LA-UR-15-25715

Approved for public release; distribution is unlimited.

Title:	Analyzing task-based user study data to determine colormap efficiency
Author(s):	Ashton, Zoe Charon Maria Wendelberger, Joanne Roth Ticknor, Lawrence O. Turton, Terece Samsel, Francesca
Intended for:	Technical Exchange with Colleagues Report
Issued:	2015-07-23

Analyzing task-based user study data to determine colormap efficiency

Zoe Ashton
Florida Institute of Technology and
CCS-6, Los Alamos National Laboratory

Joanne Wendelberger
CCS-6, Los Alamos National Laboratory

Lawrence Ticknor
CCS-6, Los Alamos National Laboratory

Terece Turton
Center for Agile Technology, University of Texas

Francesca Samsel
Center for Agile Technology, University of Texas

Colormaps are widely used among domain scientists to visualize and understand their data, but traditional colormaps such as the cool/warm and rainbow are inefficient. This inefficiency stems from the small perceptual range of these colormaps which makes it difficult to discern an area of interest. To resolve this issue, new colormaps have been designed to correct these areas of visual deficiency. We conducted task-based user studies to determine the efficiency of these colormaps. This paper will discuss the methods of exploratory data analysis we used to compare the new and traditional colormaps as well as their results. Through graphical analysis and statistical tests, the traditional colormaps are shown to be less efficient than the new colormaps. A model has been created that allows insight into the effectiveness of different colormaps as well as the influence of age, gender, and education on colormap perception. This model provides a better overall idea of what affects colormap perception.

I. INTRODUCTION

Domain scientists need colormaps to visualize their data and are especially useful for identifying areas of interest, like in ocean data to identify eddies or characterize currents. However, traditional Rainbow colormap performs poorly for understanding details, because of the small perceptual range [1]. In order to assist domain scientists in recognizing and identifying important details in their data, different colormaps need to be applied to allow higher perceptual definition. Visual artist Francesca Samsel used her understanding of color theory to create new colormaps to improve perception. While domain scientists find the new colormaps to be useful, we implemented a rigorous and quantitative study to determine whether or not the new colormaps have perceptually more colors [2]. Color count data from one of these studies will be analyzed in depth in order to determine whether or not the new colormaps have more perceivable colors and what affects the number of perceivable colors.

II. STUDY DESIGN

We created the study using Qualtrics software. Qualtrics is a survey creation tool that allows for different kinds of survey questions. We distributed the survey through email solicitation and the University of Texas Psychology PSY301 Subject Pool. Later, we expanded the subject pool using Amazon Mechanical Turk. Amazon Mechanical Turk is a crowdsourcing online subject pool created by Amazon. People registered with the service participate in short tasks and are then compensated for their time. This study showed users a colormap, and then asked to click on as many distinct colors they could find in the colormap. Eight different colormaps were tested in this study where three are the traditional colormaps (Rainbow, Heat Map, and Cool/Warm) and four are the new colormaps (Gold/Grey, Autumn, Blue/Green Asymmetric Divergent, and Extended Cool/Warm). The four new colormaps were designed using perceptual theory concepts with the goal of being more effective. Effective maps are ones where users can identify more colors. The final colormap was the validation panel. This panel had a distinct number of colors and we used this map to scrub the data for bad participants.

Because the task in this study was simple and there were only eight colormaps, each respondent was asked to do every colormap. The resulting data included an integer count of the colors in each colormap, and integer values for the respondents' age, gender, and education level. Participants self-identified their demographic information. We binned this demographic information according to categories (Appendix). We used an earlier data set for these analyses. This data set included 77 of the final participants. This number was cut down to 63 after validation. We used the validation panel to ensure that every participant included in the final dataset was capable and willing to provide reliable data. Any respondent who was not within +/-2 colors of the validation panel was determined to be unreliable and was removed from the data set.

