bar charts and line graphs.

Response accuracy, search time and perceived aesthetics are three salient factors that influence the effectiveness of a highlight color against other data point colors in a visualization. I tested for these factors by performing an edge to edge sampling of the color space in CIELAB to extract highlight colors that have even characteristics of the three aspects of color: luminance, saturation and hue. Although CIELAB space was used for sampling I used the color axes $L*$, $c*$ and $h*$ in CIELCH for analysis because, in practice, color preference and harmony scores are rated against lightness, chroma and hue.

5.1 Results

I collected data from 302 participants from the United States with an approval rating of 95% or greater. 37 participants were excluded from analysis for self-reporting a color vision deficiency or abnormal vision and 17 for failing an engagement check, resulting in 248 participants total ($\mu_{age} = 36.3$, $\sigma_{age} = 10.36$, 111 female, 136 male, 1 did not report) and 18,600 total trials in my analysis.

5.1.1 Analysis of Factors

I analyzed the effects of a color ramp on a highlight color across experimental factors(color axes and color ramp) using a four-level two factorial ANCOVA with interparticipant variation and question order as random co-variates. Post-hoc analysis used Tukey's Honest Significant Different Test (HSD, $\alpha = .05$).

I analyzed the effects of color ramp and highlight color across experimental factors (highlight color L^*, C^*, and h^* and color ramp) using a four-way two-factorial ANCOVA with interparticipant variation and question order as random covariates. Post-hoc analysis used Tukey's Honest Significant Different Test (HSD, $\alpha = .05$) with Bonferroni correction. Tables 5.1, 5.2, 5.3, and 5.4 summarize the significant results for all dependent variables.

For transparency into the ANCOVA results [16] I discuss the overall results and provide descriptive statistics (means and bootstrapped 95% CIs) for relevant results. Figures 5.1(a) and 5.1(b) summarize the objective measures. Figures 5.2(a), 5.2(b) and 5.2(c) summarize subjective preference, harmony and discriminability.

Table 5.1: Summary of results for response time and accuracy (grey indicates non-significant results)

Factors	**Response Time (ms)**	**Response Accuracy**
Ramp	$F(24, 223) = 8.65$, $p < .0001$	$F(24, 223) = 15.65$, $p < .0001$
L	$F(1, 246) = 191.47$, $p < .0001$	$F(1, 246) = 373.77$, $p < .0001$
C	$F(1, 246) = 102.80$, $p < .0001$	$F(1, 246) = 11.6$, $p = .0007$
h	$F(1, 246) = 57.25$, $p < .0001$	$F(1, 246) = 95.43$, $p < .0001$
Ramp*L	$F(24, 223) = 4.46$, $p < .0001$	$F(24, 223) = 10.88$, $p < .0001$
Ramp*C	$F(24, 223) = 9.94$, $p < .0001$	$F(24, 223) = 8.66$, $p < .0001$
Ramp*h	$F(24, 223) = 32.53$, $p < .0001$	$F(24, 223) = 49.08$, $p < .0001$
L*C	$F(1, 246) = 12.52$, $p < .0005$	$F(1, 246) = 32.29$, $p < .0001$
L*h	$F(1, 246) = 5.81$, $p < .19$	., .
C*h	$F(1, 246) = 5.81$, $p < .0001$	

Table 5.2: Summary of results for preference rating (grey indicates non-significant results)

Factors	**Preference**
Color Ramp	$F(24, 223) = 6.66$, $p < .0001$
L	$F(1, 246) = 125.16$, $p < .0001$
C	$F(1, 246) = 67.75$, $p < .0001$
h	$F(1, 246) = 25.42$, $p < .0001$
Color Ramp*L	$F(24, 223) = 3.29$, $p < .0001$
Color Ramp*C	$F(24, 223) = 5.16$, $p < .0001$
Color Ramp*h	$F(24, 223) = 15.15$, $p < .0001$
L*C	$F(1, 246) = 0.003$, $p = 0.9518$
L*h	$F(1, 246) = 0.026$, $p = 0.8703$
C*h	$F(1, 246) = 0.888$, $p = 0.346$

Table 5.3: Summary of results for harmony rating (grey indicates non-significant results)

Factors	**Harmony**
Color Ramp	$F(24, 223) = 2.72$, $p < .0001$
L	$F(1, 246) = 87.73$, $p < .0001$
C	$F(1, 246) = 165.32$, $p < .0001$
h	$F(1, 246) = 18.87$, $p < .0001$
Color Ramp*L	$F(24, 223) = 5.47$, $p < .0001$
Color Ramp*C	$F(24, 223) = 10.04$, $p < .0001$
Color Ramp*h	$F(24, 223) = 61.47$, $p < .0001$
L*C	$F(1, 246) = 0.15$, $p = 0.6984$
L*h	$F(1, 246) = 6.789$, $p = 0.0092$
C*h	$F(1, 246) = 1.098$, $p = 0.2947$

Table 5.4: Summary of results for discriminability rating (grey indicates non-significant results)

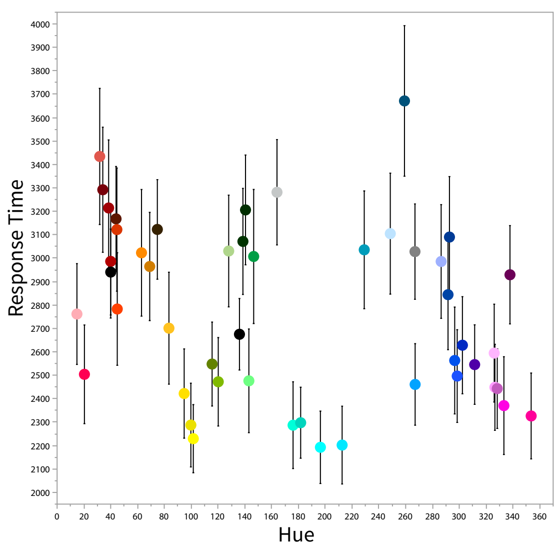
Factors	**Discriminability**
Color Ramp	$F(24, 223) = 12.65$, $p < .0001$
L	$F(1, 246) = 611.83$, $p < .0001$
C	$F(1, 246) = 398.96$, $p < .0001$
h	$F(1, 246) = 110.31$, $p < .0001$
Color Ramp*L	$F(24, 223) = 9.72$, $p < .0001$
Color Ramp*C	$F(24, 223) = 22.7$, $p < .0001$
Color Ramp*h	$F(24, 223) = 104.53$, $p < .0001$
L*C	$F(1, 246) = 0.104$, $p = 0.7468$
L*h	$F(1, 246) = 1.269$, $p = 0.259$
C*h	$F(1, 246) = 10.55$, $p = 0.0012$

5.1.2 H1: Effects of Luminance

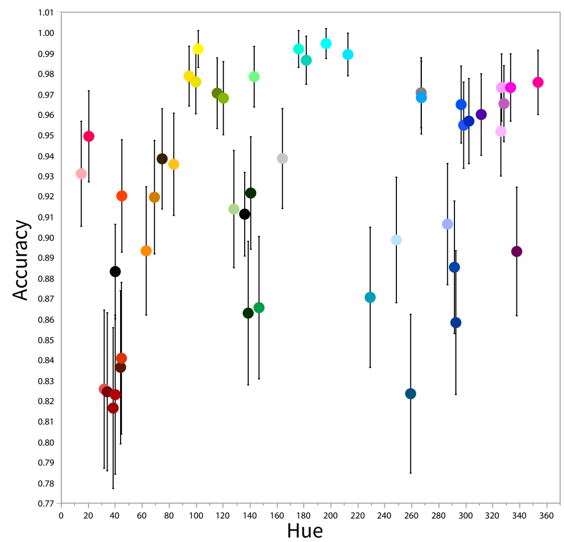
Objective Results: High luminance colors led to faster and more accurate target identification overall. People identified highlight colors with high luminance ($L^* > 82$, $\mu_{accuracy} = 97\% \pm 0.54\%$) significantly more accurately than colors with low luminance ($L^* < 33$, $\mu_{accuracy} = 88\% \pm 0.98\%$). I found that response accuracy increased with L^* and had significantly less variance as colors grew lighter, supporting **H1**.

