Universität Innsbruck

Somewhere over the Rainbow

How to Make Effective Use of Colors in Statistical Graphics

Achim Zeileis

https://eeecon.uibk.ac.at/~zeileis/

Introduction

- Zeileis A, Hornik K, Murrell P (2009).
  "Escaping RGBland: Selecting Colors for Statistical Graphics."
  Computational Statistics & Data Analysis, **53**(9), 3259–3270.
  `doi:10.1016/j.csda.2008.11.033`
- Stauffer R, Mayr GJ, Dabernig M, Zeileis A (2015).
  "Somewhere over the Rainbow: How to Make Effective Use of Colors in Meteorological Visualizations."
  Bulletin of the American Meteorological Society, **96**(2), 203–216.
  `doi:10.1175/BAMS-D-13-00155.1`
- Ihaka R, Murrell P, Hornik K, Fisher J, Stauffer R, Zeileis A (2016).
  colorspace: Color Space Manipulation. R package version 1.3-0.
  `https://CRAN.R-project.org/package=colorspace`

- Overview:
  `https://hclwizard.org/`

Introduction

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.

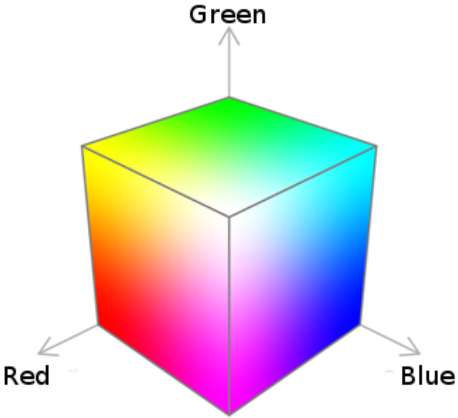
Question: What are useful color palettes for coding qualitative and quantitative variables?

Introduction

Main goal of our work:

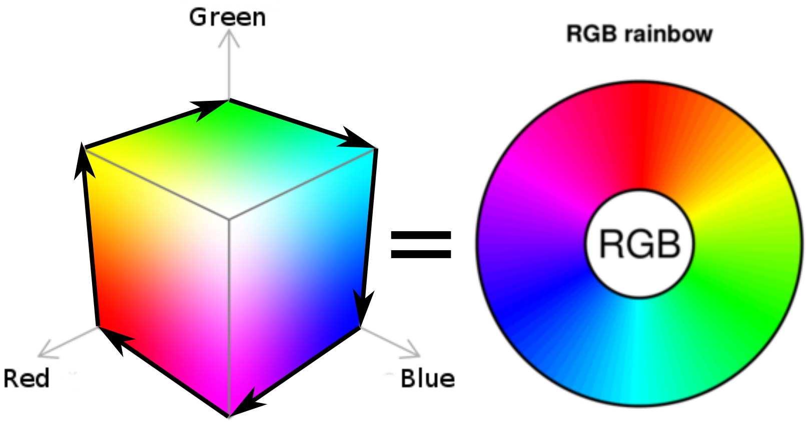
- Raise awareness of the issue.
- Introduce Hue-Chroma-Luminance (HCL) model.
  - Based on human perception.
  - Better control for choosing color palettes.
- Provide convenient software for exploring and assessing HCL-based palettes.

RGB rainbow



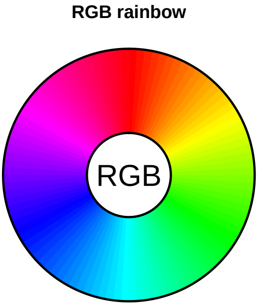
RGB color space: And the (in)famous rainbow color palette.

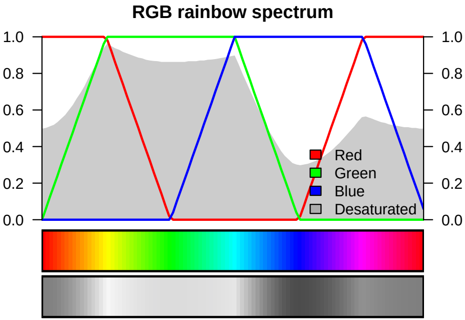
RGB rainbow



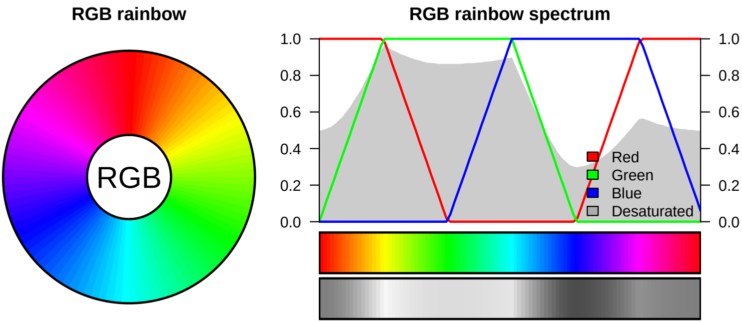
RGB color space: And the (in)famous rainbow color palette.

RGB rainbow


RGB rainbow
RGB


RGB rainbow spectrum
1.0
0.8
0.6
0.4
0.2
0.0
Red
Green
Blue
Desaturated

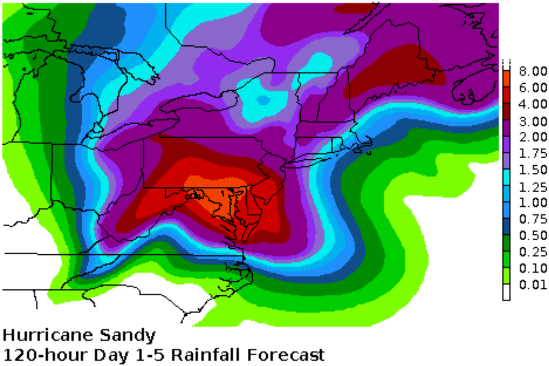
RGB rainbow



- The default color in many software packages.
- Conveniently used by many practitioners.
- Defaults only change slowly (if at all).

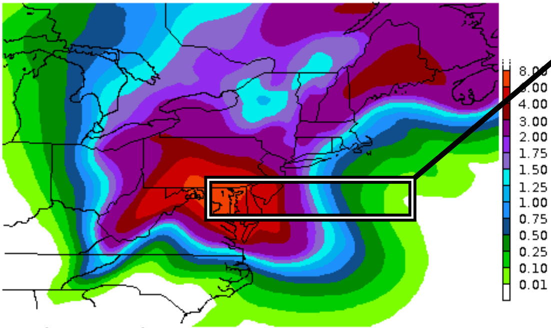
Question: Everybody does it – why should it be wrong?

What’s wrong?



Original figure as published by the NOAA (2012-10-27).

What’s wrong?

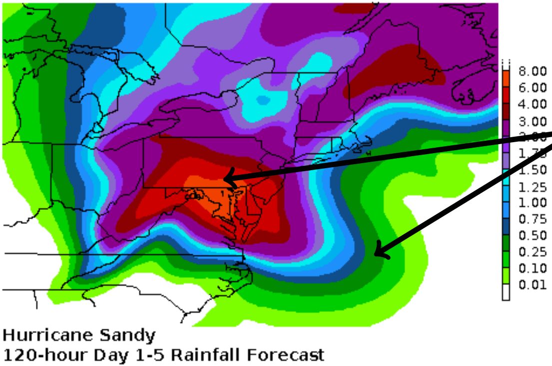


Hurricane Sandy
120-hour Day 1-5 Rainfall Forecast

Original figure as published by the NOAA (2012-10-27).

Gradients:
Very strong.

What’s wrong?

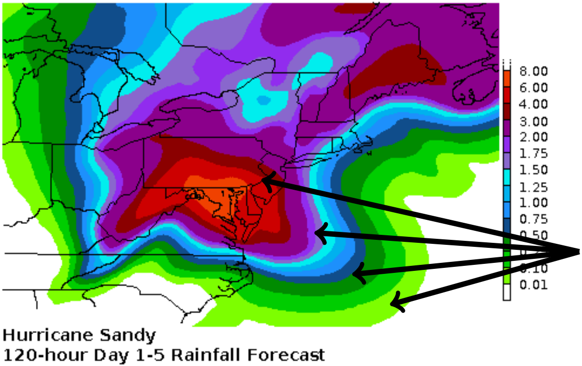


Original figure as published by the NOAA (2012-10-27).

Gradients:
Very strong.

Saturation:
Highly-saturated colors.

What’s wrong?



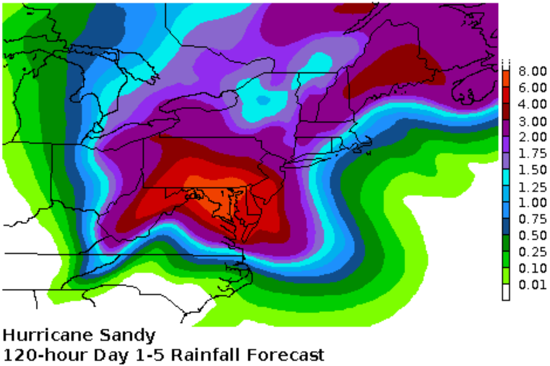
Original figure as published by the NOAA (2012-10-27).

Gradients:
Very strong.

Saturation:
Highly-saturated colors.

Discontinuous:
Bright, dark, bright, dark, . . .