III. COMPARING COLORMAP COUNTS

We first started our analysis by comparing the colormap counts to each other graphically. We created both boxplots and density functions to perform this graphical analysis. This allowed us to identify the colormaps that seemed to have the highest number of perceivable colors. From the boxplot (Figure 1), the new colormaps appear to have higher medians than the traditional colormaps. We reach similar conclusions by looking at the density functions (Figure 2). It is especially clear that the Blue/Green Asymmetric Divergent has the higheestcounts in this graph.

It stands out against the others with the highest densities at larger counts of perceivable colors. We see that most of the new colormaps do have higher densities at higher counts. The traditional colormaps also peaked at lower counts than the new colormaps. This is a promising start to prove that the newer colormaps have more perceptual colors.

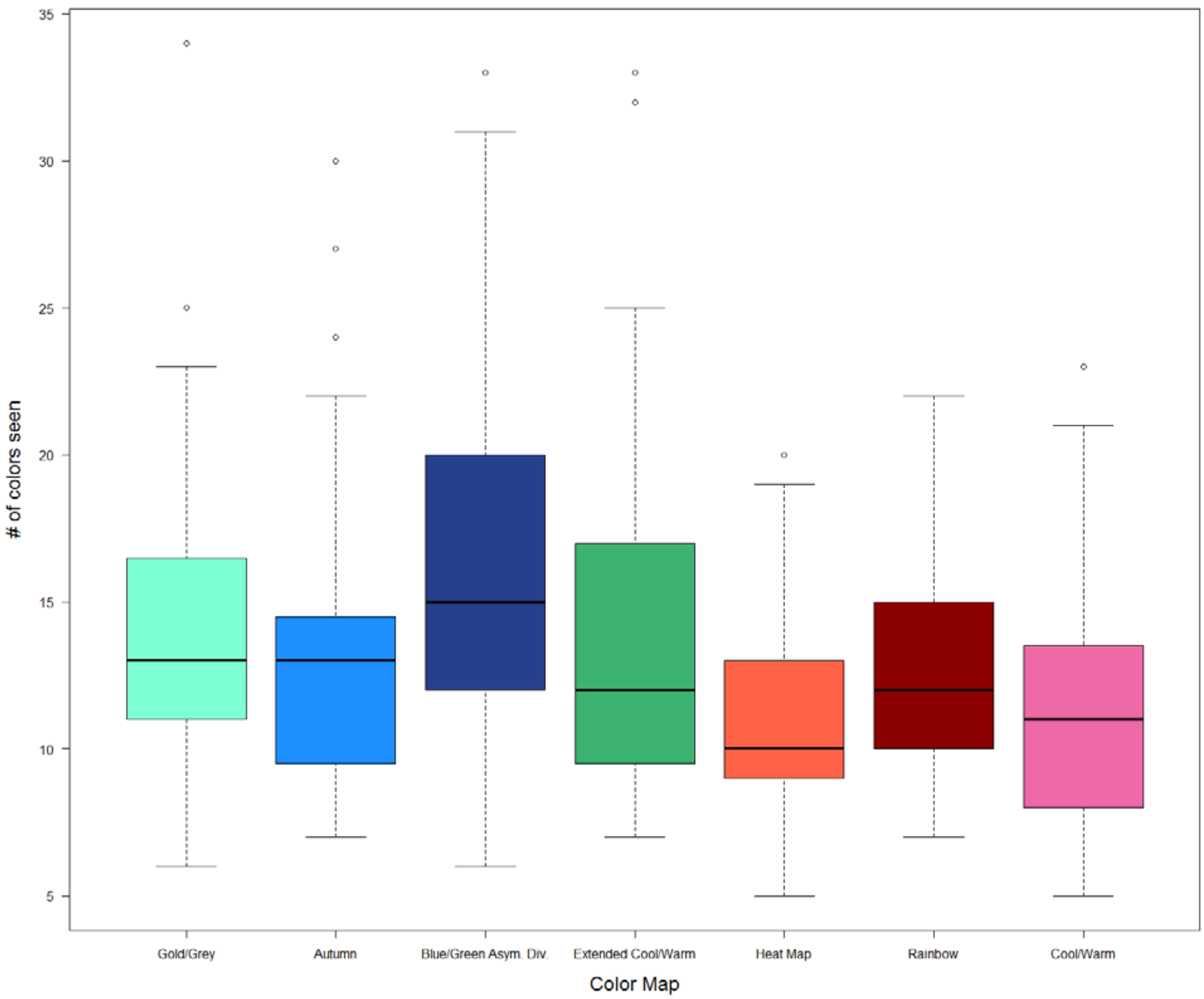

Figure 1: Boxplot of number of colors seen. Blue-toned boxes represent new colormaps. Red-toned boxes show traditional colormaps.

