Choosing Color Palettes for Statistical Graphics

Achim Zeileis  Kurt Hornik

http://statmath.wu-wien.ac.at/~zeileis/
Overview

• Motivation
  – Statistical graphics and color
  – Color vision and color spaces

• Palettes (in HCL space)
  – Qualitative
  – Sequential
  – Diverging

• Software
Motivation: Statistical graphics

Information in statistical graphics is typically coded by:

- **length**
  - easy to decode for humans
  - best for aligned common scales

- **area, volume**
  - more difficult to decode
  - dependence on shape: long/thin is seen larger than compact/convex
  - dependence on color: lighter areas seen larger

- **angle, slope**
  - problematic for humans
  - dependence on orientation

- **color**
  - omni-present in statistical graphics
Motivation: Statistical graphics

- particularly important for shading areas (e.g., bar plots, pie charts, mosaic displays, heatmaps, …)
- avoid large areas of saturated colors
- powerful for encoding categorical information
- care needed for coding quantitative information

More often than not: Only little guidance about how to choose a suitable palette for a certain visualization task.

Question: What are useful color palettes for coding qualitative and quantitative information?

Currently: Many palettes are constructed based on HSV space, especially by varying hue.
Motivation: Statistical graphics

Examples:

- heatmap of bivariate kernel density estimate for Old Faithful geyser eruptions data,
- map of Nigeria shaded by residuals from a model for childhood mortality,
- pie chart of seats in the German parliament *Bundestag*,
- mosaic display of votes for the German Bundestag,
- model-based mosaic display for hair and eye color data,
- scatter plot with three clusters (and many points).
Motivation: Statistical graphics
Motivation: Statistical graphics
Motivation: Statistical graphics
Motivation: Statistical graphics

CDU/CSU  FDP  SPD  Gr Li

Schleswig–Holstein
Hamburg
Niedersachsen
Bremen
Nordrhein–Westfalen
Hessen
Rheinland–Pfalz
Bayern
Baden–Württemberg
Saarland
Mecklenburg–Vorpommern
Brandenburg
Sachsen–Anhalt
Berlin
Sachsen
Thüringen
Motivation: Statistical graphics

Hair

Black
Brown
Red
Blond

Eye

Brown
Black
Green
Hazel
Blue
Brown
Red
Blond

Pearson residuals:

p-value =
< 2.22e−16

7.05
4.00
2.00
0.00
-2.00
-4.00
-5.85
Motivation: Statistical graphics
Motivation: Statistical graphics

Problems:

• Flashy colors: good for drawing attention to a plot but hard to look at for a longer time.

• Large areas of saturated colors: can produce distracting after-image effects.

• Unbalanced colors: light and dark colors are mixed; or “positive” and “negative” colors are difficult to compare.

• Quantitative variables are often difficult to decode.
Motivation: Statistical graphics

Solutions:

- Use pre-fabricated color palettes (with fixed number of colors) designed for specific visualization tasks: **ColorBrewer.org** (see Brewer, 1999).

  Problem: little flexibility.

- Selecting colors along axes in a color space whose axes can be matched with perceptual axes of the human visual system.

  Leads to similar palettes compared to **ColorBrewer.org** but offers more flexibility via a general principle for choosing palettes.
Human color vision is hypothesized to have evolved in three distinct stages:

1. **light/dark** (monochrome only)
2. **yellow/blue** (associated with warm/cold colors)
3. **green/red** (associated with ripeness of fruit)
Due to these three color axes, colors are typically described as locations in a 3-dimensional space, often by mixing three primary colors, e.g., RGB or CIEXYZ.

Physiological axes do not correspond to natural perception of color but rather to polar coordinates in the color plane:

- **hue** (dominant wavelength)
- **chroma** (colorfulness, intensity of color as compared to gray)
- **luminance** (brightness, amount of gray)

Perceptually based color spaces try to capture these three axes of the human perceptual system, e.g., HSV or HCL.
Color vision and color spaces

**HSV space** is a standard transformation of RGB space implemented in most computer packages.

**Specification:** triplet \((H, S, V)\) with \(H = 0, \ldots, 360\) and \(S, V = 0, \ldots, 100\), often all transformed to unit interval (e.g., in \(R\)).