What’s wrong?



Original figure as published by the NOAA (2012-10-27).

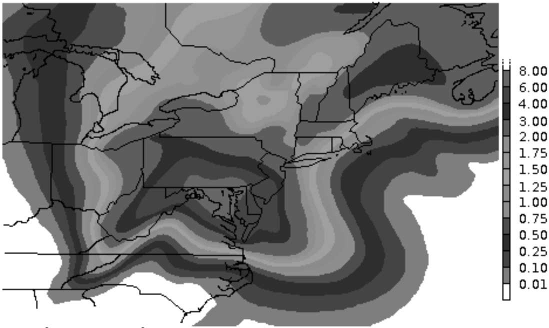
Gradients:
Very strong.

Saturation:
Highly-saturated colors.

Discontinuous:
Bright, dark, bright, dark, . . .

Basic guidelines: Colors should be assisting, simple, clear, appealing.

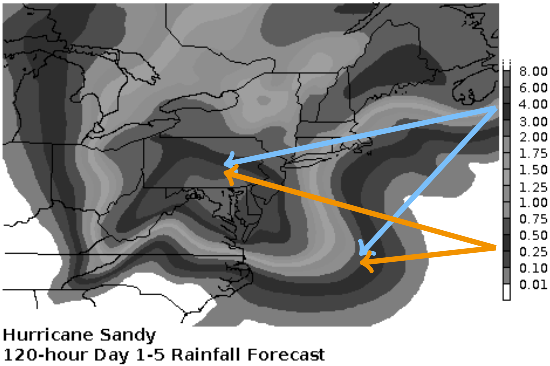
What’s wrong?



Hurricane Sandy
120-hour Day 1-5 Rainfall Forecast

Desaturated version of the original figure.

What's wrong?

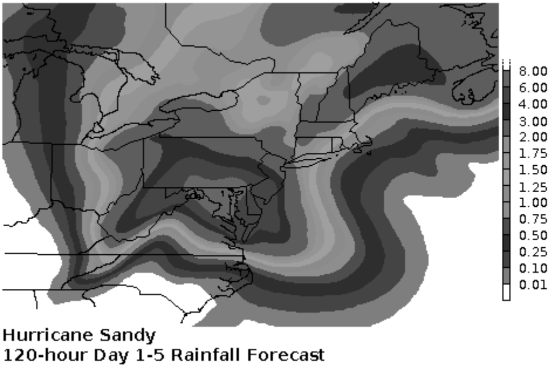


Desaturated version of the original figure.

Assignment:

No longer unique.

What’s wrong?

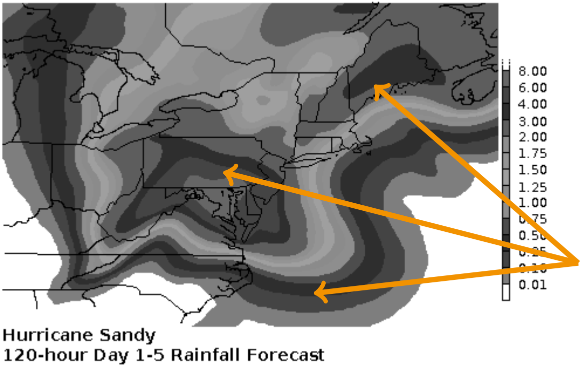


Desaturated version of the original figure.

Assignment:
No longer unique.

Interpretation:
Where is the maximum?

What's wrong?



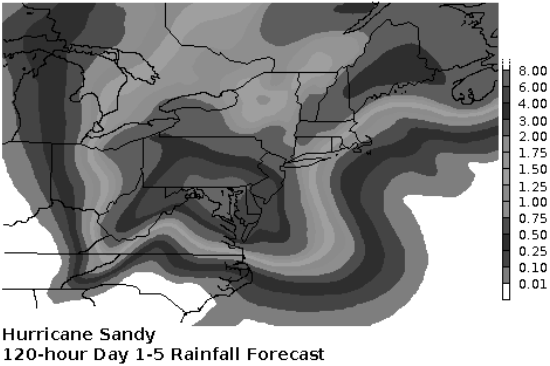
Desaturated version of the original figure.

Assignment:
No longer unique.

Interpretation:
Where is the maximum?

Focus:
On dark artefacts.

What's wrong?



Desaturated version of the original figure.

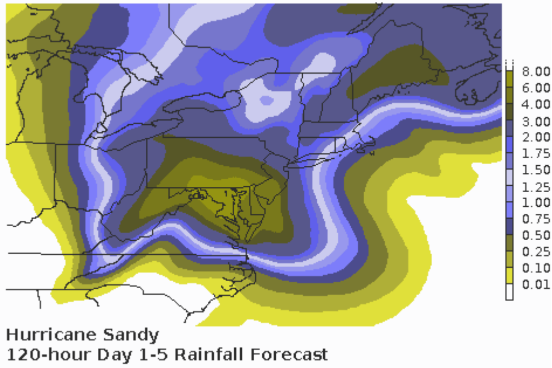
Assignment:
No longer unique.

Interpretation:
Where is the maximum?

Focus:
On dark artefacts.

Color should: Work everywhere, guide to important information.

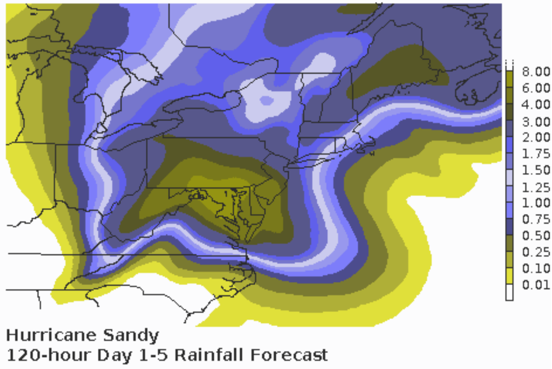
What’s wrong?



What color-blind people see (red-green weakness).

About 5% of all Europeans are affected.

What's wrong?



What color-blind people see (red-green weakness).

About 5% of all Europeans are affected.

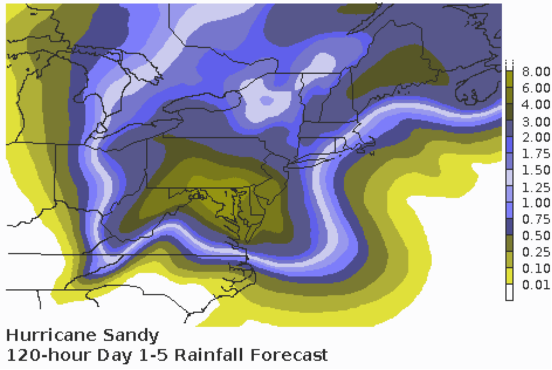
End user:

Who is it?

Consider:

Visual constraints?

What's wrong?



What color-blind people see (red-green weakness).

About 5% of all Europeans are affected.

End user:
Who is it?

Consider:
Visual constraints?

Color should: Consider end user needs.

Challenges

Summary: The colors in a palette should

- be simple and natural,
- not be unappealing,
- highlight the important information,
- not mislead the reader,
- work everywhere and for everyone.

Challenges

Summary: The colors in a palette should

- be simple and natural,
- not be unappealing,
- highlight the important information,
- not mislead the reader,
- work everywhere and for everyone.

In practice:

- People often do not think about it at all.
- …and simply use default colors.

Challenges

Summary: The colors in a palette should

- be simple and natural,
- not be unappealing,
- highlight the important information,
- not mislead the reader,
- work everywhere and for everyone.

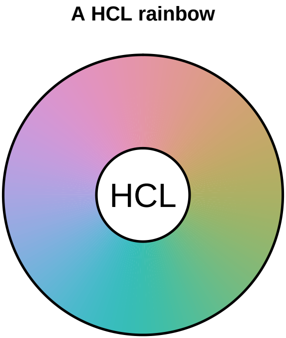
In practice:

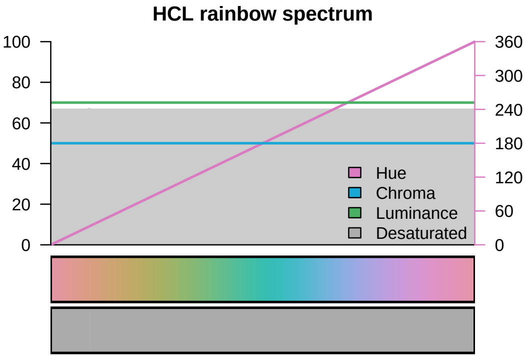
- People often do not think about it at all.
- . . . and simply use default colors.

Potential problems:

- For end users – reviewer, supervisor, collegue, customer.
- For your own day-to-day work.

HCL rainbow



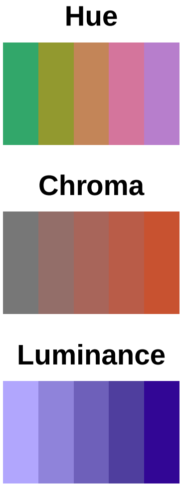


- Hue: Type of color.
- Chroma: Colorfulness.
- Luminance: Brightness.