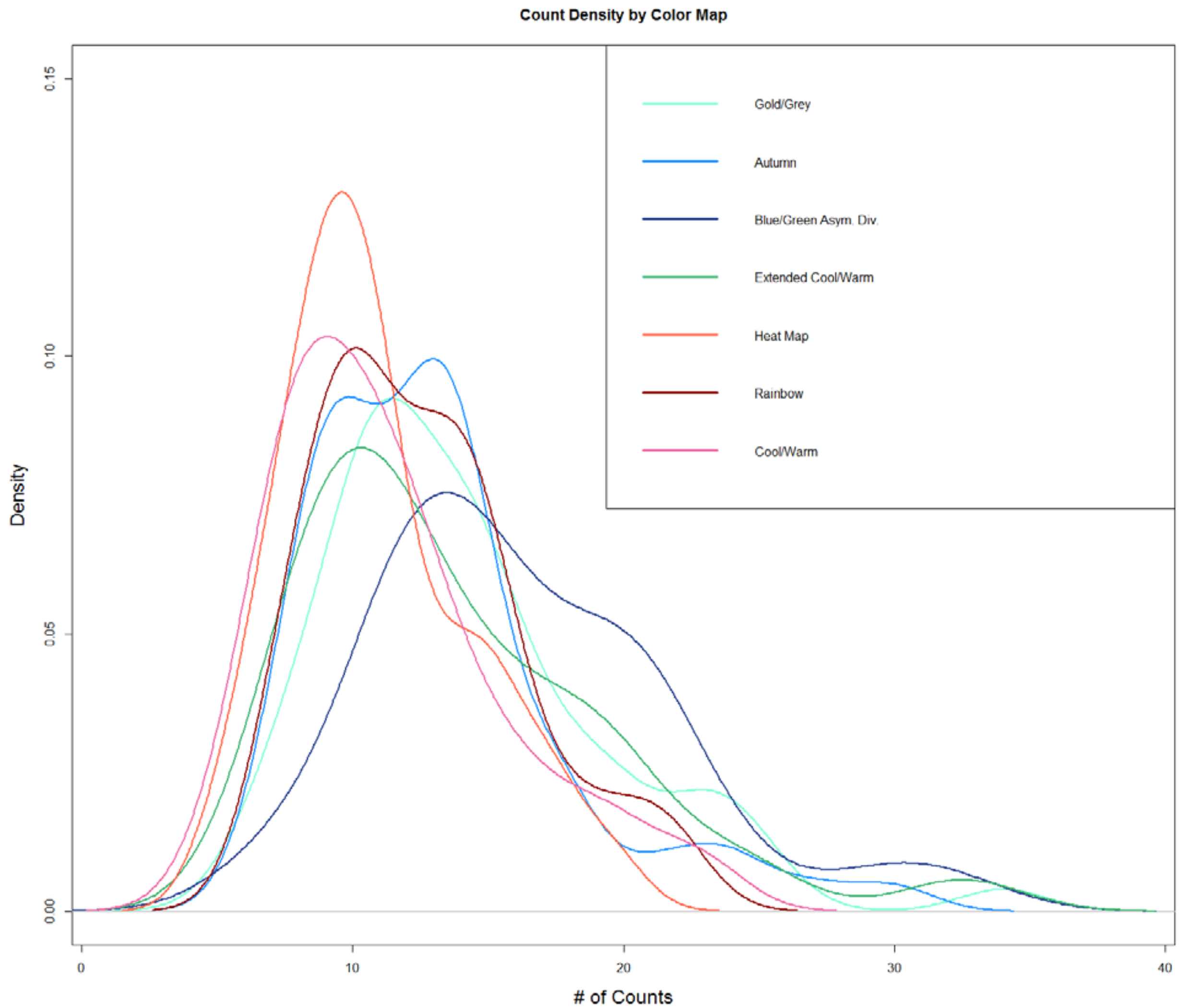

Figure 2: Color coded density graphs for each color map.

The graphical analysis showed that the new colormaps have more perceivable colors. But in order to be certain, we decided to perform sign tests. For this case, we used a a one-sided sign test, because the data is non-parametric and we wanted to determine if one colormap has more colors than the other. The null hypothesis for the sign test is that colormap X's counts are the same or lower than colormap Y's counts, where colormap X is the row name and colormap Y is the column name in Table I. The alternative hypothesis is that colormap X's counts are higher than colormap Y's counts. Table I displays the p-values for these tests; we chose to use an alpha level of 0.01. In all cases, the null hypothesis can be rejected for Blue/Green Asymmetric Divergent. Gold/Grey does better than all of the traditional colormaps. Autumn and Extended Cool/Warm both do better than Cool/Warm and Heat Map. But the p-values for Extended Cool/Warm and Autumn versus Rainbow showed that we cannot reject the null. We cannot say that the Extended Cool/Warm or the Heat Map has more perceivable colors than the Rainbow colormap. We took into account the concerns involved with multiple testing. But since the p-values in all of the cases being rejected are extremely small, we can say that there is a significant difference between the colormap counts.

Table 1: Showing p-values for the one-sided sign tests.

	Cool/Warm	Rainbow	Heat Map	Extended Cool/Warm	Autumn	Gold/Grey
Blue/Green Asym. Div.	1.59E-12	6.80E-10	8.21E-16	3.78E-07	8.37E-08	2.61E-05
Gold/Grey	1.63E-09	6.38E-05	4.53E-09	0.110	9.92E-03	
Autumn	2.05E-04	0.292	3.67E-05	0.965		
Extended Cool/Warm	1.02E-06	0.248	1.36E-08			
Heat Map	0.920	1.00				
Rainbow	7.69E-05					

IV. EFFECTS ON COLORMAP COUNTS

While we are concerned with the number of perceived colors, we want to ensure that demographics are not a factor as well.The overall goal is to have colormaps which have more perceivable colors for domain scientists. The problem is that it’s difficult to have enough domain scientists take studies repeatedly until we have a definitive answer about which colormap is better. User studies with different participant demographics