An Evaluation of Color Maps for Visual Data Exploration

Daniel Baum

This is a preprint of a manuscript that will appear in the monograph entitled Science in Color - Visualizing Achromatic Knowledge, which will be published by De Gruyter.

ZIB Report 19-42 (August 2019)
An Evaluation of Color Maps for Visual Data Exploration

Daniel Baum
Zuse Institute Berlin, 14195 Berlin, Germany

1 Introduction

Color is often used in data visualization as an additional cue to support the understanding of the data being visualized. However, care needs to be taken when applying color since it enormously influences our visual perception. Over the past sixty years, the use of color in data visualization – a subfield of computer science – has been studied in great detail. This has led to many generally accepted rules for the use of color. An important aspect when using color is the task to be carried out or the goal that is to be achieved. A different task may lead to different requirements for the usage of color. When feature detection is the main goal, for example, other color schemes may be more appropriate than those that would be beneficial for an overview of the data. The usability of color is further restricted when three-dimensional (3D) objects are the target of the visualization. In this case, for example, luminance should be avoided to represent differences in the data that is visualized on the surface of the 3D object because the luminance parameter is already required to better bring out the shape of the object (see the last example in this article). Hence, in order to use color most effectively, many considerations need to be made. In this article, we focus on visualizing continuous scalar-valued data. Such data is usually depicted with the help of color maps that assign to each scalar value a single color. The term color map originally referred to a color lookup table that was used in computer graphics to map scalar values to a specific color to be depicted on the
computer screen.\textsuperscript{1} While this article focusses on continuous color maps, different aspects of color for data visualization are considered in other articles of this volume (see, for example, the article by Jana Moser and Philipp Meyer, and the article by Bettina Bock von Wülfingen). Also note that the term color map might be used differently throughout this volume.

Let us consider a first example. The data depicted in figure 1 is a three-dimensional (3D) image of a marine sediment core containing coral and sea shell fragments. This 3D image was acquired using computed tomography (CT). The 2D visualizations depicted in figure 1 show the same cross-section of this CT image. The differences are only due to the varying color mappings used to visualize it. As a result of applying different color maps, the data might be perceived quite differently, although, at least in this example, the general structure of data can be recognized in all images. Yet, subtle differences can be observed, for example, in the close environment of the coral fragments. These become particularly obvious when we compare the images in the top row with those in the bottom row. In the bottom row, the corals seem to be made up of layers of materials whereas in the top row, we can perceive a continuous variation of the data values. This single example already raises some questions: What is the best color map to visualize some given data? What has to be considered when using color to visualize some data? And what pitfalls might appear when the “wrong” coloring is used? This is the type of questions that we want to address in this article. Since the ultimate goal of visual data exploration is to support observers in understanding their data, the effective use of color needs to be measured in terms of two performance characteristics: (i) the time that is needed to understand the data; (ii) the accuracy of the answers that were derived from the visualization of the investigated data. We have carefully studied the previous work on this topic and give a dense summary of the most important conclusions. To underline these statements using examples, we mainly resort to artificially created data since this allows us to clearly demonstrate the effects. Furthermore, we have chosen the examples such that they are simple and reproducible.

The “correct” use of color in computer-generated data visualization has concerned researchers from the beginning of computer graphics and visualization.\textsuperscript{2} It was even questioned whether color should be used at all for data visualization.\textsuperscript{3} But already in 1980, Meyer and Greenberg made a clear statement in favor of using color: “Color has been used to encode the variations of parameters such as temperature and stress. The accuracy with which an observer interprets these images has been questioned,\textsuperscript{4} but they still remain an effective means for showing data trends.”\textsuperscript{5} Until today, color remains an important means for


\textsuperscript{3}Booth et al. 1977, Morse 1979 (s. fn. 2)

\textsuperscript{4}Booth et al. 1977, Morse 1979 (s. fn. 2)

\textsuperscript{5}Meyer et al. 1980 (s. fn. 2)
visualizing data. And despite the discussions about the accuracy of data perception when using color, even in conservative fields like medical imaging and image analysis, it has been shown that the use of color is effective and improves the performance of data analysis. Thus, the question is not so much whether color should be used at all for data visualization but what the most effective way is to use it. And what aspects have to be considered when using color in order to be effective? As result of the research done over the last 60 years, many rules have been proposed for the use of color.

