Color Do's and Don'ts

This article explains the difference between color hue and color value and describes how and how not to use the two in data visualization. Color, in data visualization, can be a very good thing. It can add dimensionality, emphasize outliers and create focus. Used well, color makes visualizations more interesting and meaningful; but applied badly, it can render them misleading and dysfunctional eye candy.

Good color design has to take a variety of issues into consideration, including:

1. Existing color conventions and metaphors. A good visualization design will leverage existing color semantics when appropriate and avoid those that could lead to mixed messages and confusion. The designer also has to understand that color conventions are not universal. Red, for example, does not mean the same thing in the west as in the east.
2. The choice of media (print versus digital).
3. Color theory and perception. The remainder of this article introduces a number of basic color theory and perception principles to help the data visualization designer use color meaningfully and effectively, while avoiding common color pitfalls.
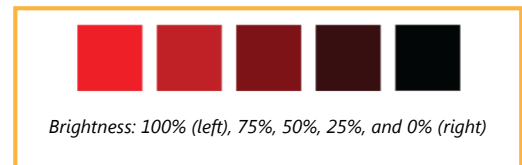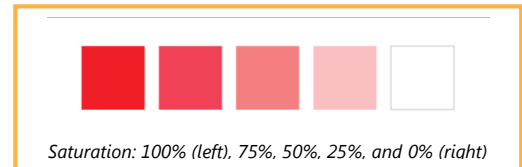
Color Theory Basics

Any given color can be described as a combination of *hue*, *saturation* and *brightness*.

Hue refers to the color itself (i.e. a particular shade of green vs blue vs orange).
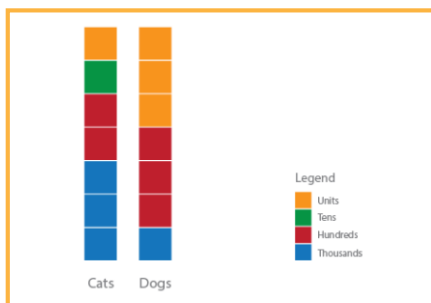

Different Color Hues

Saturation refers to the amount of color present. Think of it as the amount of pigment in a can of white paint. In the illustration below, the square on the left has 100% *saturation* and the one on the right has 0% saturation (in other words, there is a complete absence of color).

Saturation: 100% (left), 75%, 50%, 25%, and 0% (right)

Brightness refers to the lightness or darkness of the color.
Think of it as the amount of black pigment in a can of paint. In the illustration below, the square on the left has 100% *brightness* (i.e. no black) and the one on the right has 0% brightness (i.e. all black).

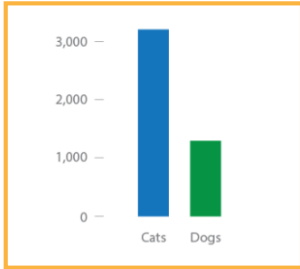*Brightness: 100% (left), 75%, 50%, 25%, and 0% (right)*

Color Rule #1: Don't use color *hue* to represent quantity. Color can be used to indicate the quality of a good number (such as good, neutral, bad)

For a number of reasons, it is bad practice to map color *hue* to quantitative data. First, we do not naturally associate color with quantity. For example, we don't think of orange as inherently bigger or smaller than, say, green. To illustrate this point, consider the following chart. Quick, which is more, cats or dogs?
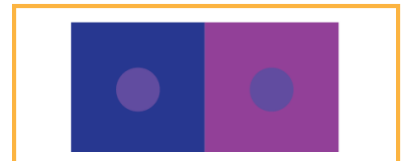
While not, technically, incorrect, this chart would not be an effective data visualization. On first sight, we tend to perceive boxes of equal area and bars of equal height as being of equal magnitude. Size governs over color. As a result, determining whether there are more cats or dogs requires a great deal of mental math to add up the Cats column, commit the total to memory, then add up the Dogs column and compare the two numbers. And all that cognitive load defeats the purpose of data visualization. A better representation would be this straightforward bar chart, which clearly shows the difference between the two animals.

Different color *hues*, therefore, can be used to distinguish categories—cats and dogs, profit and loss, political parties—but should not be used to represent quantity.

Also if you give meaning to a color within a visualization (such as to represent the quality of a number (good, bad, great) or a category (dogs, cats) then it should be noted that one should not redefine the meaning of that color in other parts of the visualization. So, don't use red in one part to mean 'bad' and use red somewhere else to mean 'cats' or remove the meaning by just using as a color of a line in a line chart. If red represents 'bad' in one area then it should mean 'bad' for all other areas within a visualization.

