Use of Color in Data Visualization

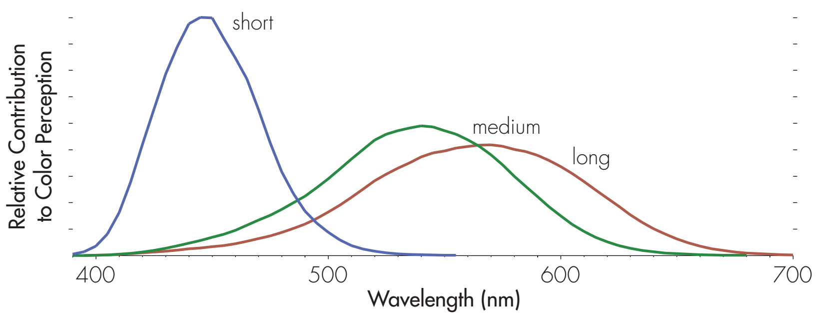
Robert Simmon robert.simmon@nasa.gov

Color selection in data visualization is not merely an aesthetic choice, it is a crucial tool to convey quantitative information. Properly selected colors convey the underlying data accurately, in contrast to many color schemes commonly used in visualization that distort relationships between data values. Judicious use of color also allows multiple datasets to be layered together, which helps to create graphics that tell stories of cause and effect. I will explain theories of color use in visualization, and show methods for selecting effective color palettes.

Introduction to Color

Visible light is the portion of the electromagnetic spectrum visible to the human eye, ranging from wavelengths of roughly 400 to 700 nm. Differences in wavelength are perceived as the familiar colors of the rainbow. From short to long wavelengths: violet, blue, green, yellow, orange, and red.

In the eye, cells called cones are responsible for our ability to discriminate colors. There are three varieties of cones, sensitive to short, medium, and long wavelengths. Each type of cone is responsive to a range of wavelengths, with peak sensitivities at 420, 530, and 560 nm. Color is determined by the relative number of photons detected by each type of cone. Because of this, two combined lights with different wavelengths are indistinguishable from a single color. Cone response is not linear across the spectrum: some colors (green and red in particular) extend over a broad range of wavelengths, while others (yellow and blue) occupy narrow bands.

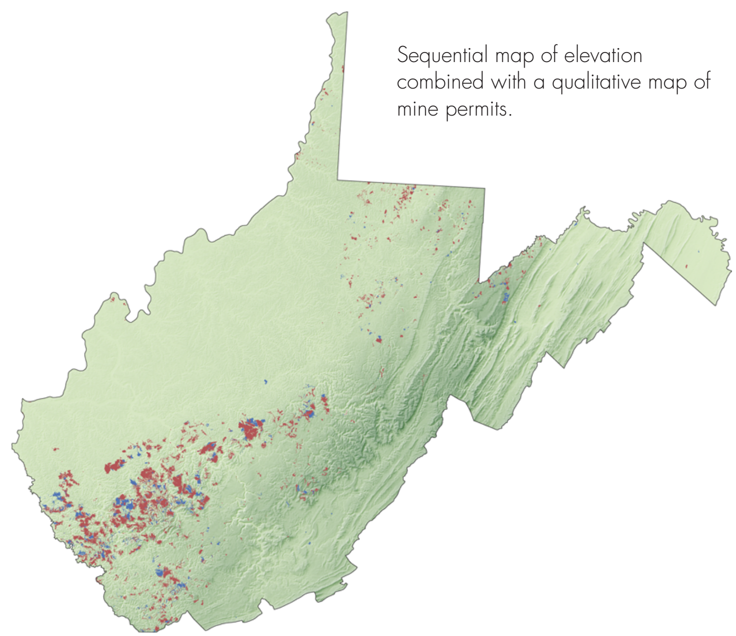


RGB Color

Televisions and computer screens generate a spectrum of colors by combining pixels of separate primary colors that roughly correspond to the three types of cones—red, green, and blue. The wavelengths of the three primaries do not exactly match the peak wavelengths of cones in the eye, and emit at narrow wavelengths vs. the broad response of cones. Combined, these effects result in a gamut of colors on a display that is smaller than the full range of colors humans can distinguish. Furthermore, pure red, green, and blue are not equal in brightness, and changes in their intensities can result in nonlinear changes in perceived color.

