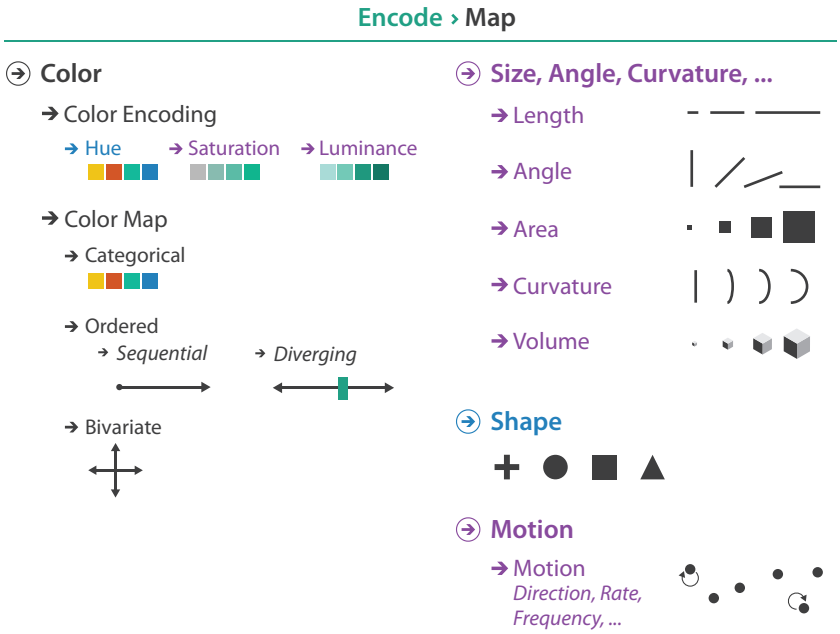


Figure 10.1. Design choices for mapping color and other visual encoding channels.

Chapter 10

Map Color and Other Channels

10.1 The Big Picture

This chapter covers the mapping of color and other nonspatial channels in visual encoding design choices, summarized in Figure 10.1. The colloquial term *color* is best understood in terms of three separate channels: luminance, hue, and saturation. The major design choice for colormap construction is whether the intent is to distinguish between categorical attributes or to encode ordered attributes. Sequential ordered colormaps show a progression of an attribute from a minimum to a maximum value, while diverging ordered colormaps have a visual indication of a zero point in the center where the attribute values diverge to negative on one side and positive on the other. Bivariate colormaps are designed to show two attributes simultaneously using carefully designed combinations of luminance, hue, and saturation.

The characteristics of several more channels are also covered: the magnitude channels of size, angle, and curvature and the identity channels of shape and motion.

10.2 Color Theory

Color is a rich and complex topic, and here I only touch on the most crucial aspects that apply to vis.

10.2.1 Color Vision

The retina of the eye has two different kinds of receptors. The **rods** actively contribute to vision only in low-light settings and provide low-resolution black and white information. I will thus not discuss them further in this book. The main sensors in normal lighting conditions are the **cones**. There are three types of cones, each with peak sensitivities at a different wavelength within the spectrum of

visible light. The visual system immediately processes these signals into three **opponent color channels**: one from red to green, one from blue to yellow, and one from black and white encoding luminance information. The luminance channel conveys high-resolution edge information, while the red–green and blue–yellow channels are lower resolution. This split between luminance and **chromaticity**—what most people informally call would normally call *color*—is a central issue in visual encoding design.

The theory of opponent color channels explains what is colloquially called **color blindness**, which affects around 8% of men in its most common form. A more accurate term is **color deficiency**, since a "colorblind" person's ability to differentiate color along the red–green channel is reduced or absent but their blue–yellow channel is still in full working order.*

★ There is a type of color deficiency where the blue–yellow channel is impaired, **tritanopia**, but it is extremely rare and not sex linked. The two common forms of red–green color blindness are **deuteranopia** and **protanopia**; both are sex linked.

10.2.2 Color Spaces

The **color space** of what colors the human visual system can detect is three dimensional; that is, it can be adequately described using three separate axes. There are many ways to mathematically describe color as a space and to transform colors from one such space into another. Some of these are extremely convenient for computer manipulation, while others are a better match with the characteristics of human vision.

The most common color space in computer graphics is the system where colors are specified as triples of red, green, and blue values, which is called the **RGB** system. Although this system is computationally convenient, it is a very poor match for the mechanics of how we see. The red, green, and blue axes of the RGB color space are not useful as separable channels; they give rise to the integral perception of a color. I do not discuss them further as channels in my analysis of visual encoding.

Another color space, the hue–saturation–lightness or **HSL** system, is more intuitive and is heavily used by artists and designers. The **hue** axis captures what we normally think of as pure colors that are not mixed with white or black: red, blue, green, yellow, purple, and so on. The **saturation** axis is the amount of white mixed with that pure color. For instance, pink is a partially desaturated red. The **lightness** axis is the amount of black mixed with a color. A common design for color pickers is a disk with white at the center and the hue axis wrapped around the outside, with separate linear control for the amount of darkness versus lightness, as shown in Figure 10.2. The HSV space is very similar, where V stands for grayscale value and is linearly related to L.