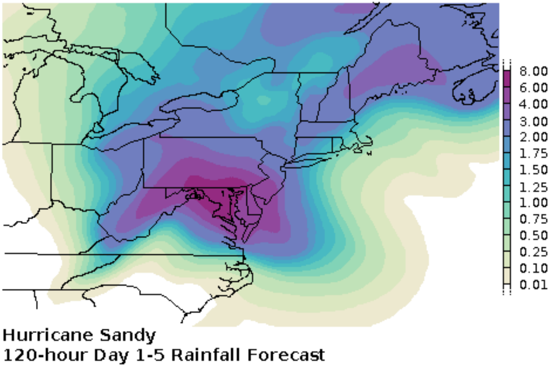
HCL color space

Perceptually-based color model:

- Hue: Type of color.
- Chroma: Colorfulness.
- Luminance: Brightness.

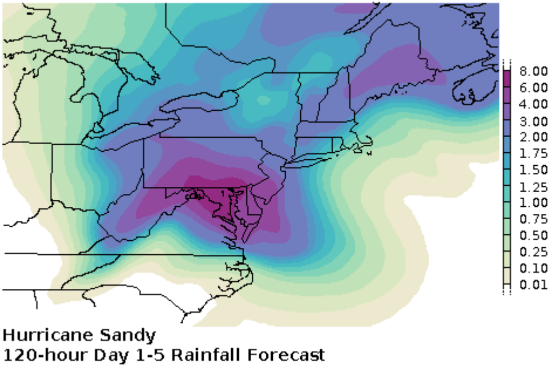


HCL version



Same information, HCL-based color palette.

HCL version



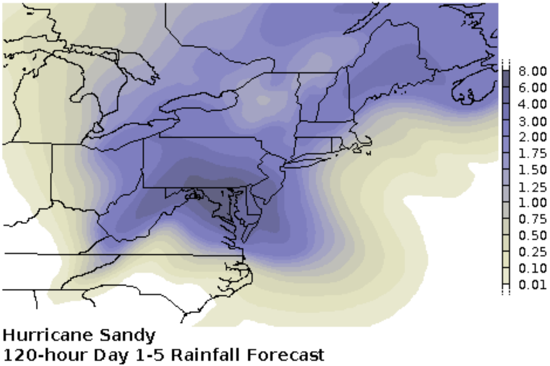
Same information, HCL-based color palette.

Colors:
Smooth gradients.

Information:
Guiding, no hidden information.

Works:
Screen, projector, gray-scale device.

HCL version



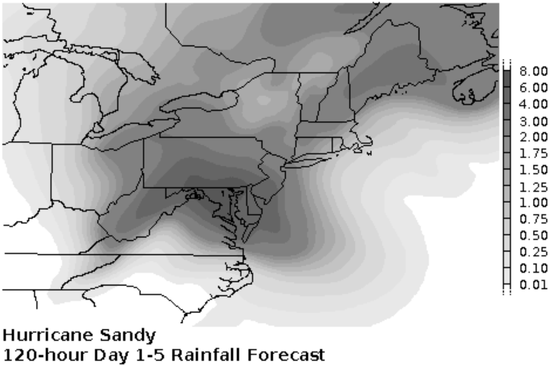
Colors:
Smooth gradients.

Information:
Guiding, no hidden information.

Works:
Screen, projector, gray-scale device.

Same information, HCL-based color palette (red-green weakness).

HCL version



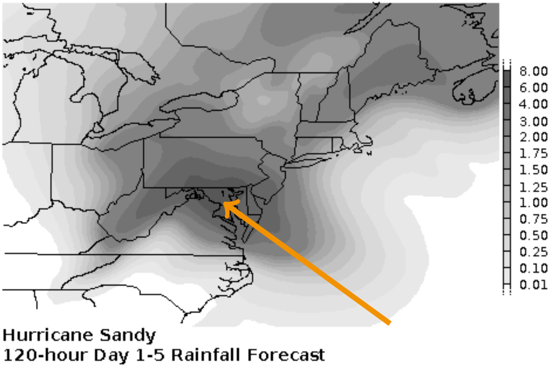
Same information, HCL-based color palette (desaturated).

Colors:
Smooth gradients.

Information:
Guiding, no hidden information.

Works:
Screen, projector, gray-scale device.

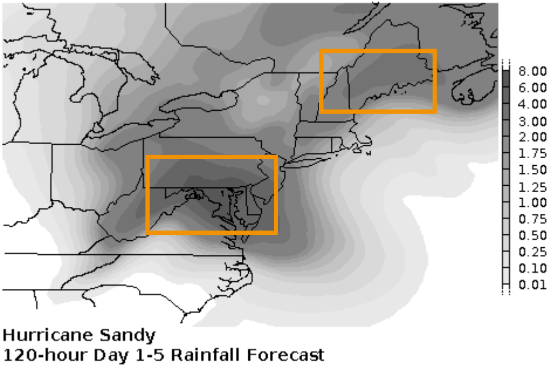
HCL version



Same information, HCL-based color palette (desaturated).

Assignment:
Higher values
(more precipitation)
$\rightarrow$ lower luminance.

HCL version

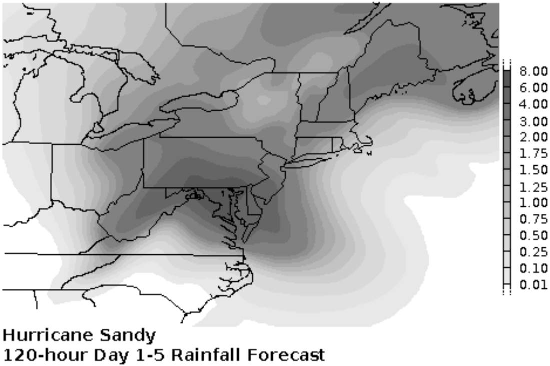


Same information, HCL-based color palette (desaturated).

Assignment:
Higher values (more precipitation) → lower luminance.

Focus:
Lead readers to most important areas.

HCL version



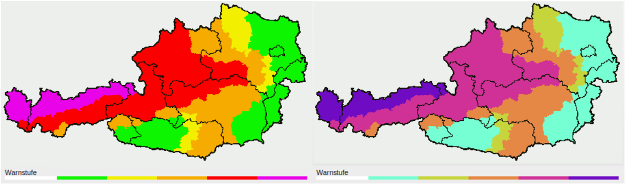
Same information, HCL-based color palette (desaturated).

Assignment:
Higher values
(more precipitation)
$\rightarrow$ lower luminance.

Focus:
Lead readers to most
important areas.

Summary: Solved a lot of problems by changing the color palette.

Warning map example



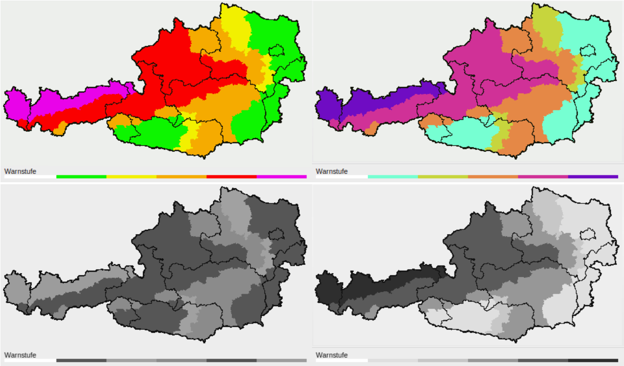
Colorized

Original (left)

HCL idea (right)

UBIMET GmbH, www.uwz.at, 2013-05-31.

Warning map example



Colorized

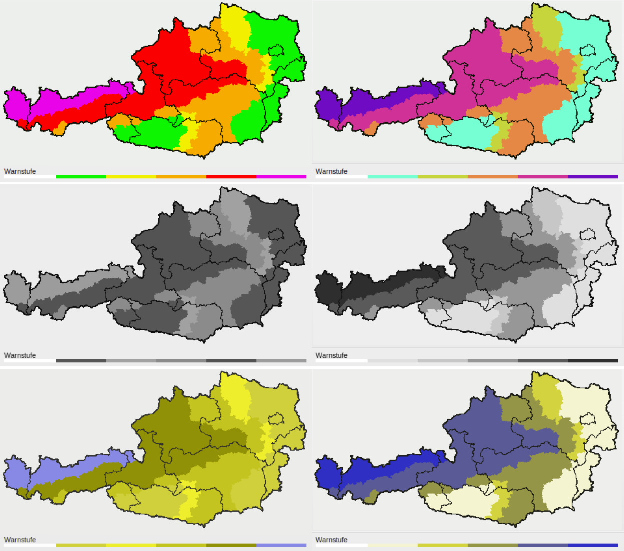
Original (left)

HCL idea (right)

Gray-scale

UBIMET GmbH, www.uwz.at, 2013-05-31.

Warning map example



Colorized

Original (left)

HCL idea (right)

Gray-scale

Deuteranopia

Red-green weakness

UBIMET GmbH, www.uwz.at, 2013-05-31.

Experiences with practitioners

In the beginning:

- Hesitation of colleagues.
- “Not necessary!”
- “Why should we change existing products?”
- “Everybody does it like this…”

Experiences with practitioners

In the beginning:

- Hesitation of colleagues.
- “Not necessary!”
- “Why should we change existing products?”
- “Everybody does it like this. . . ”

A few days later:

- Mainly positive feedback
- Decrease of misinterpretations in classroom (“Weather and Forecast”).
- “Much easier to interpret. . . ”
- “How can I make use of those palettes (in my software)?”

