

Escaping RGBlanc: Selecting Colors for Statistical Graphics

Achim Zeileis, Kurt Hornik, Paul Murrell

<http://statmath.wu.ac.at/~zeileis/>

Overview

- Motivation
 - Color in statistical graphics
 - Challenges
 - Illustrations
- Color vision and color spaces
- Color palettes
 - Qualitative
 - Sequential
 - Diverging
- Color blindness
- Software

Color in statistical graphics

Color:

- Integral element in graphical displays.
- Easily available in statistical software.
- Omnipresent in (electronic) publications: Technical reports, electronic journal articles, presentation slides.

Problem: Little guidance about how to choose appropriate colors for a particular visualization task.

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

Challenges

Colors should not be unappealing:

- Basic principles: Colors should be intuitive, avoid large areas of saturated colors.
- Control of perceptual properties: Hue, brightness, colorfulness.
- Employ a *color model* or *color space*.
 - RGB (Red-Green-Blue): Corresponds to generation of colors on computer, unintuitive for humans.
 - HSV (Hue-Saturation-Value): Simple transformation of RGB, easily available. *But:* Maps poorly to perceptual properties, encourages use of highly saturated colors.
 - HCL (Hue-Chroma-Luminance): Transformation of CIELUV space, mitigates problems above.

Challenges

Colors in a statistical graphic should cooperate with each other:

- Purpose: Distinguish different elements of a statistical graphic depending on the levels of some variable.
- *Palette* of colors needed.
- Natural idea: Vary one or more perceptual properties.
- In perceptual color space: Traverse paths along dimensions.
- Problem: Often based on HSV space with rather poor results.
- Solution: HCL. Not only “location” but also “motion” is intuitive.

Colors should work everywhere:

- Ideally: Screen, projector, (grayscale) printer, color-blind viewers.
- Cannot always (but often) be attained.

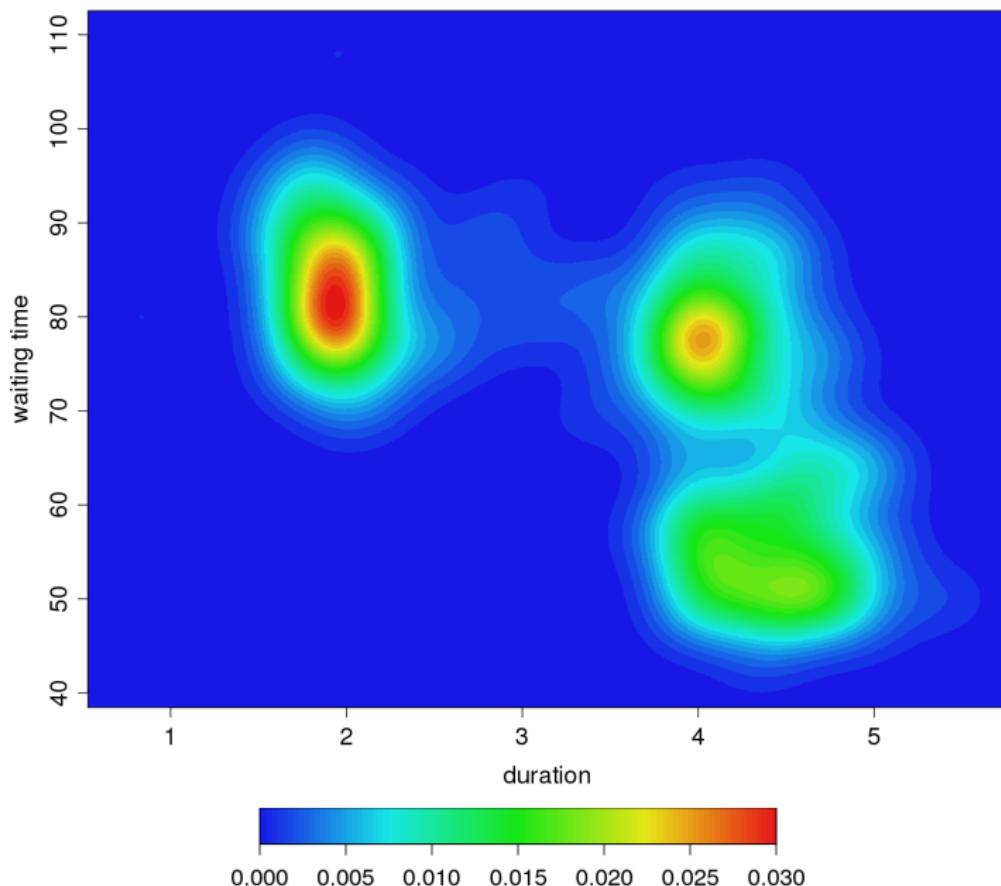
Illustrations

Examples:

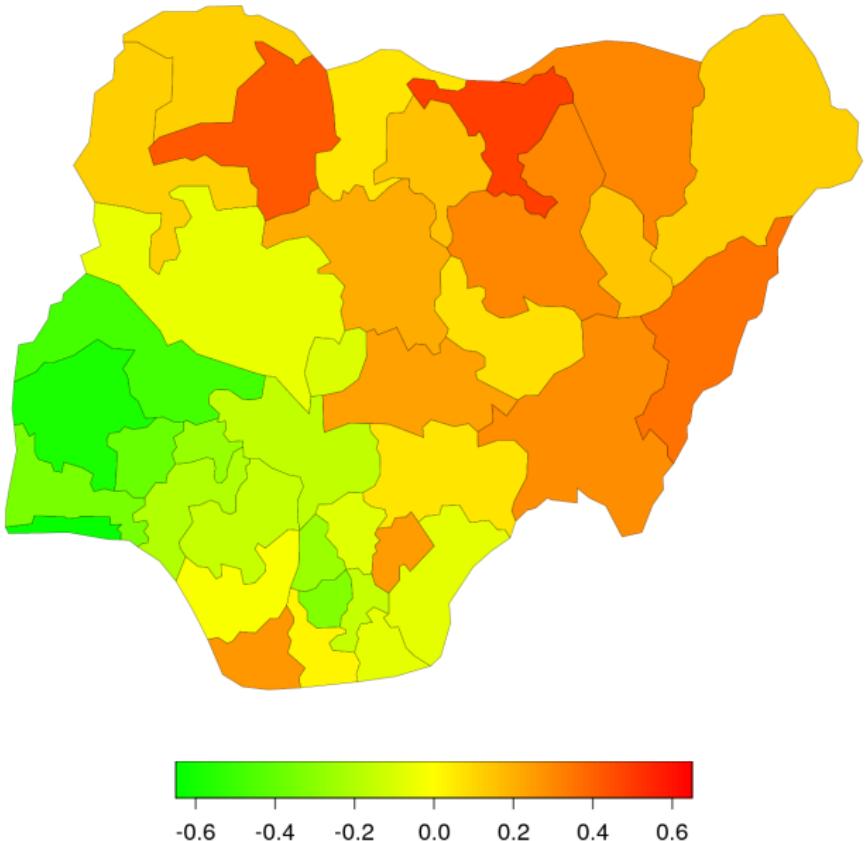
- Heatmap of bivariate kernel density estimate for Old Faithful geyser eruptions data.
- Map of Nigeria shaded by posterior mode estimates for childhood mortality.
- Pie chart of seats in the German parliament Bundestag, 2005.
- Mosaic display of votes for the German Bundestag, 2005.
- Model-based mosaic display for treatment of arthritis.
- Scatter plot with three clusters (and many points).

Colors: Palettes are constructed based on HSV space, especially by varying hue.

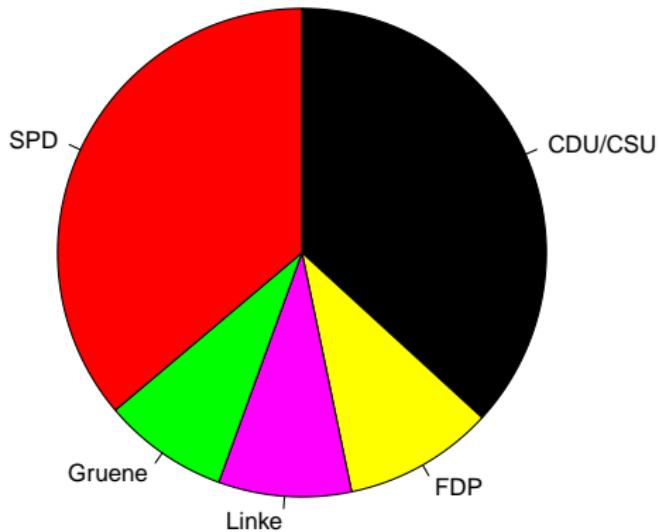
Illustrations: Old Faithful heatmap



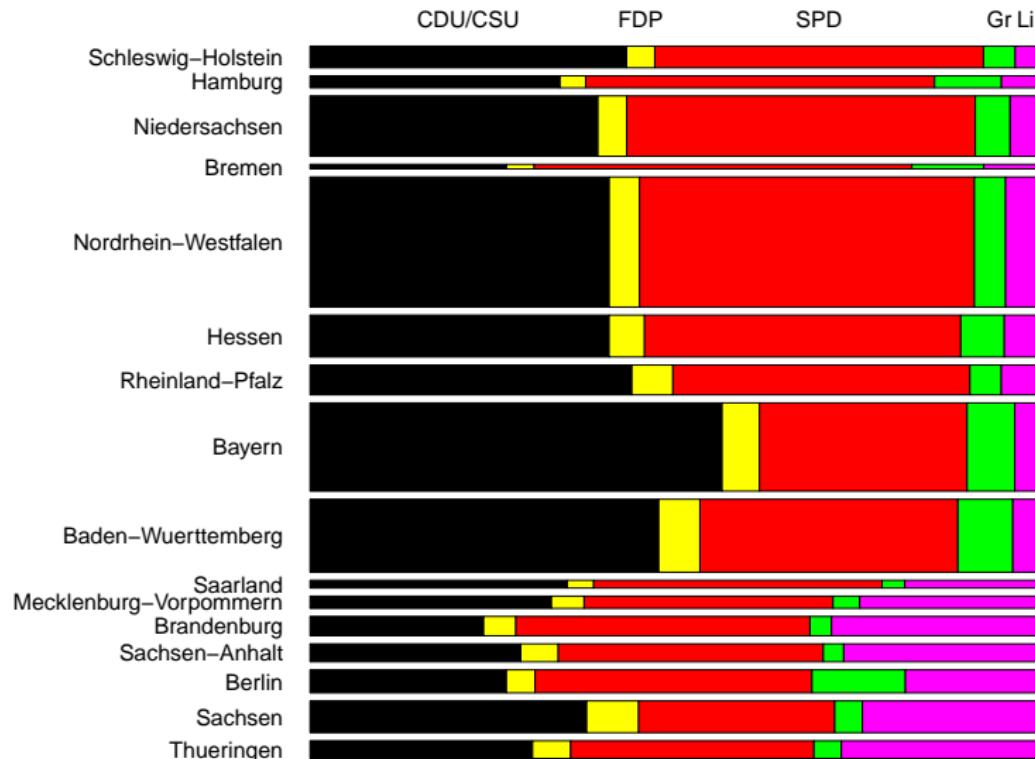
Illustrations: Nigeria childhood mortality map



