Choosing Color Palettes for Statistical Graphics

Achim Zeileis and Kurt Hornik
Wirtschaftsuniversität Wien

Abstract

Statistical graphics are often augmented by the use of color coding information contained in some variable. When this involves the shading of areas (and not only points or lines)—e.g., as in barplots, pie charts, mosaic displays or heatmaps—it is important that the colors are perceptually based and do not introduce optical illusions or systematic bias. Here, we discuss how the perceptually-based HCL color space can be used for deriving suitable color palettes for coding categorical data (qualitative palettes) and numerical variables (sequential and diverging palettes).

Keywords: qualitative palette, sequential palette, diverging palette, HCL colors, HSV colors, perceptually-based color space.

1. Introduction

First, we motivate why HCL colors (Ihaka 2003) are suitable for choosing color palettes by contrasting them with the more commonly implemented HSV colors.

Then, we describe strategies how color palettes for categorical and numerical data can be chosen in this space. Following Brewer (1999), we distinguish three types of palettes: qualitative, sequential and diverging. All palettes described are available in R (R Development Core Team 2006) in the package **vcd** (Meyer, Zeileis, and Hornik 2006) using the HCL color implementation from **colorspace** (Ihaka 2004). Technical documentation to the R implementations along with a large collection of examples is available via `help("rainbow_hcl")` that provides more comparisons between existing R palettes (based on HSV colors) and the HCL color palettes.

2. Color spaces

For choosing color palettes, it is imperative to have an understanding how colors are perceived. For this it is helpful to have an idea how human color vision evolved. It has been hypothesized that it developed in three distinct stages: 1. perception of *light/dark* contrasts (monochrome only), 2. *yellow/blue* contrasts (usually associated with our notion of warm/cold colors), 3. *green/red* contrasts (helpful for assessing the ripeness of fruit). See Ihaka (2003) for more details and references.

Due to these three color axes, colors are typically described as locations in a 3-dimensional spaces. However, human perception of color does not correspond to the physiological axes above, but rather to polar coordinates in the color plane (yellow/blue vs. green/red) plus a third light/dark axis. Thus, perceptually-based color spaces are defined by three dimensions that try to capture

1. **hue** (dominant wavelength)
2. **chroma** (colorfulness)
3. **luminance** (brightness, 'amount of gray')

A popular implementation of such a color space, available in many graphics and statistics software packages, are *HSV* (hue, saturation, value) colors. They are a simple transformation of *RGB* (red,

