When Red Means Good, Bad, or Canada: Exploring People’s Reasoning for Choosing Color Palettes

Jarryullah Ahmad†
University of Toronto
Elaine Huynh‡
University of Toronto
Fanny Chevalier‡
University of Toronto

ABSTRACT
Color palette selection is an essential aspect of visualization design, influencing data interpretation and evoking emotions in the viewer. Rules of thumb grounded in perceptual science and visual arts generally form the basis of recommendation tools to support color assignment, but palette design is more nuanced than optimizing for perceptual tasks. In this work, we investigate how the general public reconciles the varied facets of color design in visualization. Does their decision-making align with established rules of thumb? What factors do they take into consideration? Through a crowdsourced study with 63 participants, we find that the majority of palette choices are perceptually motivated, but other factors such as semantic associations and bias also play a role. We identify some flaws in participant reasoning, highlight clashes in opinions, and present some implications for future work in this space.

1 INTRODUCTION
Color is an essential element of visual communication, but from a designer’s standpoint, there are a multitude of factors to consider in color palette choice [1]. Often carefully considered for its aesthetic qualities, color can also be used to invoke thoughts of certain objects or concepts through semantic association [17, 19, 33], as well as impact viewer cognition, affect [2], and even behavior [8]. These properties are not lost when applied in the context of data visualization. By playing with contrast and hues, color can be used to draw user attention to certain data points or to evoke particular emotional responses from viewers. These notions, however, only capture the effects of good color usage — a poor choice of colors may detract from or diminish the effectiveness of the intended data storytelling, or perhaps evoke unintended feelings or meanings. Choosing an appropriate color palette, then, can be a daunting task.

While prior research allows people to derive guidelines and best practices for color usage—a lot of which is operationalized in recommendation tools [12, 21] and models [34, 35]—it has mostly been concerned with how viewers perceive and interpret color. Comparatively, little is known about the reasoning process behind color palette choice when creating a visualization. While experts in visualizations are cognizant of the many implications of color selection, someone not as well-versed in the field may not make the same choices. By developing a more comprehensive understanding of novice thought processes, we may better identify gaps in understanding and consider ways of addressing them, such as through informed design changes or additions to color recommendation tools.

In our work, we explore novices and their knowledge of color in visualizations, specifically in the process of color palette selection. What factors do they consider crucial in their understanding? How do they determine what is a reasonable choice? Why might they consider certain color choices to not be as “fit” or “strong” in comparison to others? In a crowdsourced study, we present participants with two types of visualizations (pie charts and choropleth maps) encoding categorical or ordered data. Given a selection of categorical, sequential, and diverging color palettes, we ask participants to identify the “most appropriate” color palette for each visualization, and to explain their decision to determine what factors influence each response. In particular, we observe (1) whether people can apply sensible palettes to data types as defined by perceptually-motivated rules of thumb, and (2) the role that factors such as semantic, emotional, and familiarity associations play in palette choice.

Our results show that novices often arrive at a choice that aligns with perceptually-motivated guidelines (e.g., a categorical scheme for categorical data), but their rationale is not always the same and sometimes even in direct opposition with what is expected. In cases where participants expressed a consideration for the type of data presented, some still selected a palette type that did not support this reasoning. We find that participant choices are affected by factors such as semantic association, aesthetics, and bias, causing them to choose (or even avoid) certain palettes, or make decisions to facilitate storytelling, revealing what novices find important to prioritize. We reflect on the implications for design for future studies and tools supporting palette design in visualization.
2 Related Work

In visualization design, color—more specifically, the channels that constitute color (most commonly hue as an identity channel, and saturation or luminance as magnitude channels)—are used as visual variables mapped to data attributes [4, 23]. Color perception and perceptual science provide guidelines for approaching color selection, proposing sets of best practices to adhere to—e.g., a staple rule of thumb is to use different color hues to represent categorical data, and use different shades of the same hue by varying saturation or luminance, when conveying ordered data [12, 23, 37]. Other advances include the development of perceptually-driven models [34, 35] and algorithms (e.g., ranking metrics for color [11], or looking at interactions of opacity and other variables on perception [36]).