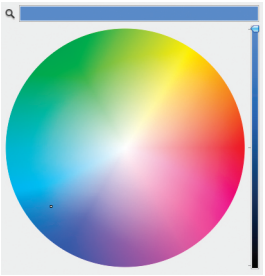


Figure 10.2. A common HSL/HSV colorpicker design, as in this example from Mac OS X, is to show a color wheel with fully saturated color around the outside and white at the center of the circle, and a separate control for the darkness.

Despite the popularity of the HSL space, it is only pseudoperceptual: it does not truly reflect how we perceive color. In particular, the lightness L is wildly different from how we perceive luminance. Figure 10.3 shows six different hues, arranged in order of their luminance. The corresponding computed L values are all identical. The true luminance is a somewhat better match with our perceptual experience: there is some variation between the boxes. However, our perception of luminance does not match what

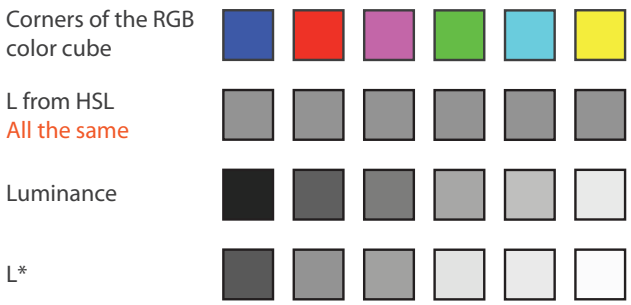


Figure 10.3. Comparing HSL lightness, true luminance, and perceptually linear luminance L^* for six colors. The computed HSL lightness L is the same for all of these colors, showing the limitations of that color system. The true luminance values of these same six colors, as could be measured with an instrument. The computed perceptually linear luminance L^* of these colors is the best match with what we see. After [Stone 06].

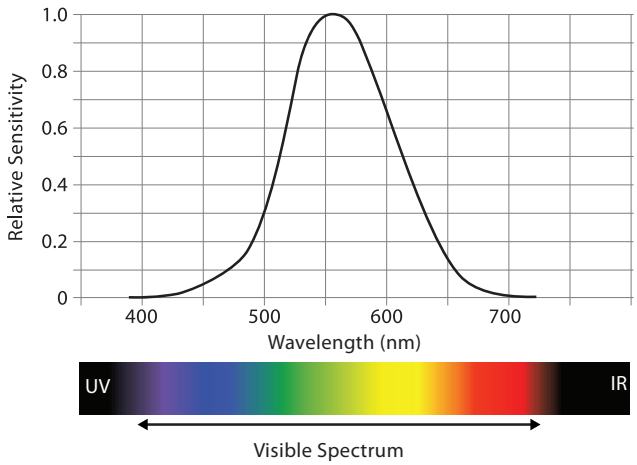


Figure 10.4. The spectral sensitivity of our eyes to luminance depends on the wavelength of the incoming light. After [Kaiser 96], http://www.yorku.ca/eye/photopik.htm.

an instrument would measure: the amount of luminance that humans perceive depends on the wavelength. Figure 10.4 shows the roughly bell-shaped **spectral sensitivity** curve for daylight vision. We are much more sensitive to middle wavelengths of green and yellow than to the outer wavelengths of red and blue.

There are several color spaces that attempt to provide a perceptually uniform space, including one known as $L^*a^*b^*$. This space has a single black and white luminance channel L^*, and the two color axes a^* and b^*. Figure 10.3 also shows the L^* values for the same six boxes, which is even a better match with what we see than true luminance. The L^* axis is a nonlinear transformation of the luminance perceived by the human eye. Our perception of luminance is compressed, as shown by the exponent of $n = 0.5$ for the brightness power law curve in Figure 5.7. The L^* axis is designed to be perceptually linear, so that equally sized steps appear equal to our visual systems, based on extensive measurement and calibration over a very large set of observers. The color axes have also been designed to be as perceptually linear as possible. The $L^*a^*b^*$ space is thus well suited for many computations, including interpolation and finding differences between colors, that should not be executed in the perceptually nonlinear RGB or HSL spaces.

► Accuracy of perception is discussed in Section 5.5.1.

In this book, I use the term *luminance* as an evocative way to describe the black and white visual channel.* I use *saturation* and *hue* for the other two chromaticity channels.*

★ I avoid the term *perceptually linear luminance* as inaccurate for visualization analysis because very few visual encoding idioms carry out this computation. Similarly, I avoid the term **brightness**, the technical term for the human perceptual experience of luminance, because it is affected by many factors such as illumination levels and surrounding context; again, visual encoding idioms typically manipulate luminance rather than attempting to deliver true brightness.

★ This hybrid usage of **luminance**, **saturation**, and **hue** does not correspond exactly to any of the standard color spaces used in computer graphics.

10.2.3 Luminance, Saturation, and Hue

Color can be confusing in vis analysis because it is sometimes used as a magnitude channel and sometimes as a identity channel. When I use the precise terms *luminance*, *hue*, and *saturation*, I mean the three separable visual channels that pertain to color in the analysis of visual encoding. Luminance and saturation are magnitude channels, while hue is a identity channel. When I use the generic term **color**, I mean the integral perception across all three of these channels at once, and I am analyzing it as an identity channel.