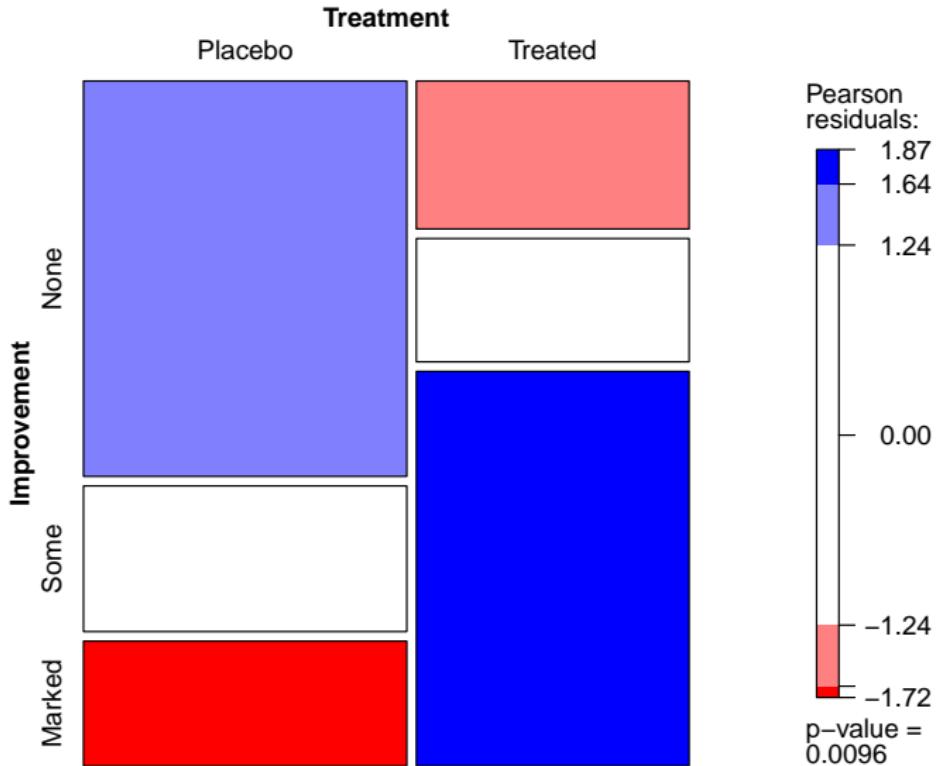
Illustrations: Bundestag seats pie chart



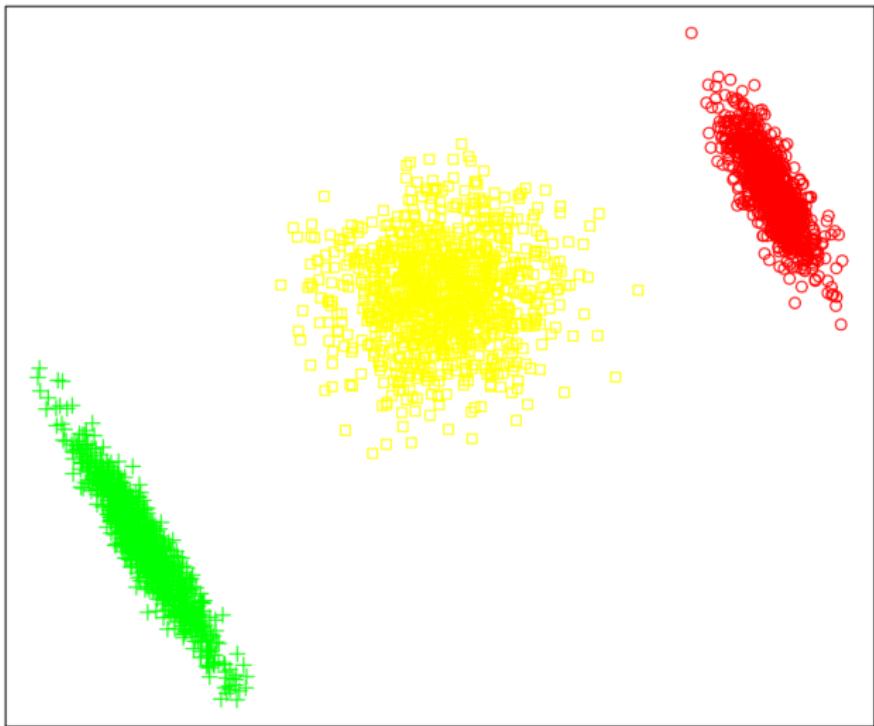
Illustrations: Bundestag votes mosaic



Illustrations: Arthritis treatment mosaic



Illustrations: Scatter plot with clusters



Illustrations

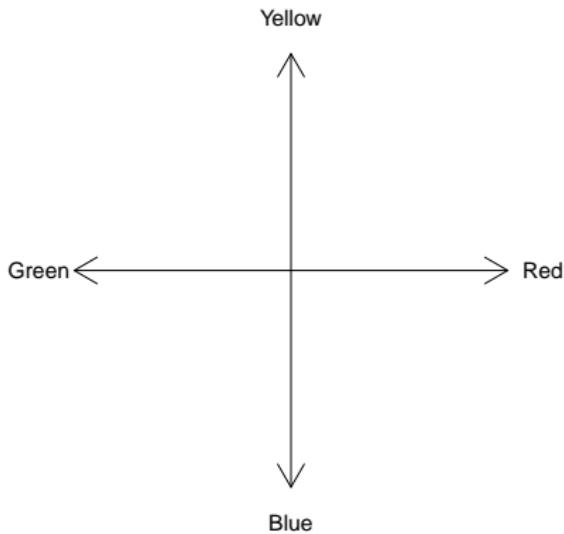
Problems:

- 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.

Color vision and color spaces

Human color vision is hypothesized to have evolved in three stages:

- ① **Light/dark** (monochrome only).
- ② **Yellow/blue** (associated with warm/cold colors).
- ③ **Green/red** (associated with ripeness of fruit).



Color vision and color spaces

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 space: HSV

HSV space: Standard transformation of RGB space implemented in most computer packages.

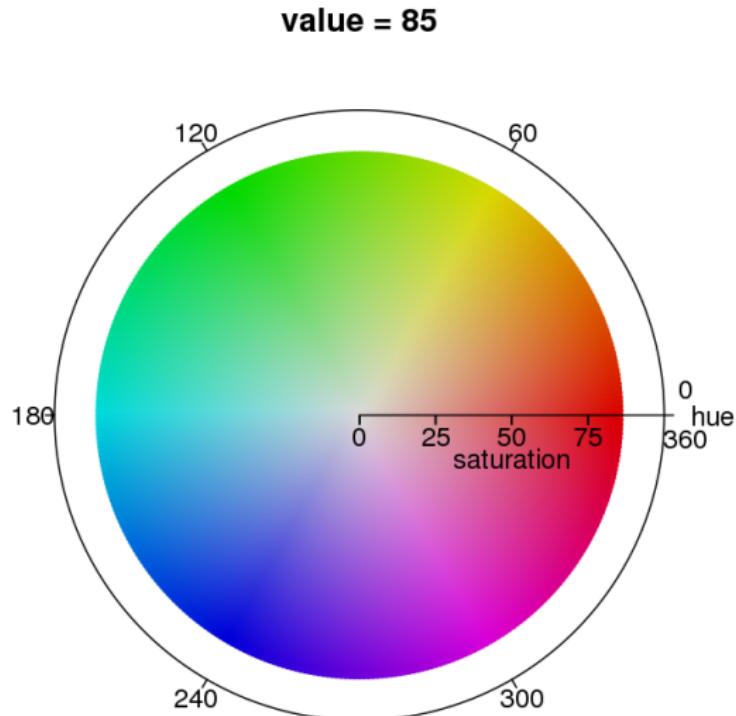
Specification: Triplet (H, S, V) with $H = 0, \dots, 360$ and $S, V = 0, \dots, 100$, often all transformed to unit interval (e.g., in R).

Shape: Cone (or transformed to cylinder).

Problem: Dimensions are confounded, hence not really perceptually based.

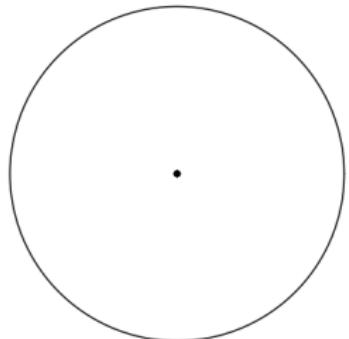
In R: `hsv()`.

Color space: HSV

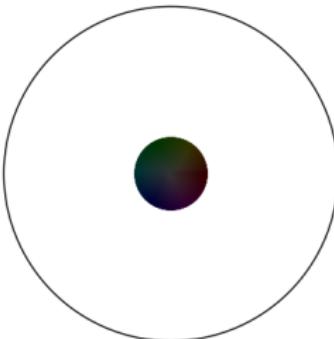


Color space: HSV

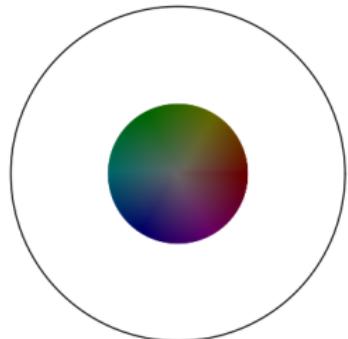
value = 0



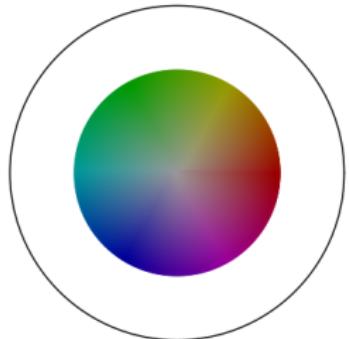
value = 20



value = 40



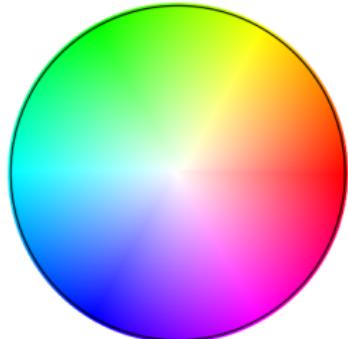
value = 60



value = 80

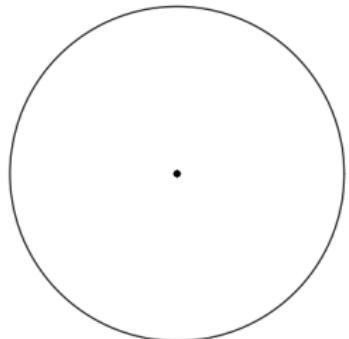


value = 100

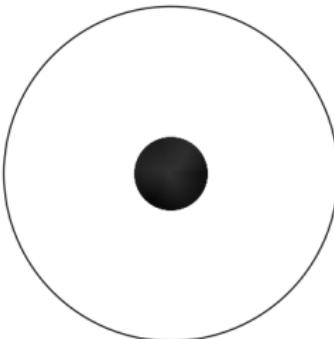


Color space: HSV

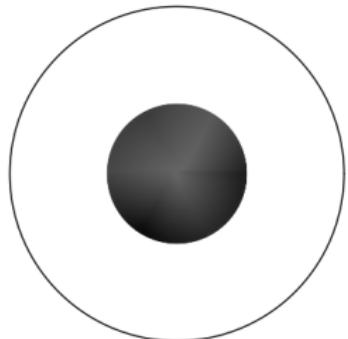
value = 0



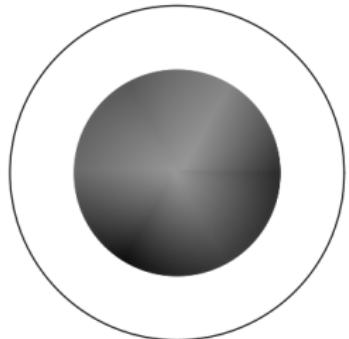
value = 20



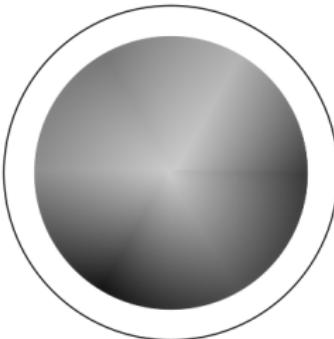
value = 40



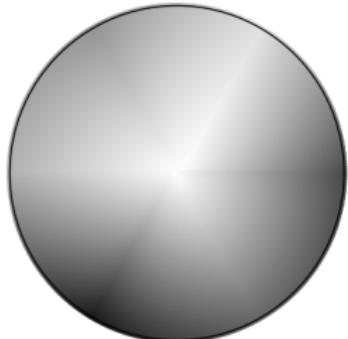
value = 60



value = 80



value = 100



Color space: HCL

HCL space: Perceptually based color space, polar coordinates in CIELUV space.

