Insights from World Color Survey

Elizabeth Hau, Christina Michaels, Sravani Yajamanam Kidambi*

MIT, 6.859 Interactive Data Visualization

ABSTRACT

The World Color Survey (WCS) was an experimental initiative conducted in the late 1970s in which approximately 2,640 male and female speakers of 110 societies around the globe that employ unwritten languages were interviewed to observe how they would name 330 Munsell color chips. Established in the early 1900s, the Munsell color system has been used in many color studies and focuses on three primary attributes of color: the color intensity or chroma, the hue, and the lightness value. Our team analyzed a single experiment conducted during the WCS and constructed three distinct, yet connected, visualizations including a word cloud of survey responses, a tree map showing a geographic distribution of terms, and a world choropleth map. 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.

1 INTRODUCTION

The World Color Survey (WCS) was conducted in the late 1970s to evaluate a color terminology system across different languages[1]. Berlin and Kay at UC Berkeley wanted to investigate universal crosslinguistic constraints on naming colors and whether 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 during each interview. 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.

The speakers were expected to participate in two tasks: a naming task and a focus task. The naming task required the speaker to associate a single term, t, with a single color chip and repeat the process for all 330 Munsell chips presented. The chips were ordered randomly, but were presented in the same order to each survey participant. We assume this random order was decided to ensure the same term was not applied to similar color presented in quick succession. In the focus task, the survey participant was asked to indicate the best examples of a color term, t, for a set of similar colors. For example, a small collection of reddish tinted Munsell cchips were presented and the respondent was asked to select which of those chips best represented the term "red". Considering the survey respondents were members of societies in which languages were unwritten, those conducting the experiments used the closest available translation into the English language as possible. It is important to note that this adds an element of subjectivity to the term data resulting from WCS and also explains why some elements of the dataset are termed "undefined" or include numbers; it is likely in those cases, the provided term did not align with the Englishh alphabet and standards of pronunciation.

Our team decided to focus on the data collected from the first experiment to illustrate the objectives of the WCS. Our goal is to analyze this data across the languages, geographic regions, and speakers to show patterns relating to interpretations of common perceived colors.

* ehhau@mit.edu, ckm2022@mit.edu, sravaniy@mit.edu

2 RELATED WORK

Human 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 1900s[2]. 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 languages[3]. They prioritized the use of color because physiological aspects of color vision are considered universal, so it would be easy to achieve uniform test results as opposed to using words associated 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 work from Berlin and Kay at UC Berkeley, which is the source material our team used for these visualizations.

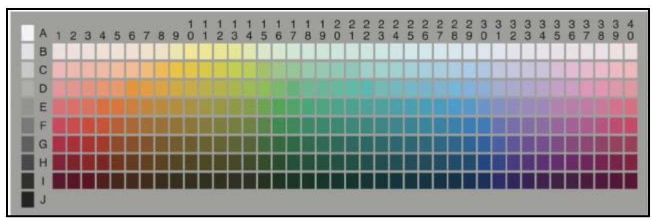

Figure 1. Munsell color chart as used by the World Color Survey[4].

3 METHODS

Our team obtained the WCS dataset from Kaggle[5], which contained 9 separate files. We began with extensive data cleansing and identifying the relations between various tables, as shown in Figure 2. The associations between tables are shown in various colors, representing the unique ID in each table. We explicitly did not edit the different language term abbreviations for each Munsell chip; these represent as much nuance from the experiment as was possible to capture and we felt that additional abbreviations would only increase the inherent subjectivity in the experimental results. For example, what we may consider to be red is "cajan23" in one language and "zeli(ku)" in another language. If we modified 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 and explains why it is possible to find several unfamiliar terms in the associated visualizations.