The magnitude channel of **luminance** is suitable for ordered data types. However, one consideration with luminance is our low accuracy in perceiving whether noncontiguous regions have the same luminance because of contrast effects. Thus, the number of discriminable steps is small, typically less then five when the background is not uniform: Ware suggests avoiding grayscale if more than two to four bins are required [Ware 13].

Moreover, a fundamental problem with using it for encoding a specific data attribute is that luminance is then "used up" and cannot be used for other purposes. A crucial consideration when visual encoding with color is that luminance contrast is the only way we can resolve fine detail and see crisp edges; hue contrast or saturation contrast does not provide detectable edges. In particular, text is not readable without luminance contrast. The standard guidelines are that 10:1 is a good luminance contrast ratio for text, with 3:1 as a minimum [Ware 13]. If it's important that fine-grained features are legible, ensure that you provide sufficient luminance contrast.

The magnitude channel of **saturation** is also suitable for ordered data. Saturation shares the problem of low accuracy for noncontiguous regions. The number of discriminable steps for saturation is low: around three bins [Ware 13].

Moreover, saturation interacts strongly with the size channel: it is more difficult to perceive in small regions than in large ones. Point and line marks typically occupy small regions, so using just two different saturation levels is safer in these cases. Finally,

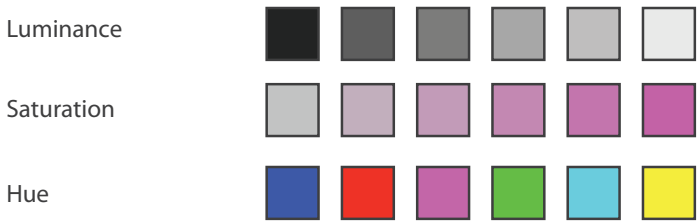


Figure 10.5. The luminance and saturation channels are automatically interpreted as ordered by our perceptual system, but the hue channel is not.

saturation and hue are not separable channels within small regions for the purpose of categorical color coding.

For small regions, designers should use bright, highly saturated colors to ensure that the color coding is distinguishable. When colored regions are large, as in backgrounds, the design guideline is the opposite: use low-saturation colors; that is, pastels.

The identity channel of **hue** is extremely effective for categorical data and showing groupings. It is the highest ranked channel for categorical data after spatial position.

However, hue shares the same challenges as saturation in terms of interaction with the size channel: hue is harder to distinguish in small regions than large regions. It also shared the same challenges as saturation and luminance for separated regions: we can make fine distinctions in hue for contiguous regions, but we have very limited discriminability between separated regions. The number of discriminable steps for hue in small separated regions is moderate, around six or seven bins.

Unlike luminance and saturation, hue does not have an implicit perceptual ordering, as shown in Figure 10.5. People can reliably order by luminance, always placing gray in between black and white. With saturation, people reliably place the less saturated pink between fully saturated red and zero-saturation white. However, when they are asked to create an ordering of red, blue, green, and yellow, people do not all give the same answer. People can and do learn conventions, such as green–yellow–red for traffic lights, or the order of colors in the rainbow, but these constructions are at a higher level than pure perception.

10.2.4 Transparency

A fourth channel strongly related to the other three color channels is **transparency**: information can be encoded by decreasing the opacity of a mark from fully opaque to completely see-through. Transparency cannot be used independently of the other color channels because of its strong interaction effects with them: fully transparent marks cannot convey any information at all with the other three channels. In particular, transparency coding interacts strongly with luminance and saturation coding and should not be used in conjunction with them at all. It can be used in conjunction with hue encoding with a very small number of discriminable steps, most frequently just two. Transparency is used most often with superimposed layers, to create a foreground layer that is distinguishable from the background layer. It is frequently used redundantly, where the same information is encoded with another channel as well.

► Superimposing layers is covered in Section 12.5.

10.3 Colormaps

A **colormap** specifies a mapping between colors and data values; that is, a visual encoding with color.* Using color to encode data is a powerful and flexible design choice, but colormap design has many pitfalls for the unwary.

★ *Colormapping* is also called **pseudocoloring**, especially in earlier literature. Another synonym for *colormap* is **color ramp**.

Figure 10.6 shows the taxonomy of colormaps; it is no coincidence that it mirrors the taxonomy of data types. Colormaps can be **categorical** or **ordered**, and ordered colormaps can be either **sequential** or **diverging**. Of course, it is important to match colormap to data type characteristics, following the expressiveness principle. Colormaps for ordered data should use the magnitude channels of luminance and saturation, since the identity channel of hue does not have an implicit ordering.

► The principle of expressivness is covered in Section 5.4.1.

Colormaps can either be a **continuous** range of values, or **segmented** into discrete bins of color.* Continuous colormaps are heavily used for showing quantitative attributes, especially those associated with inherently spatial fields. Segmented colormaps are suitable for categorical data. For ordinal data, segmented colormaps would emphasize its discrete nature, while continuous would emphasize its ordered nature. Bivariate colormaps encode two attributes simultaneously. While bivariate colormaps are straightforward to understand when the second attribute is binary, with only two levels, they are more difficult for people to interpret when both attributes have multiple levels.

