



Color Aesthetic Enhancement for Categorical Data Visualization

Wei Chuan Lim¹(✉), Chee Onn Wong¹, and Lai Kuan Wong²

¹ Faculty of Creative Multimedia, Multimedia University, Cyberjaya, Malaysia
cowong@mmu.edu.my

² Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia
lkwong@mmu.edu.my

Abstract. Data visualizers are usually experts in statistics and the field they involved. Most of them have lack of knowledge in arts. This leads to the condition where visualization created by them most of the time are not pleasant in terms of color, or they will be limited to the palette choices provided in the platform if they want to get a better output. This reduced the potential of visualization to act as an effective medium for marketing or awareness-raising purposes. In this paper, we study on the coding of colors in the way that is closer to human perception, together with the concept of color harmonization based on existing research. By integrating them, we get a framework that can retrieve the range of colors that looks harmony based on any request color. Our aim is to enhance the aesthetics and beauty of data visualization diagram through color modification. In the process of harmonizing the colors, our approach uses a distance scaling method on the hue dimension. This approach can better preserve the intended relationship between different colors from the original visualization. In most cases, the scaling process would be scale down, decreasing the distance between colors. Therefore, we need to take additional precautions to make sure that the scaled colors can still be perceived differently. We conducted a color difference calculation on all colors with the colors that are closest to them. Through the numerical method, we can set a minimum value and computationally identify that whether does the two colors are safe enough to distinguish. The visualization can perform an entire hue shifting process by adding a constant value to the hue of all colors after being harmonized through our approach. Our proposed approach helps data visualization artist or automated program to create a more color harmonized output, as well as providing the ability to freely change its entire color theme, which is useful to match the presenting environment, without needing to consider the problem of color difference.

Keywords: Visualization · Color harmony · Human visual perception · Aesthetic

1 Introduction

The technology nowadays is matured enough to generate data visualization diagrams computationally with minimal human effort, and the outcome has reached a certain level of satisfaction. Apparently, the demands of data visualization skills will go on for the following decades as data processing and analysis is continuing to emerge in many new fields. Apart from the professionals that highly depending on data visualization in their work, business companies also started to use data visualization as a persuasive method while approaching their clients. Besides that, organizations are starting to show data visualization diagram to the public to raise awareness regarding some specific topics too. Therefore, the beauty of data visualization has become more important, as it needs to ensure the visualization is effective, in such a way that it can attract the target audiences' attention and enhance their interest to read it.

One of the key factors to evaluate the aesthetic value of visualization is the color. Based on Burchett [1], colors seen together to produce a pleasing affective response are said to be in harmony. In data visualization, color is often use as a notation for data classification, or representing one or multiple dimensions of the data. Meanwhile, visualization that is constructed with harmony colors will be more appealing.

Visualization tools such as ColorBrewer and the charts function in Microsoft Excel provides several choices of preset color palettes that seems pleasant. However, choices are still limited, and fine adjustment process will be challenging when the creator is required to follow certain criteria in color for specific data category. Besides, color palettes generated through algorithmic method such as the linguistic approach [2, 3] are also lacking in harmonious value when they are viewed together.

In this paper, we are going to improve the pleasantness of categorical data visualization diagram by modifying its colors, so that it will be more attractive, and readers will feel more comfortable while looking at it, under the premise that the modified colors will not lose their original role. Our main intention of this work is to provide a significant improvement in terms of aesthetics for the continuous development of automated data visualization.

2 Background and Related Work

Our work consists of color transform based on harmonic schemes, with the integration of perceptually uniform color space.

2.1 Color Theory and Color Harmonization

Color theory is the set of guidelines on mixing various colors onto a visual presentation to imperceptibly leads the viewer to perceived the feeling based on the creator's intention.

Using the layout of color wheel, people starts to discover the relation between colors based on their position in terms of rotation angle and their effects. There are five widely accepted basic color schemes adopted in Adobe Color, which consists of analogous, monochromatic, triad, complementary, and split complementary.