I found similar patterns for response time: darker colors took on average 0.6 seconds longer to find than lighter colors ($\mu_{low\_L\_rt} = 3.04s \pm 0.07s$ vs. $\mu_{high\_L\_rt} = 2.41s \pm 0.06s$). Response times decreased on average as luminance increased.

Subjective Results: People generally preferred high luminance highlight colors to low luminance colors, finding the visualizations more usable ($\mu_{high\_L} = 2.76 \pm 0.21$ vs. $\mu_{low\_L} = 1.08 \pm 0.21$) and colors more discriminable ($\mu_{high\_L} = 5.77 \pm 0.18$ vs. $\mu_{low\_L} = 2.58 \pm 0.22$) when high luminance colors were used. The lightest colors led to the highest preference ($L^* = 91.11$, $\mu_{preference} = 3.92 \pm 0.59$) . However, participants also found high luminance colors less harmonious on average ($\mu_{high\_L} = -0.2 \pm 0.23$ vs. $\mu_{low\_L} = 1.33 \pm 0.21$).



(a)



(b)

Figure 5.1: Mean (a) response time (ms) and (b) accuracy for each of the 50 tested highlight colors sorted by hue (error bars represent 95% confidence intervals). I found that lighter and more saturate colors tended to lead to high accuracy and low response times.

5.1.3 H2: Effects of Chroma

Objective Results: High chroma colors tend to correspond with bright, eye-catching marks. I found that high chroma colors ($C^* > 89$) were on average faster ($\mu_{rt} = 2.5s \pm 0.06s$) and easier to find ($\mu_{accuracy} = 96\% \pm 0.6\%$) than low chroma colors ($C^* < 43$, $\mu_{rt} = 3.09s \pm 0.07s$, $\mu_{accuracy} = 89\% \pm 0.93\%$). Dark reds, greens and blues were exceptions to this pattern, generating lower overall response accuracy regardless of their chroma ($C^* : 62 - 72$, $\mu_{accuracy} = 88\% \pm 1.86\%$).

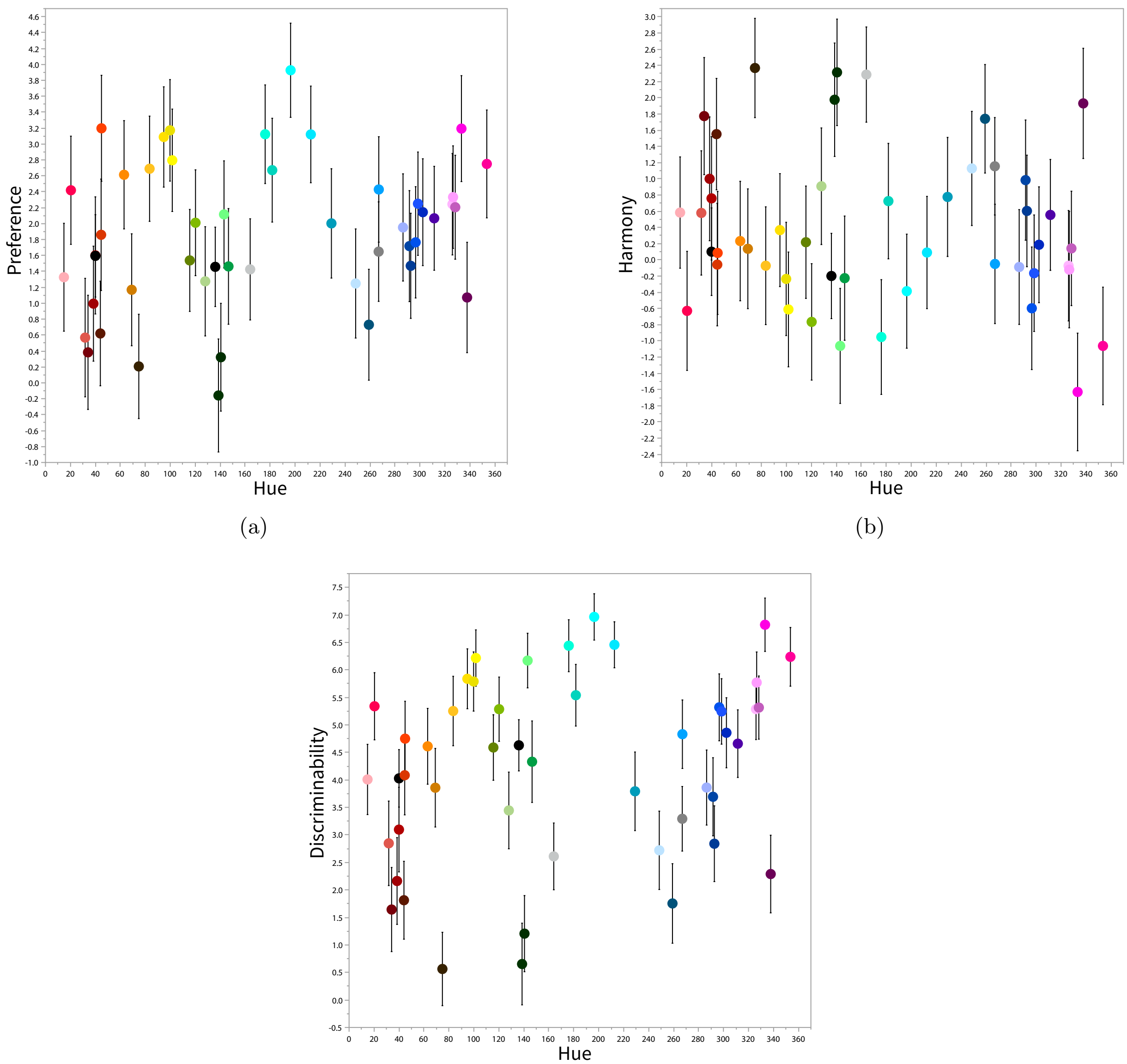
Subjective Results: High chroma colors received high preference ($\mu_{preference} = 2.38 \pm 0.22$) and discriminability scores ($\mu_{discriminability} = 5.31 \pm 0.21$) whereas colors with low chroma had low preference ($\mu_{preference} = 1.03 \pm 0.21$) and discriminability ($\mu_{discriminability} = 2.55 \pm 0.21$). However, low chroma colors received higher overall harmony scores ($\mu_{harmony} = 1.22 \pm 0.21$). This difference is likely due to brighter colors being more disruptive of the overall effect of the figure. Further exploring this hypothesis is useful future work.