★ There are many synonyms for *segmented colormap*: **quantized colormap**, **stepped colormap**, **binned colormap**, **discretized colormap**, and **discrete colormap**.

► Continuous versus discrete data semantics is discussed in Section 2.4.3.

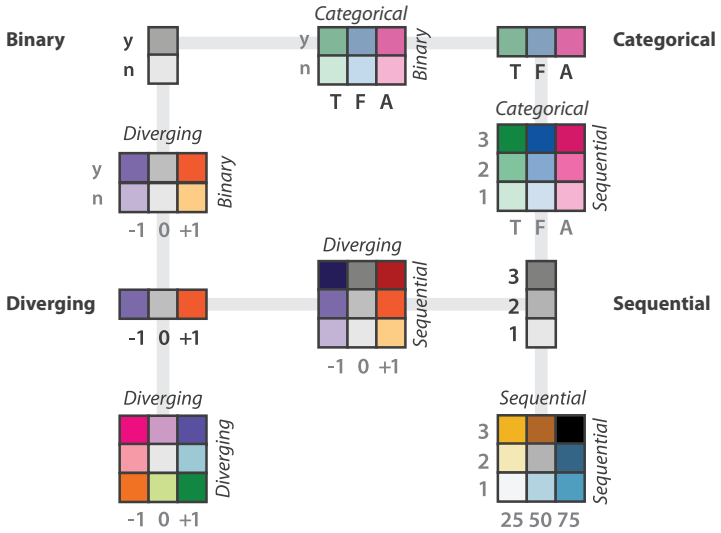


Figure 10.6. The colormap categorization partially mirrors the data types: categorical versus ordered, and sequential and diverging within ordered. Bivariate encodings of two separate attributes at once is safe if one has only two levels, but they can be difficult to interpret when both attributes have multiple levels. After [Brewer 99].

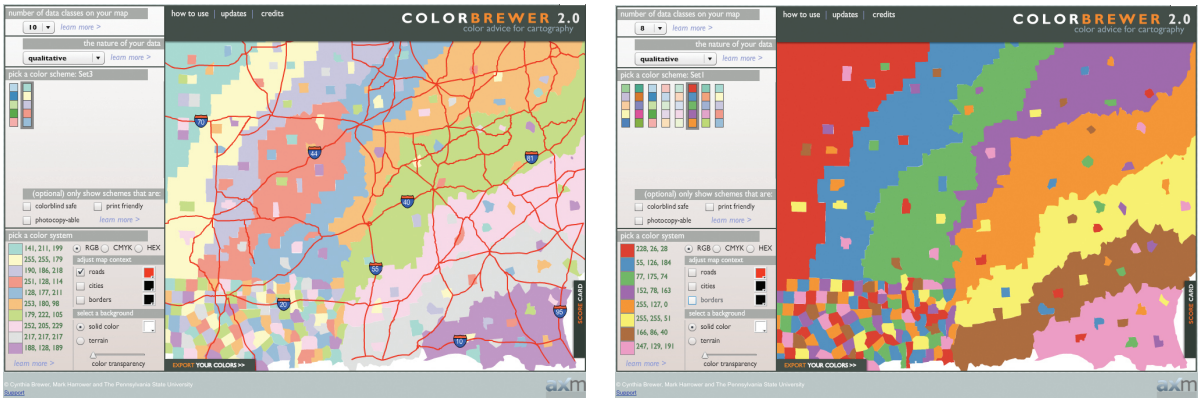
10.3.1 Categorical Colormaps

★ *Categorical colormaps* are also known as **qualitative** colormaps.

A **categorical** colormap uses color to encode categories and groupings. Categorical colormaps are normally segmented.* They are very effective when used appropriately; for categorical data, they are the next best channel after spatial position.

Categorical colormaps are typically designed by using color as an integral identity channel to encode a single attribute, rather than to encode three completely separate attributes with the three channels of hue, saturation, and luminance.

The number of discriminable colors for coding small separated regions is limited to between six and twelve bins. You should remember to include background color and any default object colors in your total count: some or all of the most basic choices of black, white, and gray are often devoted to those uses. Easily nameable colors are desirable, both for memorability and ability to discuss them using words. A good set of initial choices are the fully saturated and easily nameable colors, which are also the opponent

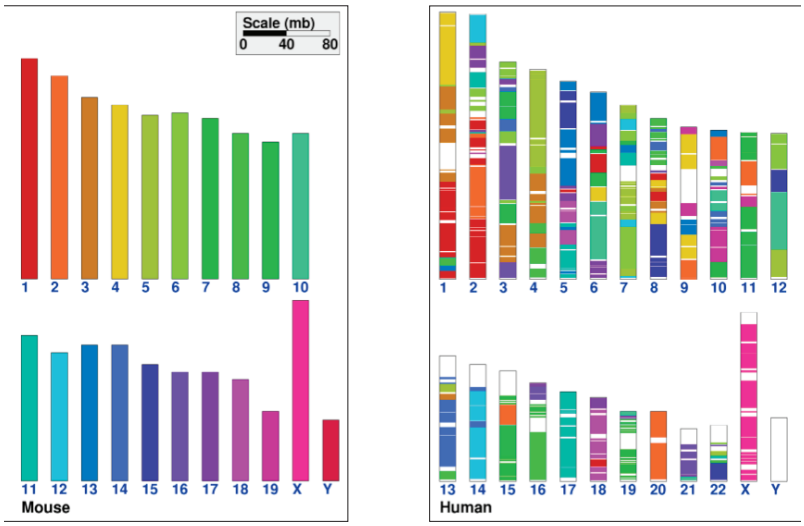


(a) (b)