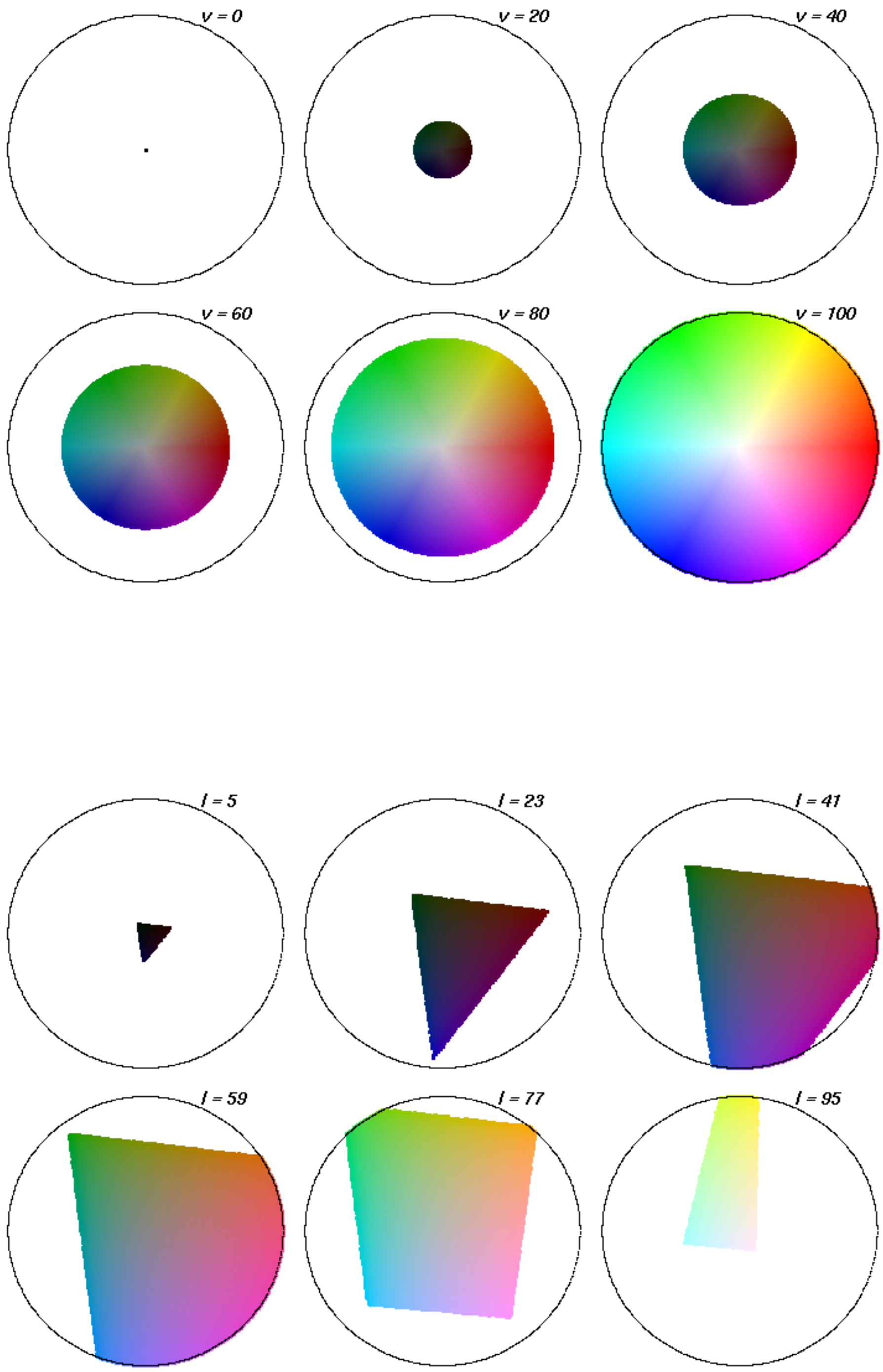

Figure 1: HSV and HCL space

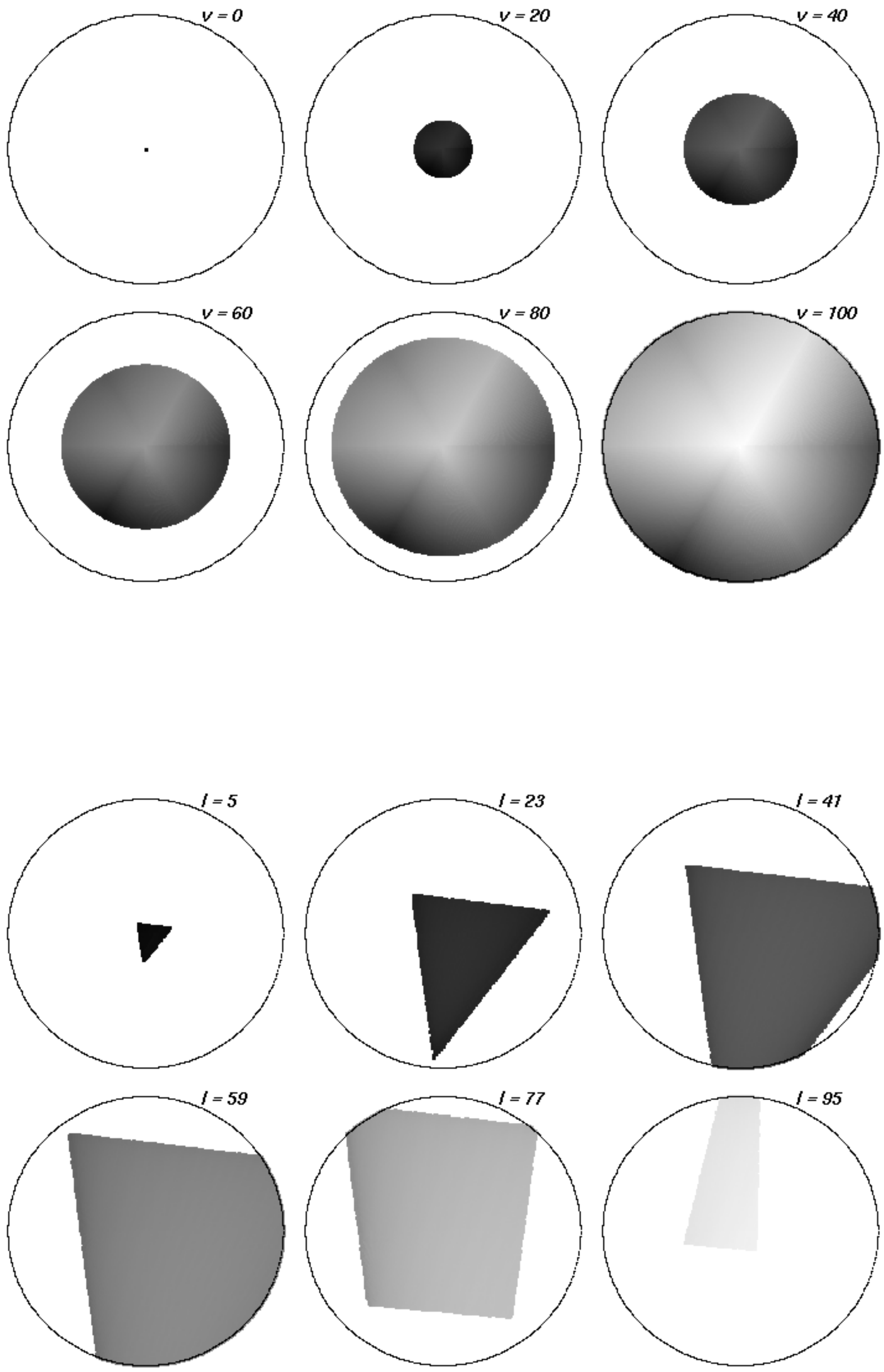

Figure 2: HSV and HCL space (in gray levels)

green, blue) colors and are defined by a triplet (H, S, V) with $H \in [0, 360]$ and $S, V \in [0, 100]$. HSV space has the shape of a single regular cone (often inflated to a regular cylinder). Vertical sections through this space are shown in the upper panel of Figure 1, depicting hue and saturation given value. Although simple to specify and easily available in many computing environments, HSV colors have a fundamental drawback: its three dimensions map to the three dimensions of human color perception very poorly. The three dimensions are confounded which is most easily seen when converting the vertical sections to gray scale images in Figure 2. Clearly, the brightness of colors is not uniform over hues and saturations (given value)—therefore, HSV colors are often not considered to be perceptually based.

To overcome these drawbacks, various color spaces have been suggested that properly map to the perception dimensions, the most prominent of which are the CIELUV and CIELAB spaces developed by the Commission Internationale de l'Éclairage (2004). Ihaka (2003) argues that CIELUV colors are typically preferred for use with emissive technologies such as computer screens which makes them an obvious candidate for implementation in statistical software packages. By taking polar coordinates in the UV plane of CIELUV, *HCL* (hue, chroma, luminance) colors are obtained, defined by a triplet (H, C, L) with $H \in [0, 360]$ and $C, L \in [0, 100]$. HCL space has the shape of a distorted double cone: the admissable chroma and luminance values depend on the hue chosen. The lower panel of Figure 1 shows vertical sections through this space: each of the resulting hue/chroma planes (given luminance) is now properly balanced towards the same gray (going from black to white with increasing luminance) which becomes obvious when converting the colors to a gray scale is in Figure 2. This balancing of HCL colors gives us the opportunity to conveniently choose color palettes which code categorical and/or numerical information by translating it to the three perceptual dimensions. However, some care is required for dealing with the irregular shape of the HCL space which will be addressed in the following sections.