5.1.4 H3: Reds & Yellows

Objective Results: As reds and yellows are common **de facto** color choices for highlighting data, I anticipated they would lead to optimal overall performance. While people could quickly and reliably find yellows ($\mu_{accuracy} = 97\% \pm 0.83\%$, $\mu_{rt} = 2.3s \pm 0.09s$), reds led to lower search accuracy and slower searches ($\mu_{accuracy} = 86\% \pm 1.05\%$, $\mu_{rt} = 3.04s \pm 0.07s$). I found that yellow, cyan, olive, royal blue, and magenta hues all led to high accuracy and fast search overall, whereas reds, purples, and muted blues tended to provide lower accuracy and slower searches.

The results also indicate that the optimal highlight color depended on the colors present in the tested ramp. I discuss this finding in more detail section 5.1.6; however, preliminary results suggest that selecting a **de facto** color can lead to suboptimal highlight performance.

Subjective Results: People preferred yellow hues ($\mu_{yellows} = 2.94 \pm 0.32$) over red hues ($\mu_{reds} = 1.48 \pm 0.21$) and found visualizations using yellows more discriminable ($\mu_{yellows} = 5.79 \pm 0.28$ vs. $\mu_{reds} = 3.35 \pm 0.22$). However, participants generally found yellow hues less harmonious than red



(a) (b)

(c)

Figure 5.2: Mean (a) subjective preference, (b) subjective harmony and (c) subjective discriminability for 50 tested highlight colors ordered by hue (error bars represent 95% confidence intervals). I found that lighter and more saturate colors led to higher preference and discriminability scores, but lower perceived harmony.

hues ($\mu_{yellows} = -0.14 \pm 0.34$ vs. $\mu_{reds} = 0.56 \pm 0.22$), in line with prior studies of the negative aesthetics of yellowish tones [69].

5.1.5 H4: Complementarity

To measure the complementary relationship between a ramp and highlight color I selected the top-performing optimal highlight color across both objective and subjective measures (more details in section 5.3) for every ramp. I then analyzed the hue angles between every color in the ramp against the optimal highlight color. Ramps containing hues between 160° and 200° from the highlight color were considered complementary. I found that 10 of the 25 tested ramps had optimal highlight colors that were complementary to a color in the ramp. This result indicates that while complementarity can provide a useful highlight color, the heuristic is not sufficient to predict overall performance. Figure 5.3 shows the ramps from the study corpus with complementary highlight colors.

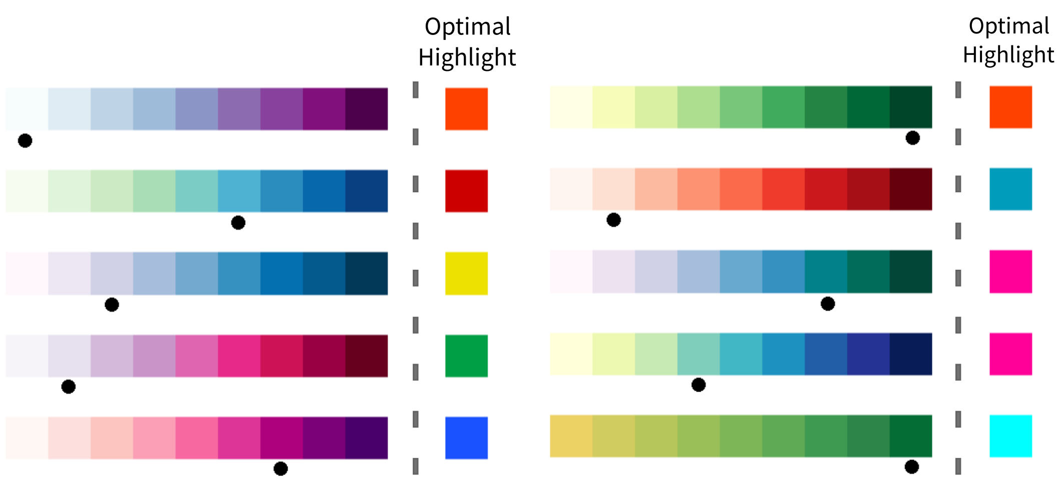


Figure 5.3: I measured the complementarity of highlight colors against all hues in each ramp. The ten ramps above had optimal highlight colors within 20° of a complementary relationship (best hue indicated by a black dot). The remaining 15 did not.

5.1.6 Ramp Dependence

While the patterns above generally held across all tested ramps, I also found that the ramp used in a scatterplot was a significant factor for all of the dependent measures and significantly interacted with all other factors. Taken collectively, the data shows that the optimal highlight color for a given visualization is a function both of the aspects of the color (e.g., high luminance, high chroma, etc.) and the ramp used.

For example, the ColorBrewer sequential green-blue ramp (figure 5.4(a)) experienced a predictable drop in accuracy for blue highlight colors (figure 5.4(c)). Mean response times for this ramp were 1 second slower on average for green highlights and 4 seconds slower for blue highlights compared to all other hues (figure 5.4(b)). Preference and discriminability ratings averaged near zero for green hues and were negative for blue hues despite having the highest overall harmony ratings.

5.2 Synthesis of Results

The results provide preliminary support for all the hypotheses:

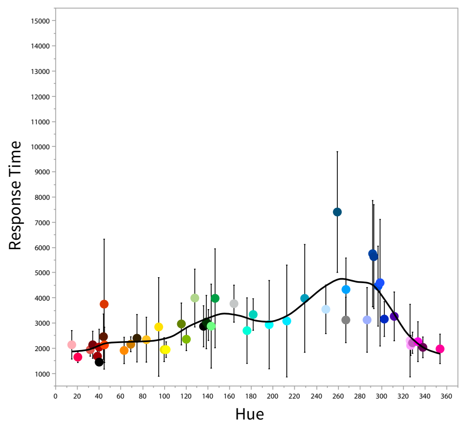
H1: As the luminance is increased, response time decreased and preference, response accuracy increased.

H2:(partial) Colors with low saturation generated a mixture of low and mid preference scores. Colors with very little saturation (chroma: 0 - 42) generated the lowest preference scores.

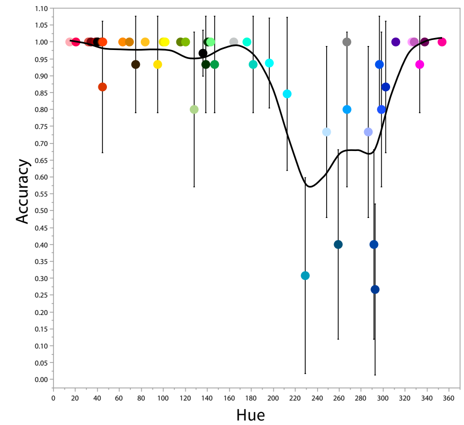
H3:(partial) Yellow hues had low response time. The results failed to support the same for red hues.