Figure 10.7. Saturation and area. (a) The ten-element low-saturation map works well with large areas. (b) The eight-element high-saturation map would be better suited for small regions and works poorly for these large areas. Made with ColorBrewer, http://www.colorbrewer2.org.

color axes: red, blue, green, and yellow. Other possibilities when more colors are needed are orange, brown, pink, magenta, purple, and cyan. However, colormap design is a tricky problem: careful attention must be paid to luminance and saturation. Luminance constrast is a major issue: for some uses, the colors should be close in luminance to avoid major differences in salience and to ensure that all can be seen against the same background. For example, fully saturated yellow and green will have much less luminance contrast against a white background than red and blue. For other uses, colors should be sufficiently different in luminance that they can be distinguished even in black and white. Moreover, colormaps for small regions such as lines should be highly saturated, but large regions such as areas should have low saturation. Thus, the appropriate colormap may depend on the mark type.

A good resource for creating colormaps is ColorBrewer at http://www.colorbrewer2.org, a system that incorporates many perceptual guidelines into its design in order to provide safe suggestions. It was used to create both the ten-element low-saturation map in Figure 10.7(a) and the eight-element high-saturation map in Figure 10.7(b). The low-saturation pastel map is well suited for large regions, leaving fully saturated colors for small road marks. In contrast, the eight-element map that uses highly saturated colors is much too bright for the large areas shown here, but would be a good fit for small line or point marks.



(a) (b)

Figure 10.8. Ineffective categorical colormap use. (a) The 21 colors used as an index for each mouse chromosome can indeed be distinguished in large regions next to each other. (b) In noncontiguous small regions only about 12 bins of color can be distinguished from each other, so a lot of information about how regions in the mouse genome map to the human genome is lost. From [Sinha and Meller 07, Figure 2].

Figure 10.8 illustrates an attempt to use categorical color that is ineffective because of a mismatch between the number of color bins that we can distinguish in noncontiguous small regions and the number of levels in the categorical attribute being encoded. Figure 10.8(a) shows that one color has been assigned to each of the 21 chromosomes in the mouse. All 21 of these colors can indeed be distinguished from each other in this view that acts as a legend and an index, because regions are large and the most subtle differences are between regions that are right next to each other.

In Figure 10.8(b), the regions of the human chromosomes that correspond to those in the mouse chromosomes have been colored to illustrate how genomic regions have moved around as the species evolved independently from each other after diverging from a common ancestor. In this case, the colored regions are much smaller and not contiguous. The 21 colors are definitely not distinguishable from each other in this view: for example, only about three bins of green can be distinguished in the human view, rather

than the full set of seven in the mouse view. Similarly, the full set of five pinks and purples in the mouse view has collapsed into about three distinguishable bins in the human view. In total, only around 12 bins of color can be distinguished in the human view.

When you are faced with the problem of discriminability mismatch, there are two good design choices. One choice is to reduce the number of bins explicitly through a deliberate data transformation that takes into account the nature of the data and task, so that each bin can be encoded with a distinguishable color. This choice to derive a new and smaller set of attributes is better than the inadvertent segmentation into bins that arises from the user's perceptual system, which is unlikely to match meaningful divisions in the underlying data. For example, the attribute may have hierarchical structure that can be exploited to derive meaningful aggregate groups, so that one color can be used per group. Another possibility is to filter the attributes to only encode a small set of the most important ones with color, and aggregate all of the rest into a new category of *other*; the Constellation system analyzed in Section 15.8 takes this approach to color coding links, where a few dozen categories were narrowed down to fit within eight bins.

► Aggregation and filtering idioms are covered in Chapter 13.

The other choice is to use a different visual encoding idiom that uses other visual channels instead of, or in addition to, the color channel alone. Figure 10.9 shows an example of systematically considering a large space of visual encoding possibilities for visualizing biological experiment workflows. The dataset has 27 categorical levels in total that are gathered into seven categories with between three and seven levels each. Seven designs were considered, using multiple channels in addition to color: shape, size, and more complex glyphs that evoke metaphoric associations. Figure 10.10 shows the final choice made, where the color channel was only used to encode the four levels in category S7, and other channels were used for the other categories.