In this paper, we focus on data represented as scalar fields \( f: \mathbb{R}^d \rightarrow \mathbb{R} \), in particular two-dimensional \((d = 2)\) and three-dimensional \((d = 3)\) scalar fields, where to each point in a spatial domain a scalar value is assigned. If the domain is a rectangle or a cuboid, we often call these scalar fields gray-scale images, but the domain can also be the surface of a three-dimensional object. Color maps are probably the most important means to visualize scalar fields. However, the design of color maps is by no means trivial. In the following, we review some findings concerning the use of color maps and rules that should be considered when designing them. We underline those findings and rules by applying several commonly used color maps to specifically created artificial test datasets and by evaluating the respective visualizations.

2 Definition of Color Maps

A color map is a function \( x: [a, b] \subseteq \mathbb{R} \rightarrow C, a < b \), that assigns to each scalar value between \( a \) and \( b \) a color \( c \in C \), where \( C \) is some color space. Usually, the colors onto which the scalar values are mapped are given as a finite set of \( n \) ordered colors \( c_1, c_2, c_3, …, c_n \), where \( n \) can be any natural number but is usually 100 or greater. Examples of color maps are shown in figure 2. In this paper, we use the term color map for both the function \( x \) and the set of ordered colors. The function \( x \) has the property that it is monotone, that is for \( d < e \in [a, b] \) with \( x(d) = c_i \) and \( x(e) = c_j \) it holds that \( i \leq j \). For most purposes, the color map \( x \) represents a linear mapping of the scalar values to the colors in the color map. Thus, for example, if the difference between two scalar values \( d \) and \( e \) is twice as large as the distance between two other scalar values \( g \) and \( h \), then the distance of the indices of the mapped colors of these values in the color map representation should also be (approximately) twice as large.

Before going into the details about how a color map should be designed, it is important to consider the possibilities in the choice of the color space \( C \), which is a set of colors for which a distance is defined. Thus, different color spaces vary either according to the set of colors, the defined color distance, or both. “All color spaces are based on the tristimulus theory,

---


which states that any perceived color can be uniquely represented by a 3-tuple. This result is a side effect of the fact that there are exactly 3 different types of color receptors in the human eye.”

Examples are the RGB color space, the XYZ color space, and the perceptual color spaces like CIELAB, CIELUV. For a color space to be perceptual, the difference of two colors perceived by an observer should be proportional to the Euclidean distance between the two points representing the colors in the color space. For data visualization, perceptual color spaces have many favorable properties, for example, they allow to easily create a perceptually linear color map by drawing a line in the perceptual color space and using the colors along the line as color map. However, it should be noted that when creating perceptual color spaces, several parameters are kept constant like the size of the color sample, the spacing between color samples, luminance and chromaticity of the background, as well as luminance and chromaticity of the ambient light. Thus, “to maintain the perceptual uniformity of the space, the settings of these parameters in the final application must match the conditions existing when the data was originally recorded.” This is only rarely the case. Nevertheless, perceptually uniform color spaces are still the best we can work with.

3 Design Rules for Color Maps

Even though the use of color and its effectiveness depends on the given task that needs to be carried out, for example, identifying maxima or minima, rapid changes in the data, borders, or in general, the identification of irregularities or particularities, it is often desirable to have good general-purpose color maps that can be applied to many tasks instead of being very specific to just one task. Therefore, many researchers have identified common requirements that need to be fulfilled to make a good color map.

Herman and Levkowitz named three important requirements. Note that they use the term “color scale” instead of “color map”, but it has the same meaning.

- Order: “The colors used to represent the values in the scale should be perceived as preserving the order of the values.” This means that all people should be able to order the colors in the same way and this ordering should be the same as the naturally defined order of the scalar values.
- Uniformity and representative distance: “The colors used should convey the distances between the values they represent. Colors representing values equally different from each other along the scale should be perceived as equally different.”
- Boundaries: “The color scale should not create perceived boundaries that do not exist in the numerical data. That is, it should be able to continuously represent continuous scales.”

Light and Bartlein added further requirements, in particular regarding the creation of color maps that are suitable also for people with color deficiencies:

- Spectral schemes: “Avoid the use of spectral schemes to represent sequential data because the spectral order of visible light carries no inherent magnitude message.

---

9 Moreland 2009 (s. fn. 7)
10 Meyer et al. 1980 (s. fn. 2)
11 Meyer et al. 1980 (s. fn. 2)
12 Moreland 2009, Herman and Levkowitz 1992, Light and Bartlein 2004, Fortner and Meyer 2012 (s. fn. 7)
13 Herman and Levkowitz 1992 (s. fn. 7)
14 Light and Bartlein 2004 (s. fn. 7)
Readers do not automatically perceive violet as greater than red even though the two colors occupy opposite ends of the color spectrum.