H4: Ramps with optimal highlight colors that were complementary to a color in the ramp generated high preference and discriminability ratings.

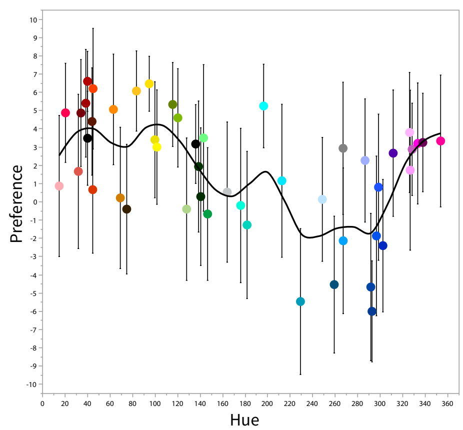
(a)



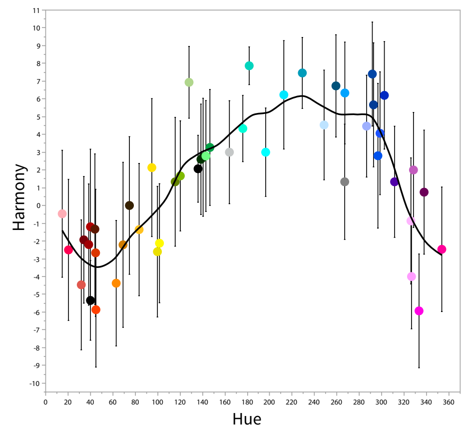
(b)



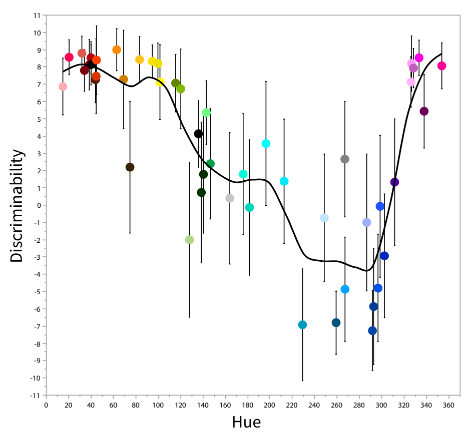
(c)



(d)



(e)



(f)

Figure 5.4: The averaged results for all dependent variables along the *hue* axis for ramp in (a). Colors with hues perceptually similar to the colors on the ramp generate high response time and harmony ratings while producing low response accuracy, preference and discriminability ratings.

5.3 Empirical Modeling

While the best performing color for each ramp tended to be sufficiently distinct from the ramp colors (Figure 5.5), I found no patterns between optimal highlight colors and ramp colors that suggested an actionable heuristic. The data can instead inform data-driven highlight color selection for ramps similar to those in the study corpus, which constitute a large proportion of current visualizations due to the popularity of ColorBrewer and Tableau, and other smoothly-varying sequential ramps.

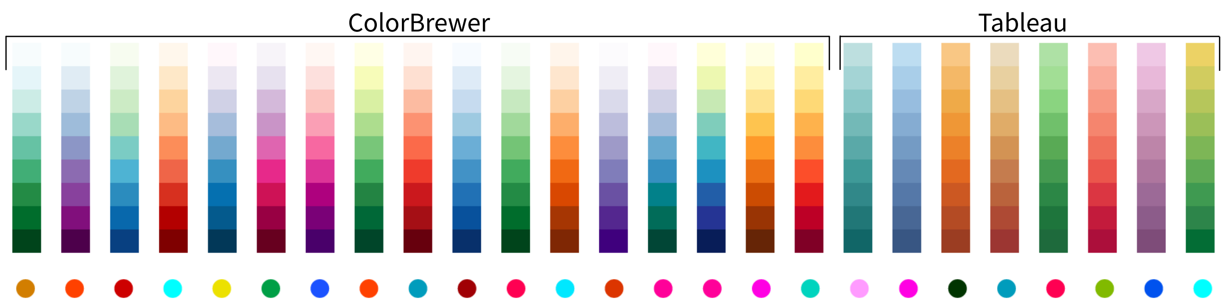


Figure 5.5: The optimal highlight colors assuming a uniformly weighted sum of all dependent metrics are indicated below each ramp used in the study.

Prior studies have used data from crowdsourced experiments to construct statistical models for color and graphical perception [79, 71]. Similar steps can be used with the data from this study to construct empirical models to automate highlight color selection. I achieve this goal by providing statistical models of the experiment data that recommend effective highlight colors.

I computed a weighted sum of the mean objective and subjective scores over the 18,600 collected samples to create a **performance score**:

$$score = w_{rt}\mu_{rt} + w_{acc}\mu_{acc} + w_{pref}\mu_{pref} + w_{har}\mu_{har} + w_{disc}\mu_{disc}$$

Designers can tune the weights w_i to emphasize specific desired properties of the resulting highlight color. The recommended highlight colors for a given ramp are determined by computing the reweighted performance score across the provided samples. Figure 4.2 shows top colors for $w_i = 0.2$.

I used these performance scores to isolate the top three highlight colors to train three Support Vector Regression models with linear kernels: one for each channel of the highlight color (L^*, a^*, and b^*). Each regression model uses the CIELAB coordinates of a ramp to generate a corresponding color coordinate for a target highlight color. I use the top three highlight colors per ramp to train the model to account for variance introduced by the empirical data and avoid overfitting, allowing for a more representative sample of the ideal luminance, chroma, and hue for a given ramp configuration.

Large hue shifts between nameable colors are not well captured by CIELAB distances, making standard error measures generally used in regression analysis difficult to assess for this data [38]. As the aim is to inform design, I instead rely on visual assessment for this work, leaving a formal validation of the regression models to future work. Figure 5.6 shows the top three colors for five ramps using uniform weights ($w_i = .2$) alongside the predicted highlight colors from the regression models. Figure 5.7 shows the results of this model applied to four ramps from outside of the study corpus. Visual inspection of these results suggests that the model generally predicts colors close to those recommended by the data, a pattern shared across the corpus. However, a detailed evaluation of the predictive model on a broader corpus of ramps is important future work.

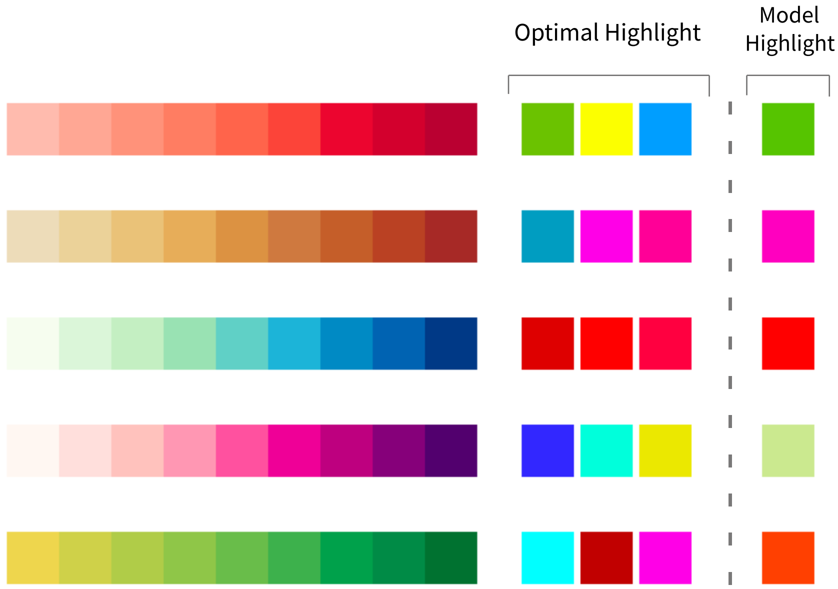


Figure 5.6: Optimal and predicted highlight colors based on the weighted sum of objective and subjective scores.