Specification: Triplet (H, C, L) with $H = 0, \dots, 360$ and $C, L = 0, \dots, 100$.

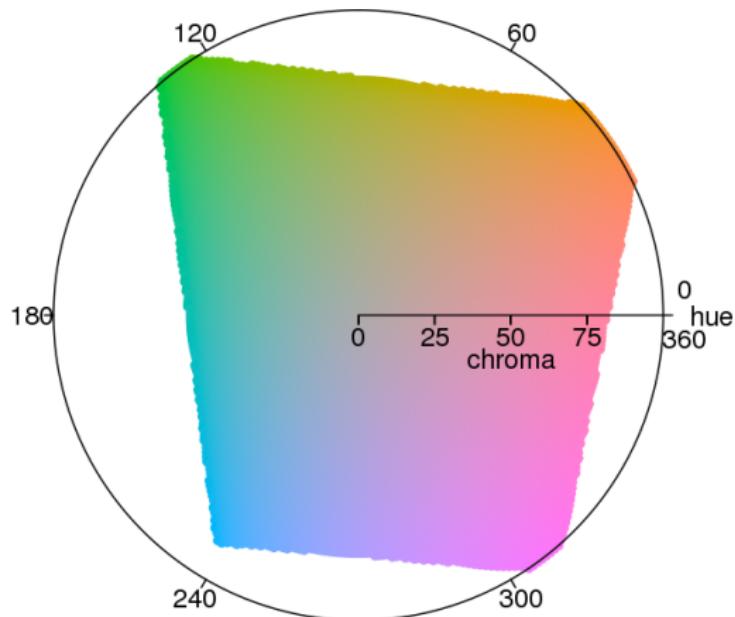
Shape: Distorted double cone.

Problem: Care is needed when traversing along the axes due to distorted shape.

In R: `hcl()`.

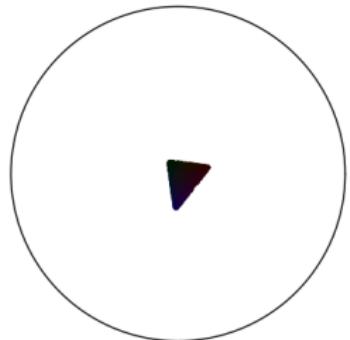
Color space: HCL

luminance = 70

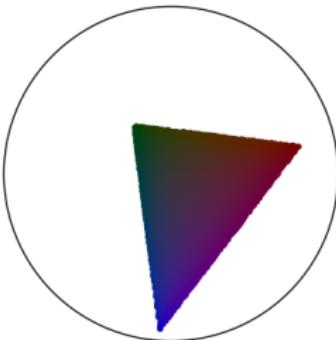


Color space: HCL

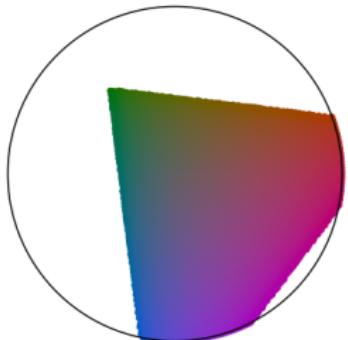
luminance = 5



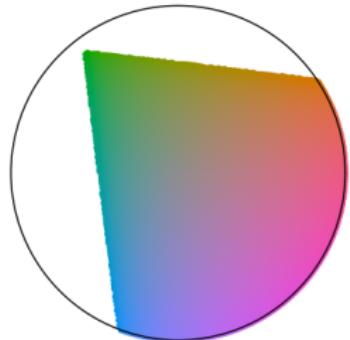
luminance = 23



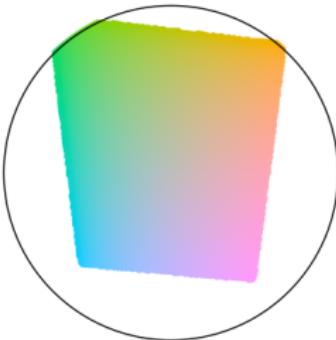
luminance = 41



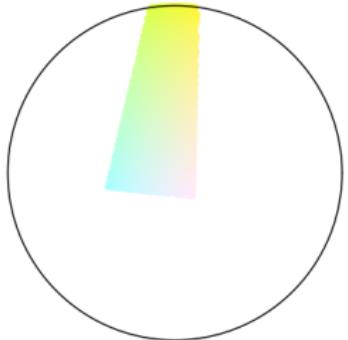
luminance = 59



luminance = 77

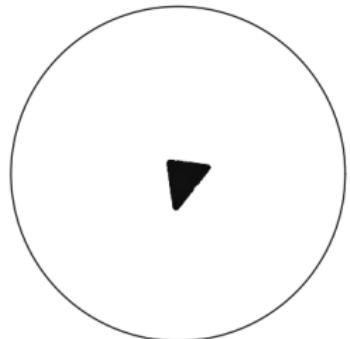


luminance = 95



Color space: HCL

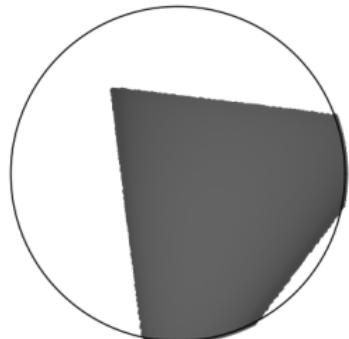
luminance = 5



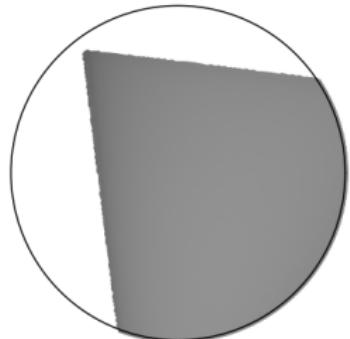
luminance = 23



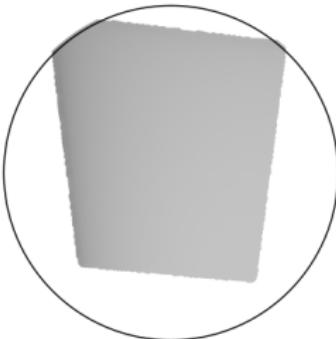
luminance = 41



luminance = 59



luminance = 77



luminance = 95



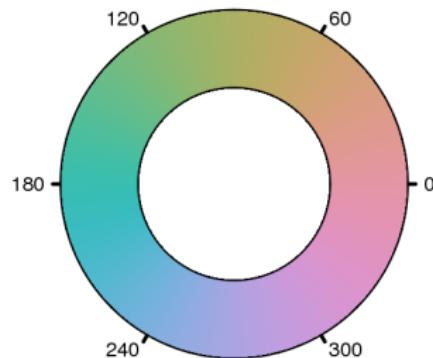
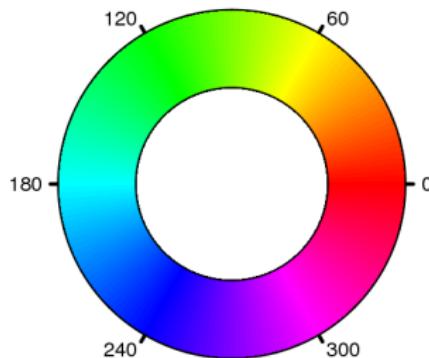
Color palettes: Qualitative

Goal: Code qualitative information.

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

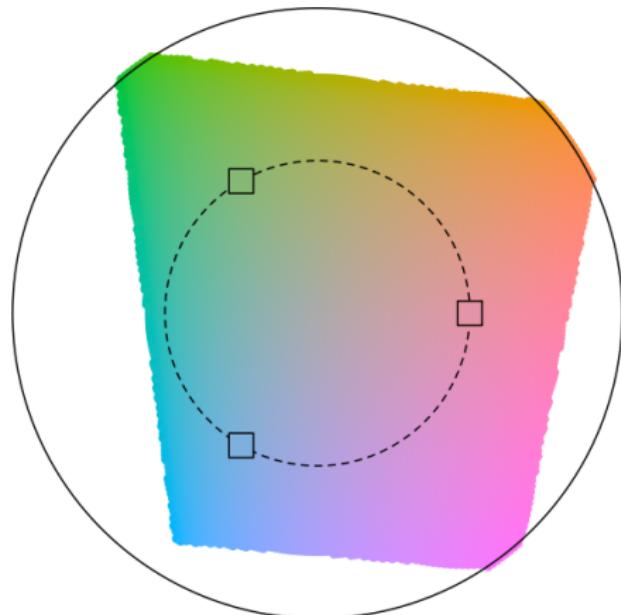
$$(H, 50, 70)$$

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



Color palettes: Qualitative

luminance = 70



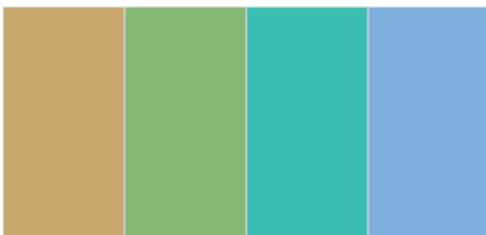
Color palettes: Qualitative

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

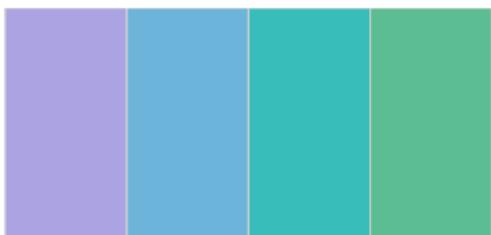
dynamic [30, 300]



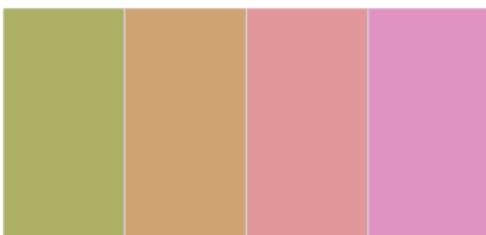
harmonic [60, 240]



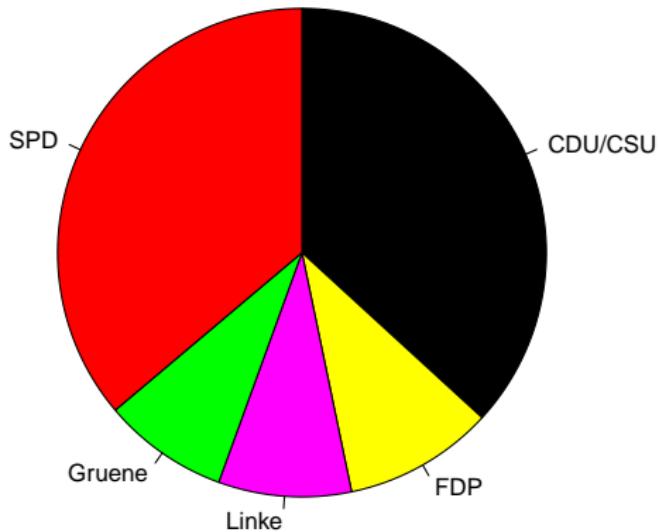
cold [270, 150]



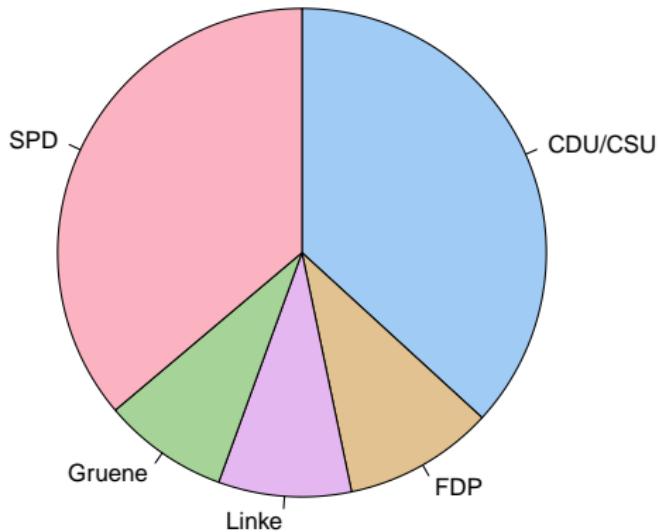
warm [90, -30]