are the best option available. In order to use these, though, we need to know how much of an effect the respondents’ education, age, and gender make on their color counts. First, we used graphical analysis to understand which participants perceived colors in a similar way. Then, we created a linear model. This model allowed us to further understand the interaction between perceived color count, colormap type, age, gender, and education of participants.

Much like in the first case, we started with graphical analysis to obtain an understanding of the data. In this case, dendrograms were the most appropriate graph. Dendrograms are a type of graph where participants are clustered together hierarchically to show which participants are the most similar. Clusters near the bottom of the graph are the most similar and clusters that occur at the top of the tree are the least similar. We used euclidean distance for these dendrograms to calculate how different each participant was from every other participant. To calculate this distance, only the seven colormaps of interest were used. The validation panel as well as age, gender, and education were left out of the distance calculation. We removed these values because we only wanted to see how the actual responses clustered together. If age, gender, or education were left in, they would create clusters that were close in those aspects too. We then applied the average clustering method to create the dendrograms which is the most robust method for clustering. We displayed the age, gender, and education on separate dendrograms so clustering trends would be most obvious. Difficulties arise when looking at the gender dendrogram (Figure 3) where we see the 2’s cluster together. The problem is there are more 2’s than 1’s, limiting how much the dendrogram can show. But the other two dendrograms are different (Figure 4 and Figure 5), because there is a good mixture of clusters. For almost every cluster that's homogeneous, there's another cluster - at about the same height measure - that has a variety of different participant types. This provides us a good idea of how the participants' color counts match up with each other. Participants' demographics appear to have little effect on the final perceived color counts.