Color harmony is one of the topics that always being discussed and generally adopted by designers for arts creation. Matsuda [4] proposed that a creation looks harmonious if all the colors exist in the creation follows certain patterns on a hue plane. Based on that, he developed eight harmonic patterns as shown in Fig. 1. These patterns are mostly named using the alphabets that look like their appearance (i, V, L, I, T, Y, X) on the plane, and the only exception (N) is a Neutral type that is used for the grayscale (between white and black) color sets. In Matsuda’s research, the patterns can be rotated to any angle based on user preferences, and the harmonization value should be preserved. These harmonic patterns are widely used in the later research till now [5–8]. Cohen-Or [5] showed the harmonized output of photographs and posters, [6] emphasize the use of those patterns to generate color palette for fashion design, [7, 8] referred to the patterns in the creation and assessment of data or information visualization.

Carlos [6] categorized five out of the eight patterns into three basic color schemes – analogous (i, V), complementary (I, X), and triad (Y). However, we preferred to consider Type-i as monochromatic, which works better with the changes in tone (saturation and lightness) due to the potential ambiguity of minimal hue difference. Type-Y also should be better considered as split-complementary based on the shape of its pattern, and Type-X should be a combination of analogous with complementary.

In this paper, we use Matsuda’s harmonization approach to enhance the aesthetic value by recoloring the existing visualization into one that adheres to the harmonization pattern.

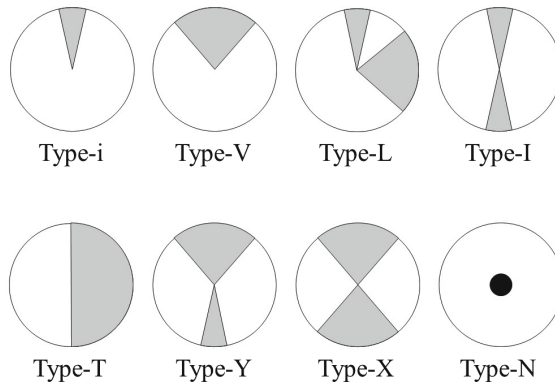


Fig. 1. Matsuda’s color harmonic templates on a hue wheel.

2.2 Perceptually Uniform Color Space

During the process of choosing colors for arts creation, various kinds of color wheels will be used by different artists. As the common ones will be the RGB and RYB color wheels. The visible difference between those color wheels are the distance between hues. For example, distance between red and green in RGB is 120° , while the distance of that in RYB is 180° . Therefore, the color palette output will be different when the same color

scheme is applied on these color wheels, and we need to consider the best color wheel for human perception.

In the first decade of 20th century, Albert H. Munsell introduced the Munsell color system. The Munsell system defined colors in a three-dimensional space, the hue value changes along the angle on a hue circle plane, contrast value depending on the distance from the center of circle, and lightness changes following the height of the hue circle. As compared to RGB and RYB, Munsell system divided the hue circle into five reference hues which are red, yellow, green, blue, and purple. Approximately, the angle distance between red and green is 144° . It is the first color space that take human perception into consideration [9]. Matsuda's color harmonic templates are defined based on this color system.

In this paper, we attach to the importance of user intuition on the relation between colors. The CIELAB color space defined later by International Commission of Illumination (CIE) is intended to be a perceptually uniform color space. By showing the hues on a color circle or hue plane, humans should perceive the same difference between two colors in any angle rotation when their angle distance is the same.

2.3 Visualization of Categorical Data and Continuous Data

In data visualization, the type of data determines how it should be presented. Color can be used in various ways to represent data.

For categorical data, different color will be representing different data group for readers to compare between the underlying value. The data group is distinct and there is a finite number of colors exist in the visualization (usually less than 10). For example, we use a pie chart to show the number of different fruits in a collection. Red will be suitable to represent the apple category whereas yellow will suit for banana. Based on the difference in size of the red pie and yellow pie, we can compare the percentage of both fruits in the entire collection. The choice of color for categories is important that it should highly achieves the color-concept associations [10], and research such as [2, 3] are relying on machine learning methods to assign the most suitable color to the respective terms.