**Shape:** cone (or transformed to cylinder).

**Problem:** dimensions are confounded, hence not really perceptually based.
Color vision and color spaces
Color vision and color spaces
**Color vision and color spaces**

**HCL space** is a perceptually based color space, polar coordinates in CIELUV space.

**Specification:** triplet \((H, C, L)\) with \(H = 0, \ldots, 360\) and \(C, L = 0, \ldots, 100\).

**Shape:** distorted double cone.

**Problem:** Care is needed when traversing along the axes due to distorted shape.
Color vision and color spaces
Color vision and color spaces
Palettes: Qualitative

**Goal:** Code qualitative information.

**Solution:** Use different hues for different categories. Keep chroma and luminance fixed, e.g.,

\[(H, 50, 70)\]

**Remark:** The admissible hues (within HCL space) depend on the values of chroma and luminance chosen.

Hues can be chosen from different subsets of \([0, 360]\) to create different “moods” or as metaphors for the categories they code (see Ihaka, 2003).
Palettes: Qualitative
Palettes: Qualitative
Palettes: Qualitative

dynamic [30, 300]

harmonic [60, 240]

cold [270, 150]

warm [90, −30]
Palettes: Qualitative

- SPD
- CDU/CSU
- Grüne
- Linke
- FDP
Palettes: Qualitative

CDU/CSU | FDP | SPD | Gr Li
--- | --- | --- | ---
Schleswig-Holstein
Hamburg
Niedersachsen
Bremen
Nordrhein-Westfalen
Hessen
Rheinland-Pfalz
Bayern
Baden-Württemberg
Saarland
Mecklenburg-Vorpommern
Brandenburg
Sachsen-Anhalt
Berlin
Sachsen
Thüringen
Palettes: Qualitative
Palettes: Qualitative
Palettes: Sequential

Goal: Code quantitative information. Intensity/interestingness $i$ ranges in $[0, 1]$, where 0 is uninteresting, 1 is interesting.

Solution: Code $i$ by increasing amount of gray (luminance), no color used, e.g.,

$$(H, 0, 90 - i \cdot 60)$$

The hue $H$ does not matter, chroma is set to 0 (no color), luminance ranges in $[30, 90]$, avoiding the extreme colors black and white.

Modification: In addition, code $i$ by colorfulness (chroma). Thus, more formally:

$$(H, 0 + i \cdot C_{\text{max}}, L_{\text{max}} - i \cdot (L_{\text{max}} - L_{\text{min}}))$$

for a fixed hue $H$. 
Palettes: Sequential
Palettes: Sequential

Modification: To increase the contrast within the palette even further, simultaneously vary the hue as well:

\[
(H_2 - i \cdot (H_1 - H_2), \quad \text{C}_{\text{max}} - i^{p_1} \cdot (\text{C}_{\text{max}} - \text{C}_{\text{min}}),
\]
\[
L_{\text{max}} - i^{p_2} \cdot (L_{\text{max}} - L_{\text{min}}).
\]

To make the change in hue visible, the chroma needs to increase rather quickly for low values of \(i\) and then only slowly for higher values of \(i\).

A convenient transformation for achieving this is to use \(i^p\) instead of \(i\) with different powers for chroma and luminance.
Palettes: Sequential
Palettes: Sequential
Palettes: Sequential
Palettes: Sequential
Palettes: Diverging

**Goal:** Code quantitative information. Intensity/interestingness \( i \) ranges in \([-1, 1]\), where 0 is uninteresting, \( \pm 1 \) is interesting.

**Solution:** Combine sequential palettes with different hues.

**Remark:** To achieve both large chroma and/or large luminance contrasts, use hues with similar chroma/luminance plane, e.g., \( H = 0 \) (red) and \( H = 260 \) (blue).
Palettes: Diverging
Palettes: Diverging
Palettes: Diverging
Palettes: Diverging
Palettes: Diverging

Hair

Black
Brown
Red
Blond

Eye

Black
Brown
Red
Blond

Hair residuals:
p-value = < 2.22e−16

Pearson residuals:
−5.85
−4.00
−2.00
0.00
2.00
4.00
7.05

Eye

Blue
Brown
Green
Hazel

p-value = < 2.22e−16
Palettes: Diverging

Pearson residuals:
p-value = < 2.22e−16
Implementing HCL-based palettes is not difficult:

- If HCL colors are available, our formulas are straightforward to implement.
- If not, HCL coordinates typically need to be converted to RGB coordinates for display. Formulas are available, e.g., in Wikipedia (2006ab).

R has an implementation of various color spaces (including HCL) in Ross Ihaka’s `colorspace` package. Based on this, our `vcd` package provides `rainbow_hcl()`, `sequential_hcl()`, `heat_hcl()`, and `diverge_hcl()`.

For documentation and further examples, see `?rainbow_hcl` and `vignette("hcl-colors", package = "vcd")`. 