Experiences with practitioners

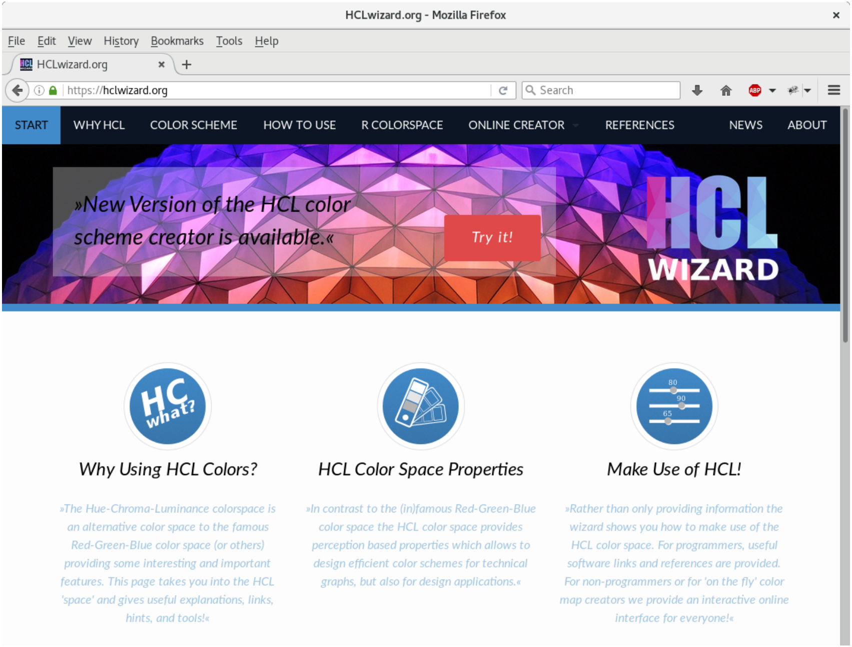
In the beginning:

- Hesitation of colleagues.
- “Not necessary!”
- “Why should we change existing products?”
- “Everybody does it like this…”

A few days later:

- Mainly positive feedback
- Decrease of misinterpretations in classroom (“Weather and Forecast”).
- “Much easier to interpret…”
- “How can I make use of those palettes (in my software)?”
  $\rightarrow$ Visit `https://hclwizard.org/`

Experiences with practitioners



Experiences with practitioners

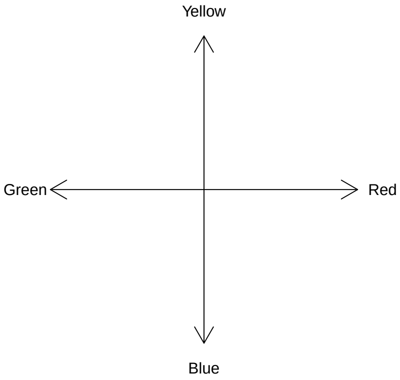
In R:

- HCL color space: `hcl()` or `polarLUV()` in *colorspace*.
- HCL-based palettes in *colorspace*: `rainbow_hcl()`, `heat_hcl()`, `sequential_hcl()`, `diverge_hcl()`, ...
- Interactive exploration: `choose_palette()` (in Tcl/Tk) or `hclwizard()` (in *shiny*).
- Further useful packages: *RColorBrewer* (fixed palettes from *ColorBrewer.org*), *ggplot2*, ...

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.

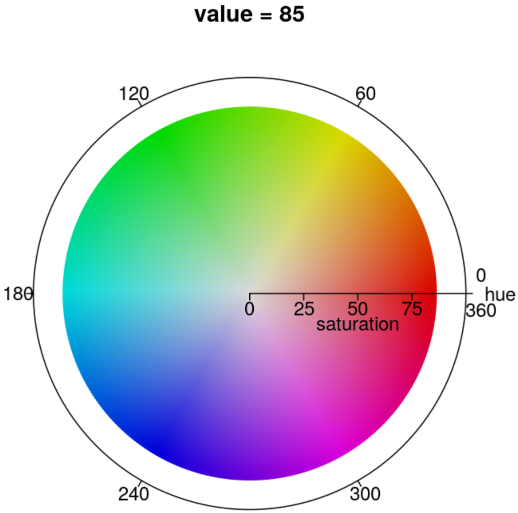
Specification: Triplet (H, S, V) with $H = 0, \dots, 360$ and $S, V = 0, \dots, 100$, often all transformed to unit interval (e.g., in R).

Shape: Cone (or transformed to cylinder).

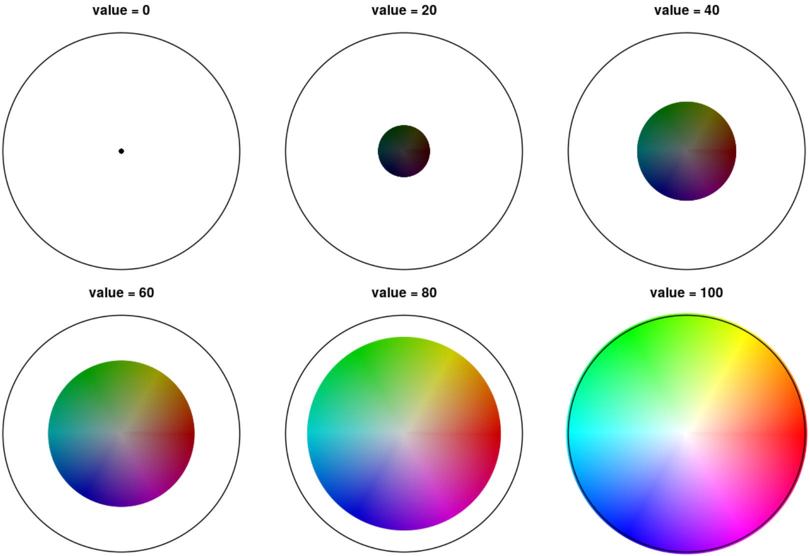
Problem: Dimensions are confounded, hence not really perceptually based.

In R: `hsv()` or `HSV()` in *colorspace*.

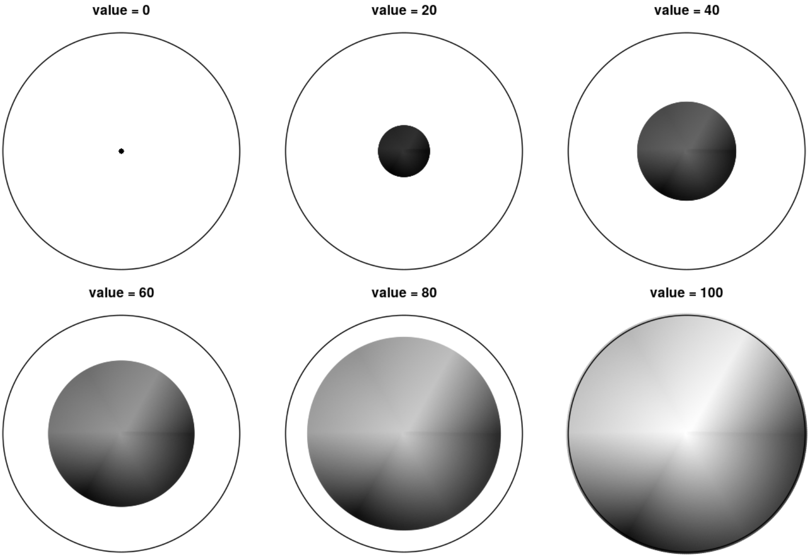
Color space: HSV



Color space: HSV



Color space: HSV



Color space: HCL

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

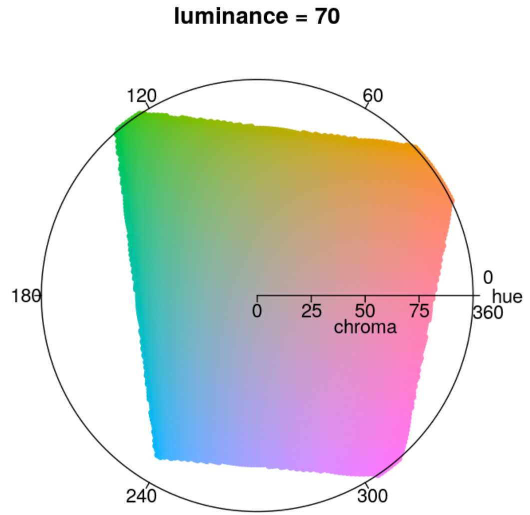
Specification: Triplet (H, C, L) with $H = 0, \dots, 360$ and $C, L = 0, \dots, 100$.

Shape: Distorted double cone.

