Colorific: A Mixed Initiative Model for Choosing the Right Color

Chinmay Kulkarni
Stanford University HCI Group
Computer Science Department
Stanford, CA 94305
chinmay@cs.stanford.edu

Julie Fortuna
Stanford University HCI Group
Computer Science Department
Stanford, CA 94305
jfortuna@stanford.edu

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? We envision a tool that automatically creates aesthetically pleasing and topic-relevant palettes for a large class of topics. In order to do this, we must first extract palettes from color pixel values of images from Google Images via clustering and topic models.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Design, Human Factors, Experimentation

Keywords: Information visualization, colors, crowdsourcing, user study, mixed initiative

INTRODUCTION
TODO Julie/Chinmay. 300 words. Example based– Stanford Cardinal? Anger? Sadness? Understanding is aided, if supported

RELATED WORK
CEK 400 words. One para each: Color, mixed initiative, emotions and color (http://socrates.berkeley.edu/ plab/).

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 [?]. 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 [?], 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 fo-cuses on optimizing visual properties, such as color saliency and perceptive color distance, both manual or rule-based, as pioneered by Brewer [?]; and with varying degrees of automation [?, ?]. 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. Statistical work on color saliency is valuable, even if it hasn't been directly applied as a optimization objective; color saliency in the context outside data-visualization in [?, ?].

Topic models have been shown to be effective in information retrieval. Latent semantic analysis (and later, LDA), for instance, has been used to find "latent" similarities between concepts [?, ?]. Similar similarity-measures have been computed for nodes in a graph [?]. While these similarity measures may help to better cluster color-values, they don't target the domain of color recommendations directly.

SYSTEM DESCRIPTION
550 Words

Query System
Google Images is queried for images from the category.

Statistical summarization
We assume that the images from the category are a random sampling from the concept-space of the category. Taking this assumption further, we look at the *average* frequencies of the different colors as a metric of how concepts are shared across the values in a category.

$$Old = \alpha * average + (1 - \alpha) * new \qquad (1)$$

$$new = \frac{(old - \alpha * average)}{(1 - \alpha)} \qquad (2)$$

Since we are interested in the colors specific to a category value, we substract a fraction of the average color frequency.

Clustering
We cluster the result to get relevant colors in LAB space. We found that low saturation colors are less likely to be relevant, so we reweight more saturated colors to be more relevant.

SYSTEM EVALUATION
TODO Julie. NOTE: Use APA format. Relevance, Likability (and Concrete/abstract)

We conducted surveys on Amazon Mechanical Turk to judge the quality of colors and palettes generated by Colorific. The study evaluated the following hypotheses:

H1: Machine-generated colors are perceived to be more topic-relevant than colors chosen randomly from the Protovis palette, a palette available in an industry standard data visualization toolkit.

H2: Machine-generated palettes are perceived to be more likable and aesthetically pleasing than palettes composed of random Protovis colors.

Method
To test *topic relevance*, we showed 50 participants from Mechanical Turk six options for each topic. Four of them were colors generated by Colorific. One was a random color picked from the Protovis palette, an industry standard in data visualization. The last choice was none of the above. Participants picked one choice that best suited the topic.

To test *palette likability*, we wanted to test if we can combine colors from different football teams, like Cardinal, Cal Bears, and others to create a single, cohesive palette. We hypothesized that the algorithm can choose palettes that are liked better than random palettes. We werent sure of the best way to pick one color, so we tried four variations. One tried to pick the most frequent color, another optimized for saturation, one maximized perceptual distance between the colors, and the last picked colors from the topic at random. We asked 49 participants from Mechanical Turk to rate our four variations and one randomly generated palette, from Protovis on a scale of 1-7.

Results
Topic Relevance   We found that the Colorific colors were preferred in an overwhelming majority ($\chi^2 = 83.7562$, df $= 2$, $p < 0.001$). 947 times out of 1200, Turkers picked a Colorific color. So, Colorific is able to pick at least some good colors for a topic.

Palette Likability

DISCUSSION
JMF Quotes and stuff. Include pilot if any. (500-600 words)

At the end of the test, participants filled an optional feedback form. Like we predicted, some of them had trouble picking the best color where more than one color was acceptable. I thought this was HARD! Some of the time I wanted to pick more than one color (for CAL there was a gold and a blue option). It definitely forced to me to think in a way I don't normally think. This was a very interesting task and it really made me think about which colors are associated with certain items.

Apparently, while Colorific got the right set of colors, it failed to pick the best one for the palette. I thought it was interesting. I wish I could have chosen my own colors for a few things. It was sometimes difficult to decide because usually none of the color palettes were colors I would have chosen.

So, we thought we could use a mixed initiative strategy in which the computer offers the set of appropriate colors, and the human chooses the final color for the visualization.

SCENARIO: DESIGNING WITH COLORIFIC
Colorific is designed to help novices with little experience in designing information visualizations quickly choose appropriate and compelling colors for their visualizations. Let's examine David, a football fan, as he makes a bar chart of the number of games won by a selection of Pac-10 football teams.

David begins making his chart in Microsoft Excel, but Excel has automatically colored each bar from a random set of colors. David wants to color each bar a representative color for the team. However, both the Cal Bears and the UCLA Bruins have the colors blue and gold. Which should be blue, and which should be gold? He loads Colorific to find out. David enters the specific topics he is plotting, such as the Stanford Cardinal and the Cal Bears, as well as their values. Colorific plots his values on a bar chart and automatically generates a selection of four potential palettes he can choose from. David selects one automatically generated palette, which looks almost the way he wants it. However, after seeing the results in context, he thinks the Cal Bears bar should be gold instead of blue. He clicks on the bar and is presented with a selection of other Colorific generated colors for the Cal Bears, as well as a color picker that allows him to tweak Colorific colors or select his own.
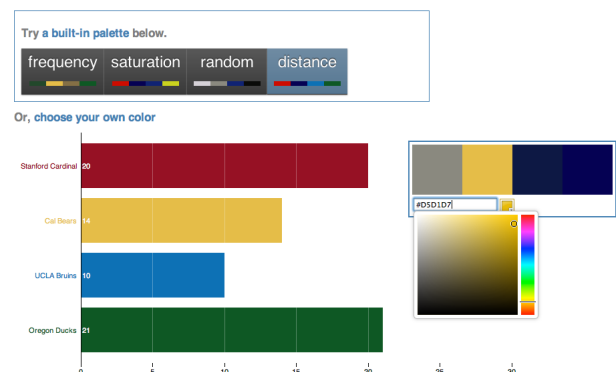
DESIGN SPACE
Colorific provides one point in the space of many data visualization tools. We discuss limitations in our design and propose avenues for future work.

Task Type
Creating data visualizations requires many component tasks. People must clean the data, then plot the data in a reasonable chart. Finally, aesthetic changes to the chart must be made, including coloring the chart appropriately. Colorific currently only supports the task of coloring a given chart. In the future, it would be beneficial to incorporate Colorific into a larger data visualization tool, so that people could clean, plot, and color their data all from within one tool.

Expertise

Colorific is designed to be beneficial to both novice and expert data visualization designers. For novices, the main advantage of Colorific is the pre-made palettes, which can be tweaked and edited while giving the designer real-time feedback on what the color will look like in context of the visualization. For experts, Colorific's main advantage is saving time and effort. Trying to find the exact right color via a color picker is difficult and time-consuming. Colorific proposes appropriate colors, and makes it easy to make slight adjustments to a color to fit a theme.

Time Scale

blah blah blah time

CONCLUSION

CEK Make up stuff