WIRTSCHAFTS UNIVERSITÄT WIEN

Escaping RGBland: Selecting Colors for Statistical Graphics

Achim Zeileis, Kurt Hornik, Paul Murrell

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

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

- Basic principles: Colors should be intuitive, avoid large areas of saturated colors.
- Purpose: Distinguish different elements of a statistical graphic depending on the levels of some variable.
- 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.
- Ideally, colors should work for: Screen, projector, (grayscale) printer, color-blind viewers, . . .

Tools (in R)

Basic color spaces: `rgb()`, `hsv()`, `hcl()`, ...

HSV-based palettes: `rainbow()`, `heat.colors()`, ...

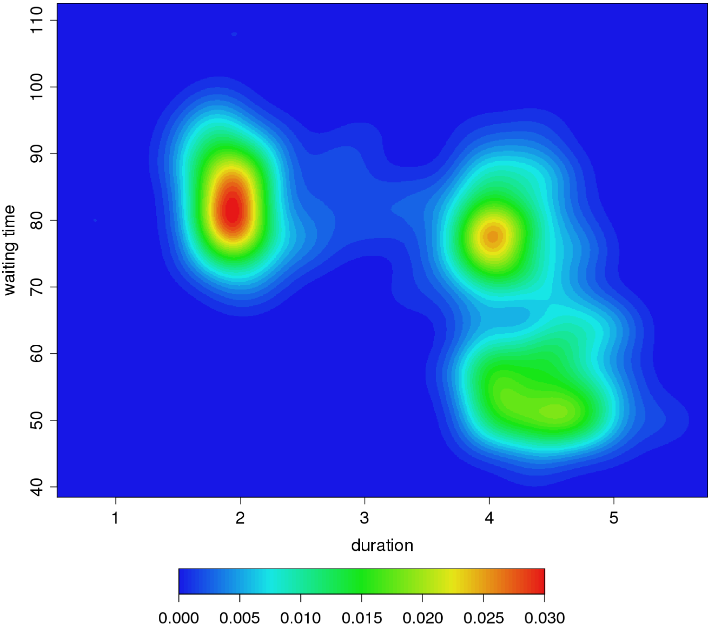
More suitable tools: *RColorBrewer* (fixed palettes from *ColorBrewer.org*), *ggplot2*, *plotrix*, `colorRamp()` (based on RGB and CIELAB), ...

Here: *colorspace* with `RGB()`, `polarLUV()`, ..., and `rainbow_hcl()`, `heat_hcl()`, `sequential_hcl()`, `diverge_hcl()`, ...

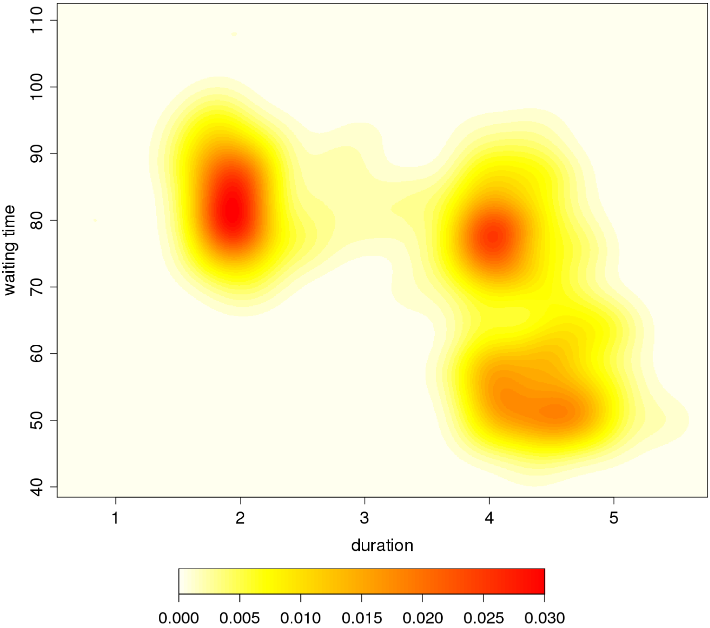
Result: Similar to *ColorBrewer.org* but with more flexibility and more insight into underlying ideas.