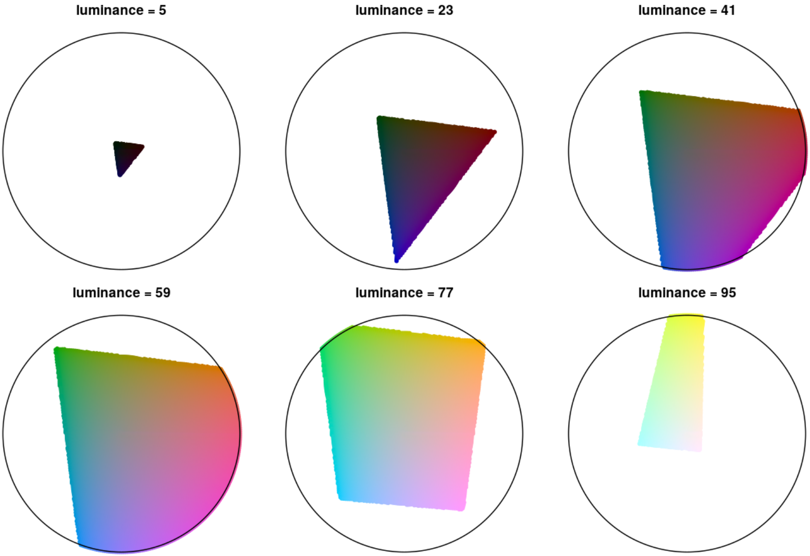
Problem: Care is needed when traversing along the axes due to distorted shape.

In R: `hcl()` or `polarLUV()` in *colorspace*.

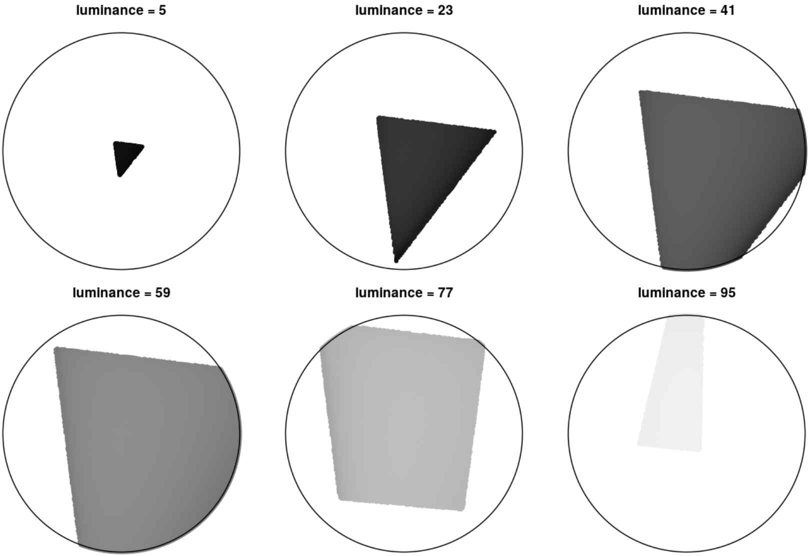
Color space: HCL



Color space: HCL



Color space: HCL



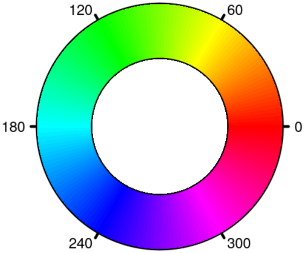
Color palettes: Qualitative

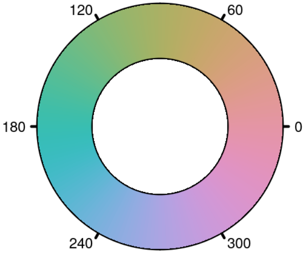
Goal: Code qualitative information.

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

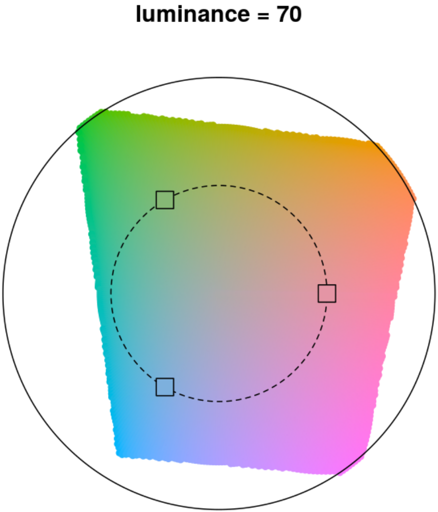
$$(H, 50, 70)$$

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



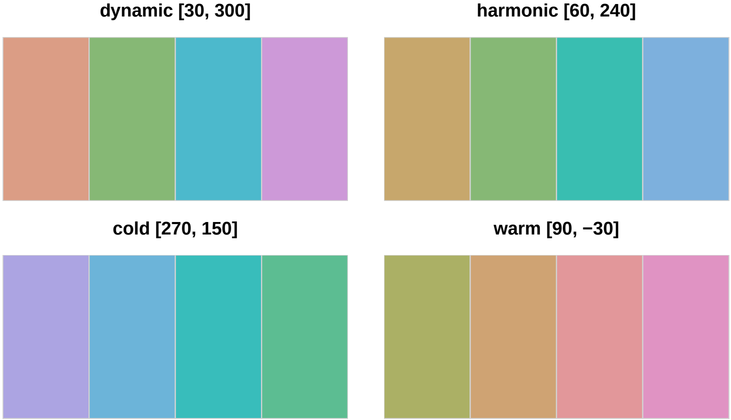


Color palettes: Qualitative



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).



Color palettes: Qualitative

Illustrations:

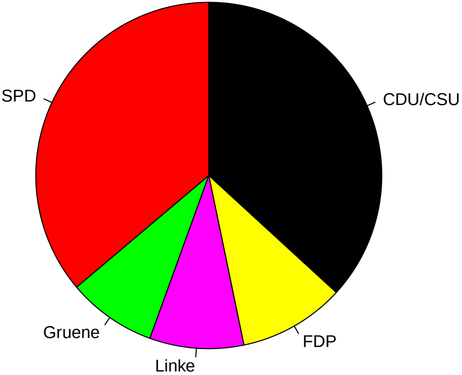
- Pie chart of seats in the German parliament Bundestag, 2005.
- Mosaic display of votes for the German Bundestag, 2005.
- Scatter plot with three clusters (and many points).

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

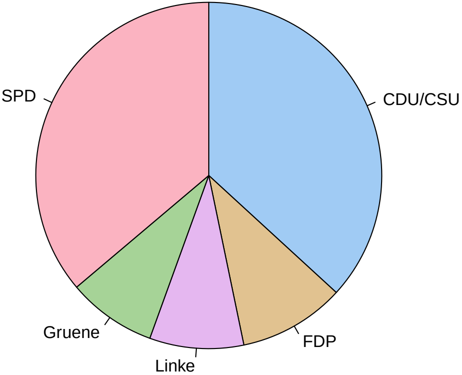
Problems:

- 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.