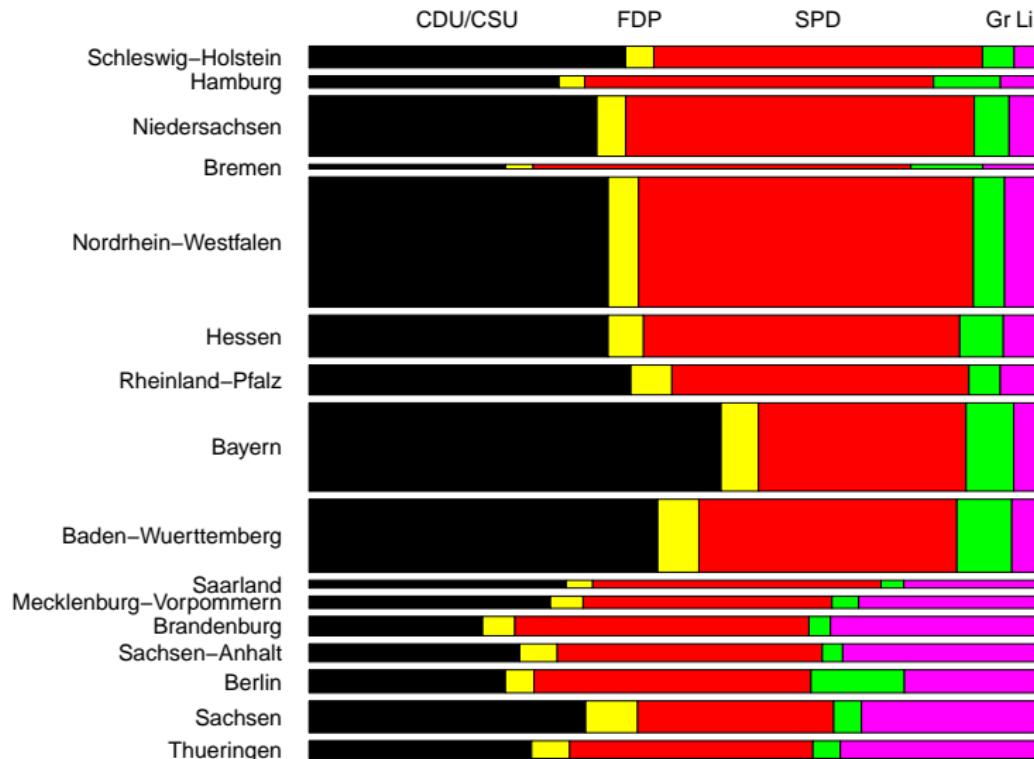
Color palettes: Qualitative



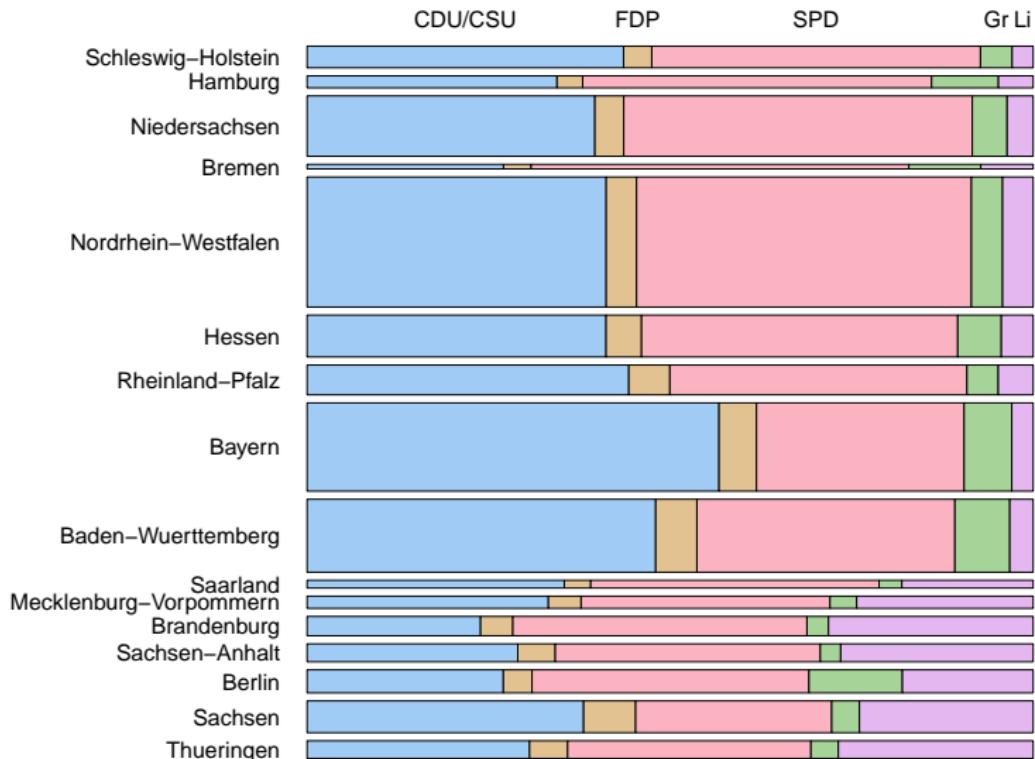
Color palettes: Qualitative



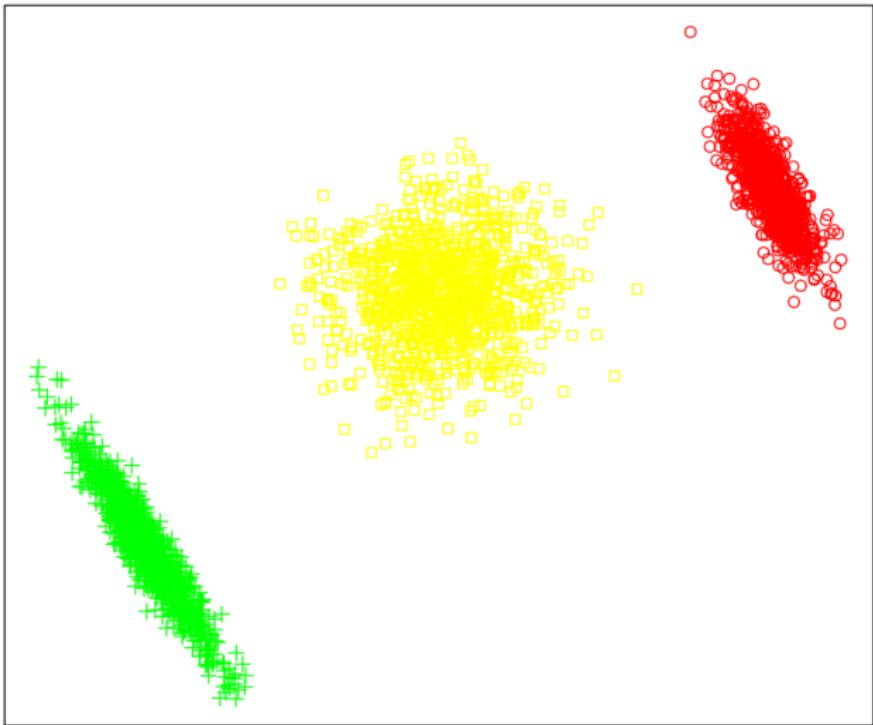
Color palettes: Qualitative



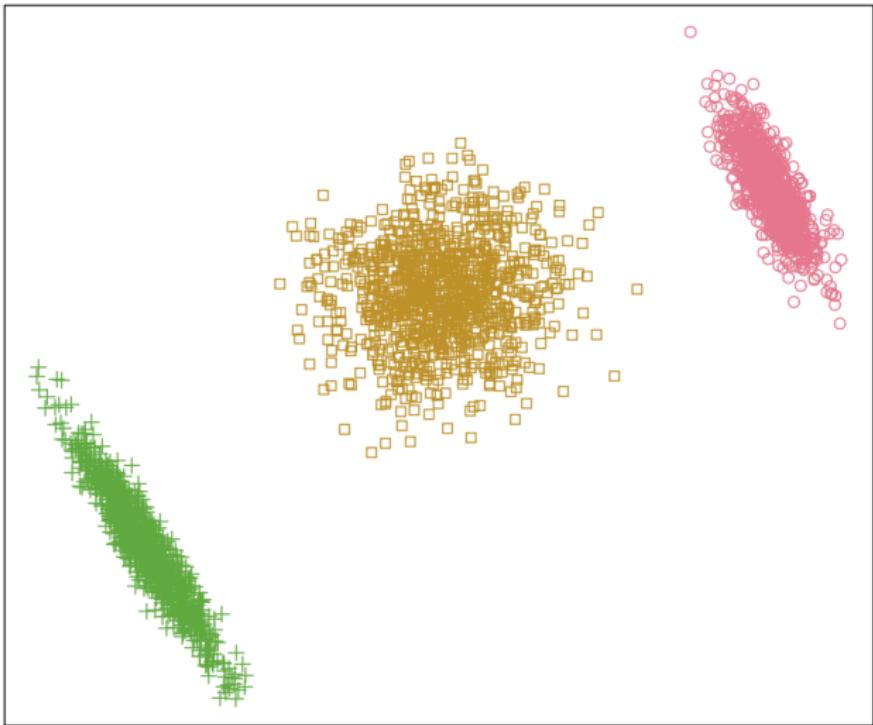
Color palettes: Qualitative



Color palettes: Qualitative



Color palettes: Qualitative



Color 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.,

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

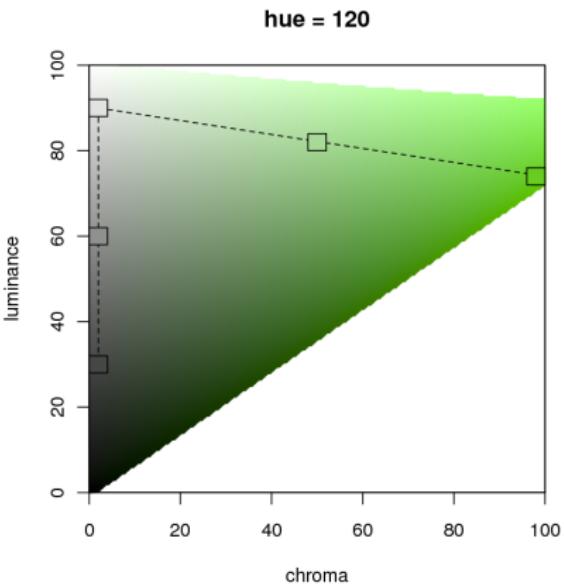
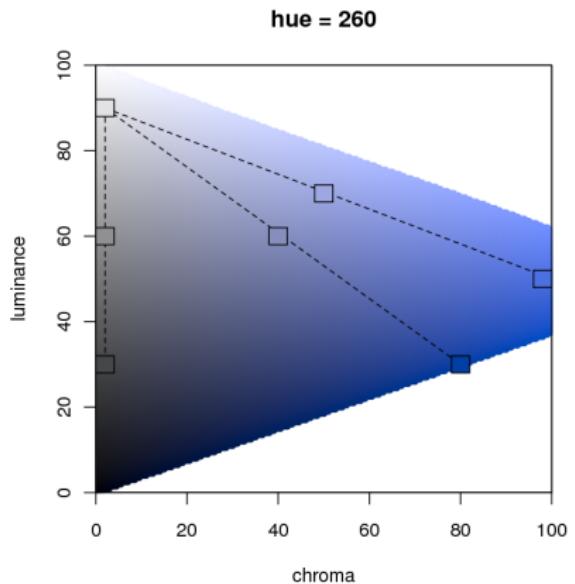
The hue H does not matter, chroma is set to 0 (no color), luminance ranges in $[30, 90]$, avoiding the extreme colors black and white.