The use of colors to represent continuous data usually will make use of tone (chroma and lightness) changes. Normally a high tone indicates a positive or active value of data and vice versa, and the range between the preset highest and lowest value consists of all the values available. Theoretically, the available values are potentially infinite. However, due to the limitation of colors in all kinds of displays and ease for mathematical calculation, the value will be rounded onto a specified precision. But it can still be hundreds of available colors between the highest and lowest tones. For continuous data, as in Fig. 2 [11], the actual value of each single data sometimes is not important. The purpose of this kind of diagram is more to visualize the collection as a whole to show the trend of data changes.

Our proposed approach is focusing on categorical data, where each color usually symbolizes a different class or group. In this type of data, hue variation possesses a higher importance compared to tone.

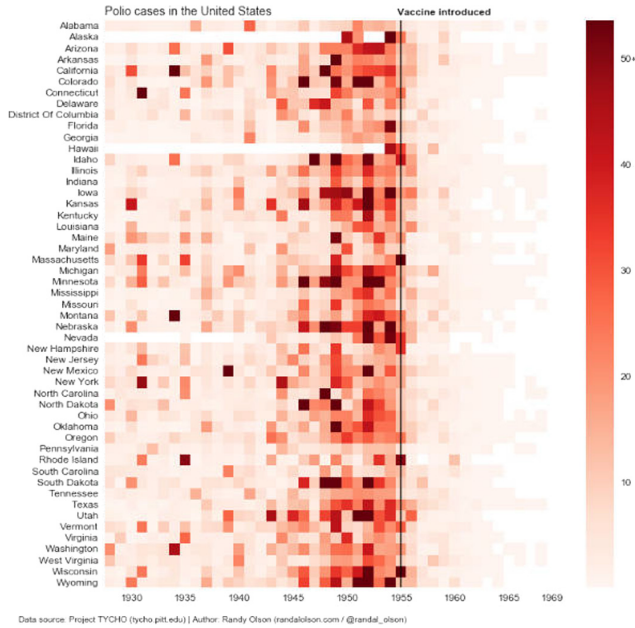


Fig. 2. Example of continuous data visualization which focus on the whole view rather than providing the ability to look for the precise value.

2.4 Color Discriminability

A main concern of constructing palette for visualization is the ability to identify different data group immediately based on the colors. Color appeared differently to human when the color angle changes [12]. A simple way of color discriminability evaluation can be done by calculation based on the Euclidean distance between two colors in a color space, known as the color difference.

Besides the standard calculation, CIE had introduced some formulas based on their CIE color spaces, more concerning on perceptual uniformity, and named their distance metric as ΔE^* . $\Delta E^* = 1$ supposed to be the Just Noticeable Differences (JND), which is the minimum difference that people able to distinguish between two different colors at 50% of all the time. [13] found out the 50% noticeable difference for colors varies under different visualization type and colored areas. This paper then has conducted tests to calculate the 50% noticeable difference for points, bars, and lines type of data visualization in different sizes.

Maji [14] proposed an Equilibrium Distribution Model (EDM) to achieve a higher color discriminability between different sections in visualization. This approach has shown a higher effectiveness on image that consists of more than 30 features to be differentiated, and is theoretically proven will be applicable on image with 100 features. However, a perceivable color difference for usual data visualization diagram need to be guaranteed when compared with any color in the whole diagram. This is because we are not just considering the discriminability between two colors when comparing side

by side. We also need to make sure that the readers will not be confused when they are referring different colors from the legend.

3 Methods

Our enhancement methodology adopts the idea in [15, 16], where the harmonized palette is generated by extracting color from the original visualization and perform adjustment based on specific algorithms. This approach can be better to retain the original color characteristics.

The colors in an existing visualization image can be retrieved through k-means algorithm. This algorithm is able to look for the set of most representative colors in an image. Then, on a hue circle each color will move towards the nearest segment of the harmony pattern, decided by which segment has the least hue difference from the original color. The colors are listed in lightness, contrast, hue form according to the CIELCH (polar form of CIELAB space) measurement.