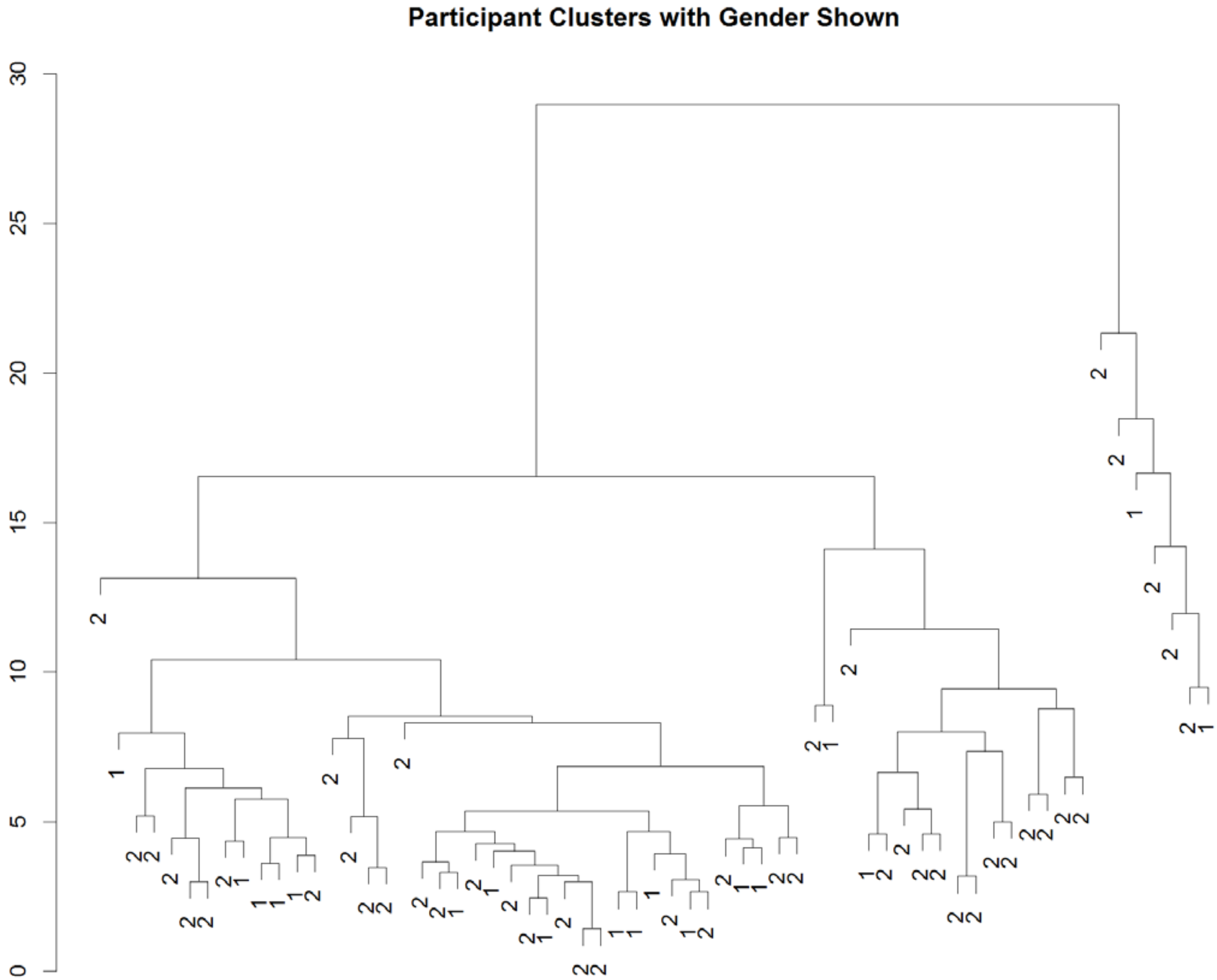

Figure 3: Dendrogram of participant clusters showing the gender of participants.