While perceptual properties are often the primary consideration in visualization, other cognitive aspects have been studied. People tend to associate quantities with qualities such as luminance (dark-is-more bias [6, 20, 25, 30, 32]) and opacity (opaque-is-more bias [32]). These encodings exemplify an overlap between color’s perceptual features and semantic associations aspects. Such pre-established connections, along with other symbolic relationships between colors and objects or concepts, can be successfully factored in while preserving perceptually-relevant properties such as discriminability, in what is referred to as semantically-resonant palettes [10, 17]. These have been found to impact understanding [17], recall [13, 29], and decoding [32]. Associations with emotions have also demonstrated how color properties may be manipulated to achieve affective expressiveness in visualizations [2]. Many of these works look at how readers understand or are affected by the colors in visual media—most of which were crafted by experts. To our knowledge, it remains to be seen to what extent semantic associations are taken into consideration by novices when tasked with designing visualizations. Despite the importance of perceptual properties and semantic associations, it is not simply enough to make color choices based solely on these factors. For instance, in spite of its poor perceptual properties, rainbow maps are still preferred by scientists [7]. The aesthetic appeal of a visualization can affect how useful viewers perceive visualizations [27] and other constructs [22]. Like with perceptual properties, models and guidelines exist to increase aesthetic appeal [3, 9, 15, 31], but it can be difficult (if not impossible) to make a design that appeals to all audiences as a result of differences on personal, cultural, and demographic levels [14, 18, 24].

3 Study Design

To better understand non-experts, we set out to answer the following:

RQ1. Do people tend to follow perceptually-driven guidelines in visualization—that is, choose a palette best suited to encode data of a particular type (categorical, ordered sequential, ordered diverging)? We call this data type fit.

RQ2. Do other considerations such as semantic associations, aesthetic considerations or personal preference weigh significantly in the choice of a color palette?

RQ3. What other factors do people consider important when choosing a color palette to encode data?

We conducted an online experiment (see https://viz-exp-dev.herokuapp.com/) where we presented participants with coloring tasks (see Fig. 2). Given a visualization with initially unfilled visual marks and the corresponding legend for the data attribute (Sect. 3.1), participants were asked to color the visualization by selecting the “most appropriate color palette” to encode a data attribute (Sect. 3.2).

3.1 Data and Visualization Types

We designed six coloring tasks: one for each of two visualization types commonly found in news media (pie chart or choropleth map) and three data types (categorical, ordered sequential, ordered diverging). Considering data type fit, each task has a clear preferred type of colormap: a categorical colormap for categorical data, a sequential one for ordered sequential data, and a diverging colormap for ordered diverging data. To balance tasks, we fixed the number of encoded values to 5, so quantitative data on a continuous scale would be discretized through binning (see e.g., Fig. 3d).

3.2 Color Palettes

The set of color palettes used in our study can be seen in Fig. 2. We constrained the task to choosing between a preset collection of palettes as opposed to having the participants create their own, to allow them to devote most of their time thinking about why a palette is most reasonable to them. By asking participants to justify why they chose one palette over other options, we hoped that participants could more concretely articulate their reasoning via contrast (e.g., “I chose this one, because the other one is less”).

We chose an equal number of categorical, sequential and diverging palettes. Initial palette sets were chosen from ColorBrewer [12] and arranged in both ascending and descending order. This mirroring was meant to capture perception-based reasoning and semantics-based reasoning (e.g., “Red means bad so it shouldn’t be in this order”). We also chose complementary hues for the sequential and diverging palettes to further capture aesthetics- and semantics-based reasoning (e.g., “Red and blue are associated with politics in the US and should be avoided because it is not the point of this data”).