Based on the chosen color harmony pattern on a hue plane, we look for the two furthest colors that will be grouped into the same region. Then, we scale the hue distance down until the two colors reach the highest acceptable distance of that region. The scale then will be applied to all the other colors with the point next to it in anticlockwise direction. The output of this process will remain the color differences under the same scale and all colors are able to be fully covered in the harmony segments.

3.1 Hue Values Normalization

Before starting the harmonization, we need to take a preliminary step to make sure the hue is valid for calculation. For a set of colors that going to be merged into the same segment, if the $0^\circ/360^\circ$ border lies between the two furthest hue point (in terms of degree), the hue value of colors after the border will start from 0° , thus will not be continuous following the anticlockwise direction. Therefore, for each color we need to get the continuous hue values H_c based on the following condition.

$$H_c = \begin{cases} H + 360, & \text{if } H < H_0 \\ H, & \text{otherwise} \end{cases}$$

H represents the original hue and H_0 is the first hue of the set according to anticlockwise direction.

3.2 Hue Transformation

In data visualization, the similarity level between colors that representing data categories may consists of some underlying intentions. Colors that looks nearer to each other is expected to have a closer relationship. This information should not be lost after the aesthetic enhancement. For the hue transformation process, we use the distance scaling approach so that we can preserve the distance ratio between different colors.

For each segment, given a hue range limit l . We can get the hue scale s using the following equation:

$$s = \frac{l}{H_n - H_0} \quad (1)$$

where the set $H = \{H_0, H_1, H_2, \dots, H_n\}$ is the list of hues (adopt H_c if it is different from the original hue value) of all colors to be grouped, in ascending order. H_n represents the highest hue and H_0 is the lowest hue.

The first hue H'_0 will remain the same while applying scale transformation, then for all following hues, the harmonization value, H' can be calculated by.

$$H'_i = H'_{i-1} + (H_i - H_{i-1})s \quad (2)$$

After the harmonization scaling is done, we need to shift the colors to match the pattern by adding a constant value to all the hues in H' . This step can be skipped for those patterns that only have one segment (Type-i, Type-V and Type-T).

By moving along the hue plane with a constant distance from the origin, only the hue value will be changed, the chroma and lightness remain the same as the original color.

Additionally, by adding the same shift value to all the hues on the plane, we can get the same effect as rotating all colors on a hue plane, giving the effect of shifting the hue of the entire output.

3.3 Minimum Color Difference

In 2001, CIE published the CIEDE2000 formula after the CIE76 and CIE94 for distance metric calculation. CIEDE2000 is the to date formula that best resolve the perceptual uniformity issue [17, 18]. The distance matric value can be calculated with the formula.

$$\Delta E_{00}^* = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{k_C S_C}\right) \left(\frac{\Delta H'}{k_H S_H}\right)} \quad (3)$$

$\Delta L'$, $\Delta C'$, $\Delta H'$ are the lightness, contrast and hue difference of the two colors for comparison. k_L , k_C , k_H are the correction factors of the observation environment. S_L , S_C , S_H are the compensation for visual perception. R_T is representing the hue rotation term to overcome the blue region problem.

In this research, we adopt a minimum color difference (ΔE_{min}^*) of 10 between each color to ensure that the color is obviously distinct when referring to the legend. This value is an integer round up based on the noticeable difference found out in [13] for bar charts with the visual angle of 2° . In case of low color difference detected, manual shifting process should be done by human action.

3.4 Color Theme Modification

Based on the CIEDE2000 formula, we acknowledge that in the CIE color space, the calculated color difference will always be the same when a constant value is added to (or subtracted from) either the lightness, contrast, or hue of all the colors.

In the perspective of hue, which ranged between 0° and 360° , adding or subtracting a constant value of all colors is same as the process of rotating all colors together on a color hue wheel. This rotation will change the theme and feel of the visualization, while ensuring that all colors can still be perceived differently.