- **Yellow/green**: “Use yellow with care and avoid yellow-green colors altogether in spectral schemes. Readers with color-deficient vision often confuse yellow-green with orange colors. Yellow appears brightest among the primary colors and stands out visually for color-impaired and normally sighted readers alike.”

- **Intensity**: “Use color intensity (or value) to reinforce hue as a visual indicator of magnitude. [...] Using intensity as well as hue also makes the quality of color reproduction less critical to the presentation and helps make even gray scale photocopies somewhat legible.”

Moreland suggests some more general rules that include many of the above-mentioned rules:¹⁵

- “The map yields images that are aesthetically pleasing.
- The map has a maximal perceptual resolution.
- Interference with the shading of 3D surfaces is minimal.
- The map is not sensitive to vision deficiencies.
- The order of the colors should be intuitively the same for all people.
- The perceptual interpolation matches the underlying scalars of the map.”

Of particular interest among these requirements is the first one, that is, that the color map should be “aesthetically pleasing”. One problem with this requirement is certainly that it is difficult to quantify. “Nevertheless, aesthetic appeal is important as users will use that as a criterion in selecting visualization products and generating images.”¹⁶ Another interesting point is the mentioning of 3D surfaces. Meyer and Greenberg support this by giving a clear advice how this can be achieved:¹⁷ “For three-dimensional data presentations only chromaticity can be used to encode the parameter since the length of the color vector [...] is used for color intensity variations which convey the shape of the object.” While most of the stated requirements are intuitive for most people, it is not obvious how to quantify them in an objective way, particularly, if expensive user evaluations should be avoided. In order to overcome this deficiency, Bujack et al. have tried to translate these, sometimes vague, rules into the mathematical language of equations.¹⁸ In addition to this, they have tried to resolve the many ambiguities in the naming of properties that a color map should or should not have. “For example, order as used by Levkowitz and Herman has a local and directed meaning: a color \( c_i \) should perceptually precede the next color.¹⁹ In contrast, the term order as used by Ware has a global but undirected meaning.²⁰ He requires that a user be able to sort colors picked from anywhere in the color map but does not distinguish whether they are sorted from low to high or vice versa.”²¹ As result of their study, Bujack et al. identify 16 design rules that occur in the literature.²² Three of those are stated to be most important: order (of the colors),

---

¹⁵ Moreland 2009 (s. fn. 7)
¹⁶ Moreland 2009 (s. fn. 7)
¹⁷ Meyer and Greenberg 1980 (s. fn. 2)
²⁰ Ware (s. fn. 2)
²¹ Bujack et al. 2018 (s. fn. 18)
²² Bujack et al. 2018 (s. fn. 18)
discriminative power, and uniformity. For those properties, they derive a mathematical framework that helps evaluate the properties for a given color map.

4 Application of Color Maps to Test Datasets

In this section, we apply eight color maps (fig. 2) to several test datasets that are described below. The color maps have been taken from visualization tools and other sources. The gray scale, the hot iron, and the temperature color maps have been taken from the Amira software. The viridis and the plasma color maps have been created by the Berkeley Institute for Data Science and are freely available. The rainbow color map has been taken from the Matlab software, and the diverging blue-red as well as the diverging brown-green color maps have been generated using the ColorBrewer software.

4.1 Two-dimensional (2D) Data Visualization

In order to support the statements that are made in this paper and the references mentioned herein, several datasets were created that are described in detail such as to allow the user to better understand the results as well as to reproduce the findings below. Each of these datasets will be visualized with the same set of color maps to allow an easy comparison of the color maps and the perception of the data as a result of applying the color maps.

Dataset 1 (fig. 3) contains values that are constant along the vertical axis and only vary across the horizontal axis. It shows several sinus curves with varying frequency parameters and thus lengths of the period. The first period has a length of $2\pi$, the next one has a length of $\pi$, then $2/3 \pi$, $1/2 \pi$, $2/5 \pi$ etc. In mathematical terms, the $i$th period has a length of $2\pi$ divided by $i$. All of the color maps in the first row seem to represent the data in a similar way that allows to grasp the underlying mathematical function of the scalar fields, that is, sinus curves. Both viridis and plasma seem to do a slightly better job than the gray scale and the hot iron color maps which have problems in both very high and very low values to allow a differentiation of the values. All of the color maps in the bottom row make it difficult to grasp the variation of the scalar values and to get an intuition of the underlying mathematical formula. This is particularly bad in case of the rainbow and temperature color maps. The white in the middle