After graphical analysis, we created a linear model that allowed us to have a more robust understanding of the impact of colormap type, age, gender, and education on color counts. Each of the colormaps, except validation, was made its own categorical variable. We also made age, gender, and level of education their own variable, but they kept the parameterization from the study definition (Appendix). We used the least squares method to estimate a linear model based on 10 variables.

$$Y = 14.012X_1 + 12.9728X_2 + 16.1502X_3 + 13.5695X_4 + 10.6663X_5 + 12.3760X_6 + 11.1663X_7 + 0.2335X_8 - 0.0941X_9 + 0.0682X_{10} + e$$

Y is the perceived number of colors and the Xs match up to different variables. X_1 is the Gold/Grey colormap, X_2 is the Autumn, X_3 is Blue/Green Asymmetric Divergent, X_4 is Extended Cool/Warm, X_5 is the Heat Map, X_6 is the Rainbow, X_7 is Cool/Warm. Each of these was 0 if the count wasn't for that colormap or 1 if it was. X_8 was gender with original value options 1-3. X_9 is education with original values 1-8. X_{10} is the participant's age which could take values 1-7. We performed a t-test on the coefficients of this model in order to understand what effect each of these variables has on the final count. The null hypothesis is that the true coefficient value is zero. The alternative is that the true coefficient is not zero. If the true coefficient can be non-zero, then the variable has a significant effect of the perceived number of colors. Table II displays the t-values and p-values that result from this test; we used an alpha level of 0.01. The test results confirmed what was seen in the earlier graphical analysis. We can't reject the null hypothesis for the demographic coefficients. This shows that the coefficients for gender, education, and age could be zero. We can, however, reject the null hypothesis for all of the colormaps. Age, gender, and education didn't significantly effect the number of perceived colors. The colormap type did make a significant effect on the color count.