Color and Data Display

Color is one of the most effective ways to encode two-dimensional data. Differences in color can distinguish different categories (for example cropland, forest, or urban areas in a land cover map) or indicate quantity (percent forest cover or population). Color schemes for these two types of maps are described as qualitative and sequential.


Sequential map of elevation combined with a qualitative map of mine permits.

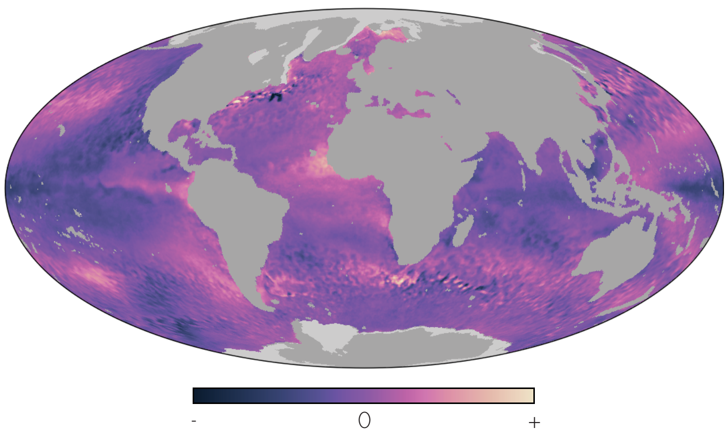
Divergent Schemes

A subset of sequential color schemes, used for data that depart from an average or neutral quantity (temperature anomaly, electric charge, or pH), is called a divergent scheme.

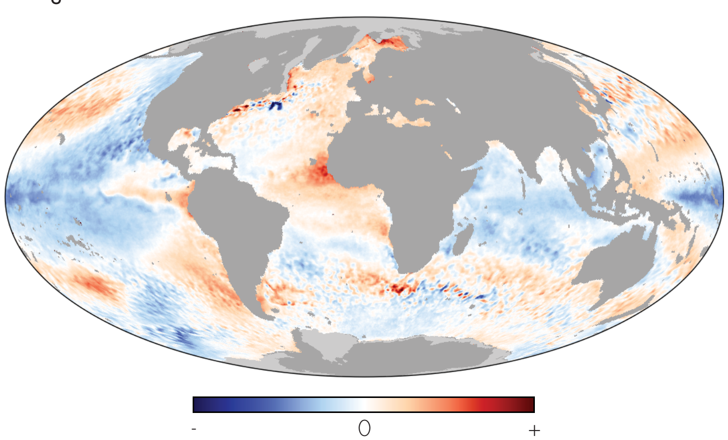
Qualitative Schemes

Colors in qualitative maps should be easily distinguishable from one another. They should also be similar in lightness and saturation to prevent classes from being over or under emphasized. Unfortunately, humans are only able to reliably distinguish 5–10 colors simultaneously, so the number of class must be small. Using saturated, medium-bright "named" colors is a good approach: red, blue, green, purple, orange, etc.

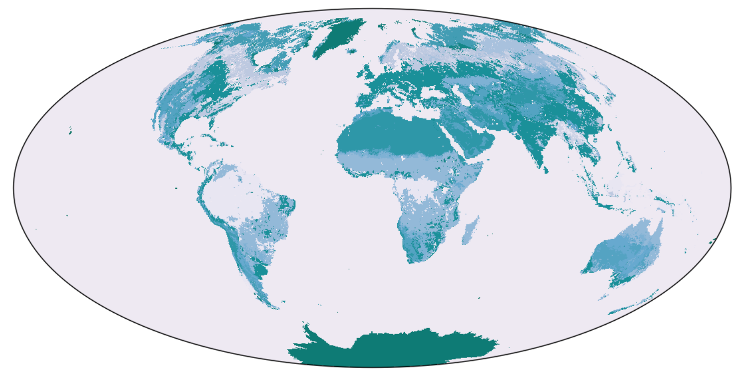
Sequential scheme on a divergent dataset