---


3: Dataset 1 in which the scalar values vary from left to right according to a sinus signal with decreasing period length starting with a period length of $2\pi$, then $\pi$, $2/3 \pi$, $1/2 \pi$, etc.

of the diverging color maps allows an easy distinction between negative and positive values but makes it hard to understand the data. Furthermore, for all of the color maps in the bottom row, the frequency of the pattern seems to be much higher when looking at the very right side. It seems that our eyes are not able to distinguish the colors very well anymore at this high frequency, resulting in the perception of a much higher frequency in the data that there really is.
Dataset 2 (fig. 4) starts off from dataset 1 but modifies it along the vertical axis. The values of a linear ramp starting with 1 at the bottom and falling off to 0 at the top are multiplied with the values of dataset 1, resulting in the final dataset 2. A similar dataset was also used by Moreland.\textsuperscript{27} In all the visualizations except for those using the rainbow and the temperature color maps, one can clearly observe the linear decay of the amplitude of the sinus curves. The rainbow color map seems to perform worst. While for the sinus curves with short period almost up to the top one can distinguish between the sinus curves, this distinction is less clear the further to the left we go.

\textsuperscript{27} Moreland 2009 (s. fn. 7)
Dataset 3 (fig. 5) was created by multiplying two sinus curves that have a fixed period, one along the horizontal axis, the other one along the vertical axis. A linear horizontal ramp, starting with a value of -0.5 at the left border which increases linearly up to 0.5 at the right border, is added to the double sinus pattern. As a result, the values range from -1.5 to 1.5. The troughs are deepest on the left side with lower hills. On the right side, the hills are highest with low troughs. A similar dataset was also used by Moreland,\textsuperscript{28} which, however, is stretched in vertical direction. Again, the color maps in the top row seem to perform best when the task is to give an overview about the whole dataset. However, if the task is to identify minima and maxima, the rainbow color map draws our attention quickest to the respective points. In particular, the maxima are very clearly visible. The temperature color map also performs well in this respect, as well as the hot iron and the diverging blue-red color map. For this task, the white color in the diverging color maps functions as a separator of the extrema, which supports the perception of the extrema.

Dataset 4 (fig. 6) again is a combination of two patterns. The first pattern is a simple linear ramp going from 0 at the left border to 1 at the right border. This first pattern is overlaid by a second pattern that is a diagonal sinus curve which decreases non-linearly from bottom to top. This dataset was created in analogy to the test pattern described by Ware et al.\textsuperscript{29} It was designed to test the uniformity property in conjunction with the discriminative power. With respect to this pattern, a color map can be considered to be uniform if the diagonal pattern can be perceived equally well from left to right. The gray scale and the hot iron work quite well for this task except for the very dark region on the left. In this region, the viridis and plasma color maps seem to perform slightly better, while the distinction of the pattern in the

\textsuperscript{28} Moreland 2009 (s. fn. 7)

\textsuperscript{29} Colin Ware, Terece L. Turton, Francesca Samsel, Roxana Bujack, and David H. Rogers: \textit{Evaluating the Perceptual Uniformity of Color Sequences for Feature Discrimination}. In: \textit{Proceedings of EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV3)}. The Eurographics Association, 2017.
violet region of the plasma is not as good as for the viridis color map. For all of the color maps in the bottom row, the perception of the pattern varies enormously along the horizontal axis. Especially in the middle, the sinus pattern is very hard to discern. For both the rainbow and the temperature color maps, it varies greatly from left to right, supporting the statement that it is not uniform. Particularly in the cyan-green-yellow region in the middle, the pattern is hard to see. Similarly, for the temperature color map, the pattern is worst visible in the blue-violet region.

Dataset 5 (fig. 7) is a mixture of two simple patterns, a linear ramp that increases from 0 to 1 horizontally from left to right, and a linear ramp that increases from 0 to 1 vertically from top to bottom. These two patterns are combined using a chessboard pattern, where the values of pattern 1 are used for the white squares of the chessboard, and pattern 2 is taken for the black squares of the chessboard. Only pattern 1 is visualized by all the color maps, whereas pattern 2 is always visualized using the gray scale color map. This combination of patterns was designed to show the influence of the environment on the perception of colors. In all the images of figure 7, the colored squares in each column are identical but since the environments of these squares change from top to bottom, we also perceive a change in the color, which, however, is not there. This means that our cognition is tricked due to the changing environment, which is a well-known phenomenon called simultaneous contrast.