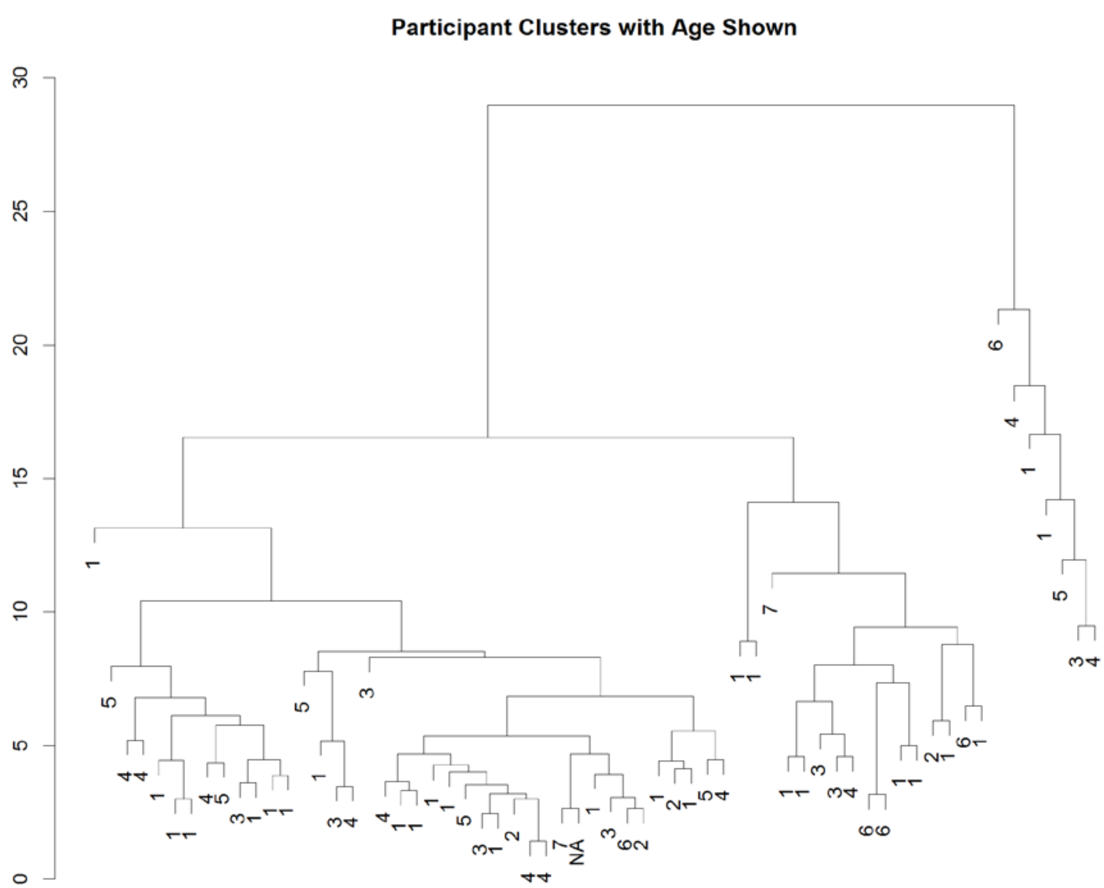

Figure 4: Dendrogram of participant clusters showing the ages of participants.

