Insights from World Color Survey

Elizabeth Hau, Christina Michaels, Sravani Yajamanam Kidambi\*

MIT, 6.859 Interactive Data Visualization

Abstract

TBD

Our team aims to show that the language of color is not black and white – there are striking similarities and differences across communities when describing color.

**Keywords**: Color theory, Munsell chart.

# Introduction

The World Color Survey (WCS) was conducted in the late 1970s to evaluate a color terminology system across different languages1. Berlin and Kay at UC Berkeley wanted to investigate universal crosslinguistic constraints on naming colors and that color terminology systems develop in a fixed order. WCS used 330 Munsell chips of 40 equally spaced hues, 8 levels of lightness and 10 achromatic chips to show the interviewees. Using this color palette, the team interviewed an average of 24 speakers of each of 110 unwritten languages globally. These interviews were conducted on approximately 2,640 individuals in total.

LEAVE 0.5 INCH SPACE AT BOTTOM OF LEFT COLUMN ON FIRST PAGE FOR COPYRIGHT BLOCK

The speakers were expected to participate in two tasks: naming task and focus task. In the naming task, the speaker was to indicate all the chips that they would call a color term t. In the focus task, the speaker was to indicate the best examples of t for each basic color term t.

Our team aims to analyze this data across the languages, geographic regions, and speakers to show patterns on interpretations of color.

# Related work

Our perception of color is extremely subjective. What you may consider to be teal may be someone else’s dark green. The Munsell color system was developed by an art professor Albert Munsell in the early 1900s2. As shown in Figure 1, the chart classifies color numerically based on three categories: hue, value (lightness/darkness), and chroma (intensity of color). Munsell wanted to standardize the color classification to avoid subjectivity. Since then, the Munsell color system has given scientists a gold standard in any color-related study. For example, the US Department of Agriculture’s official soil-research relies on the Munsell color system.

In the 1950s, Lenneberg and Roberts published about how anthropologists can compare sensory descriptions in different languages3. They wanted to use color because physiological aspects of color vision are universal, so it would be easy to test compared with words for smells, for example. Lenneberg and Roberts studied color terms in English and Zuni, using the Munsell book of color (1942). This book mapped all color samples onto the color solid to show all human visible colors. They were the first to use the Munsell Book of Color for the experiments. Afterwards, in 1969, the most famous example of color linguists came through in the

\* [ehhau@mit.edu](mailto:ehhau@mit.edu), [ckm2022@mit.edu](mailto:ckm2022@mit.edu), [sravaniy@mit.edu](mailto:sravaniy@mit.edu)

work from Berlin and Kay, which our team used for the visualizations.

A picture containing table

Description automatically generated

Figure 1. Munsell color chart as used by the World Color Survey4.

# Methods

Our team obtained the WCS dataset from Kaggle5, which contained 9 separate files. We started off with some data cleansing and pieced together the relations between various tables, as shown in Figure 2. The associations between tables are shown using the highlighted terms, which are in various colors representing the unique ID in each table. We explicitly did not clean up the different language term abbreviations for each Munsell chip. For example, what we may consider to be red is “cajan23” in one language and “zeli(ku)” in another language. If we cleaned up these terms, it would diminish our team’s goal to analyze different color interpretations. Thus, we decided to not clean up the term abbreviation field.

Diagram

Description automatically generated

Figure 2. Diagram of field relationships in the WCS dataset.

# Results

Taking learnings from the A4 journey, our team tackled three major visual interactions on our dashboard: word cloud with Munsell color palette, tree map, and choropleth map.

## Word Cloud with Munsell Color Palette

We randomized the colors in the Munsell chart and took learnings from our color lecture on the CIELAB color space. The CIELAB expresses color using three values: L\* for perceptual lightness, and a\* and b\* for the four unique colors of human vision: red, green, blue and yellow. Thus, we implemented a tooltip to show the L\*, a\* and b\* values of each color.