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

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

for a fixed hue H .

Color palettes: Sequential



Color palettes: Sequential

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

$$(H_2 - i \cdot (H_1 - H_2), \quad C_{\max} - i^{p_1} \cdot (C_{\max} - C_{\min}), \\ L_{\max} - i^{p_2} \cdot (L_{\max} - L_{\min})).$$

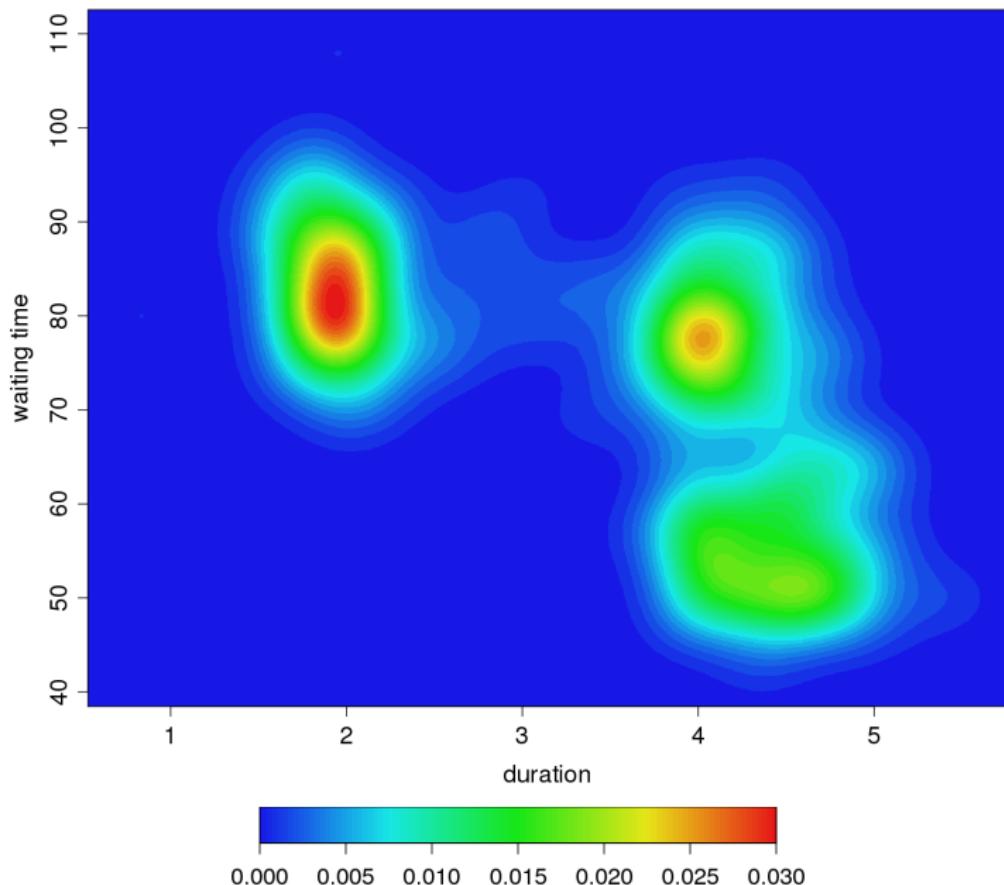
To make the change in hue visible, the chroma needs to increase rather quickly for low values of i and then only slowly for higher values of i .

A convenient transformation for achieving this is to use i^p instead of i with different powers for chroma and luminance.