Example: Heatmap of bivariate kernel density estimate for Old Faithful geyser eruptions data

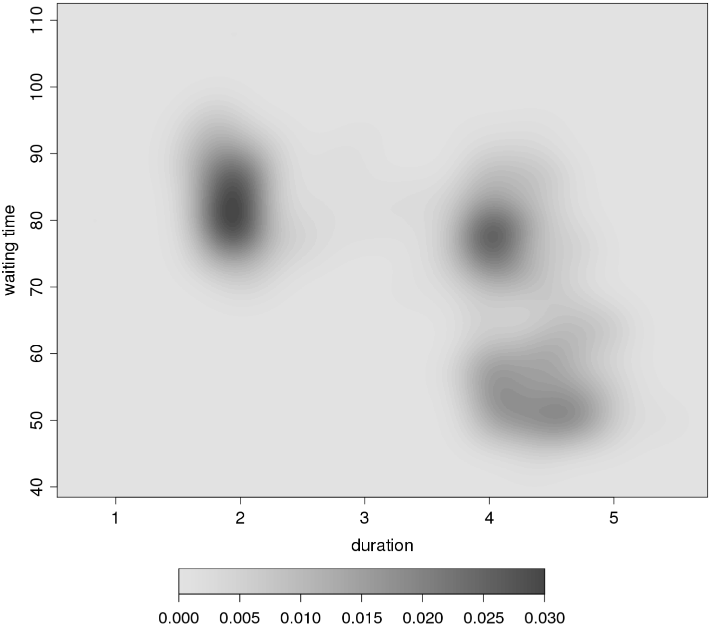
Example: Heatmap



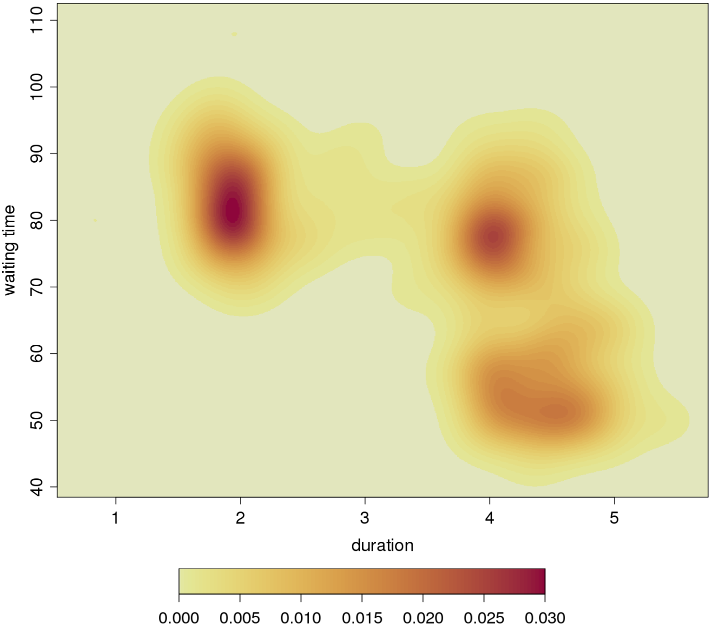
Example: Heatmap



Example: Heatmap



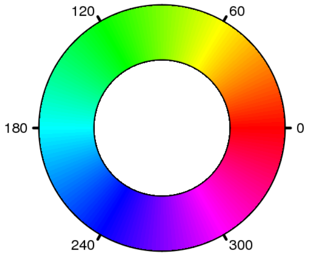
Example: Heatmap

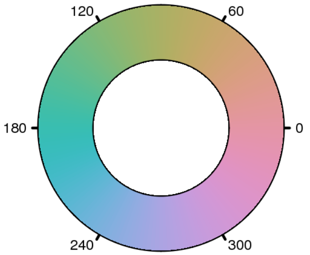


Types of palettes

Qualitative:

- Code categorical information.
- Examples: Barplot, mosaic display, ...
- Use different hues, keeping chroma and luminance fixed: e.g., $(H, 50, 70)$.





Types of palettes

Sequential:

- Code numerical information ranging from “uninteresting” to “interesting”.
- Increase luminance along with interestingness.
- Additionally increase chroma. Potentially vary hue.
- When interestingness i is standardized to $[0,1]$:
  $(H_2 - i \cdot (H_1 - H_2), C_{\max} - i' \cdot (C_{\max} - C_{\min}), L_{\max} - i'' \cdot (L_{\max} - L_{\min}))$.