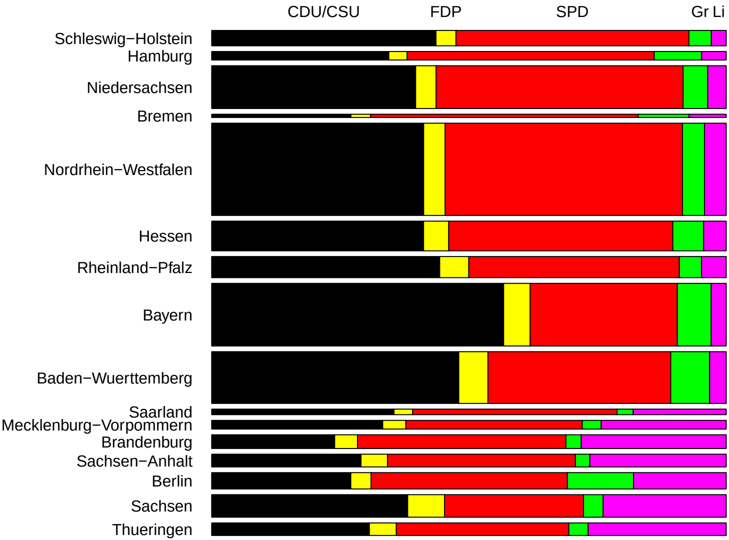
Color palettes: Qualitative



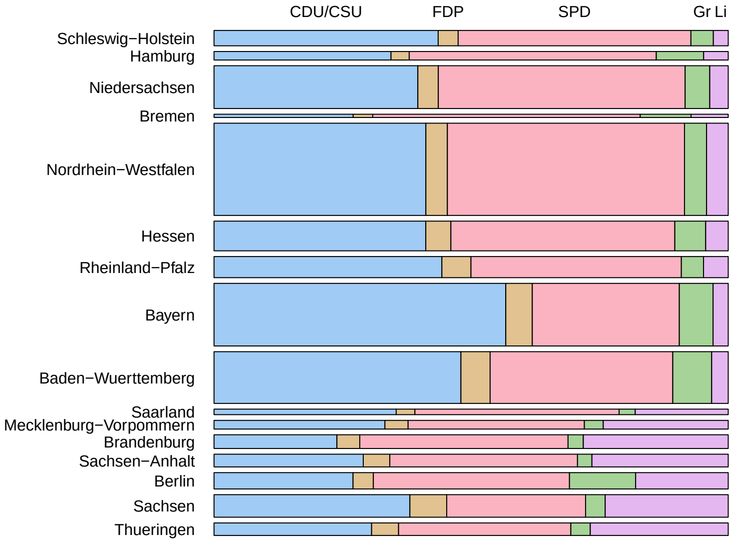
Color palettes: Qualitative



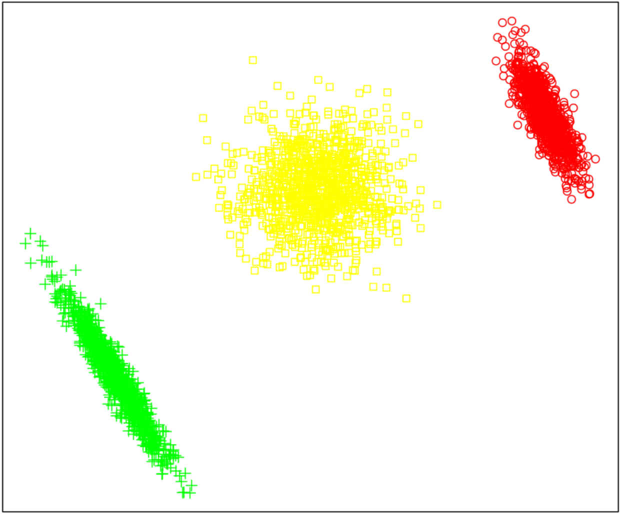
Color palettes: Qualitative



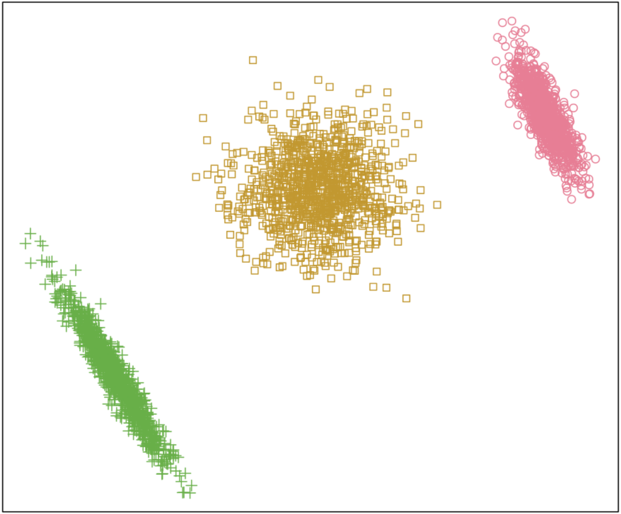
Color palettes: Qualitative



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.,

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

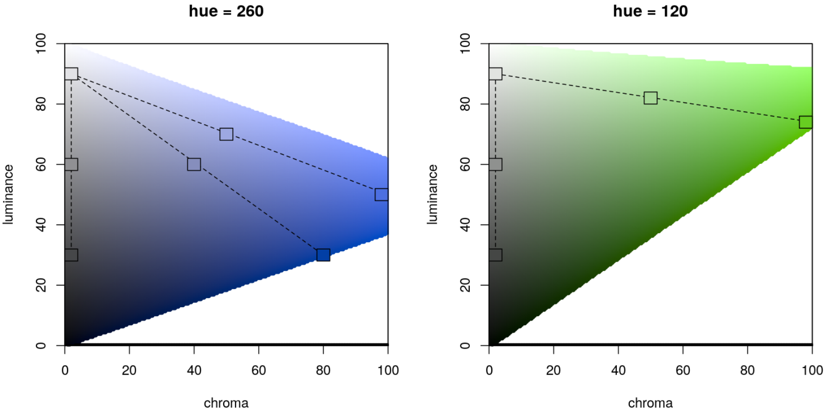
The hue H does not matter, chroma is set to 0 (no color), luminance ranges in $[30, 90]$, avoiding the extreme colors black and white.

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

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

for a fixed hue H.

Color palettes: Sequential



Color palettes: Sequential

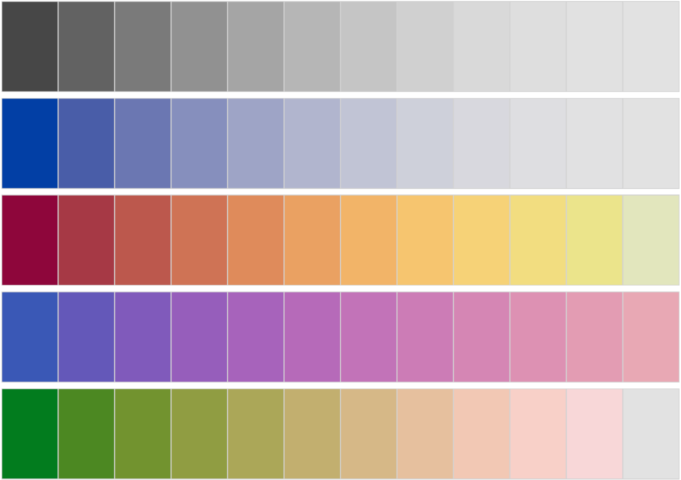
Modification: To increase the contrast within the palette even further, simultaneously vary the hue as well:

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

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



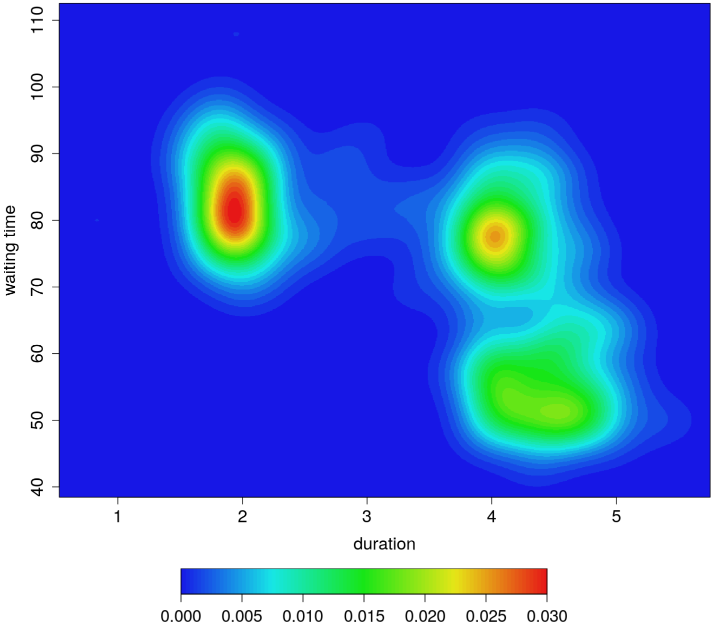
Color palettes: Sequential

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

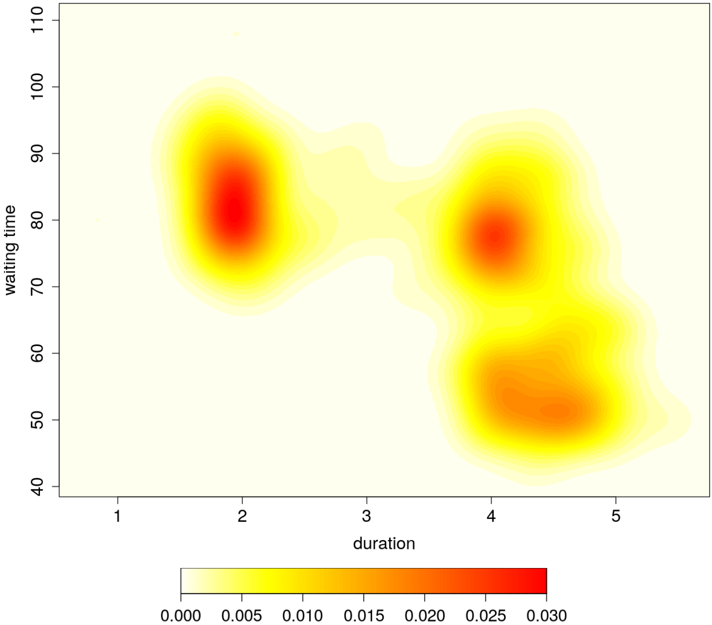
Palettes:

- `rainbow()`
- `heat.colors()`
- `heat_hcl()` (color and desaturated)