10.3.2 Ordered Colormaps

An **ordered** colormap is appropriate for encoding ordinal or quantitative attributes. The two major variants of continuous colormaps for ordered data have expressiveness charactistics that should match up with the attribute type. A **sequential** colormap ranges from a minimum value to a maximum value. If only the luminance channel is used, the result is a grayscale ramp. When incorporating hue, one end of the ramp is a specific hue at full saturation and brightness. If saturation is the variable, the other end is pale

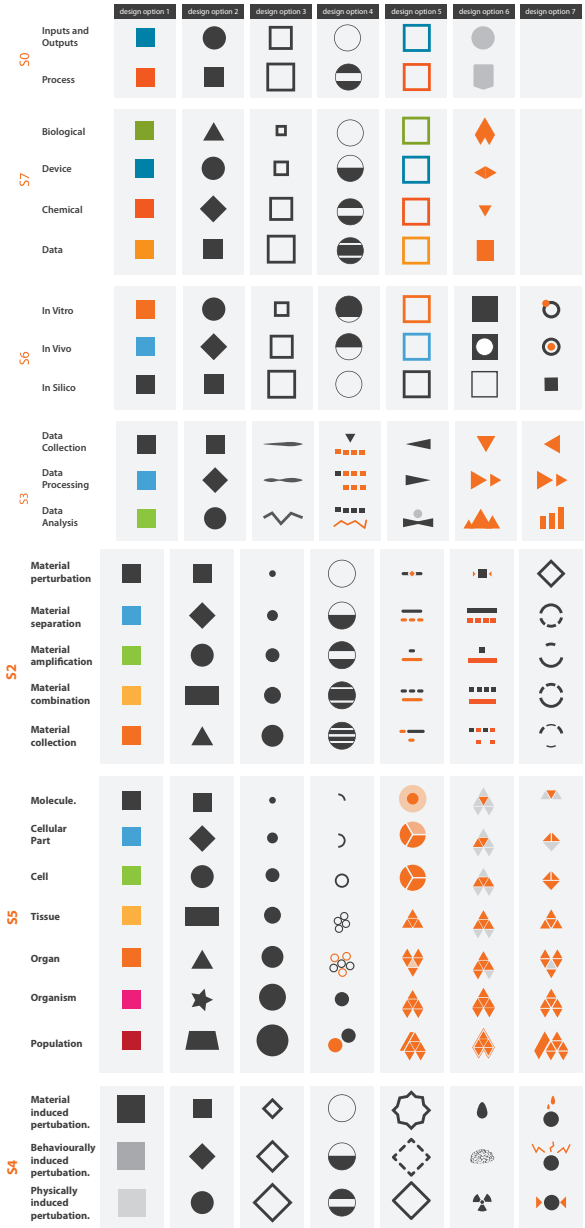


Figure 10.9. Effective categorical colormap use: A large space of visual encoding possibilities for 27 categories was considered systematically in addition to the color channel, including size and shape channels and more complex glyphs. From [Maguire et al. 12, Figure 5].

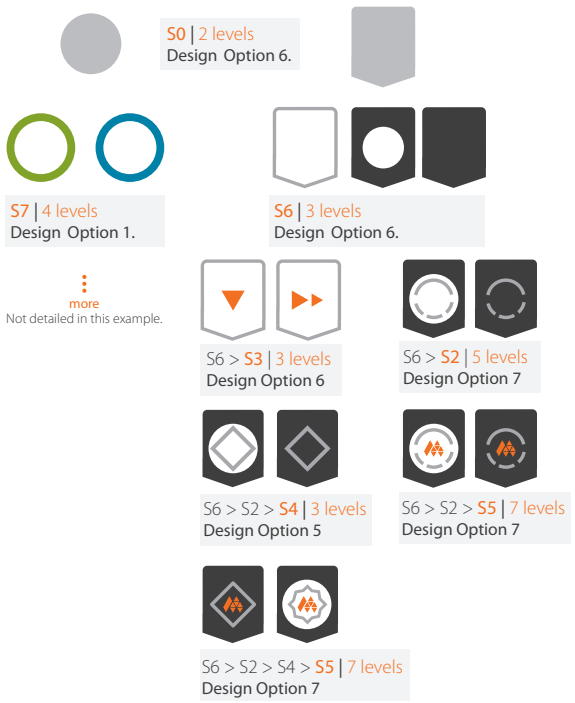
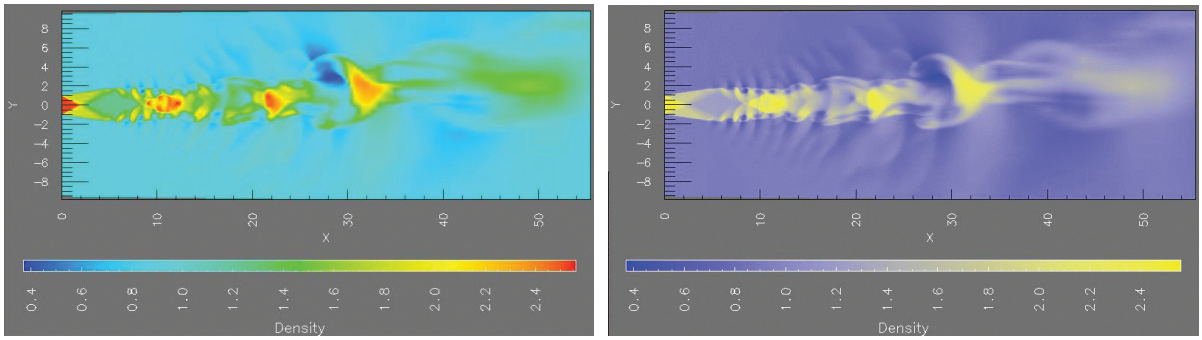


Figure 10.10. Effective categorical colormap use: The final design uses the color channel for only four of the categories. From [Maguire et al. 12, Figure 6].

or white; when luminance is the varying quantity, the other end is dark or black. A **diverging** colormap has two hues at the endpoints and a neutral color as a midpoint, such as white, gray, or black, or a high-luminance color such as yellow.