3.3 Procedure

The experiment was run on the Amazon Mechanical Turk (mTurk) crowdsourcing platform, using a strict inclusion criteria [26, 28]: all participants were Canadian or US residents with a prior 99% task approval rate. Approximate completion time was 15 minutes and participants were compensated 2.50 USD upon completion. Participants were asked demographics-related questions and prompted to indicate their familiarity with reading and creating visualizations via 5-point Likert scales, then shown six coloring tasks in a randomized order. For each prompt, participants had to select a palette that they felt best suited the visualization—clicking on a palette automatically colored the visualization accordingly, allowing participants to experiment with palette choices until satisfied. Participants were also asked to explain their choice, as well as answer a basic multiple-choice comprehension question to test their understanding of the visualization. See our supplemental materials and site for details.

4 Results and Discussion

A total of 63 participants (22F, 40M, 1 undisclosed) completed the study. Participants varied in age (24–69, mean = 40.6 std = 11.3) and level of education (16 secondary school, 39 undergraduate studies, 8 graduate studies). Using a 5-point Likert scale (1=not familiar
We sought to understand whether people tend to follow perceptually-commonly-cited factors and motivations behind applying certain reader and coloring in a way most beneficial to data interpretation. Tended to optimize for readability, with many adopting the lens of a express consideration for the type of data being represented. Their rationale: 299 out of 378 responses (79%) were found to not stray far from perceptually-motivated practices when exploring CI: 0.51-0.64) of the participants’ chosen palettes follow perceptual driven rules of thumb in visualization, i.e., select a color palette to visualizations, and then consolidated a set of codes for each of the coloring tasks; 57.5% (bootstrap that matches with the data type (RQ1). Fig. 3 summarizes the participant responses for each of the coloring tasks. The participant ascribes meaning to colors or palettes. The participant made a decision based on something previous knowledge. The participant’s choice is influenced by political sensi- tivities or a positive/negative effect on readers. None There is no clear motivation behind why the participant chose their selected color palette.

Table 1: Commonly-cited factors affecting novice color palette choice. at all, 5=very familiar), all indicated some familiarity with viewing visualizations (mean = 3.78, std = 0.8) and less with creating them (mean = 2.60, std = 1). Two of the authors separately conducted a round of open coding on a subset (30%) of participant responses to determine the most commonly-cited factors and motivations behind applying certain color palettes to visualizations, and then consolidated a set of codes as found in Table 1. The two coders then performed closed coding separately on all responses (with the possibility of each response being given to them. We posit that this could be an effect of the chart type—participants may just associate pie charts with a need for distinction between encoded values, as suggested by the many comments in line with: “This palette offers colors that allow you to easily tell apart the different sections in the pie graph.” Another possibility could be that the use of words (as opposed to numbers) may cause people to view the encoded values as wholly separate categories, so they miss the connection that the values fall on a scale—rather than viewing “daily”, “few times a week”, and “weekly” as part of a sequence, they see each as being unrelated to one another.

4.1 Data Type Fit (and Misfits)

We sought to understand whether people tend to follow perceptually-driven rules of thumb in visualization, i.e., select a color palette type that matches with the data type (RQ1). Fig. 3 summarizes participant responses for each of the coloring tasks; 57.5% (bootstrap CI: 0.51-0.64) of the participants’ chosen palettes follow perceptual guidelines, with higher rates in the case of categorical data (Fig. 3-a, b) and the map featuring ordered sequential data (Fig. 3-d).

Perhaps reassuringly, we found that people’s mental models did not stray far from perceptually-motivated practices when exploring their rationale: 299 out of 378 responses (79%) were found to express consideration for the type of data being represented (Data Type Fit in Table 1). Akin to visualization practices, our non-experts tended to optimize for readability, with many adopting the lens of a reader and coloring in a way most beneficial to data interpretation. Participants who indicated data types influenced their choice often referenced a desire to accomplish one of the following goals:

- **Provide distinction between categories**, e.g., “The colors are all very distinct from one another, which makes it visually simpler to separate the different categories.”
- **Use luminance to represent sequential data**, e.g., “The most rainfall should be in the darkest shade available so you know just by looking at the pic that those states were harder hit.”
- **Use diverging scales to represent extreme values with neutral midpoints**, e.g., “I think this is good because the blue and dark blue represent positive change, but the red and dark red easily represent a negative change in a way that is intuitive.”