4 Results

Based on the methodology, we apply our approach on a sample data visualization chart with random picked colors. All the colors are collected and ordered based on the hue value. Type-V pattern is used for this harmonization. Due to the V segment crossing the $0^\circ/360^\circ$ of hue circle, we added 360 to hue value of the colors after the 0° border to make the value continuous.

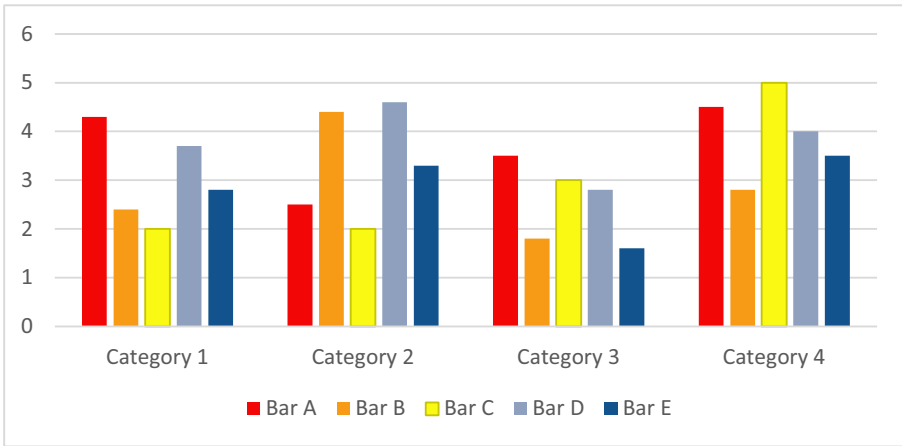


Fig. 3. Sample of categorical data visualization diagram to be harmonized.

Using Eq. (1), we set the hue range limit l to 60° , and we will get the value of s as 0.312, rounded to 3 decimal places. Based on the s value we get, we apply harmonization with Eq. (2). The harmonized hue values for each color (H') are listed in Table 2. For this example, we take an additional step by adding a constant shift of 88.548 to all hues to remain the third color (red) constant after harmonization. By referring that as the base point, all other hues will move towards it after transformation.

Based on harmonized color Table 3, we calculate the ΔE^* between each color and the color that is nearest to it in anticlockwise direction using Eq. (3). Hue value is retrieved from Table 2 while lightness and contrast is referring from Table 1 because our harmonization method did not make any modifications on the lightness and contrast value.

All color differences are higher than 10. This means that the harmonized outcome of this visualization can be perceived clearly and will not confuse the readers in terms of color discriminability.

Table 1. Lightness, contrast, hue and continuous hue values of the data representation colors in Fig. 3.

Bar	Color	Lightness	Contrast	Hue (H)	Continuous hue (H_c)
D	Light Blue	65	17.047	270.682	270.682
E	Dark Blue	35	38.283	274.999	274.999
A	Red	51	99.541	39.471	399.471
B	Orange	72	77.742	70.435	430.435
C	Yellow	95	93.574	102.709	462.709

Table 2. Original hue and transformed hue values after harmonization of the data representation colors in Fig. 3.

Bar	Original hue (H)	Harmonized hue (H')	Transformed hue after harmonization	Final color
D	270.682	270.682	359.230	Light purple
E	274.999	272.031	0.579	Dark purple
A	39.471	310.923	39.471	Red
B	70.435	320.598	49.146	Orange
C	102.709	330.682	59.230	Light orange

Table 3. Color difference of each harmonized color with the color next to it in anticlockwise direction on a hue plane.

Base Bar	Bar with color next to the base bar color in anticlockwise direction on hue plane	Color difference (ΔE^*)
D	E	31.570
E	A	29.445
A	B	20.173
B	C	17.734
C	D	38.451

Figure 4 shows the outcome after harmonization. As compared to Fig. 3 the original version, this outcome is presented with a more comfortable look and feel, and still able to perform well as a data visualization diagram, showing the classification of data clearly.