3. Qualitative palettes

Qualitative palettes are sets of colors for depicting different categories, i.e., for coding a categorical variable. Usually, these should give the same perceptual weight to each category so that no group is perceived to be larger/more important than any other one. Typical applications of qualitative palettes in statistics would be bar plots, pie charts (see Figure 10) or highlighted mosaic displays (see Figure 11).

Ihaka (2003) describes a simple strategy for choosing such palettes: chroma and luminance are kept fixed and only the hue is varied for obtaining different colors which are consequently all balanced towards the same gray. If colors from the full color wheel (i.e., $H \in [0, 360]$) should be used, not all combinations of chroma and luminance are feasible. Figure 3 depicts how three colors are chosen, given $C = 50$ and $L = 70$.

Various strategies for choosing the hues in a certain palette are conceivable. A simple and intuitive one is to use colors as metaphors for categories (e.g., for political parties). Figure 4 shows a few further examples for generating qualitative sets of colors $(H, 50, 70)$. In the upper left panel colors from the full spectrum are used ($H = 30, 120, 210, 300$) creating a 'dynamic' set of colors. The upper right panel shows a 'harmonic' set with $H = 60, 120, 180, 240$. Warm colors (from the blue/green part of the spectrum: $H = 270, 230, 190, 150$) and cold colors (from the yellow/red part of the spectrum: $H = 90, 50, 10, 330$) are shown in the lower left and right panel, respectively.

In **vcd**, these palettes are available in the function

```
rainbow_hcl(n, c = 50, l = 70, start = 0, end = 360*(n-1)/n, ...)
```

4. Sequential palettes

Sequential palettes are used for coding numerical information that simply ranges in a certain interval where low values are considered to be uninteresting and high values are interesting. Without