While data type fit was the most prevalent factor cited in participant responses (299/378), note that we observed 97 instances where participants selected a palette that did not align with the data type, yet their rationale suggests that the intent was to achieve a perceptually meaningful encoding with regard to the data.

When it came to pie charts (Fig. 3a,c,e), we saw that many participants opted for a categorical palette, regardless of the data type given to them. We posit that this could be an effect of the chart type—participants may just associate pie charts with a need for distinction between encoded values, as suggested by the many comments in line with: “This palette offers colors that allow you to easily tell apart the different sections in the pie graph.” Another possibility could be that the use of words (as opposed to numbers) may cause people to view the encoded values as wholly separate categories, so they miss the connection that the values fall on a scale—rather than viewing “daily”, “few times a week”, and “weekly” as being part of a sequence, they see each as being unrelated to one another.

While participants mostly selected categorical palettes for categorical data (Fig. 3a,b), there were a few cases where they chose otherwise. Those who chose a diverging palette believed such palettes were more effective at showing contrast between values. This was also seen in cases where diverging palettes were chosen for sequential data, e.g., “I like this because the white in the middle is good break for weekly, whereas the red for less and blue for more provides a good split. It’s easy to distinguish the colors.” Participants who favored sequential palettes for pie charts with categorical data, however, generally did so for a different reason. Slices in a pie chart are typically ordered by size and participants wanted to use darker shades to draw emphasis to the largest slice in the pie chart, e.g., “I think this is the most appropriate color palette because it is a good choice since the dark color is good to represent the majority.”

In choosing sequential palettes for diverging data, we saw many participants wished to highlight higher values with darker shades—in Fig. 3e, respondents seemed to associate darker shades with...
stronger agreement, whereas in Fig. 3f, they associated darker shades with more positive change. In the latter, some saw the progression from negative to positive values (−10% to +10%) as a sequence as opposed to values diverging in two opposite ways from the zero baseline. Building on this, we observed that several participants may carry misconceptions about amounts of change and assume that an increase in population is to be considered as “more of a change” than an equal amount of decrease in population, and chose a palette that is prone to induce the same misconception in the reader (e.g., “The deeper the color the more drastic the change is for that state.”)

4.2 Semantic Associations
In RQ2, we sought to determine whether or not participants would consider non-perceptual factors of color in their decision making process. While some participants mentioned aesthetic appeal (7%) and personal preferences (6%) in their responses, we did not find many recurring trends among these responses. More interestingly, however, was the subject of semantic associations. We anticipated that semantic association would be the most commonly-cited reason—that novices might not be aware of perception-based guidelines and would instead be swayed by symbolic and mental associations. To our surprise, only 76 of the 378 (20%) coded responses referenced some semantic association, the common ones of which included:

- Matching a color name in the data to its color, e.g., “White people are white in the pie chart and that just makes the most sense to me”, though the mapping was only partial, as other ethnicities were mapped to shades of green (see Fig. 3-a).
- Associating the topic of the visualisation with a color palette, e.g., “I think red stands out a lot and since it’s the USA, it’s appropriate to go with red, white, and blue.”
- Associating a color with the potential outcomes of the data, e.g., “Green matches symbolically, with more rain yielding greener landscape. Furthermore, it makes sense to use shades of a single hue to represent low to high rainfall amounts.”
- Considering implicit meanings of colors, e.g., “I think there’s really nothing wrong with playing video games, nor is there anything wrong with not playing them, so colors which don’t make any particular feelings come to mind are appropriate.”