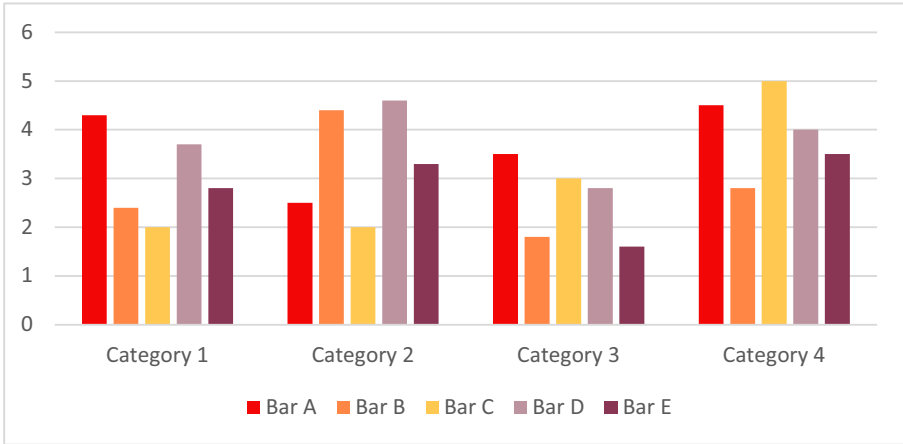


Fig. 4. Harmonized data visualization diagram of Fig. 3.

Using the harmonized visualization in Fig. 4, we can perform a color theme modification process by shifting all the hues of the colors in the entire image. Figure 5 is the outcome after applying a hues rotation of 180°. This rotation can be interpreted as an action to convert the color temperature from a warm color set (as in Fig. 4) to a cool palette.

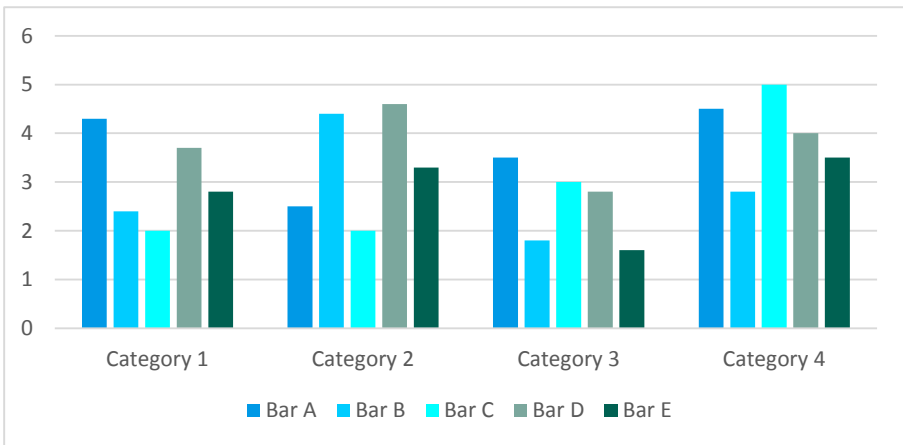


Fig. 5. Hues shifted visualization outcome after harmonization process.

5 Conclusion and Future Works

The result shows that our method can enhance the aesthetic value of categorical data visualization. By grouping the hues based on a certain pattern, we can improve the harmonization of visualization. Choosing the appropriate hue pattern also will help to indicate the relation between data categories, presenting more information to the viewer. However, this selection process still requires human action because humans can comprehend the relation of the categories better.

The proposed method does not take lightness and contrast into consideration, maintaining their original value in harmonization process. The utilization of lightness and contrast dimension has the potential to apply a monochrome scheme (Type-i) to show a different kind of data relationship. This will be helpful when showing numerical data. Moreover, by using hue and tone in the same visualization, we can increase the dimension of data representation. For example, different hue may represent different data category, while tone symbolize the numerical value. Further research can be done to identify the practicality of this concept and improve our model.

Beside the aspect of harmonization, it is also important that the readers need to be able to discriminate colors that represent different categories. Therefore, the transformation methodology needs to consider the minimum color difference.

For our future work, we plan to construct an algorithm that can be applied on the colors to computationally adjust them until all the colors adhere to the minimum perceptual difference, while preserving the aesthetic value as a whole. This will make our approach more integrated to achieve the realization of automated data visualization.