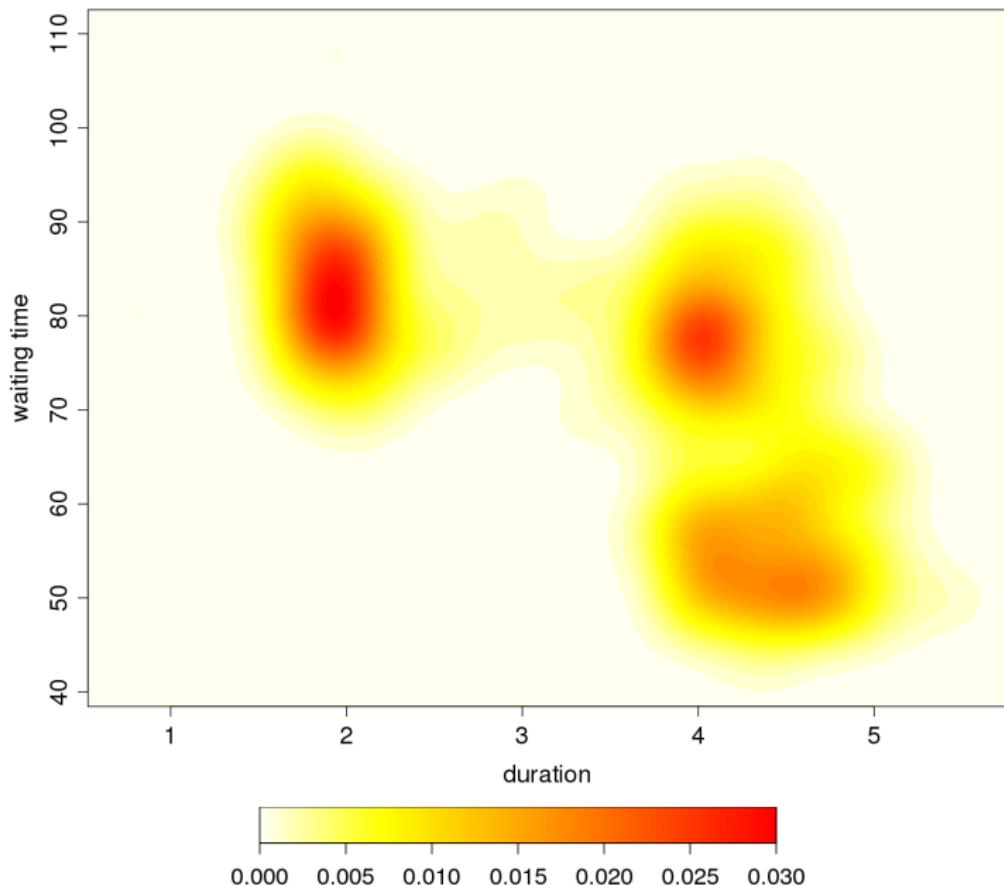
Color palettes: Sequential



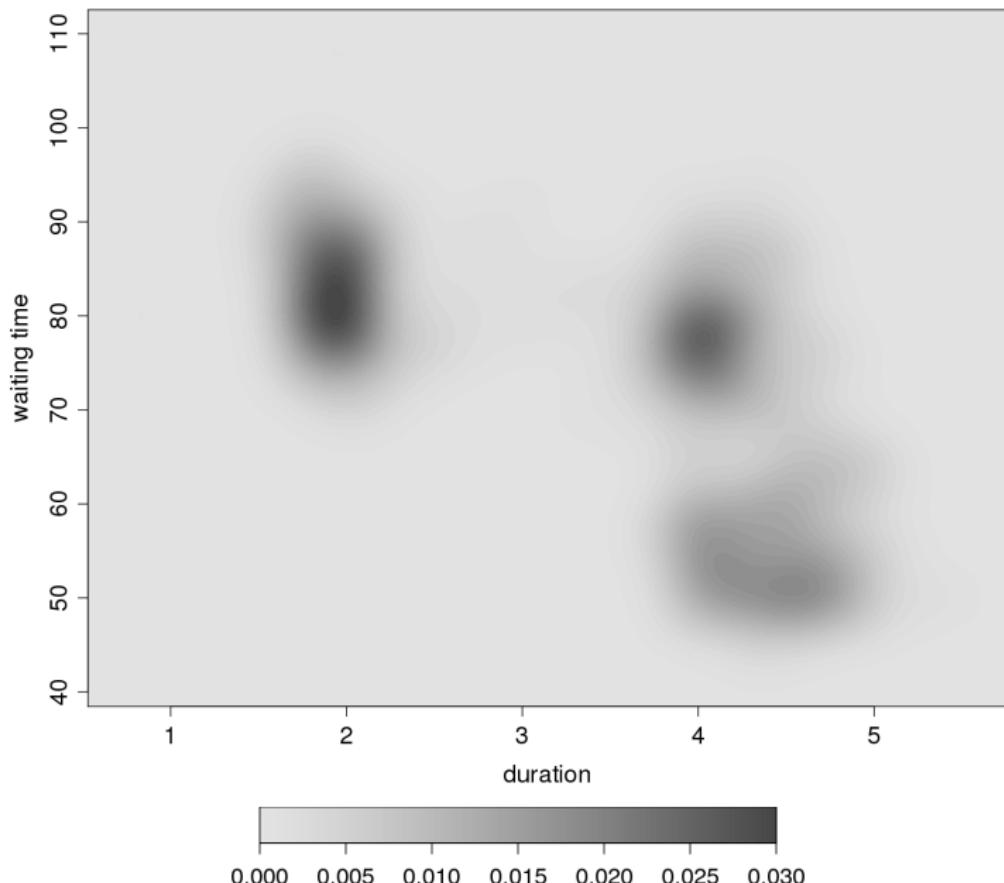
Color palettes: Sequential



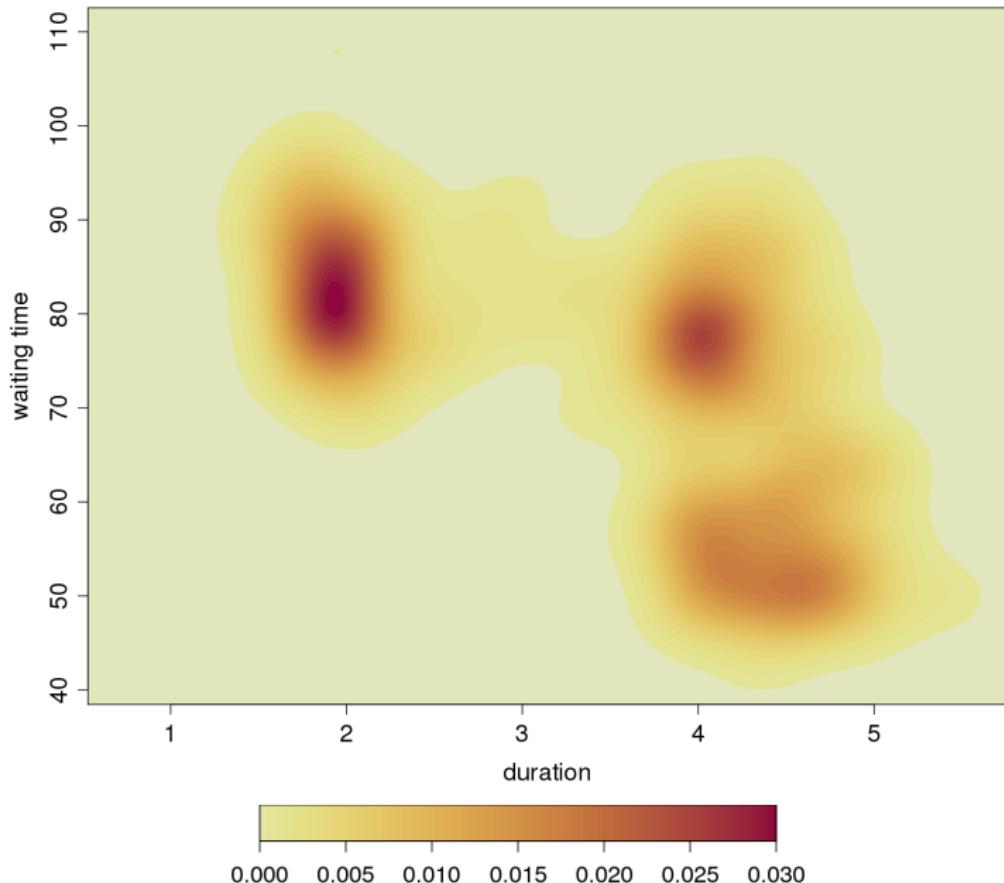
Color palettes: Sequential



Color palettes: Sequential



Color palettes: Sequential



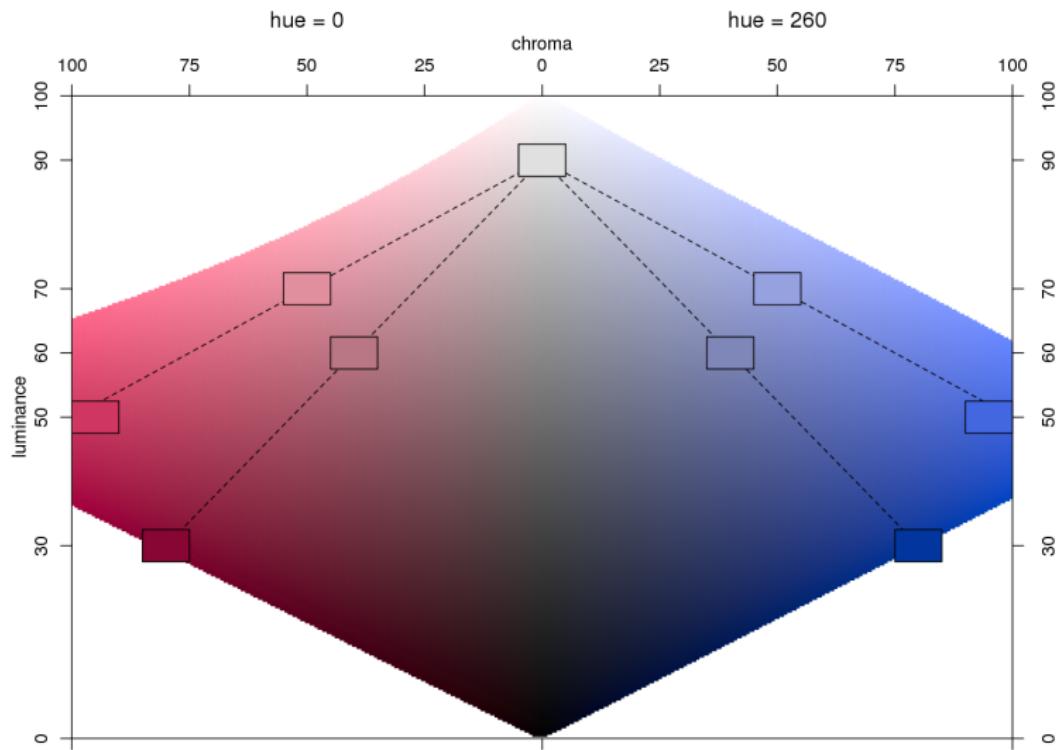
Color palettes: Diverging

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

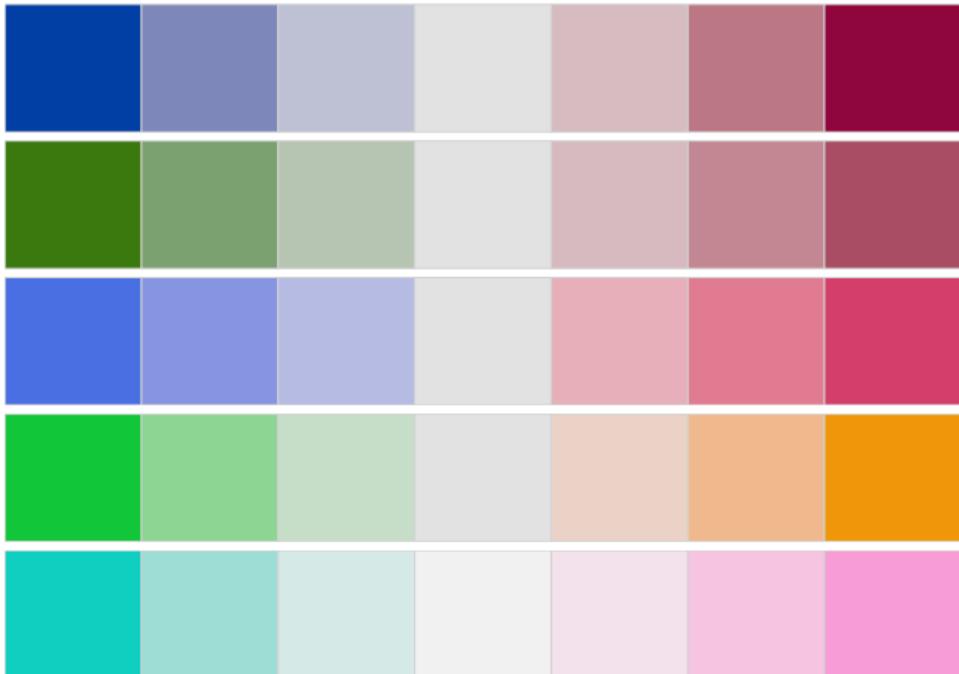
Solution: Combine sequential palettes with different hues.

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).