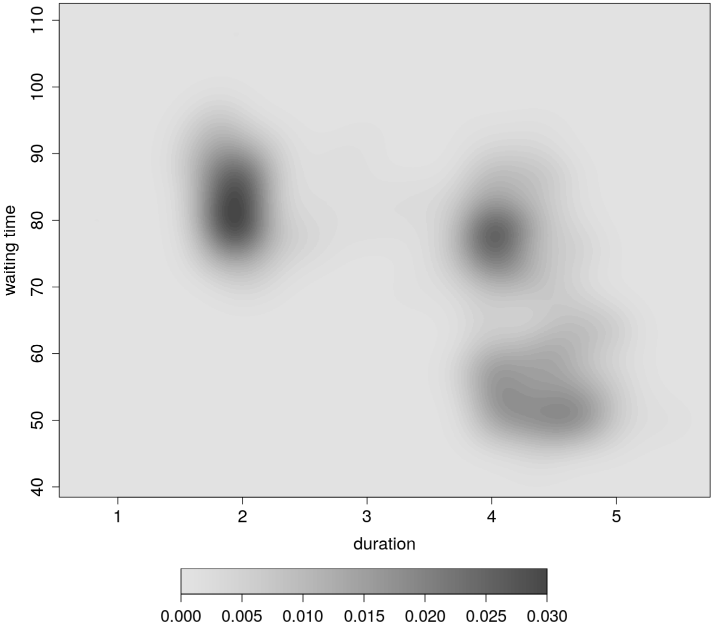
Color palettes: Sequential



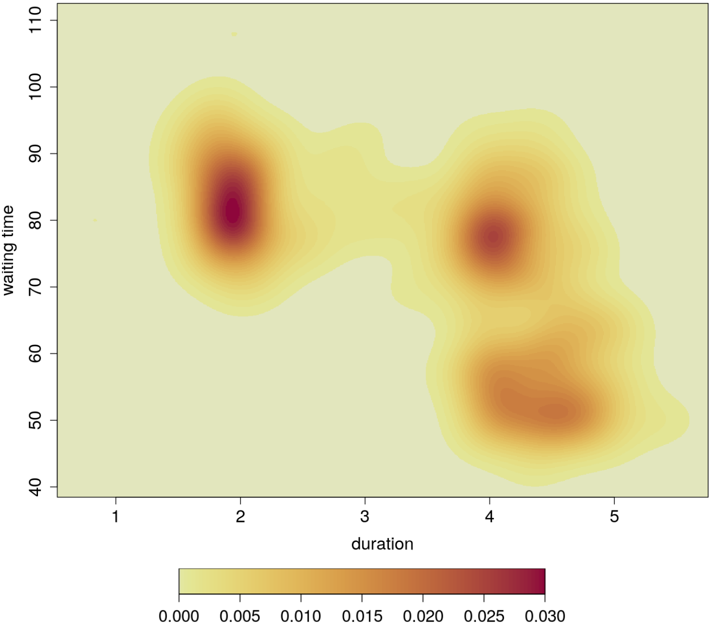
Color palettes: Sequential



Color palettes: Sequential



Color palettes: Sequential



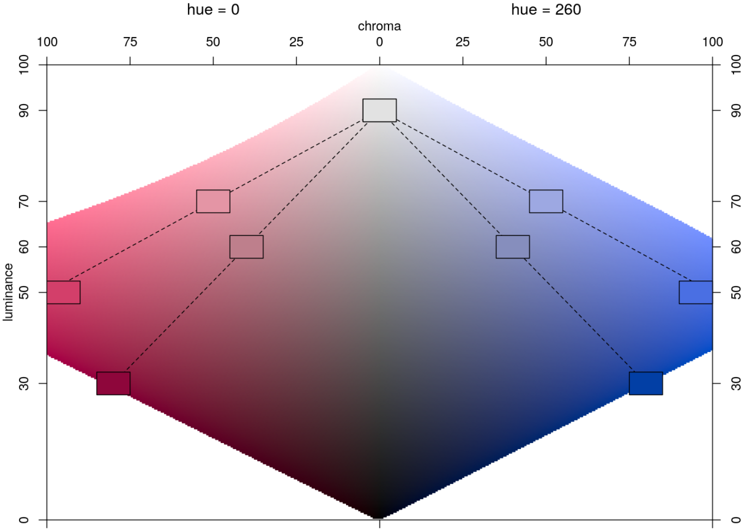
Color palettes: Diverging

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

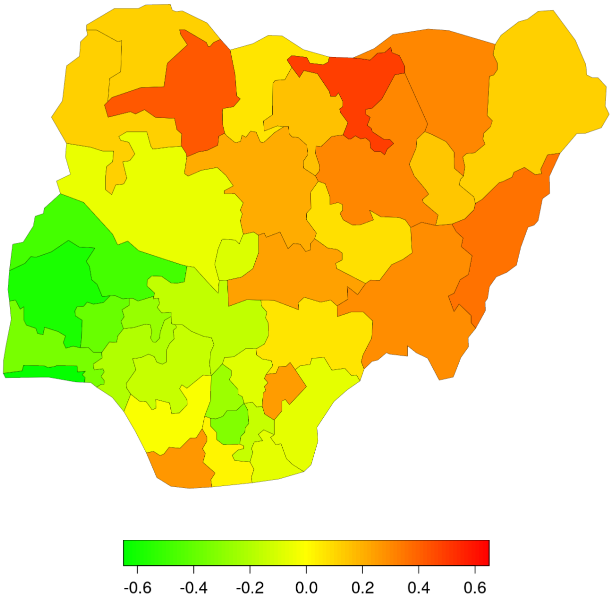
Illustrations:

- Map of Nigeria shaded by posterior mode estimates for childhood mortality.
- Model-based mosaic display for treatment of arthritis.

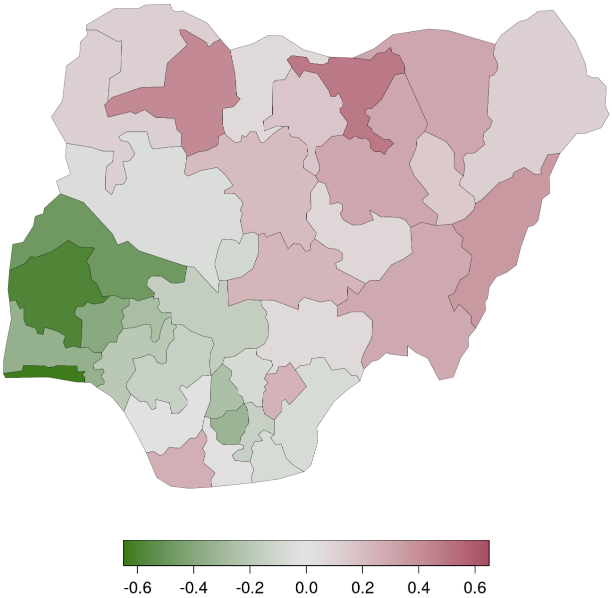
Palettes:

- `rainbow()` vs. `diverge_hcl()`
- `diverge_hsv()` vs. `diverge_hcl()`

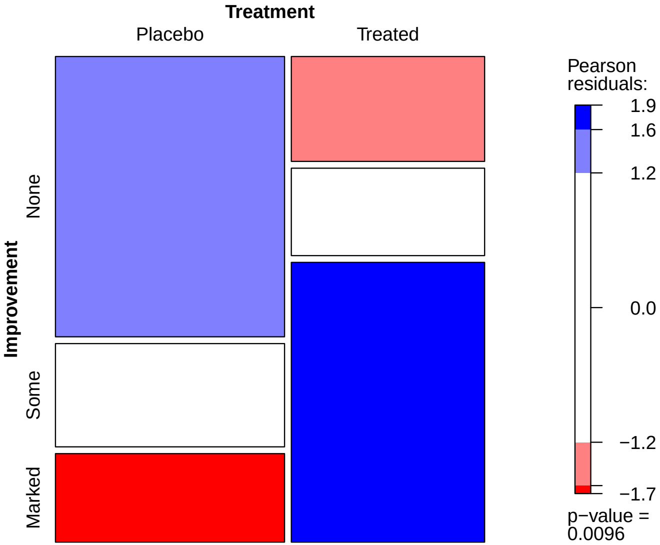
Color palettes: Diverging



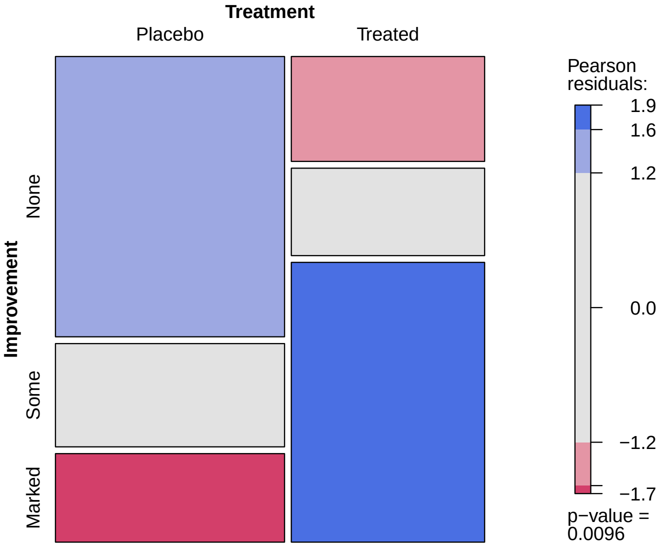
Color palettes: Diverging



Color palettes: Diverging



Color palettes: Diverging



Color blindness

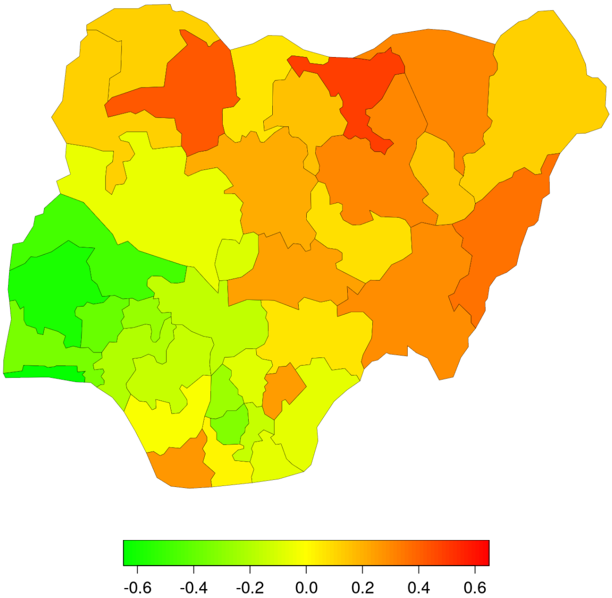
Problem: A few percent of humans (particularly males) have deficiencies in their color vision, typically referred to as *color blindness*.

