

# Coloring Right

## 1 Abstract

Is it possible to automatically create a color palette relevant to a topic? Could such a palette be used to guide color choices while visualizing data? Prior work has shown that when data is presented in colors that are relevant to the data-topic, it leads to better and faster understanding. This project explores the possibility of automatically creating topic-relevant palettes for a large class of topics. This automated system finds color palettes based on a corpus of images (from either ImageNet, Flickr or Google Images). The hypothesis is that generating palettes based on images in these corpora that are labeled with the topic will be topic-relevant. While we haven't narrowed down on the exact technique to extract palettes from color pixel values, clustering and topic models seem promising. The system will be evaluated on three metrics: how well users like the generated palettes, how topic-relevant they are perceived to be, and if they affect data understanding. For all three metrics, the algorithmically generated palettes will be compared against a randomly generated palette, and one generated by experts. For the likability and understanding metrics, the random palette will be chosen from the set of palettes generated for other topics by our system (so that only the relevance, not the base quality of the colors is considered). Likability will be measured using a Likert scale. For relevance, an association task is used: given a topic (e.g. "US Politics") and one of the topic terms (e.g. "Democrat"), the participant chooses which color, among a set of displayed swatches, is relevant to it. For understanding, users will be shown differently-colored infographics, and participants will be timed while they answer conceptual questions related to the infographic. Since the three metrics may interact strongly, they will be studied in a within-subjects design. Participants will be drawn from workers on Mechanical Turk, and non-crowd study participants will be recruited through the class and through dorm mailing lists.

## 2 Related work

Prior work exists on automatic creation of color palettes. This work falls broadly in two categories. The first focuses on finding representative colors from images, that can be used as color palettes. The most recent of these is [1]. This line of research has so far focused only on extracting colors from a single image. This project extends this work by extracting colors from multiple, related images. I believe that some of the techniques used by [1], such as a weighted histogram that uses color saturation and neighborhood color coherence, can be adapted for multiple images too. Depending on constraints of time, I plan to explore some of these techniques.

The second category of research on palette generation focuses on optimizing visual properties, such as color saliency and perceptive color distance [2]. I believe most such optimization research is complementary to this project, and can be used as a post-extraction step to optimize the colors chosen.

Topic models, though not strictly related to colors, have been shown to be effective in information retrieval. Latent semantic analysis.