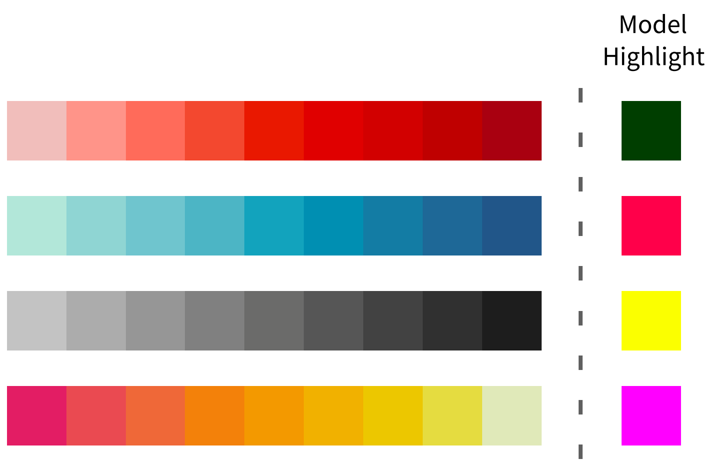


Figure 5.7: Predicted highlight colors from the models for untrained ramps from Tableau and R.

Chapter 6

Discussion

6.1 Discussion

I measured the effect of highlight color in visualizations on search time, response accuracy and perceived aesthetics across 50 highlight colors applied to 25 ramps in color-coded scatterplots. The results show that:

- High luminance and high chroma colors lead to better performance and higher subjective preference.
- Yellows are more effective highlight colors than reds, but no single color works well for all ramps.
- Complementary hues to those of the encoding can provide strong highlight colors, but these colors are not always the optimal choices.

Highlight colors should help people quickly and precisely locate target data. However, simply selecting colors that are easy to find or that "pop-out" is insufficient in many applications: aesthetics are key to creating engaging visualizations [20, 59] and can also increase factors like perceived usability [85]. The results show a correlation between objective performance and subjective aesthetics in a directed search task. However, this correlation is imperfect. For example, darker colors often had higher perceived harmony despite lower objective performance. Measuring highlight efficiency as a function of both factors allows designers to weight these factors based on their target needs to generate informed highlight color selection.

Studies in psychology [83, 97] and visualization [90] show how high luminance can make marks stand out (**H1**). The results support these conclusions: objective performance, subjective preference, and discriminability all increased with luminance. Perceived harmony did not show the same correlation; however, I anticipate that careful manual crafting can lead to high perceived harmony for light colors, as in Samsel et al. [67]. A broader exploration of aesthetic factors in highlighting, including the role of interaction in framing these aesthetic judgments, is important future work.

One proposed highlighting strategy is to use muted colors that stand out against an otherwise colorful visualization [64]. The study suggests this design recommendation may not be ideal in practice: grey and low-chroma colors were generally considered harmonious but led to lower performance and preference, partially supporting **H2**. However, chroma effects were overridden by luminance: darker colors led to lower performance regardless of chroma. This finding suggests that luminance is likely more critical, but further study should disentangle the effects of chroma and luminance, especially in cases where the two may not be readily separable [15].

Many visualization systems default to yellows or reds as highlight colors. While yellows and reds are often both high luminance and high chroma, I found limited support for this approach: yellows generally performed well, but reds led to moderate performance and preference, partially supporting **H3**. In contrast to prior studies, I found that desaturated reds did not offer significant performance improvements over other muted colors. [49].

The candidate colors in the ramp corpus influenced highlight color performance: colors close to ramp colors tended to perform poorly. While there were slightly more reddish tones than yellows in the tested ramps, these ramps represent a sampling of popular sequential ramps used in data visualization. The high performance of shades of blue and the interaction between highlight colors and ramp types suggests that a fixed-color strategy is suboptimal overall. These results also suggest new questions for prior multichannel highlight studies by showing how the default reds and yellows may not reflect optimal color choices for highlighting.

A better strategy based on the data is to tailor color selection based on the encoding ramp.

One method for doing this is selecting colors that run complementary to hues in the encoding. Ten of the 25 tested ramps had optimal highlight colors complementary to a ramp hue. However, the position of that hue in the ramp varied significantly as did the ramps for which complementarity was an optimal strategy. This variance suggests that heuristic approaches may be insufficient to choose optimal highlight colors without substantial design expertise.

The results offer a data-driven alternative to heuristic color selection. By computing a weighted performance score from the provided data, we can determine which colors on average perform best for a given ramp. My selection of popular ramp sources means that the results generalize to many existing systems. However, the set of available ramps extends beyond the tested corpus and new techniques offer increasing means for people to create their own color ramps [5, 55, 72]. This thesis offers a preliminary model for predicting highlight colors given a set of weights and a target ramp; however, the model is built on predictions from only a subset of possible ramps and emphasizes sequential encodings. I anticipate that more data will improve the quality of the recommendations and that additional models predicting factors like performance scores can offer insight into selecting multiple highlight colors or colors for diverging encodings.

Chapter 7

Conclusion

In this work, I measure the efficiency of different highlight colors in color-coded visualizations across both objective performance and subjective aesthetic metrics. I conducted a crowdsourced study with 302 participants to determine effective highlight colors for 25 popular sequential ramps. The results enable data-driven recommendations for effective highlight colors through preference scores and predictive models. This work offers the first empirical insight into selecting effective highlight colors for color-coded visualizations.

7.1 Limitations

This study provides preliminary insight into how to select effective highlight colors in color-coded visualizations. While I anticipate the results offer actionable insight to support effect color selection, there are several considerations that should be made when using these results.

Foe example, I utilized a simple and familiar search task with a synthetic dataset to fit the purpose of the experiment but this is not usually the case in real-world visualizations. I use scatterplots with uniformly sized points as the stimuli. Factors like mark proximity [8, 56], size [79], shape [71], and spatial frequency [62] all affect color perception in visualizations. Additionally, distractor and target mark sizes were at a constant 14-pixel diameter. These constraints may lead to unreliable results for some visualization factors.

Designers frequently use sequential ramps to encode only continuous data and divergent or categorical ramps are used to encode other types of data. The color ramps in this study were limited to sequential ramps as they vary smoothly in hue and lightness and do not include divergent or categorical ramps. While the process I used in this study may work for divergent ramps (two sequential ramps attached together) in some cases, the effects found are unlikely to translate well for categorical encodings because categorical ramps don't have a continuous structure and generally contain a broad diversity of hues compared to the small number found in sequential or diverging ramps. The methods employed in this study offer a starting point to better understand categorical highlighting.