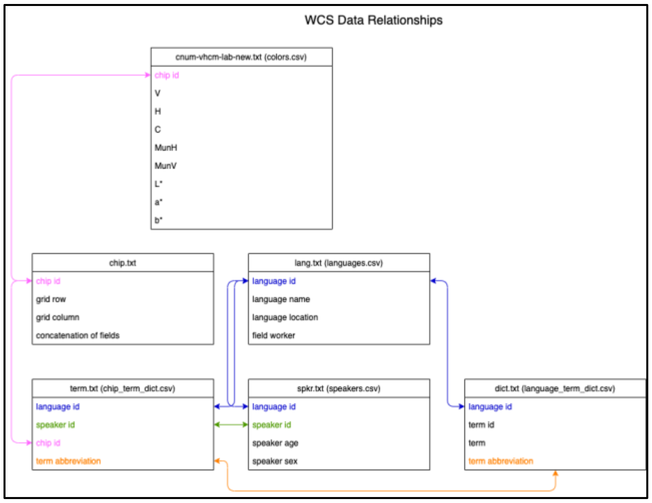

Figure 2. Diagram of field relationships in the WCS dataset.

4 Results

Though we focused on a different topic for this project, our team prioritized learnings from the A4 journey and generated three major visual interactions on our dashboard: a word cloud with the Munsell color palette, a tree map, and choropleth map. The following sections describe in detail our motivations and implementation processes.

4.1 Word Cloud with Munsell Color Palette

We applied the same, randomized order of the colors in the Munsell chart that were used in the first WCS experiment 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 11 – 20, then terms with 11 – 20 occurrences will have a larger font than terms with 0 – 10 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. We also got feedback to place the word cloud and the Munsell color chart side by side to make it intuitive that the graphs are unified tools. Thus, we explicitly placed them next to each other to dynamically show how the word cloud changes as the colors change.

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. We also received feedback from the MVP that emphasizing the current color is necessary to show, so we explicitly have a “current color selection” that highlights the current Munsell chip. The user is also encouraged to interact with the Munsell chart and with the word map with small and short-phrased instructions on the screen. One could imagine the user identifying similarities in how a single color is referred to across languages that do not share any written similarities or potentially even geographic similarities.

This word cloud addresses our team’s primary goal of showing patterns of how speakers around the world interpret different colors and with what frequency members of the same society use the same term to describe a single color.

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

4.2 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, a continent, a user can drill down to the lowest level of color terms in a particular language. This inspiration for this idea was also generated from the MVP feedback session. 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 used to describe all 330 Munsell color chips. Drilling into Mexico, there are many native languages like Mazahua, which contains 60 distinct terms used to describe the chips. 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. The insight from this visualization is that the number of distinct terms represented for a single country or language gives an intuition about how varied that language is; societies that used many different distinct terms to describe colors may have a more varied vocabulary or a more descriptive language from the perspective of visual perception.

One minor note to consider 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 also 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 to zoom in and chose this design technique intentionally. Overall, this tree map allows us to meet our goal of showing patterns across languages and geographic regions. It also provides an interactive opportunity for the user to understand how geographically diverse languages are and how respondents from the same society may or may not provide the same term for the same color. As discussed in the next section 4.3, the demographic data of respondents per country was also visualized.

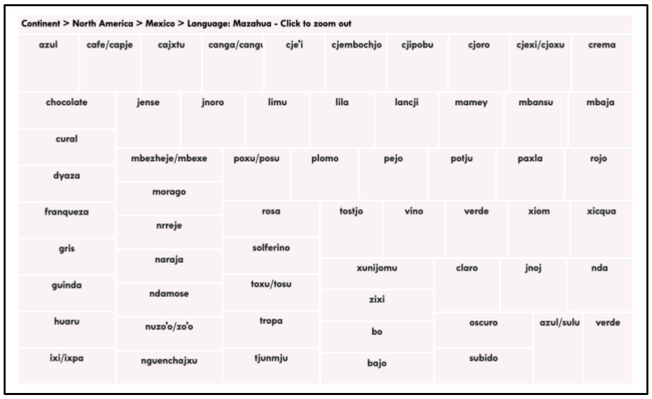

Figure 4. User drilling down to North America > Mexico > Language: Mazahua to see the term abbreviations.