Diverging:

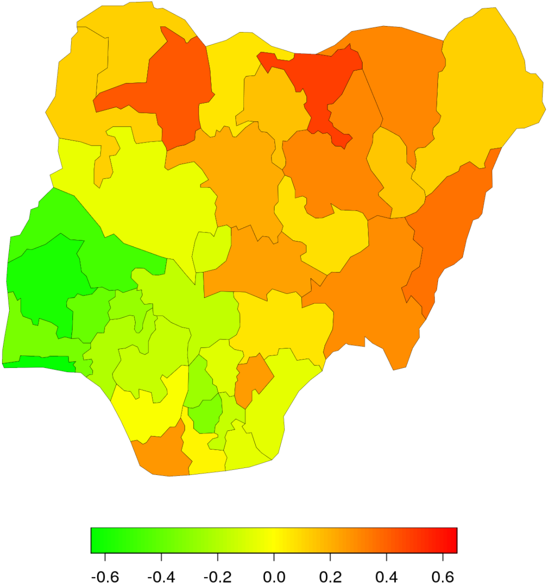
- Code numerical information diverging from neutral value into two directions of “interestingness”.
- Combine two sequential palettes with different hues.

Example: Model deviations

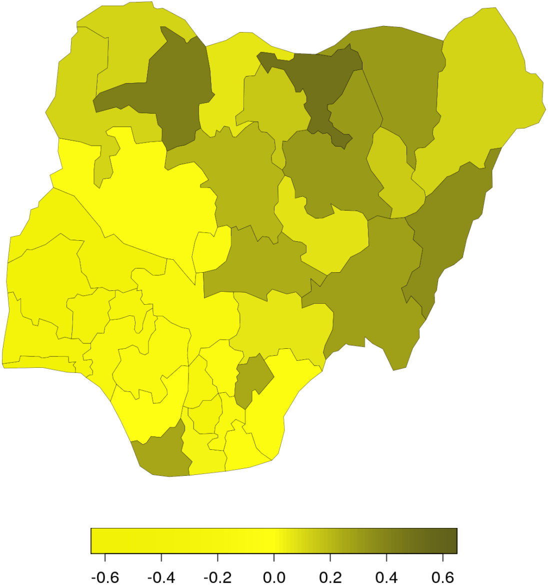
Application: Childhood mortality in Nigeria.

- Posterior mode estimates (without spatial effect).
- Map of Nigeria shaded by model deviations.
- Investigate typical HSV-based vs. HCL-based palette.
- Investigate effects of color-blindness (protanopic vision) by means of *dichromat* package.