To make sense of how each color was interpreted in different languages, we created a word cloud to show various term abbreviations across the speakers. Originally, we had the terms as the same font size and varied the directions of each term. After getting feedback on our MVP, one classmate gave us an idea to see words with a certain number of occurrences. We decided to use a binning technique to bin the font sizes by occurrence. In other words, if our bin was 10 – 15, then terms with 10 – 15 occurrences will have a larger font than terms with 0 – 5 occurrences. In this manner, a user can see how each color is represented by native speakers around the world and filter which terms are common among speakers.

As shown in Figure 3, suppose a user clicks on a particular color cell in the Munsell chart and hits the ‘Play’ button. They will notice the word cloud updating as the system slides along the Munsell chart to the next color. This gives the user an interactive way to see how the terms change from one color cell to another.

This word cloud addresses our team’s primary goal of showing patterns on how speakers around the world interpret different colors.

Text, letter

Description automatically generated

Figure 3. Word cloud updating as new Munsell chip color is shown.

## Tree Map

The tree map is a unique interactive visual that allows users to hone into a particular geographic region of interest. Starting from the highest level in the hierarchy, continent, a user can drill down to the lowest level of terms in a particular language. For example, as shown in Figure 4, a user can zoom into North America, which contains 29 languages and 678 distinct terms. Within this continent, Mexico has 15 languages represented and 313 distinct terms. Drilling into Mexico, there are many native languages like Mazahua, which contains 60 distinct terms. A user can drill into this language and see each term. Similarly, a user can easily zoom out by clicking on the header of the tree map.

One minor note on the tree map is that one language could show up in multiple countries. Although it may seem like we are double counting languages, the tree map is accurate to the given data.

When the user hovers over a region on the tree map, the box will be shaded a different shade of pink. When the user hovers over a box, they will notice an outline of the lower level in the hierarchy. Once they click into the box and zoom in, they will notice the outlines pop and become larger. We wanted to create a seamless effect for the user and chose this design technique intentionally. Overall, this tree map allows us to meet our goal of showing patterns across languages and geographic regions.

Table, calendar

Description automatically generated

Figure 4. User drilling down to North America > Mexico > Language: Mazahua to see the term abbreviations.

## World Choropleth Map

Inspired by D3 designs on the web, we created a choropleth map of the world, particularly highlighting regions in which interviews were conducted. In the WCS dataset, a CSV file shows information about the speakers, particularly their genders. We wanted to highlight the counts of females and males within each country, so we used a tooltip to highlight this information, as shown in Figure 5. As a user hovers over a country, that country will be outlined to indicate the user has already seen details about that country. Around the globe, countries that were not participating in the WCS were greyed out, although the country name is still visible if the user hovers over. This world map helps meet our goal of analyzing patterns across various speakers in the study.

During the MVP feedback session, another classmate gave us an idea to integrate the tree map with the world choropleth map, described in the next section. Essentially when a user hones into a particular geographic region on the tree map, that region will also be highlighted on the world map and additional details will be shown. We took action on this suggestion and linked the tree map with the world map.

Map

Description automatically generated

Figure 5. World choropleth map with tooltip showing details on interviewees from Mexico.

# Discussion

TBD

# Future Work

Our team collectively spent around 96 hours for this final project. If we had more bandwidth, we would’ve liked to implement the following:

1. Implement a sticky header with WCS and the logo
2. Scrolly-telling feature of each of the sections in our dashboard
3. Word cloud initializing with all the terms
4. Current color selection initializing with a color
5. Better timing on the pause and play buttons
6. Change colors of buttons

# Acknowledgements

We would like to thank our instructor and TAs for a fantastic semester, albeit virtually! Thank you for creating a fun and supportive environment for us to learn about interactive visualizations. We all hope to take these learnings to our internships and beyond!

References

1. <http://www1.icsi.berkeley.edu/wcs/>
2. <https://www.sciencehistory.org/munsell-color-charts>
3. <https://munsell.com/color-blog/color-linguistics-language-evolution-perception-part3/>
4. <https://www.researchgate.net/figure/The-Munsell-color-chart-as-used-by-the-World-Color-Survey-All-color-distribution-figures_fig1_260447669>
5. <https://www.kaggle.com/jboysen/color-survey>