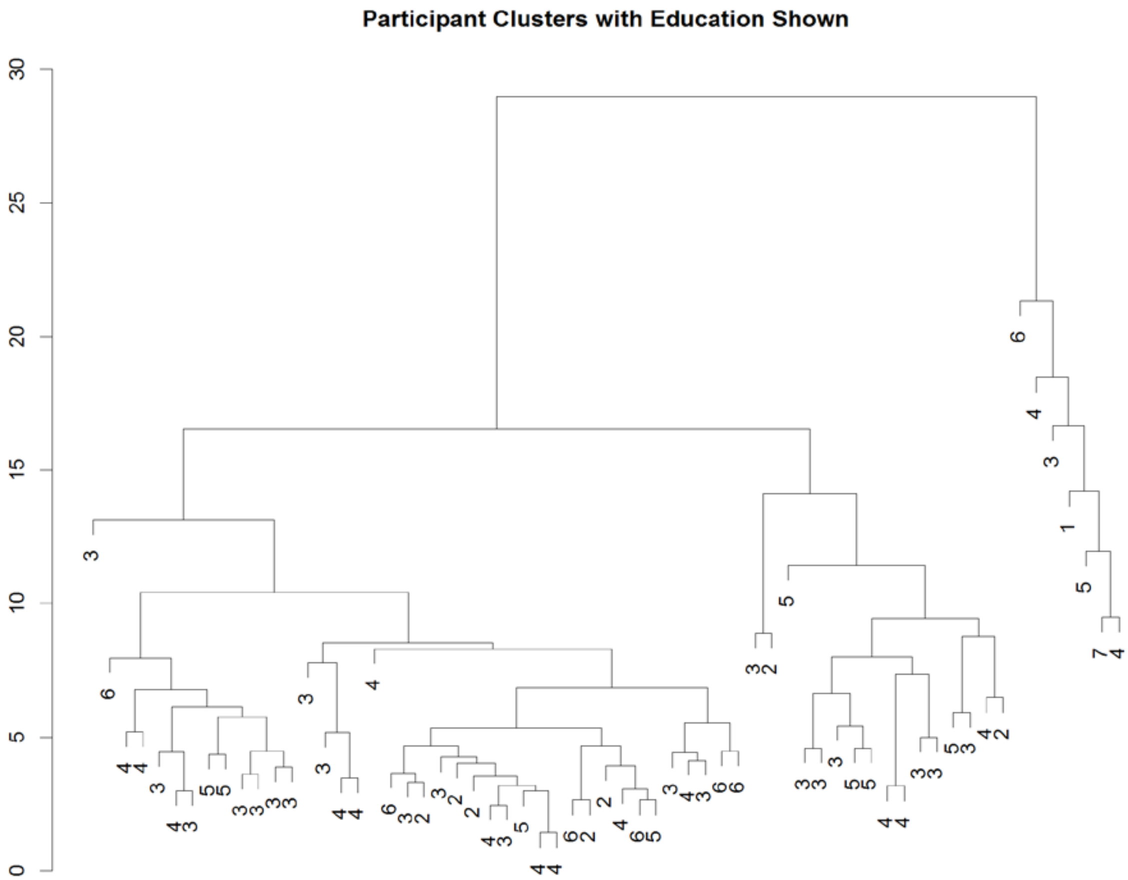

Figure 5: Dendrogram of participant clusters showing the education level of participants.

V. CONCLUSIONS AND FUTURE WORK

Data analysis for this color study answered two major questions: are the new colormaps producing more perceivable colors and what affects the number of perceived colors? By performing sign tests to compare the colormap counts, we discovered the new colormaps overall have more perceivable colors than the traditional colormaps. In addition, these tests showed that the Blue/Green Asymmetric Divergent colormap performed the best. This result showed which colormap should be investigated and developed further to be better. Creation of a linear model made it possible to understand the effects that age, gender, and education have on the perceived number of colors. From the t-tests performed on the coefficients, we determined that there is no significant effect on the perceived number of colors from the gender, age, and education of participants. This is a promising result because the ultimate target group for the new colormaps are domain scientists. If there is no significant effect from demographics, then there is little difference between perceived color count for domain scientists and the average population. Therefore, results from colormap user studies should translate well to domain scientists.

Table 2: Showing the t-value and p-value for the t-test on the coefficients.

	t Value	Pr(>|t|)
Gold/Grey	11.309	<0.0001
Autumn	10.463	<0.0001
Blue/Green Asym. Div.	13.026	<0.0001
Extended Cool/Warm	10.945	<0.0001
Heat Map	8.604	<0.0001
Rainbow	9.983	<0.0001
Cool/Warm	9.007	<0.0001
Gender	0.445	0.656
Education	-0.390	0.697
Age	0.422	0.674

Color count is not the only factor that affects the usefulness of a colormap for domain scientists. Future work will involve testing other colormaps and different tasks. Another factor to be further investigated is demographic background. While age, gender, and education proved to be insignificant in this data set, other data sets will need to be evaluated for this as well. Furthermore, while they may not affect color counts, they may affect other types of data that will be tested later. A final concern for these studies is color blindness. In the future, the goal is to see how color blindness affects colormap perception in order to create colormaps that can be useful for many different groups.

VI. ACKNOWLEDGEMENTS

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

APPENDIX: USER STUDY VARIABLE DEFINITION

Gender:

- 1: M
- 2: F
- 3: Other/Prefer not to respond

Age:

- 1: 18-25
- 2: 26-30
- 3: 31-40
- 4: 41-50
- 5: 51-60
- 6: 61-70
- 7: 71+

Highest Level of Education:

- 1: Some High School
- 2: High School Diploma or GED
- 3: Some College or Associate Degree
- 4: Undergraduate Degree
- 5: Masters Degree
- 6: Doctorate or Professional Degree
- 7: Other (please explain)
- 8: Prefer not to respond

REFERENCES

[1] Rogowitz, B., Treinish, L., “Data Visualization: the end of the rainbow.” *Spectrum, IEEE* (1998) 35:12(52-59).

[2] Francesa Samsel, Mark Petersen, Terece Geld, Greg Abram, Joanne Wendelberger, and Jim Ahrens. Colormaps that Improve Perception of High-Resolution Ocean Data. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’15, pages 703-710. 2015.