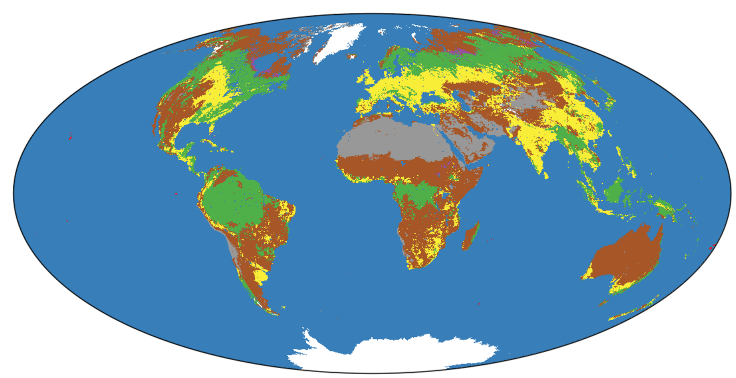
Divergent scheme



Sequential scheme on a qualitative dataset



Qualitative scheme



Sequential Displays

Sequential maps display quantities of data. To accurately display the data and relationships between data points, care must be taken to ensure that a change in the value of a parameter is perceived proportionally. Some commonly used color palettes—especially the rainbow palette—do not accurately maintain relationships, and are a poor choice for data display. Transitions between some colors, green and red, for example, occur very rapidly, leading to false contrast. Other transitions, especially green, are gradual, and there is a loss of detail. Rainbow palettes have another deficiency: because the overall brightness of the colors increases and decreases over the range of hues there is no natural progression of values. An alternative is to only use brightness, not color, to encode value, but surrounding tones can significantly alter the perceived values of pixels. Grayscale palettes are best limited to black and white reproductions. A better approach is to use a color scheme that spirals through a perceptual color space, with each step equally different in hue, saturation, and brightness.

Resources

Color Brewer
http://colorbrewer.org/

Using Color in Information Display Graphics
http://colorusage.arc.nasa.gov/

References

Brewer, Cynthia A. (2005) *Designing Better Maps: A Guide for GIS Users.* Redlands, CA: ESRI Press.

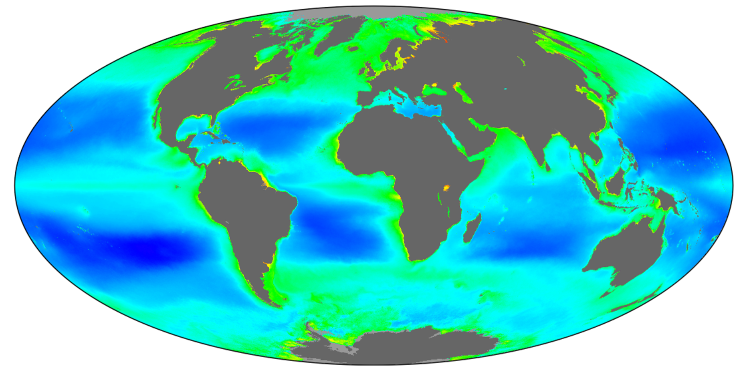
MacEvoy, Bruce. (2008) Watercolors. Accessed April 28, 2008.
http://www.handprint.com/HP/WCL/water.html

Rogowitz, B.E.; Treinish, L.A. (1998) Data visualization: the end of the rainbow," *Spectrum, IEEE* , vol.35, no.12, pp.52-59.

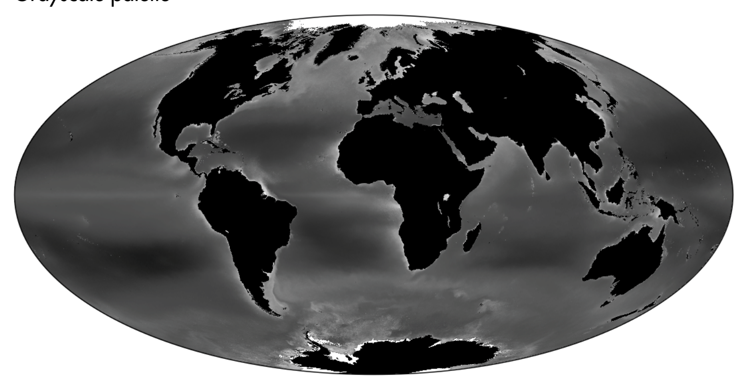
Tufte, Edward R. (1990) *Envisioning Information.* Cheshire, CT: Graphics Press.

Ware, Colin. (2005) *Information Visualization, Second Edition.* San Francisco, CA: Morgan Kaufmann.

Rainbow palette



Grayscale palette



Perceptual palette