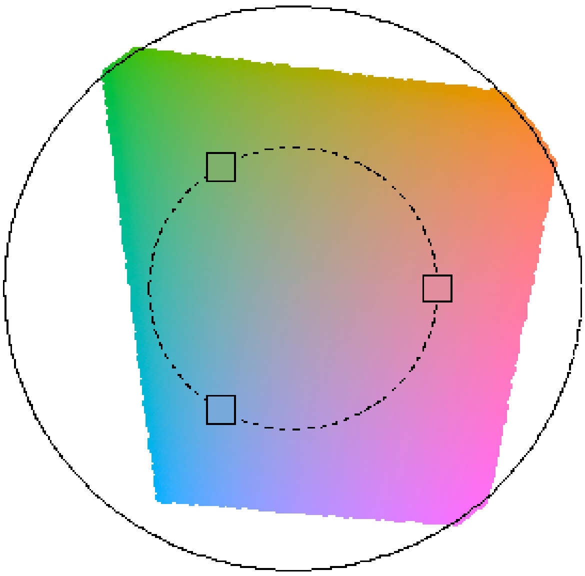

Figure 3: Constructing qualitative palettes

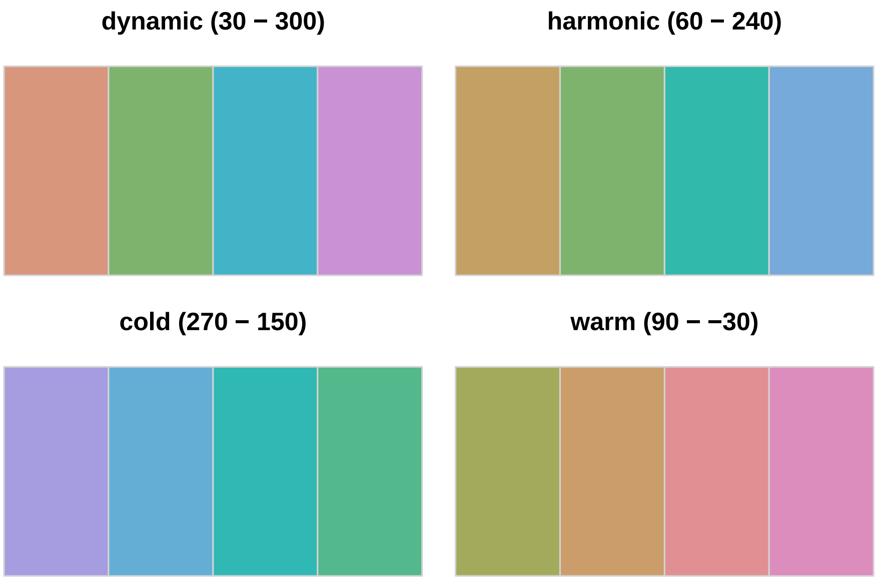

Figure 4: Examples for qualitative palettes

loss of generality, we assume that we want to visualize an intensity or interestingness $i \in [0, 1]$. A typical application in statistics are heatmaps (see Figure 9).

The simplest solution to this task is to employ light/dark contrasts, i.e., employ the oldest and simples perceptual axis. The interestingness is thus coded by an increasing amount of gray (i.e., decreasing luminance)

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

where the hue H used does not matter, chroma is set to 0 (i.e., no color), and luminance ranges in $[30, 90]$ avoiding the extreme colors white ($L = 100$) and black ($L = 0$). Instead of going linearly from light to dark gray, luminance could also be increased nonlinearly, e.g., by some function $f(i)$ that controls wether intensity/luminance is increased quickly or not. We found $f(i) = i^p$ to be a convenient transformation where the power p can be varied to achieve different degrees of non-linearity.

Furthermore, the intensity i could additionally be coded by colorfulness (chroma), e.g.,

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

This strategy is depicted in the left panel of Figure 5 for a blue hue $H = 260$ and different combinations of maximal chroma (0, 80 and 100, respecitvely) and minimal luminance (30, 30 and 50, respectively). The first two combinations are also shown in the first two rows of Figure 6. The right panel of Figure 5 shows that the exact same strategy is not possible for the green hue $H = 120$. While the gray colors without chroma can be chosen in the same way, there is a stronger trade-off between using dark colors (with low luminance) and colorful colors (with high chroma). Hence, the second path from light gray to full green ends at a much lighter color with $L = 75$.

In **vcd**, this strategy is implemented in the function

```
sequential_hcl(n, h = 260, c = c(80, 0), l = c(30, 90), power = 1.5, ...)
```

To increase the contrast between the colors in the palette even further, the ideas from the previous sequential palettes can also be combined with qualitative palettes by simultaneously varying the hue as well:

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

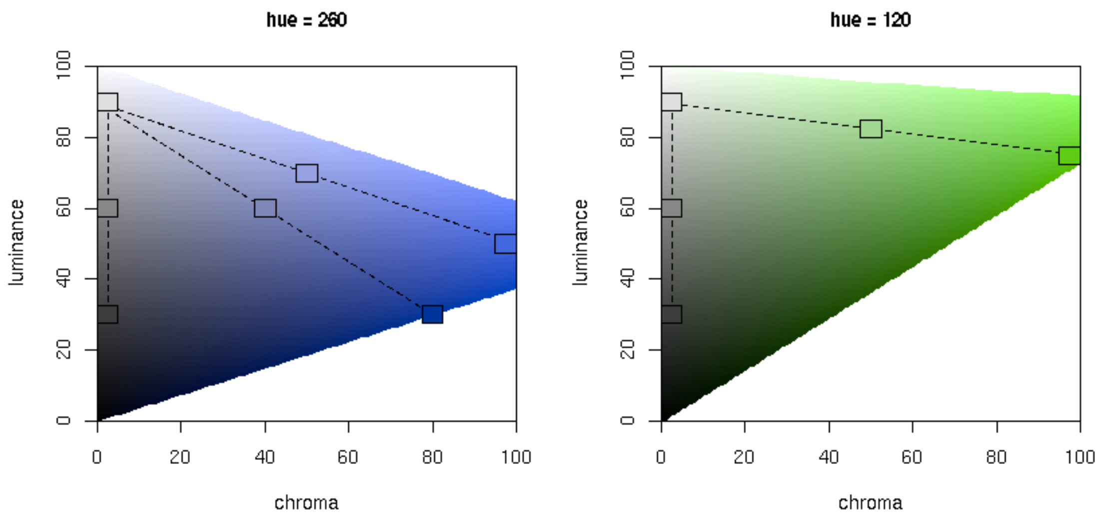

Figure 5: Constructing sequential palettes

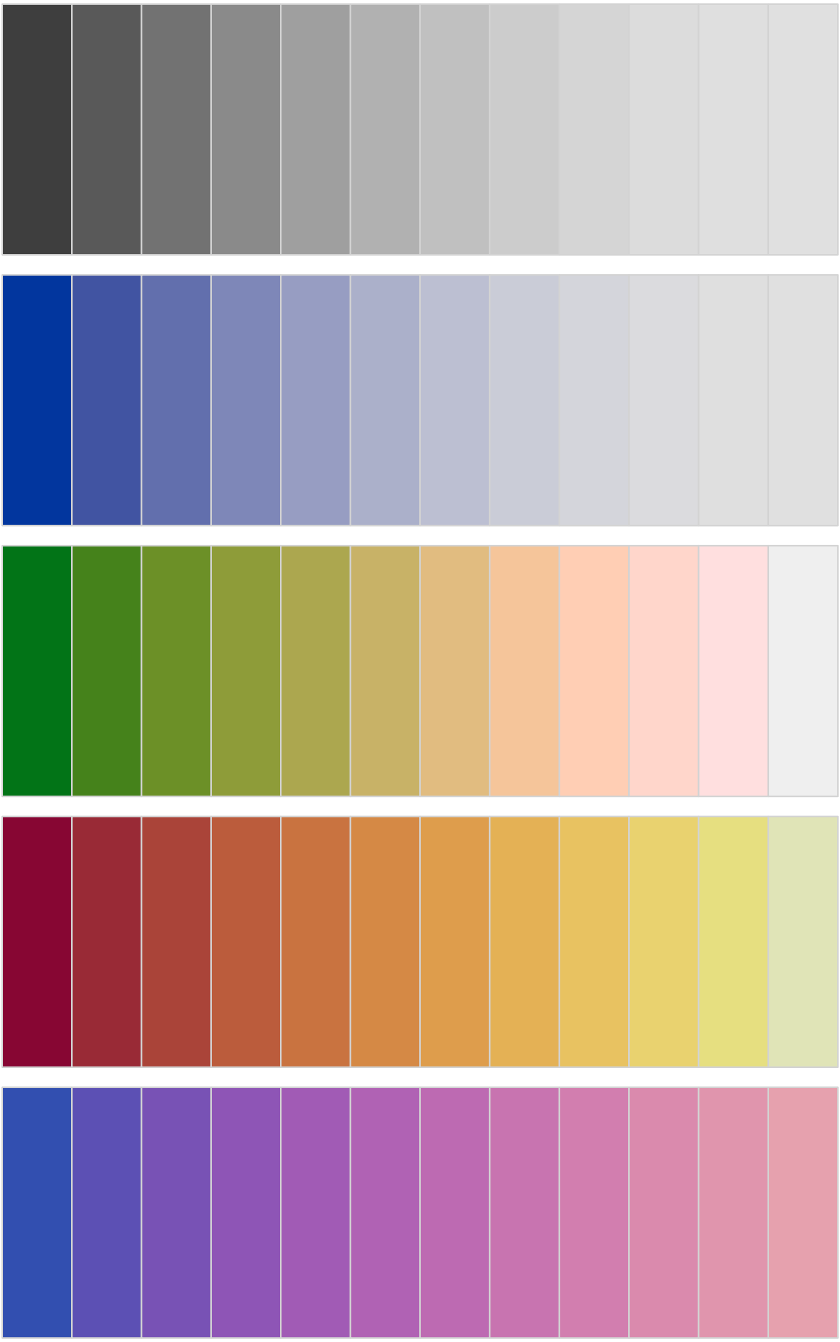

Figure 6: Examples for sequential palettes

A typical application would be heat colors that increase from a light yellow (e.g., $(90, 30, 90)$) to a full red (e.g., $(0, 100, 50)$). 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. This can be achieved by choosing a power $p_1 < 1$.