Acknowledgement. The authors would like to express appreciation for the support of the Fundamental Research Grant Scheme [FRGS/1/2019/SSI07/MMU/02/1] in providing adequate resource and guidance to complete this research.

References

1. Burchett, K.E.: Color harmony. *Color. Res. Appl.* **27**(1), 28–31 (2002). <https://doi.org/10.1002/col.10004>
2. Lin, S., Fortuna, J., Kulkarni, C., Stone, M., Heer, J.: Selecting semantically-resonant colors for data visualization. *Comput. Graph. Forum* **32**(3pt4), 401–410 (2013). <https://doi.org/10.1111/cgf.12127>
3. Setlur, V., Stone, M.C.: A linguistic approach to categorical color assignment for data visualization. *IEEE Trans. Visual Comput. Graph.* **22**(1), 698–707 (2015). <https://doi.org/10.1109/TVCG.2015.2467471>
4. Matsuda, Y.: Color design. *Asakura Shoten* **2**(4), 10 (2002)
5. Cohen-Or, D., Sorkine, O., Gal, R., Leyvand, T., Xu, Y.Q.: Color harmonization. In: *ACM SIGGRAPH 2006 Papers*, pp. 624–630 (2006). <https://doi.org/10.1145/1179352.1141933>
6. Lara-Alvarez, C.: Reyes, T: A geometric approach to harmonic color palette design. *Color Res. Appl.* **44**(1), 106–114 (2019). <https://doi.org/10.1002/col.22292>
7. Einakian, S., Newman, T.S.: An examination of color theories in map-based information visualization. *J. Comput. Lang.* **51**, 143–153 (2019). <https://doi.org/10.1016/j.cola.2018.12.003>

8. Zaeimi, M., Ghoddosian, A.: Color harmony algorithm: an art-inspired metaheuristic for mathematical function optimization. *Soft. Comput.* **24**(16), 12027–12066 (2020). <https://doi.org/10.1007/s00500-019-04646-4>
9. Cochrane, S.: The Munsell color system: a scientific compromise from the world of art. *Stud. Hist. Philos. Sci. A* **47**, 26–41 (2014). <https://doi.org/10.1016/j.shpsa.2014.03.004>
10. Rathore, R., Leggon, Z., Lessard, L., Schloss, K.B.: Estimating color-concept associations from image statistics. *IEEE Trans. Visual Comput. Graph.* **26**(1), 1226–1235 (2019). <https://doi.org/10.1109/TVCG.2019.2934536>
11. Olson, R.S.: Revisiting the Vaccine Visualizations. <http://www.randalolson.com/2016/03/04/revisiting-the-vaccine-visualizations/> (2016)
12. Fairchild, M.D.: *Color Appearance Models*. 3rd ed. John Wiley & Sons (2013). <https://doi.org/10.1002/9781118653128>
13. Szafir, D.A.: Modeling color difference for visualization design. *IEEE Trans. Visual Comput. Graph.* **24**(1), 392–401 (2017). <https://doi.org/10.1109/TVCG.2017.2744359>
14. Maji, S., Dingliana, J.: Perceptually optimized color selection for visualization. *Tetrahedron* **4**(1.63299), 1–63299 (2018)
15. Zhang, Q., Xiao, C., Sun, H., Tang, F.: Palette-based image recoloring using color decomposition optimization. *IEEE Trans. Image Process.* **26**(4), 1952–1964 (2017). <https://doi.org/10.1109/TIP.2017.2671779>
16. Tan, J., Echevarria, J., Gingold, Y.: Palette-based image decomposition, harmonization, and color transfer. *arXiv preprint arXiv:1804.01225* (2018)
17. Luo, M.R., Cui, G., Rigg, B.: The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color. Res. Appl.* **26**(5), 340–350 (2001). <https://doi.org/10.1002/col.1049>
18. Yang, Y., Ming, J., Yu, N.: Color image quality assessment based on CIEDE2000. *Adv. Multimedia* (2012). <https://doi.org/10.1155/2012/273723>