I used crowdsourcing to recruit participants as it allowed me to collect data across 1,250 combinations of ramp and highlight colors under realistic viewing conditions. While I followed best practices for conducting crowdsourced studies, the results cannot be used to gain insight into precise workings of the visual system. For example, we cannot use this data to make definitive conclusions about color pop-out as the rendered colors may vary slightly between participants and our task, while reflective of visualizations, does not afford the temporal fidelity necessary to assess search mechanisms. I anticipate the results reflect a large subset of current design practices and offer a useful starting point for improving highlighting in visualization.

7.2 Future Work

The limitations mentioned above allowed for substantial control over testing conditions in the study to foster a general understanding of experimental factors. Extending this work for future testing that involves possible solutions for these limitations may provide further understanding of the impact of the derived metrics.

In practice, visualizations use real-world data. While synthetically generated data emulates real-world data, it is not the same. It will be worthwhile to run the study conditions with real-world data and test if the resulting data derives similar guidelines.

Only one visualization type was tested in this study: scatterplots. Scatterplots are amongst the most popular methods for visualization data; however, other visualization types like bar charts and choropleth maps, interactive visualizations. Since gradient colors are extensively used by cartographers, other highlight techniques like color outlining should be evaluated against the study results.

This study only evaluated sequential ramps from two popular sources, Tableau and ColorBrewer. With the increase in color ramp and palette generation tools and models, the data-driven recommendations and derived model can be extended to sequential ramps, both designer-crafted and automated ramps, from sources not used in this experiment. Diverging ramps and categorical color palettes, as well as sequential ramps with large, irregular hue variations, offer new challenges for selecting highlight colors. I anticipate the results could help triangulate highlight colors in diverging ramps by computing performance scores over both halves of a ramp. However, categorical palettes do not exhibit smooth structures and often introduce larger-scale hue variations that may make distinguishing highlight colors from data categories difficult. Future work is needed to understand highlight color selection in these contexts.

The experiment was crowdsourced on a web-platform and allowed for diverse display and viewing conditions. However, as the aim is to inform design, I anticipate that the use of crowdsourcing offers greater benefits in terms of a robust and thorough assessment of the space of both viewing environments and design choices. Future work could use the results to design a smaller-scale in-person study to measure color pop-out in visualizations.

The model also uses data from a subset of possible ramps. As a result, it reflects a selection of popular ramps but does not cover all possible designs. Future work could build on these studies by evaluating a larger set of ramps and new designs as they become available. This preliminary model facilitates extensions that further improve the concrete recommendations for tested ramps and our preliminary predictive model.

Bibliography

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467, 2016.

[2] Lyn Bartram, Abhisekh Patra, and Maureen Stone. Affective color in visualization. In Proceedings of the 2017 CHI conference on human factors in computing systems, pages 1364–1374. ACM, 2017.

[3] Richard A Becker and William S Cleveland. Brushing scatterplots. Technometrics, 29(2):127–142, 1987.

[4] Stefanie I. Becker, Christian Valuch, and Ulrich Ansorge. Color priming in pop-out search depends on the relative color of the target. Frontiers in Psychology, 5(APR):1–11, 2014.

[5] Lawrence D Bergman, Bernice E Rogowitz, and Lloyd A Treinish. A rule-based tool for assisting colormap selection. In Proceedings Visualization'95, pages 118–125. IEEE, 1995.

[6] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. Beyond memorability: Visualization recognition and recall. IEEE Transactions on Visualization and Computer Graphics, 22(1):519–528, 2015.

[7] Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. What makes a visualization memorable? IEEE Transactions on Visualization and Computer Graphics, 19(12):2306–2315, 2013.

[8] Alžběta Brychtová and Arzu Çöltekin. The effect of spatial distance on the discriminability of colors in maps. Cartography and Geographic Information Science, 44(3):229–245, 2017.

[9] Vanessa L. Buechner, Markus A. Maier, Stephanie Lichtenfeld, and Sascha Schwarz. Red - Take a closer look. PLoS ONE, 9(9), 2014.

[10] Roxana Bujack, Terece L Turton, Francesca Samsel, Colin Ware, David H Rogers, and James Ahrens. The good, the bad, and the ugly: A theoretical framework for the assessment of continuous colormaps. IEEE Transactions on Visualization and Computer Graphics, 24(1):923–933, 2017.

[11] Sheelagh Carpendale, Qrzohgjh Dughq, Fhoohqwo Revhuyhg Dqvzhuhg, Dqglgh Exw, O H W Xv, Fxowlydwh Rxu, Bernd Meyer, Michael Bostock, Vadim Ogievetsky, Jeffrey Heer, Pat Hanrahan, a Dissertation Presented, Curran Kelleher, Benjamin Jotham Fry, Stéphane Conversy, Qi Huang, and Ken Birman. A self orgnaizing multiple view representation of 3d objects edelman and weinshall.pdf. Onward!, Ph. D(September):170, 2014.

[12] Varin Chouvatut, Wattana Jindaluang, Ekkarat Boonchieng, and Thapanapong Rukkanchanunt. Efficiency comparisons between k-centers and k-means algorithms. ICSEC 2015 - 19th International Computer Science and Engineering Conference: Hybrid Cloud Computing: A New Approach for Big Data Era, pages 1–6, 2016.

[13] Michael Correll, Dominik Moritz, and Jeffrey Heer. Value-suppressing uncertainty palettes. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, page 642. ACM, 2018.

[14] Jelmer P. De Vries, Ignace T C Hooge, Alexander H. Wertheim, and Frans A J Verstraten. Background, an important factor in visual search. Vision Research, 86:128–138, 2013.

[15] Robert L Donofrio. The helmholtz-kohlrausch effect. Journal of the Society for Information Display, 19(10):658–664, 2011.

[16] Pierre Dragicevic. Fair statistical communication in hci. In Modern Statistical Methods for HCI, pages 291–330. Springer, 2016.

[17] Michael D'Zmura. Color in visual search. Vision Research, 31(6):951–966, 1991.

[18] Howard E Egeth, Robert A Virzi, and Hadley Garbart. Searching for conjunctively defined targets. Journal of Experimental Psychology: Human Perception and Performance, 10(1):32, 1984.

[19] Niklas Elmqvist, Pierre Dragicevic, and Jean-Daniel Fekete. Color lens: Adaptive color scale optimization for visual exploration. IEEE Transactions on Visualization and Computer Graphics, 17(6):795–807, 2010.

[20] Daniel Filonik and Dominikus Baur. Measuring aesthetics for information visualization. In 2009 13th International Conference Information Visualisation, pages 579–584. IEEE, 2009.

[21] Michael Gleicher, Michael Correll, Christine Nothelfer, and Steven Franconeri. Perception of average value in multiclass scatterplots. IEEE Transactions on Visualization and Computer Graphics, 19(12):2316–2325, 2013.

[22] Connor C. Gramazio, David H. Laidlaw, and Karen B. Schloss. Colorgorical: Creating discriminable and preferable color palettes for information visualization. IEEE Transactions on Visualization and Computer Graphics, 23(1):521–530, 2017.