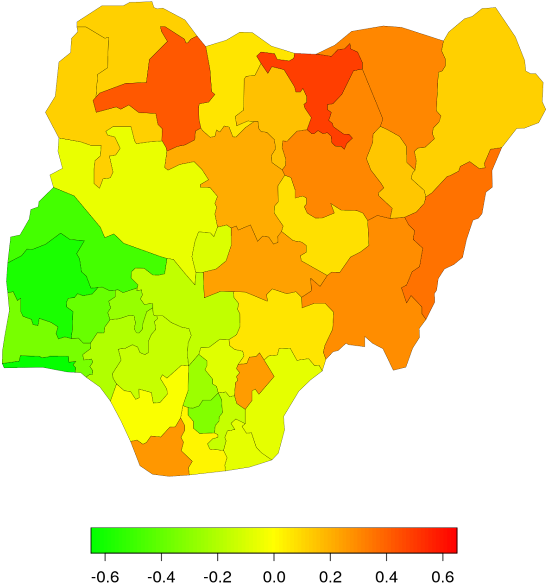
Example: Model deviations



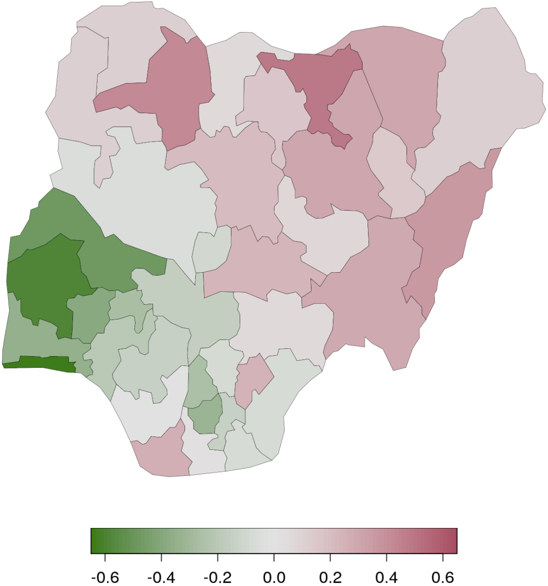
Example: Model deviations



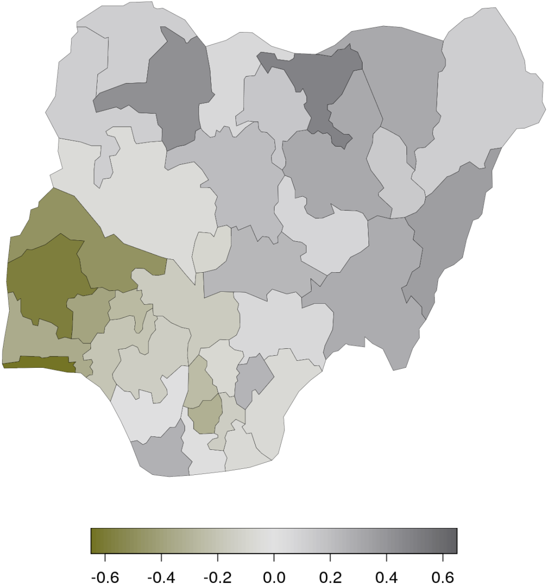
Example: Model deviations



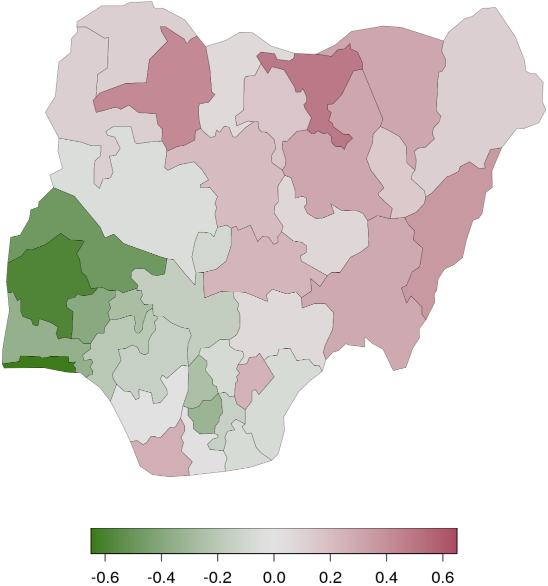
Example: Model deviations



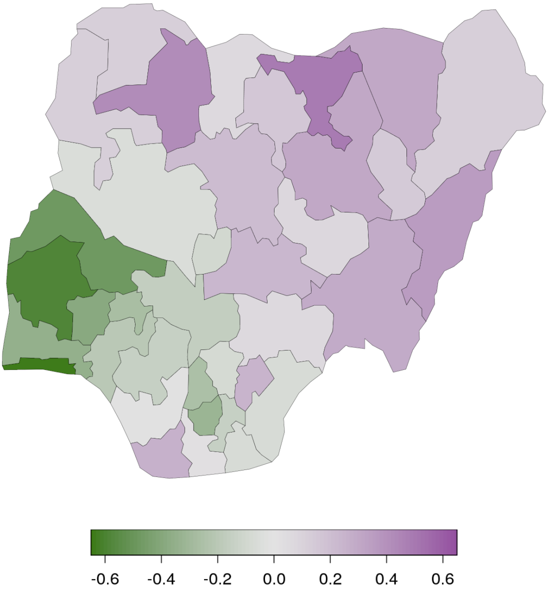
Example: Model deviations



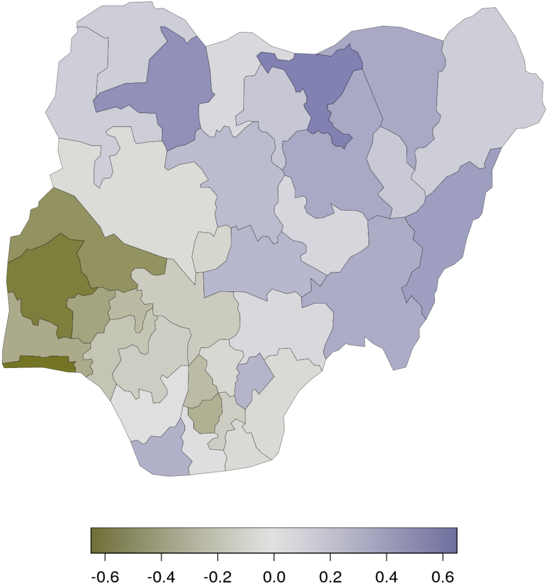
Example: Model deviations



Example: Model deviations



Example: Model deviations



Summary

- Use color with care, don’t 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 implemented 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/`.