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

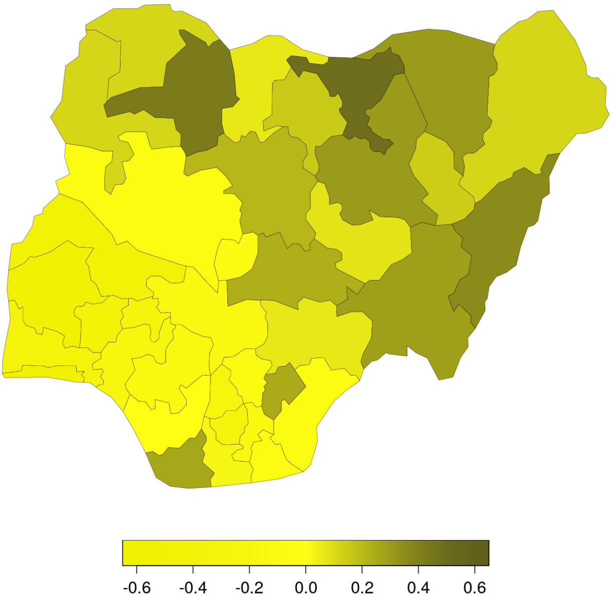
Solution: Construct suitable HCL colors.

- Use large 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).

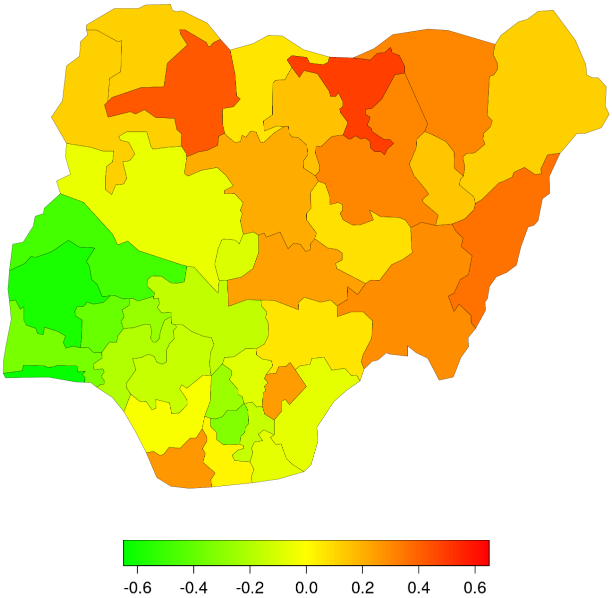
Color blindness



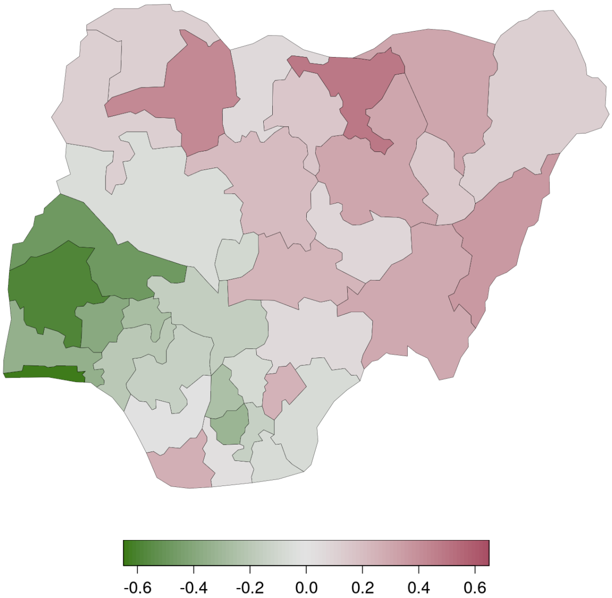
Color blindness



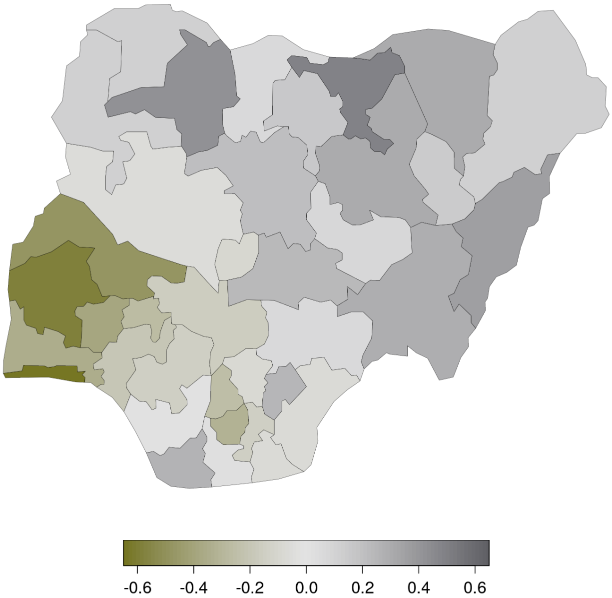
Color blindness



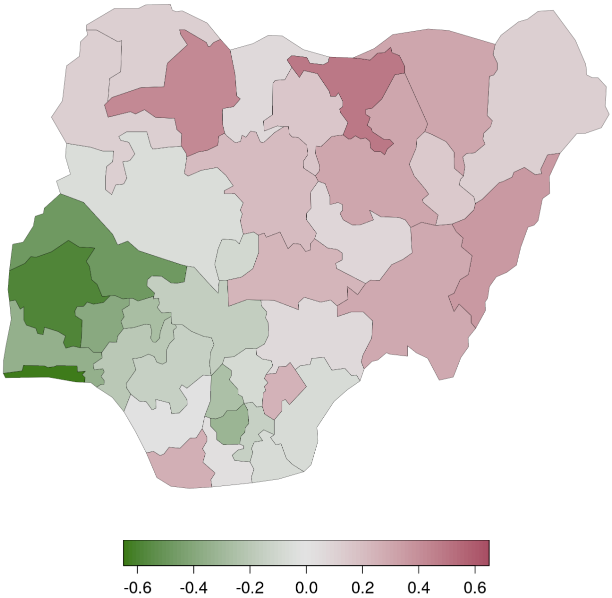
Color blindness



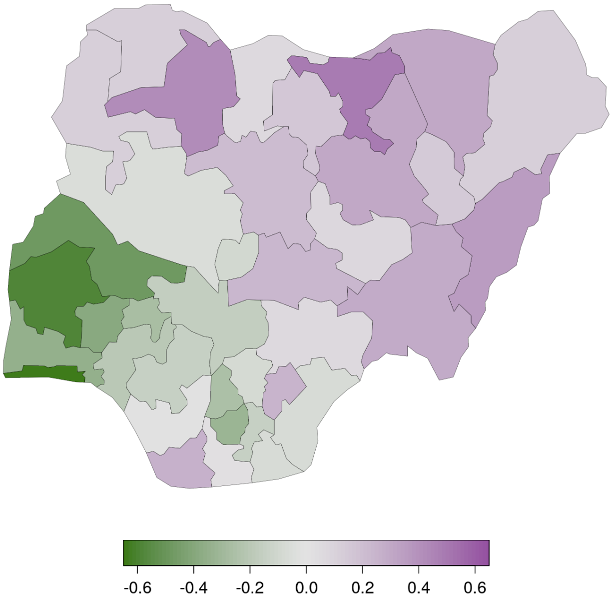
Color blindness



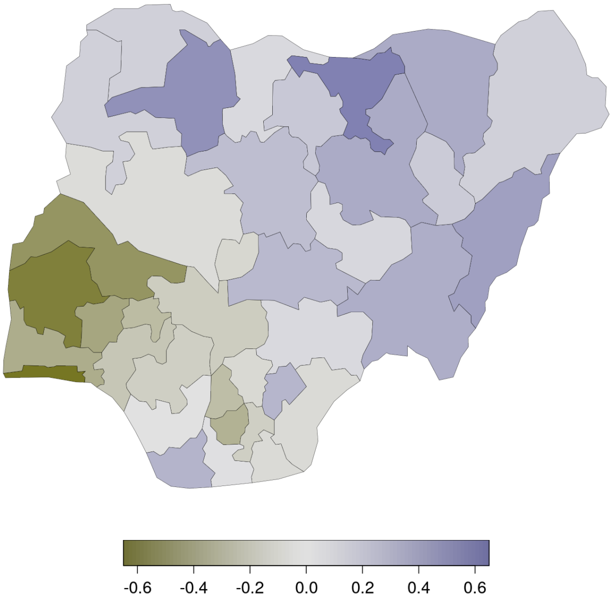
Color blindness



Color blindness



Color blindness



Summary

Choice of colors:

- Use color with care, do not overestimate power of color.
- Think about who the readers/users are.
- Avoid large areas of flashy, highly-saturated colors.
- Employ monotonic luminance scale for numerical data.

Try it yourself:

- `https://hclwizard.org`
- *colorspace* in R.

References

Stauffer R, Mayr GJ, Dabernig M, Zeileis A (2015). “Somewhere over the Rainbow: How to Make Effective Use of Colors in Meteorological Visualizations.” *Bulletin of the American Meteorological Society*, **96**(2), 203–216. `doi:10.1175/BAMS-D-13-00155.1`

Ihaka R, Murrell P, Hornik K, Fisher JC, Stauffer R, Zeileis A (2016). *colorspace: Color Space Manipulation.* R package version 1.3-0.
`https://CRAN.R-project.org/package=colorspace`

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.
`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,
`http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/`.