Second, background color affects how we perceive an object's color. For example, in the illustration below, the same purple circle appears more red against the bluer background and more blue against the redder background.

Additionally, the amount of figure ground contrast can affect our perception of the color. In the figure below, there is more contrast between the yellow circle and the black background than the white background. The opposite is true for the dark blue color. In this case, our perception is governed by background contrast more than color *hue*.
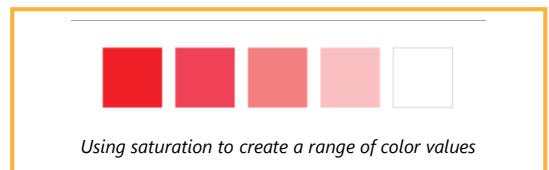

Color Rule #2: If you need the color to represent quantity, use color *value* instead of *hue*.
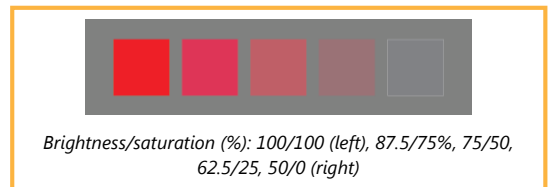
What is color *value*? Color theorists would say that *value* is just another word for *brightness* and leave it at that. While technically true, in data visualization practice, *value* can mean either *saturation*, *brightness* or a combination of the two—depending on the color of the background. For example, on a black background, *value* is, indeed, created by changing *brightness (hue* and *saturation* remaining constant).

Using brightness to create a range of color values

Against a white background, however, it is more effective to vary *saturation* than *brightness*. The figure to the right illustrates a range of *saturations* (*hue* and *brightness* remaining constant).

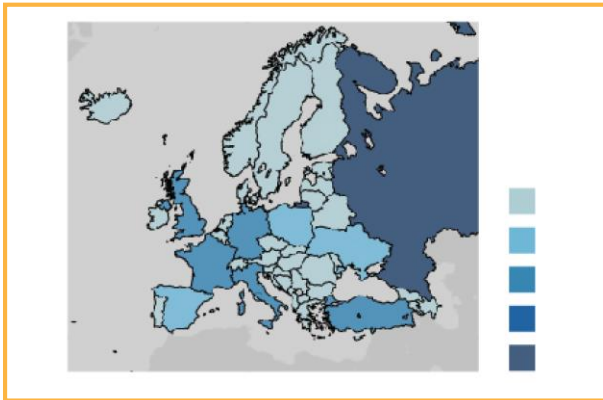*Using saturation to create a range of color values*

And, by extension, if the background were grey, value would become a slightly more complicated combination of both brightness and saturation, as below.

Brightness/saturation (%): 100/100 (left), 87.5/75%, 75/50, 62.5/25, 50/0 (right)

The important thing in all three cases is that, regardless of whether *brightness,* saturation or a combination of the two is used, the square with the maximum *value* pops out from the background; that with the minimum *value* fades into the background. In the first two cases, it would be relatively straightforward to map a quantitative variable to either *brightness* or *saturation.* In the third case, it would be more complicated because both are variable. In practice, mapping a quantitative variable to a combination of *brightness* and *saturation* should be avoided.

It is important to note that the same principal applies to other areas of visualization. For example, let's say we have a line chart showing last year sales vs. current year sales (where the focus should be on this years sales, and last year is there for reference) – the color used for last year sales should be more subtle than the color for current years sales. So, color can be used to draw ones focus to the most important parts of a visualization.

The human eye does not perceive changes in *hue*, *saturation* and *brightness* linearly. The eye's ability to distinguish subtle differences between *values* is also limited. For these reasons, *values* are typically binned, as in the choropleth map here. The number of bins should be limited to five to eight. More bins could result in reduced legibility and increased viewer error.

When it comes to color *hue* and *value*, it is also worth noting, that people with color blindness can perceive differences in a given color's *value* easier than differences between certain color *hues*. Using *value*, therefore, can improve accessibility.

Color is an important graphic variable in data visualization—provided it is applied appropriately. As a rule, use color *hue* to distinguish categorical (i.e. nominal) data; use color *value* to map quantity.

Author: Anne Stevens is a visualization designer and teacher based in Toronto.

www.dundas.com