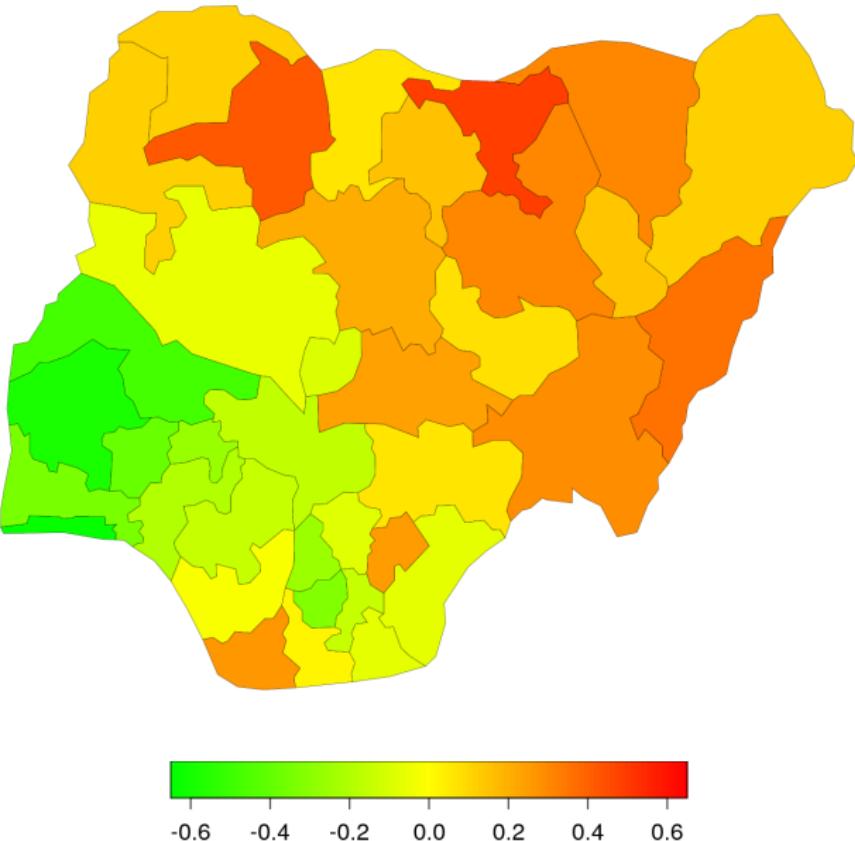
Color palettes: Diverging



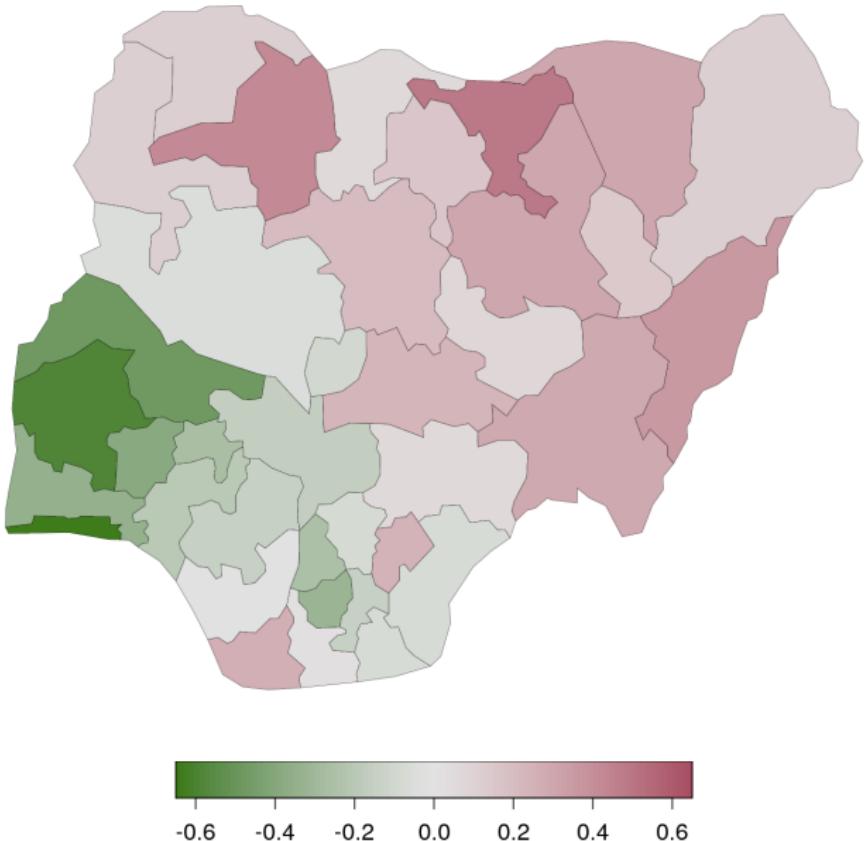
Color palettes: Diverging



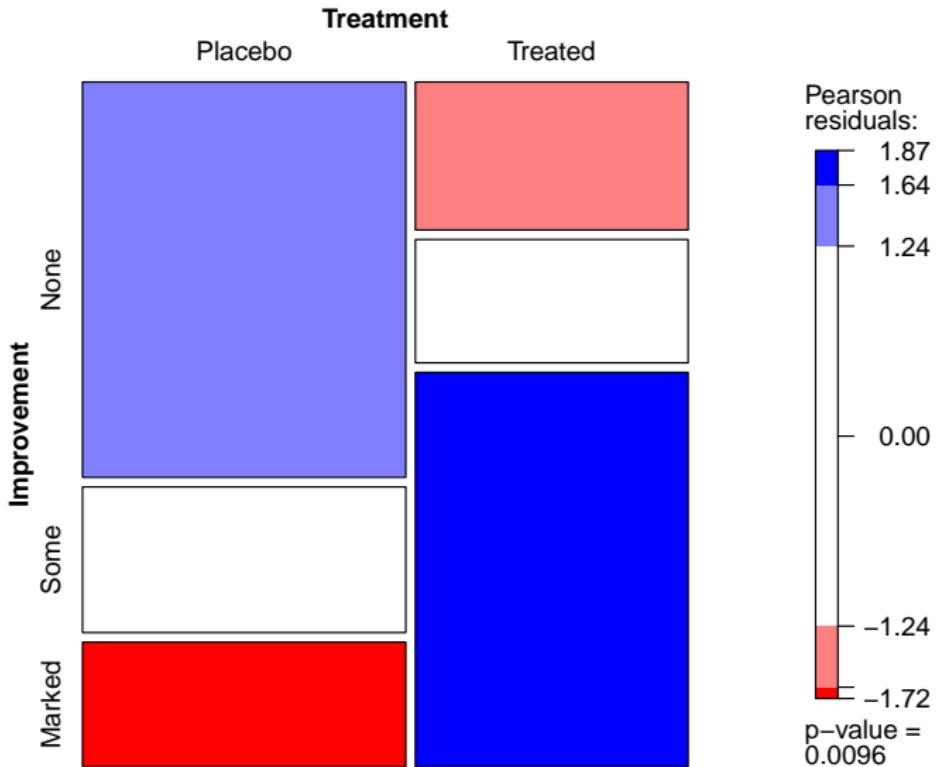
Color palettes: Diverging



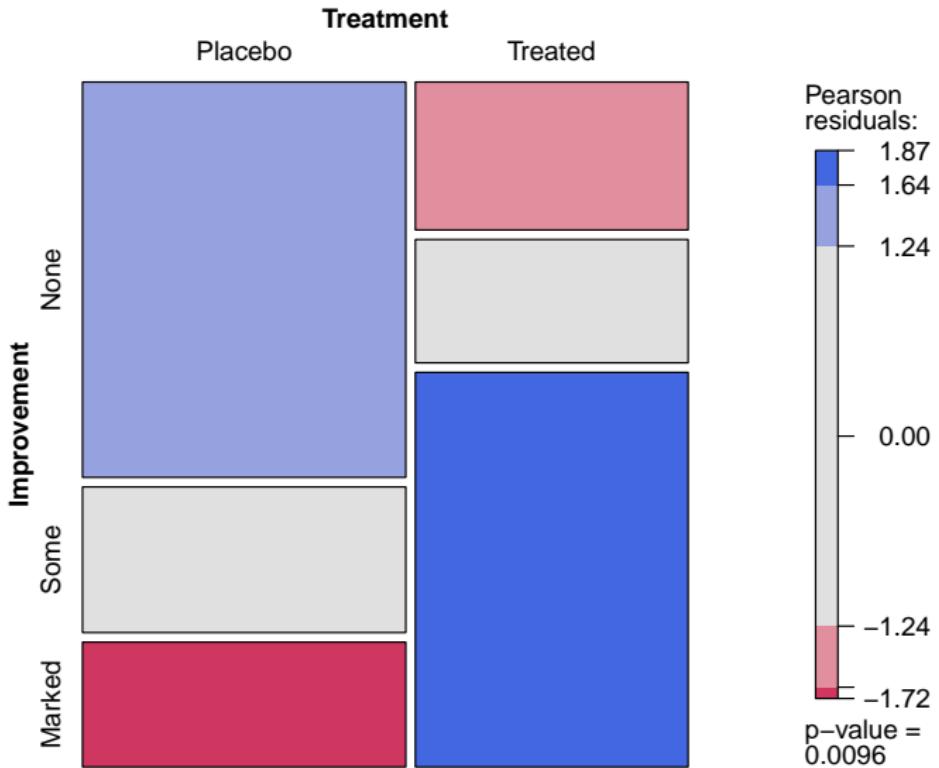
Color palettes: Diverging



Color palettes: Diverging



Color palettes: Diverging



Color blindness

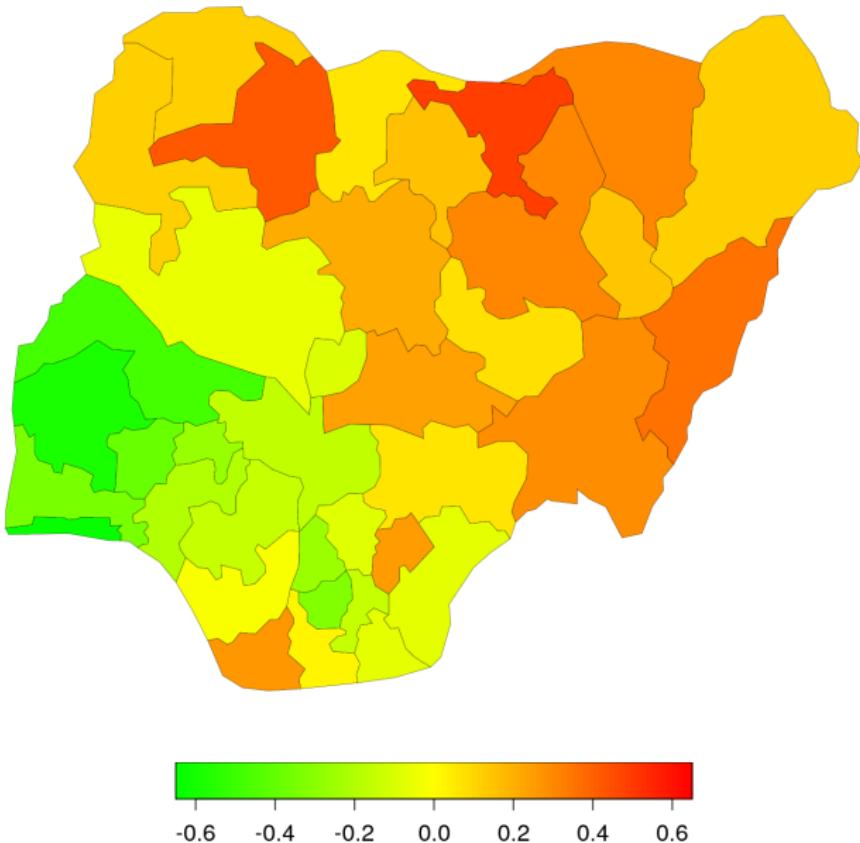
Problem: A few percent of humans (particularly males) have deficiencies in their color vision, typically referred to as *color blindness*.

Specifically: The most common forms of color blindness are different types of red-green color blindness: *deutanopia* (lack of green-sensitive pigment), *protanopia* (lack of red-sensitive pigment).

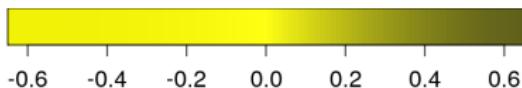
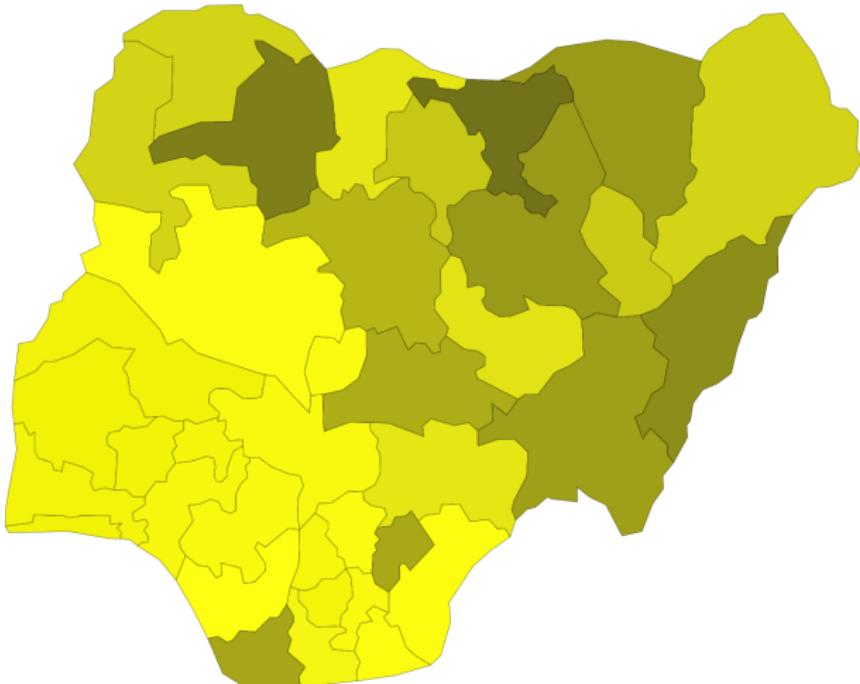
Solution: Construct suitable HCL colors.

- Use large luminance contrasts (visible even for monochromats).
- Use chroma contrasts on the yellow-blue axis (visible for dichromats).
- Check colors by emulating dichromatic vision, e.g., utilizing **dichromat** (Lumley 2006).