[23] Connor C. Gramazio, Karen B. Schloss, and David H. Laidlaw. The relation between visualization size, grouping, and user performance. IEEE Transactions on Visualization and Computer Graphics, 20(12):1953–1962, 2014.

[24] Connor C. Gramazio, Allison T. Silverman, Madeline L. Parker, Audrey S. Wang, and Karen B. Schloss. Mapping Color to Meaning in Colormap Data Visualizations. IEEE Transactions on Visualization and Computer Graphics, 25(1):810–819, 2018.

[25] Paul Green-Armytage. Complementary Colors, pages 523–535. Springer New York, New York, NY, 2016.

[26] Carl Gutwin, Andy Cockburn, and Ashley Coveney. Peripheral popout: The influence of visual angle and stimulus intensity on popout effects. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pages 208–219. ACM, 2017.

[27] K Wm Hall, Charles Perin, Peter G Kusalik, Carl Gutwin, and Sheelagh Carpendale. Formalizing emphasis in information visualization. In Computer Graphics Forum, volume 35, pages 717–737. Wiley Online Library, 2016.

[28] Steve Haroz and David Whitney. How capacity limits of attention influence information visualization effectiveness. IEEE Transactions on Visualization and Computer Graphics, 18(12):2402–2410, 2012.

[29] Mark Harrower and Cynthia A Brewer. Colorbrewer. org: an online tool for selecting colour schemes for maps. The Cartographic Journal, 40(1):27–37, 2003.

[30] Mark Harrower and Cynthia A. Brewer. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. In The Map Reader: Theories of Mapping Practice and Cartographic Representation. 2011.

[31] C.G. Healey. Choosing effective colours for data visualization. pages 263–270,. Institute of Electrical and Electronics Engineers (IEEE), dec 2002.

[32] Christopher Healey and James Enns. Attention and visual memory in visualization and computer graphics. IEEE Transactions on Visualization and Computer Graphics, 18(7):1170–1188, 2011.

[33] Christopher Healey and James Enns. Attention and visual memory in visualization and computer graphics. IEEE Transactions on Visualization and Computer Graphics, 18(7):1170–1188, 2012.

[34] Christopher G Healey. Perception in visualization. 10, 2007.

[35] Christopher G Healey, Kellogg S Booth, and James T Enns. Visualizing real-time multivariate data using preattentive processing. ACM Transactions on Modeling and Computer Simulation (TOMACS), 5(3):190–221, 1995.

[36] Jeffrey Heer and Michael Bostock. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Proceedings of the 28th Annual Chi Conference on Human Factors in Computing Systems, pages 203–212, 2010.

[37] Jeffrey Heer, Stuart K Card, and James A Landay. Prefuse: a toolkit for interactive information visualization. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 421–430. ACM, 2005.

[38] Jeffrey Heer and Maureen Stone. Color naming models for color selection, image editing and palette design. pages 1007–1016, 2012.

[39] John M Henderson and Taylor R Hayes. Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. Journal of Vision, 18(6):10–10, 2018.

[40] Johannes Itten. The Elements of Color. John Wiley & Sons, 1970.

[41] J.M.Wolfe and T.S.Horowitz. What attributes guide the deployment of visual attention and how do they do it? Nature Reviews - Neuroscience, 5(6):495–501, 2004.

[42] Younghoon Kim and Jeffrey Heer. Assessing Effects of Task and Data Distribution on the Effectiveness of Visual Encodings. Computer Graphics Forum, 37(3):157–167, 2018.

[43] N. Kita and K. Miyata. Aesthetic Rating and Color Suggestion for Color Palettes. Computer Graphics Forum, 35(7):127–136, 2016.

[44] Lars Kolbowski, Colin Combe, and Juri Rappsilber. xispec: web-based visualization, analysis and sharing of proteomics data. Nucleic Acids Research, 46(W1):W473–W478, 2018.

[45] Robert Kosara, Silvia Miksch, Helwig Hauser, et al. Semantic depth of field. 1:97–104, 2001.

[46] Andrey Krekhov and Jens Krüger. Deadeye: A novel preattentive visualization technique based on dichoptic presentation. IEEE Transactions on Visualization and Computer Graphics, 25(1):936–945, 2018.

[47] Jie Liang and Mao Lin Huang. Highlighting in information visualization: A survey. In 2010 14th International Conference Information Visualisation, pages 79–85. IEEE, 2010.

[48] Sharon Lin, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, and Jeffrey Heer. Selecting semantically-resonant colors for data visualization. In Computer Graphics Forum, volume 32, pages 401–410. Wiley Online Library, 2013.

[49] Delwin T. Lindsey, Angela M. Brown, Ester Reijnen, Anina N. Rich, Yoana I. Kuzmova, and Jeremy M. Wolfe. Color channels, not color appearance or color categories, guide visual search for desaturated color targets. Psychological Science, 21(9):1208–1214, 2010.

[50] Yang Liu and Jeffrey Heer. Somewhere over the rainbow: An empirical assessment of quantitative colormaps. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, page 598. ACM, 2018.

[51] Jasna Martinovic, Sophie M Wuerger, Steven A Hillyard, Matthias M Müller, and Søren K Andersen. Neural mechanisms of divided feature-selective attention to colour. NeuroImage, 181:670–682, 2018.

[52] Justin Matejka, Michael Glueck, Tovi Grossman, and George Fitzmaurice. The effect of visual appearance on the performance of continuous sliders and visual analogue scales. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 5421–5432. ACM, 2016.

[53] Declan J. McKeefry, Ian J. Murray, and Neil R. A. Parry. Perceived shifts in saturation and hue of chromatic stimuli in the near peripheral retina. Journal of the Optical Society of America A, 24(10):3168, 2007.

[54] Michael D'Zmura. Color in visual search. Vision Research, 31(6):951–966, 1991.

[55] Sebastian Mittelstädt, Dominik Jäckle, Florian Stoffel, and Daniel A Keim. Colorcat: Guided design of colormaps for combined analysis tasks. In EuroVis (Short Papers), pages 115–119, 2015.

[56] Sebastian Mittelstädt, Andreas Stoffel, and Daniel A Keim. Methods for compensating contrast effects in information visualization. In Computer Graphics Forum, volume 33, pages 231–240. Wiley Online Library, 2014.

[57] Nathan Moroney. Unconstrained web-based color naming experiment. In Color imaging VIII: Processing, hardcopy, and applications, volume 5008, pages 36–46. International Society for Optics and Photonics, 2003.

[58] Nathan Moroney, Mark D Fairchild, Robert WG Hunt, Changjun Li, M Ronnier Luo, and Todd Newman. The ciecam02 color appearance model. In Color and Imaging Conference, volume 2002, pages 23–27. Society for Imaging Science and Technology, 2002.

[59] Morten Moshagen and Meinald T. Thielsch. Facets of visual aesthetics. International Journal of Human Computer Studies, 68(10):689–709, 2010.

[60] Stephen E. Palmer, Karen B. Schloss, and Jonathan Sammartino. Visual Aesthetics and Human Preference. Annual Review of Psychology, 64(1):77–107, 2013.