The question of how many unique hues to use in continuous colormaps depends on what level of structure should be emphasized: the high-level structure, the middle range of local neighborhoods, or fine-grained detail at the lowest level. Figure 10.11 shows the same fluid flow data with two different colormaps. Figure 10.11(a) emphasizes mid-level neighborhood structure with many hues. Figure 10.11(b) emphasizes large-scale structure by ranging between two fully saturated hues, with saturation smoothly varying to a middle point: in this sequential case, the ends are purple and yellow, and the midpoint is gray.

One advantage of the rainbow colormap shown in Figure 10.11(a) is that people can easily discuss specific subranges because the



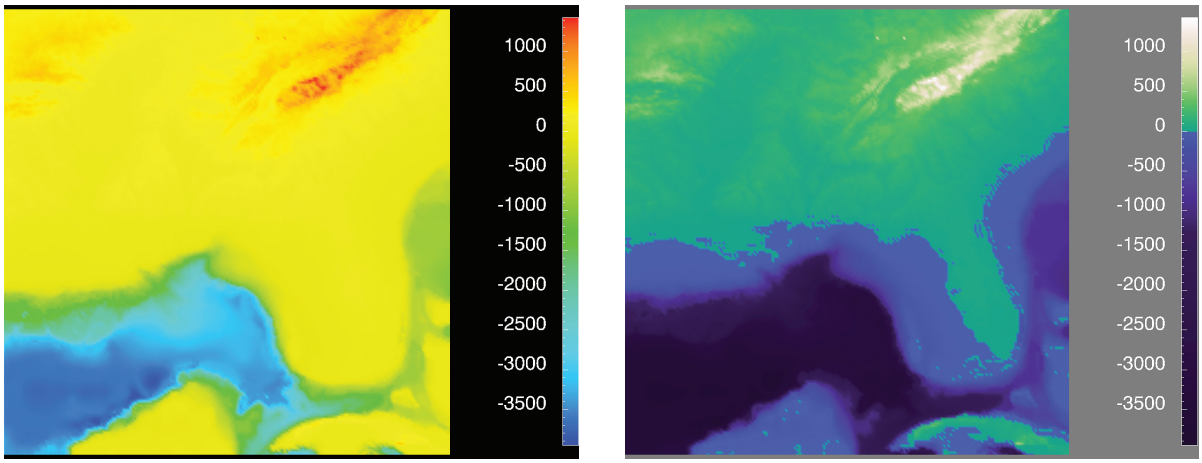
(a) (b)

Figure 10.11. Rainbow versus two-hue continuous colormap. (a) Using many hues, as in this rainbow colormap, emphasizes mid-scale structure. (b) Using only two hues, the blue–yellow colormap emphasizes large-scale structure. From [Bergman et al. 95, Figures 1 and 2].

differences are easily nameable: "the red part versus the blue part versus the green part". In colormaps that use just saturation or luminance changes, there are only two hue-based semantic categories: "the purple side versus the yellow side". It is not easy to verbally distinguish between smaller neighborhoods—we cannot easily demarcate the "sort-of-bright purple" from the "not-quite-so-bright purple" parts.

However, rainbow colormaps suffer from three serious problems at the perceptual level; it is rather unfortunate that they are a default choice in many software packages. Figure 10.12 illustrates all three problems. First, hue is used to indicate order, despite being an identity channel without an implicit perceptual ordering. Second, the scale is not perceptually linear: steps of the same size at different points in the colormap range are not perceived equally by our eyes. A range of 1000 units in Figure 10.12(a) has different characteristics depending on where within the colormap it falls. While the range from –2000 to –1000 has three distinct colors, cyan and green and yellow, a range of the same size from –1000 to 0 simply looks yellow throughout. Third, fine detail cannot be perceived with the hue channel; the luminance channel would be a much better choice, because luminance contrast is required for edge detection in the human eye.