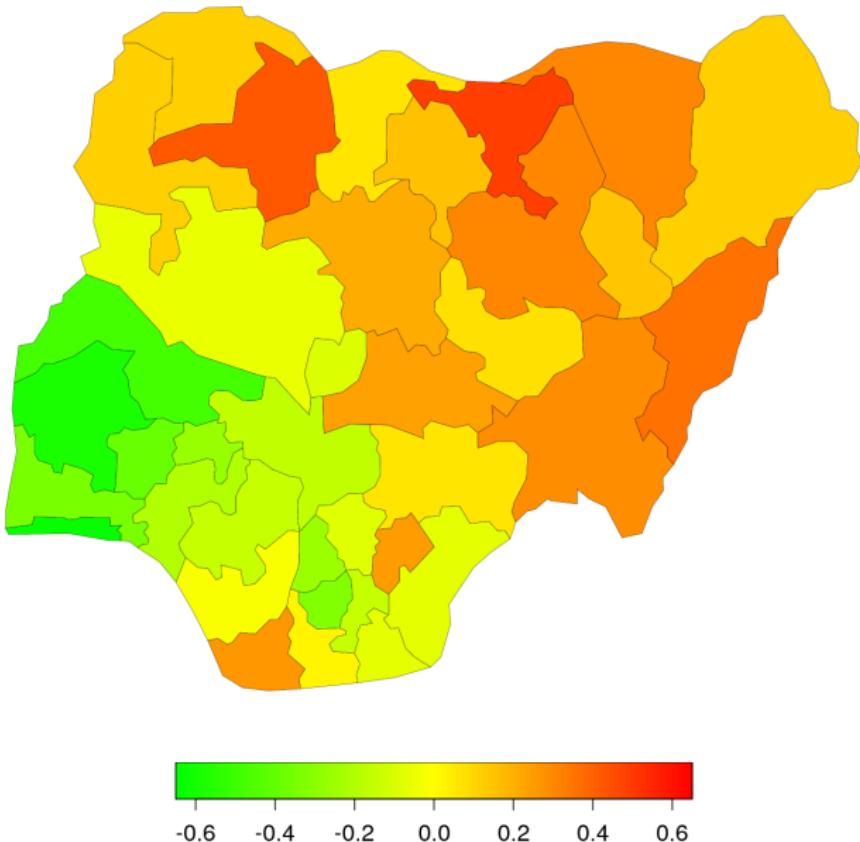
Color blindness



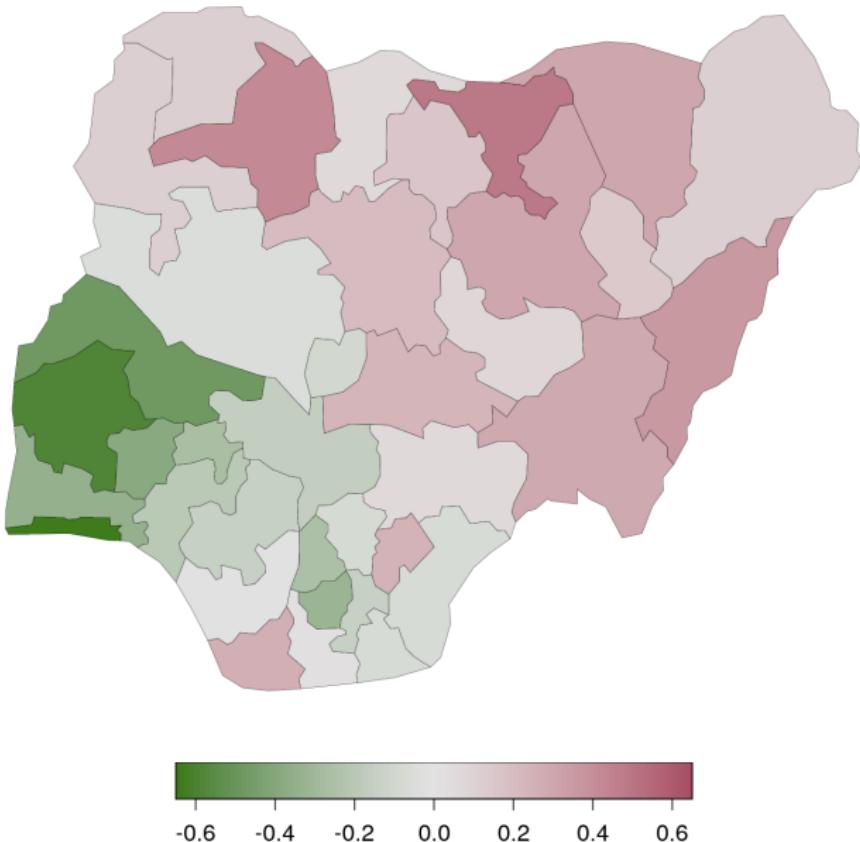
Color blindness



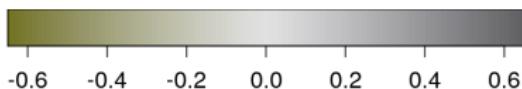
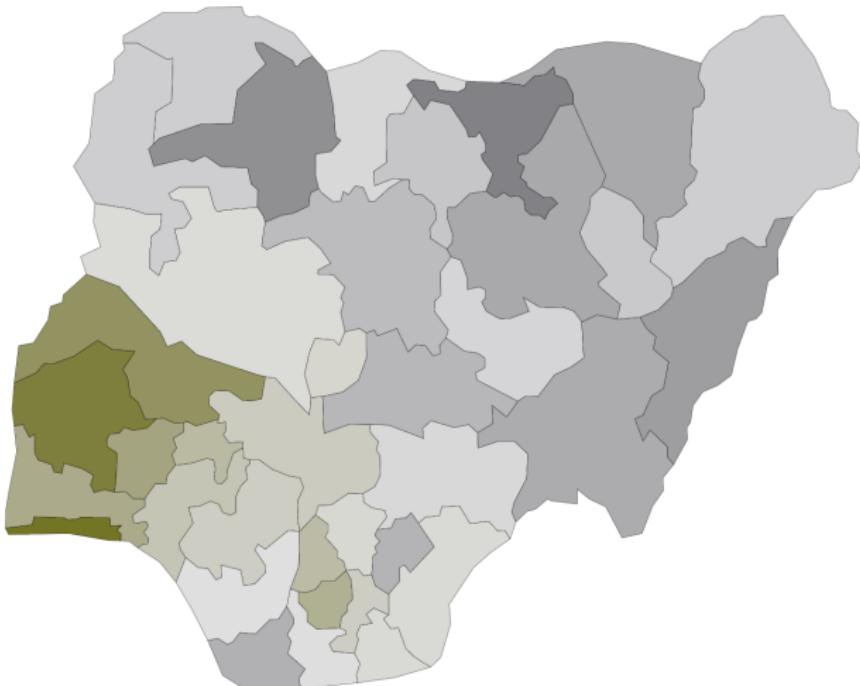
Color blindness



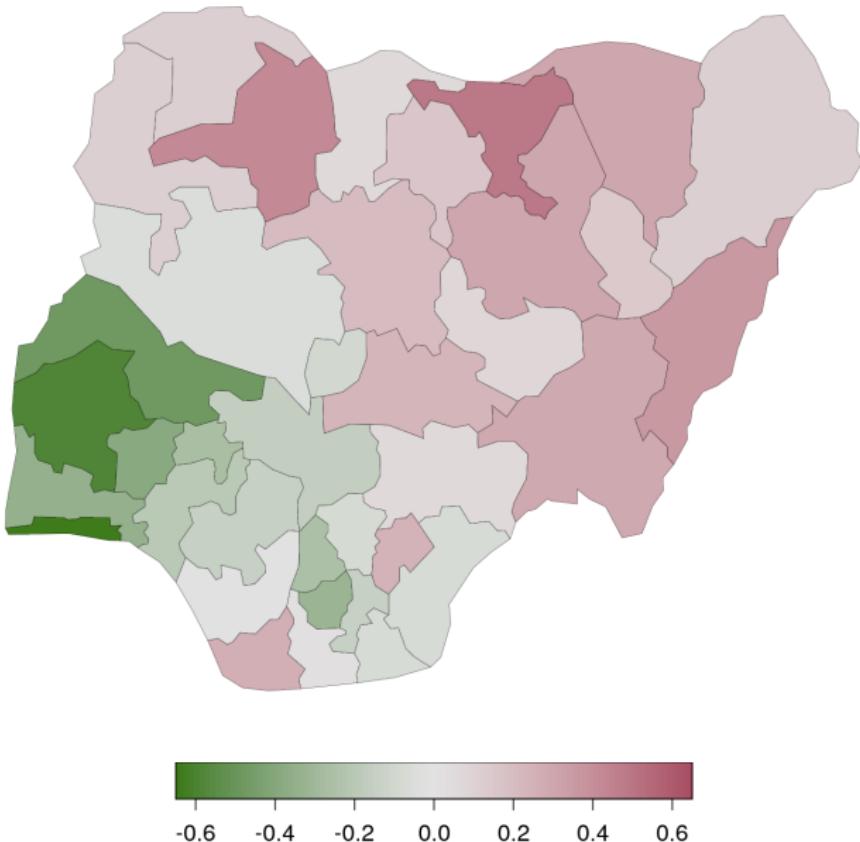
Color blindness



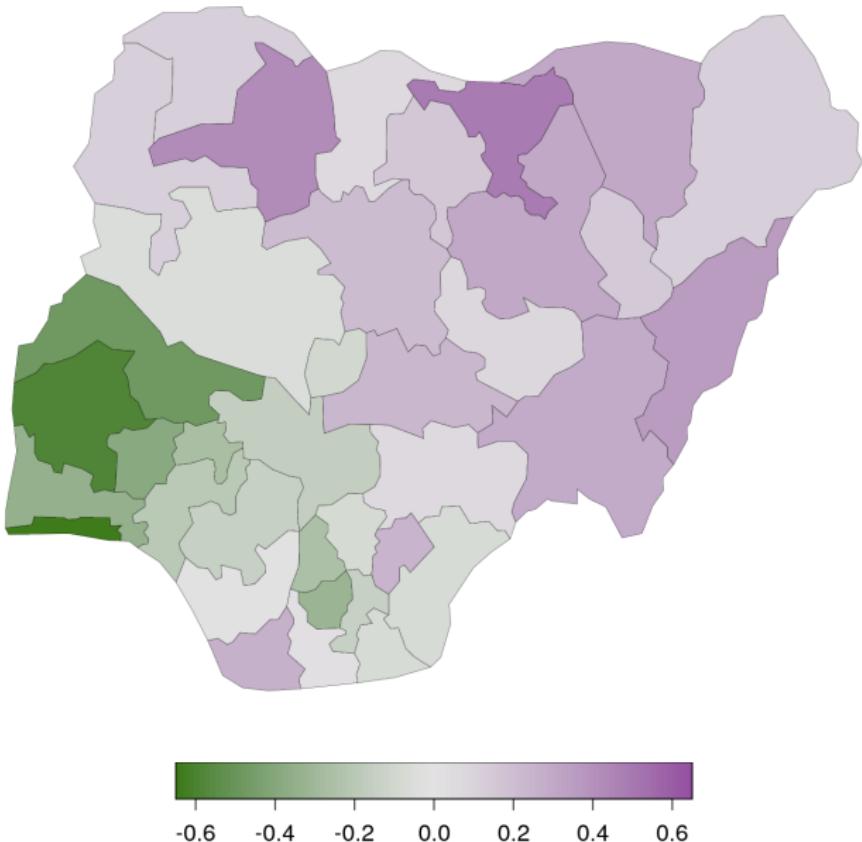
Color blindness



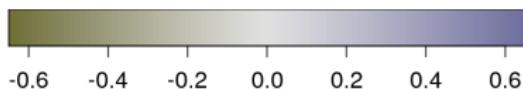
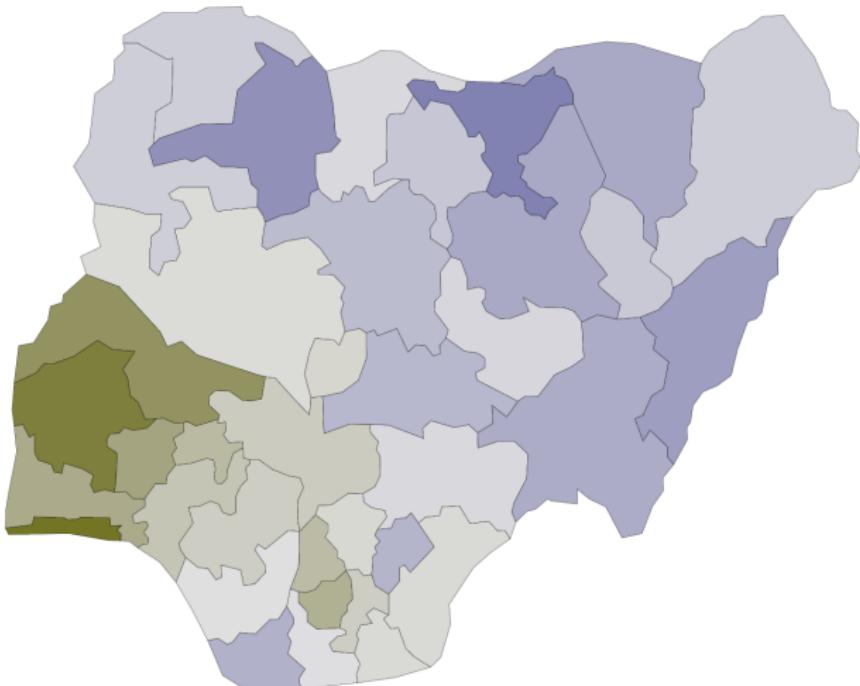
Color blindness



Color blindness



Color blindness



Software

- Given implementation of HCL colors, palette formulas are straightforward to implement.
- Otherwise, HCL coordinates typically need to be converted to RGB coordinates for display. Formulas are available, e.g., in Wikipedia or implemented in Ross Ihaka's C code underlying *colorspace*.
- The interactive online tool *ColorBrewer.org* provides many fixed palettes with similar properties. Very easy to use.

Software: R

Color spaces:

- Base system: `rgb()`, `hsv()`, `hcl()`, ...
- Package *colorspace*: `RGB()`, `HSV()`, `polarLUV()`, ...

Color palettes:

- Base system, HSV-based: `rainbow()`, `heat.colors()`, ...
- Package *colorspace*, HCL-based: `rainbow_hcl()`, `heat_hcl()`, `sequential_hcl()`, `diverge_hcl()`, ...

Examples: `?rainbow_hcl` and

```
vignette("hcl-colors", package = "colorspace").
```

Further useful packages: *RColorBrewer* (fixed palettes from *ColorBrewer.org*), *ggplot2*, *plotrix*, ...

Summary

- Use color with care, do not overestimate power of color.
- Avoid large areas of flashy, highly saturated colors.
- Employ monotonic luminance scale for numerical data.
- HCL space allows for intuitive variation of perceptual properties.
- Formulas for palettes are easy to implement in new software.
- Convenience functions (similar to base R tools) are readily provided in *colorspace*.

References

- Zeileis A, Hornik K, Murrell P (2009). "Escaping RGBland: Selecting Colors for Statistical Graphics." *Computational Statistics & Data Analysis*, **53**, 3259–3270. doi:10.1016/j.csda.2008.11.033.
- Zeileis A, Meyer D, Hornik K (2007). "Residual-Based Shadings for Visualizing (Conditional) Independence." *Journal of Computational and Graphical Statistics*, **16**(3), 507–525. doi:10.1198/106186007X237856.
- Lumley T (2006). "Color Coding and Color Blindness in Statistical Graphics." *ASA Statistical Computing & Graphics Newsletter*, **17**(2), 4–7. URL <http://www.amstat-online.org/sections/graphics/newsletter/Volumes/v172.pdf>.
- Ihaka R (2003). "Colour for Presentation Graphics." In K Hornik, F Leisch, A Zeileis (eds.), "Proceedings of the 3rd International Workshop on Distributed Statistical Computing," Vienna, Austria, ISSN 1609-395X, URL <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/>.