[61] Quang Vinh Nguyen and Mao Lin Huang. A focus+context visualization technique using semi-transparency. The Fourth International Conference on Computer and Information Technology, 2004. CIT '04., pages 101–108, 2004.

[62] Khairi Reda, Pratik Nalawade, and Kate Ansah-Koi. Graphical perception of continuous quantitative maps: the effects of spatial frequency and colormap design. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, page 272. ACM, 2018.

[63] Anthony C Robinson. Highlighting techniques to support geovisualization. 2006.

[64] Anthony C. Robinson. Highlighting in Geovisualization. Cartography and Geographic Information Science, 38(4):373–383, 2011.

[65] Bernice E Rogowitz and Alan D Kalvin. The which blair project: A quick visual method for evaluating perceptual color maps. In Proceedings of the conference on Visualization'01, pages 183–190. IEEE Computer Society, 2001.

[66] Bahador Saket, Alex Endert, and Cagatay Demiralp. Task-Based Effectiveness of Basic Visualizations. IEEE Transactions on Visualization and Computer Graphics, 25(7):2505–2512, 2019.

[67] Francesca Samsel, Trinity Overmyer, and Paul A. Navratil. Highlight Insert Colormaps : Luminance for Focused Data Analysis. (April), 2019.

[68] K. B. Schloss and S. E. Palmer. An ecological valence theory of human color preferences. Journal of Vision, 9(8):358–358, 2010.

[69] Karen B. Schloss and Stephen E. Palmer. Aesthetic response to color combinations: Preference, harmony, and similarity. Attention, Perception, and Psychophysics, 73(2):551–571, 2011.

[70] Travis Simcox and Julie A Fiez. Collecting response times using amazon mechanical turk and adobe flash. Behavior Research Methods, 46(1):95–111, 2014.

[71] Stephen Smart and Danielle Albers Szafir. Measuring the Separability of Shape, Size, and Color in Scatterplots. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19, pages 1–14, 2019.

[72] Stephen Smart, Keke Wu, and Danielle Albers Szafir. Color crafting: Automating the construction of designer quality color ramps. IEEE Transactions on Visualization and Computer Graphics, 2019.

[73] John Stasko, Carsten Görg, and Zhicheng Liu. Jigsaw: supporting investigative analysis through interactive visualization. Information Visualization, 7(2):118–132, 2008.

[74] Maureen Stone, Danielle Albers Szafir, and Vidya Setlur. An engineering model for color difference as a function of size. In Color and Imaging Conference, volume 2014, pages 253–258. Society for Imaging Science and Technology, 2014.

[75] Hendrik Strobelt, Daniela Oelke, Bum Chul Kwon, Tobias Schreck, and Hanspeter Pfister. Guidelines for effective usage of text highlighting techniques. IEEE Transactions on Visualization and Computer Graphics, 22(1):489–498, 2015.

[76] Hendrik Strobelt, Daniela Oelke, Bum Chul Kwon, Tobias Schreck, and Hanspeter Pfister. Guidelines for Effective Usage of Text Highlighting Techniques. IEEE Transactions on Visualization and Computer Graphics, 22(1):489–498, 2016.

[77] Michael J Stroud, Tamaryn Menneer, Elina Kaplan, Kyle R Cave, and Nick Donnelly. We can guide search by a set of colors, but are reluctant to do it. Attention, Perception, & Psychophysics, 81(2):377–406, 2019.

[78] D.A. Danielle Albers Szafir, Maureen Stone, and Michael Gleicher. Adapting Color Difference for Design. Color and Imaging Conference, (January):228–233(6), 2014.

[79] Danielle Albers Szafir. Modeling color difference for visualization design. IEEE Transactions on Visualization and Computer Graphics, 24(1):392–401, 2017.

[80] Danielle Albers Szafir and Michael Gleicher. Visualization-Aware Color Design. EuroVis 2016 Poster Proceedings, pages 97–99, 2016.

[81] Danielle Albers Szafir, Steve Haroz, Michael Gleicher, and Steven Franconeri. Four types of ensemble coding in data visualizations. Journal of Vision, 16(5):11, 2016.

[82] Danielle Albers Szafir, Maureen Stone, and Michael Gleicher. Adapting color difference for design. In Color and Imaging Conference, volume 2014, pages 228–233. Society for Imaging Science and Technology, 2014.

[83] Jan Theeuwes. Abrupt luminance change pops out; abrupt color change does not. Perception & psychophysics, 57(5):637–644, 1995.

[84] Christian Tominski, Georg Fuchs, and Heidrun Schumann. Task-driven color coding. In 2008 12th International Conference Information Visualisation, pages 373–380. IEEE, 2008.

[85] Noam Tractinsky, Adi S Katz, and Dror Ikar. What is beautiful is usable. Interacting with Computers, 13(2):127–145, 2000.

[86] Anne Treisman and Stephen Gormican. Feature analysis in early vision: evidence from search asymmetries. Psychological Review, 95(1):15, 1988.

[87] Massimo Turatto and Giovanni Galfano. Color, form and luminance capture attention in visual search. Vision Research, 40(13):1639–1643, 2000.

[88] Vijay V. Vazirani. Approximation Algorithms. Approximation Algorithms, 2013.

[89] M. Villarroel, P. de la Fuente, A. Pedrero, and J. Adiego. Visualizing Shared Highlighting Annotations. HCI related papers of Interacción 2004, pages 195–204, 2006.

[90] Manuela Waldner, Alexey Karimov, and Eduard Gröller. Exploring visual prominence of multi-channel highlighting in visualizations. page 8, 2017.

[91] Lujin Wang, Joachim Giesen, Kevin T McDonnell, Peter Zolliker, and Klaus Mueller. Color design for illustrative visualization. IEEE Transactions on Visualization and Computer Graphics, 14(6):1739–1754, 2008.

[92] Qinqin Wang, Patrick Cavanagh, and Marc Green. Familiarity and pop-out in visual search. Perception & Psychophysics, 56(5):495–500, 1994.

[93] Yunhai Wang, Xin Chen, Tong Ge, Chen Bao, Michael Sedlmair, Chi Wing Fu, Oliver Deussen, and Baoquan Chen. Optimizing color assignment for perception of class separability in multiclass scatterplots. IEEE Transactions on Visualization and Computer Graphics, 25(1):820–829, 2019.

[94] Colin Ware and Robert Bobrow. Motion to support rapid interactive queries on node–link diagrams. ACM Transactions on Applied Perception (TAP), 1(1):3–18, 2004.

[95] Colin Ware, Terece Louise Turton, Roxana Bujack, Francesca Samsel, Piyush Shrivastava, and David Honegger Rogers. Measuring and modeling the feature detection threshold functions of colormaps. IEEE Transactions on Visualization and Computer Graphics, 2018.

[96] Jeremy M Wolfe. Visual search. Current Biology, 20(8):R346–R349, 2010.

[97] Jeremy M Wolfe and Todd S Horowitz. What attributes guide the deployment of visual attention and how do they do it? Nature Reviews - Neuroscience, 5(6):495, 2004.

[98] Jeremy M. Wolfe and Igor S. Utochkin. What is a preattentive feature? Current Opinion in Psychology, 29:19–26, 2019.