4.3 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 and ages. We wanted to highlight the counts of females and males within each country and each language so as to indicate whether there were gender skews in the experimental set being considered; 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, one classmate gave us an idea to integrate the tree map with the world choropleth map. 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 implemented this suggestion and linked the tree map with the world map. Specifically, when the user clicks on a specific country in the tree map, information about the country on the world map shows up as text on the bottom of the world map, eliminating the additional need to move the mouse over to the world map to get additional information about the country on the world map. We placed them side by side to signal to the user that the two graphs are linked, similar to the Munsell color chart and word cloud positioning. However, we want to highlight that the linkage currently does not support continent level aggregations yet. The users must first hone in to the country level to see the two maps interact. Furthermore, there currently does not exist a link from the world map to the tree map. Although it would be sensible to have the tree map zoom in to a particular country when the user clicks on a country on the world map, we decided to leave this for future work due to the technical complexity.

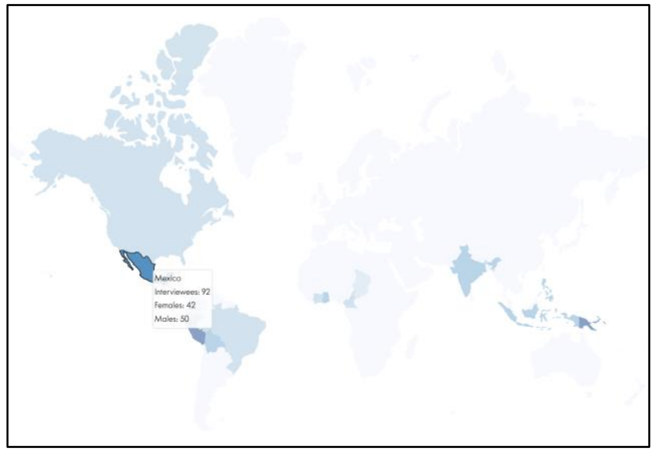

Figure 5. World choropleth map with tooltip showing details on interviewees from Mexico.

5 Discussion

Through our interactive visualizations, we hope the audience can understand our key theme: there is no fixed universal classification scheme of the Munsell color chips. Interpretations of color vary widely across languages and show similarities in some instances as well. Even members of the same society do not provide the same response for a single color, showing the breadth of language terms that can be used to describe a single visual element. Through the word cloud, a user can see how beautiful unwritten languages can be and how they are uniquely shared among speakers. The tree map shows that unwritten languages around the globe have strikingly different number of terms used to describe colors, giving us an indication that some languages might focus on visual descriptive capacities more than others. The word map also informs us that even across languages there are interesting similarities in how color is described. The world choropleth map shows that there are disparities among certain countries with an unbalanced male to female ratio.

In general, the WCS shows that Munsell chips are a fantastic way to interview people in color studies and that color naming does not have a one-fits-all answer. It is also generally interesting to note that even more than the interpretation of the colors themselves, the WCS has provided a way to analyze language through the lens of color. Much of the analysis and learnings gained while generating these visualizations were about the variety and comparability of languages, even those without a written form, that was made evident in the context of perceivable color chips. This suggests there might be a universality to color even greater than the hue, lightness, and chroma such that it represents a common thread among humans to recognize, and describe, different colors.

6 Future Work

Our team collectively spent around 150 hours for this final project. We see many possible extensions to this visualization dashboard and would've liked to implement the following suggestions, listed below (some inspired by our amazing classmates from the MVP feedback session). In addition to these, we feel that wholly separate, or integrated, visualizations could be added to the dashboard to represent the second experiment of the WCS that was described in the Introduction section.

1. Link the world map back to the tree map – if the user clicks on a country on the world map, have the tree map automatically filter / zoom to show the languages from the country the user clicked on
2. Have a visualization that summarizes the most common used terms to describe a particular color with filtering options such as by continent, country, and language

3. Add a search option for users to type in terms of interest, which in turn shows how many languages used the term and which colors were associated with the term
4. Implement a sticky header with WCS and the logo
5. Scrolly-telling feature of each of the sections in our dashboard
6. Cluster terms on the word cloud (e.g., by language family) so the user can make better sense of the terms
7. Include more information on the world choropleth map (e.g., how words from neighboring countries relate, which regions have the most color words, etc.)
8. Add in the ability to see continent level aggregated data on the world map as well when the tree map is on the continent level and a continent gets clicked on
9. Append a separate dataset to filter by different cultures, countries or language types (e.g. Indo-European)
10. Optional pop up to explain the WCS experiment and testing methodology

7 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