One way to address all three problems is to design **monotonically increasing luminance** colormaps: that is, where the multiple hues are ordered according to their luminance from lowest to high-



(a) (b)

Figure 10.12. Rainbow versus multiple-hue continuous colormap with monotonically increasing luminance. (a) Three major problems with the common continuous rainbow colormap are perceptual nonlinearity, the expressivity mismatch of using hue for ordering, and the accuracy mismatch of using hue for fine-grained detail. (b) A colormap that combines monotonically increasing luminance with multiple hues for semantic categories, with a clear segmentation at the zero point, succeeds in showing high-level, mid-level, and low-level structure. From [Rogowitz and Treinish 98, Figure 1].

est. The varying hues allow easy segmentation into categorical regions, for both seeing and describing mid-level neighborhoods. Luminance is a magnitude channel, providing perceptual ordering. It supports both high-level distinctions between one end ("the dark parts") and the other ("the light parts") and low-level structure perception because subtle changes in luminance are more accurately perceived than subtle changes in hue. Figure 10.12(b) illuminates the true structure of the dataset with a more appropriate colormap, where the luminance increases monotonically. Hue is used to create a semantically meaningful categorization: the viewer can discuss structure in the dataset, such as the dark blue sea, the cyan continental shelf, the green lowlands, and the white mountains. The zero point matches with sea level, a semantically meaningful point for this dataset.

It is possible to create a perceptually linear rainbow colormap, but at the cost of losing part of the dynamic range because the fully saturated colors are not available for use. Figure 10.13 shows an example created with a system for calibrating perceptually based colormaps [Kindlmann 02]. The perceptually nonlinear rainbow

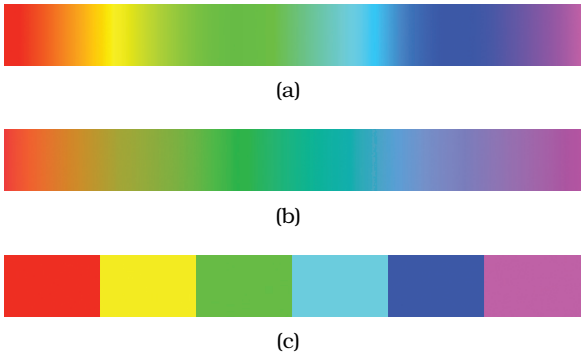


Figure 10.13. Appropriate use of rainbows. (a) The standard rainbow colormap is perceptually nonlinear. (b) Perceptually linear rainbows are possible [Kindlmann 02], but they are less bright with a decreased dynamic range. (c) Segmented rainbows work well for categorical data when the number of categories is small.

in Figure 10.13(a) can be converted to the perceptually linear one shown in Figure 10.13(b); however, it is so much less bright than the standard one that it seems almost dingy, so this solution is not commonly used.

Rainbows are not always bad; a segmented rainbow colormap is a fine choice for categorical data with a small number of categories. Figure 10.13(c) shows an example. Segmented rainbows could also be used for ordered data; while not ideal, at least the perceptual nonlinearity problem is solved because the colormap range is explicitly discretized into bins. Using a segmented colormap on quantitative data is equivalent to transforming the datatype from quantitative to ordered. This choice is most legitimate when task-driven semantics can be used to guide the segmentation into bins. The intuition behind the technique is that it is better to deliberately bin the data explicitly, rather than relying on the eye to create bins that are of unequal size and often do not match meaningful divisions in the underlying data.

10.3.3 Bivariate Colormaps

The safest use of the color channel is to visually encode a single attribute; these colormaps are known as **univariate**. Figure 10.6 includes several colormaps that encode two separate attributes, called **bivariate**. When one of the two attributes is **binary**, meaning it has only two levels, then it is straightforward to create a com-

prehensible bivariate colormap with two families of colors by fixing a base set of hues and varying their saturation, as in the bivariate categorical–binary and diverging–binary examples in Figure 10.6. This approach can also be useful for a single categorical attribute that has a hierarchical structure.

When both attributes are categorical with multiple levels, results will be poor [Wainer and Francolini 80], and thus there are no bivariate categorical–categorical maps in Figure 10.6. The case of encoding combinations of two sequential or diverging attributes with multiple levels is a middle ground. Figure 10.6 shows several examples with three levels for each attribute: sequential–sequential, diverging–diverging, diverging–sequential, and categorical–sequential. While these colormaps do appear frequently in vis systems, you should be aware that some people do have difficulty in interpreting their meaning.

10.3.4 Colorblind-Safe Colormap Design

Designers using color should take the common problem of red–green color blindness into account. It is a sex-linked inherited trait that affects 8% of males and a much smaller proportion of females, 0.5%. In the common forms of color blindness the ability to sense along the red–green opponent color axis is limited or absent. The problem is not limited to simply telling red apart from green; many pairs that are discriminable to people with normal color vision are confused, including red from black, blue from purple, light green from white, and brown from green.

► Opponent color is discussed in Section 10.2.

On the theoretical side, the safest strategy is to avoid using only the hue channel to encode information: design categorical colormaps that vary in luminance or saturation, in addition to hue. Clearly, avoiding colormaps that emphasize red–green, especially divergent red–green ramps, would be wise. In some domains there are strong conventions with the use of red and green, so those user expectations can be accommodated by ensuring luminance differences between reds and greens.

► For example, the historical and technical reasons behind red–green usage in bioinformatics domain are discussed in Section 7.5.2.

On the practical side, an excellent way to ensure that a design uses colors that are distinguishable for most users is to check it with a color blindness simulator. This capability is built into many desktop software tools including Adobe Illustrator and Photoshop, and also available through web sites such as http://www.rehue.net, http://www.color-blindness.com, and http://www.etre.com/tools/colourblindsimulator.