Clashes were observed between how participants identified and wanted to use certain associations. While some wanted to use white to represent the white ethnicity, others made choices to explicitly avoid this association: “The other palettes have white in them, and I think some people in today’s society might take offense to having the color white in a graph of people’s races.” Such contrast was also seen in cases where colors were used to represent opposing extremes. For example, one participant viewed red as a good representation of agreement: “The agrees are in a warm color, the people that disagree are indicated in a cool color and the ones that can’t decide are in white”, while another participant thought the opposite: “Blue being agree and red being disagree work best because red is usually negative and blue is usually positive. White works well for neutral.”

These examples reinforce the notion that how people apply semantic associations can be vastly different. Therefore, any tool that attempts to suggest optimal color palettes needs to take into account the context of the visualisation—how meaning will be applied to color will vary by audience and author and there is no one optimal palette that will be semantically resonant for a certain data set.

4.3 Bias and Other Factors
In response to RQ3, while other factors were not as prevalent as those previously described, there were a noticeable few—of particular note was a participant’s previous knowledge, present in 14 (3%) of the responses and captured in messages such as “I think this is the most appropriate choice because the colors are adequate with what we are used to in pictures, cartoons, etc.” and, “The most change should have the darkest color. I think that’s what we are all used to when we see news reports that have maps like this.” Prior exposure may have culminated in bias playing a bigger role in participant choices than was seen in their justifications. There is both benefit and danger in such a finding: if one is exposed to examples following guidelines, prior knowledge may help them choose sensible palettes, and vice versa. We believe this to be a potential avenue for visualization recommendation tools.

4.4 Limitations
In the real world, there are unlimited color choices available to people when deciding upon a color palette and a rich variety of visualization designs with different types of marks. This could not be captured in our limited study design wherein many compromises had to be made to remove some of the confounds, and enable quantitative analysis. In our experiment, participants were forced to choose from fixed configurations. This allowed our study to be short for participants to complete, and by pushing them to choose between palettes whose types are unambiguous, facilitated comparisons and explanations in such terms. However, the small selection of palettes and visualizations also meant that interactions between factors were likely missed. While characterizing the qualities of participant-generated palettes may not be straightforward, future study designs with more freedom of selection may yield more insights into how aspects such as semantic associations, previous exposure, and personal bias affect choice.

Another limitation arises from how to measure familiarity or prior exposure to visualization, which—as mentioned in Sect. 4.3—may impact participant choices. We asked participants to rate their self-perceived level of familiarity in designing and reading visualizations mainly to confirm they were not experts, but found these ratings were not necessarily accurate indicators of competency. Work in visualization literacy has culminated in more comprehensive tests [5, 16], but are costly in terms of time required to complete them, especially relative to the length of the rest of our study. Future work could explore compromises between such tests and quick tasks like ours, as well as other demographic data to allow stratified analysis.

5 Conclusion and Future Work
Our study confirms that factors that (consciously or subconsciously) influence data interpretation when reading visualizations are factors that novices actively reason about when making informed choices regarding color palette design when creating visualizations. Our results suggest that novices often choose palettes that encourage readability and interpretability, but we see other competing factors that may interplay with the selection of perceptually-motivated color palettes, with semantic associations being particularly noteworthy and an interesting venue for visualization authoring tools. Knowledge from prior literature on affective color [2], combined with methods to derive most commonly-used colors for concepts or objects [10, 17], could help uncover common implicit or explicit color associations, but careful steps must be taken to align findings with the expectations of the viewership, as well as avoid ethical issues such as reinforcing stereotypes or promoting unintended messages. Our study also surfaces biases in novices, indicating that tools may benefit from features that support recognition and reinforcement of “good” visualization examples in a bid to correct misconceptions formed by prior exposure to less sensible design choices.

We hoped to identify the knowledge gaps that lead to some non-experts making poor palette choices. While our work helped identify some factors, we found that novices generally employed sensible and informed reasoning aligned with visualization practices. This begs the question as to why so many poor designs exist in the wild, but which follow-up studies focusing on visualization authoring may help us understand.