For the color maps in the top row, this perceived change of color is particularly visible at the diagonal going from the left upper corner to the right bottom corner. For the two diverging color maps as well as for the rainbow color map, instead of one diagonal, we see two diagonals.

---

at which the coloring seems to change drastically. Two diagonals are also visible for the
*temperature* color map, but here the diagonals seem to point upwards. In figure 8, triples of
close-ups are shown, where the (colored) central and corner patches are identical across the
tripe of squares.

4.2 Three-dimensional Data Visualization

Dataset 6 (fig. 9) is the only 3-dimensional (3D) dataset. It shows a molecular surface (solvent
excluded surface) that was computed for the molecular structure with identifier 2RNT taken
from the Protein Data Bank. Mapped onto this surface is the electrostatic potential, which is
a scalar-valued field. This dataset is used to show the effect of 3D shading that might interfere
with coloring and hence influences the perception of the data mapped onto 3D objects. In
general, the electrostatic potential contains negative and positive values, hence, a *diverging*
color map would be ideal. And indeed, the two *diverging* color maps very well depict the
extremal negative and positive values. But since a large part of the surface has values around
zero, the surface appears very bright. This conflicts with the shading that adds visual cues to
better perceive the geometry of the molecular surface. The *rainbow* color map used here also
has very bright colors that conflict with the shading. From the color maps in the upper row,
*viridis* and *plasma* seem to work best for the 3D object since they contain relatively few light
colors. The *plasma* color map is slightly superior over *viridis* since it highlights very small values
better than its blue-green counterpart.

![ triples of close-ups showing the same central patch at different positions in the pattern shown in figure 7.](image)

---

In many publications,\textsuperscript{32} it has been stressed that the rainbow color map is generally not a good choice and should be avoided if possible. Moreland gives the following reasons:\textsuperscript{33} The rainbow color map is not uniform, that is, colors appear to change faster in the cyan and yellow regions than in the blue, green and red regions; 5% of the people cannot distinguish between red and green; viewers with color deficiencies cannot distinguish between colors that are far apart in the rainbow color map. These problems could be confirmed with our examples. It has also been shown that the use of the rainbow color map can lead to unnecessary medical diagnostic errors.\textsuperscript{34} On the other hand, Zabala-Travers et al. have shown the superiority of the rainbow color map over the gray scale and hot iron color maps for a very specific task,\textsuperscript{35} namely that of determining the image with highest maximal intensity from a pair of images. This can be explained by the fact that the rainbow color map indeed has a high discriminative power in the left and right regions.\textsuperscript{36} We have also shown that the rainbow color map works well for identifying minima and maxima (fig. 5). Nevertheless, the well-grounded criticism of the rainbow color map remains and we believe that, where possible, the use of this color map should be avoided or at least very critically questioned.


\textsuperscript{33} Moreland 2009 (s. fn. 7)


\textsuperscript{35} Silvina Zabala-Travers, Mina Choi, Wei-Chung Chen, and Aldo Badano: Effect of color visualization and display hardware on the visual assessment of pseudocolor medical images. In: Medical Physics, 42 (6), 2015, pp. 2942-2954.

\textsuperscript{36} Bujack et al. 2018 (s. fn. 18)
Diverging color maps are particularly well-suited for scalar data that have a distinguished mean value that separates opposite values with different meanings like negative and positive values. However, for 3D data, even with a distinguished mean value, they are not ideal since the very bright colors in the middle have luminance values that are too large, which conflicts with the shading parameters. From the two studied diverging color maps, the brown-green color map is certainly less favorable since it has no intuitive order. That is, it is not obvious whether green should be associated with smaller values or larger ones. Since blue and red are often associated with temperature, for the blue-red color map, there is an intuitive order.

For most of the test patterns that we used in this paper, the perceptually linear color maps, namely, gray scale, hot iron, viridis, and plasma are very good choices as general-purpose color maps when the task is to get an overview of the data. Since viridis and plasma make less use of luminance, they are also less strongly affected by simultaneous contrast, an effect that is due to the different brightness of the surrounding area causing problems for humans when trying to estimate the luminance of a color (figs. 7 and 8). Viridis and plasma are also good choices for 3D data visualization since they do not have a great variation in luminance which could otherwise cause problems due to luminance being used as visual cue for the perception of form. Hence, both the viridis and the plasma color maps make a good choice when a general-purpose color map is required.

---