In R, these are available in the function

```
heat_hcl(n, h = c(0, 90), c = c(100, 30), l = c(50, 90), power = c(1/5, 1), ...)
```

with which the lower three rows in Figure 6 are produced.

5. Diverging palettes

Diverging palettes are also used for coding numerical information ranging in a certain interval—however, this interval includes a neutral value. Examples for this include residuals or correlations (both with the neutral value 0) or binary classification probabilities (with neutral value 0.5) that could be visualized in mosaic plots (see Figure 12) or classification maps (see Figure 13). Without loss of generality, we assume that we want to visualize an intensity or interestingness $i \in [-1, 1]$.

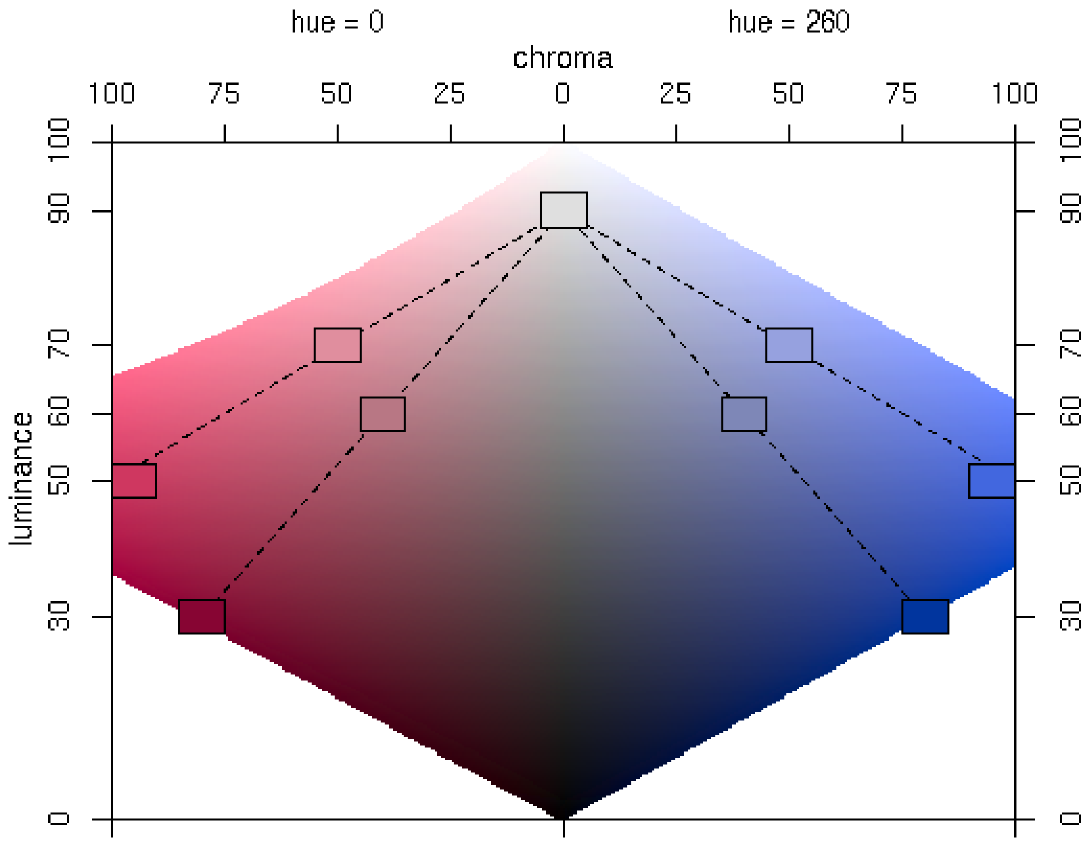

Figure 7: Constructing diverging palettes

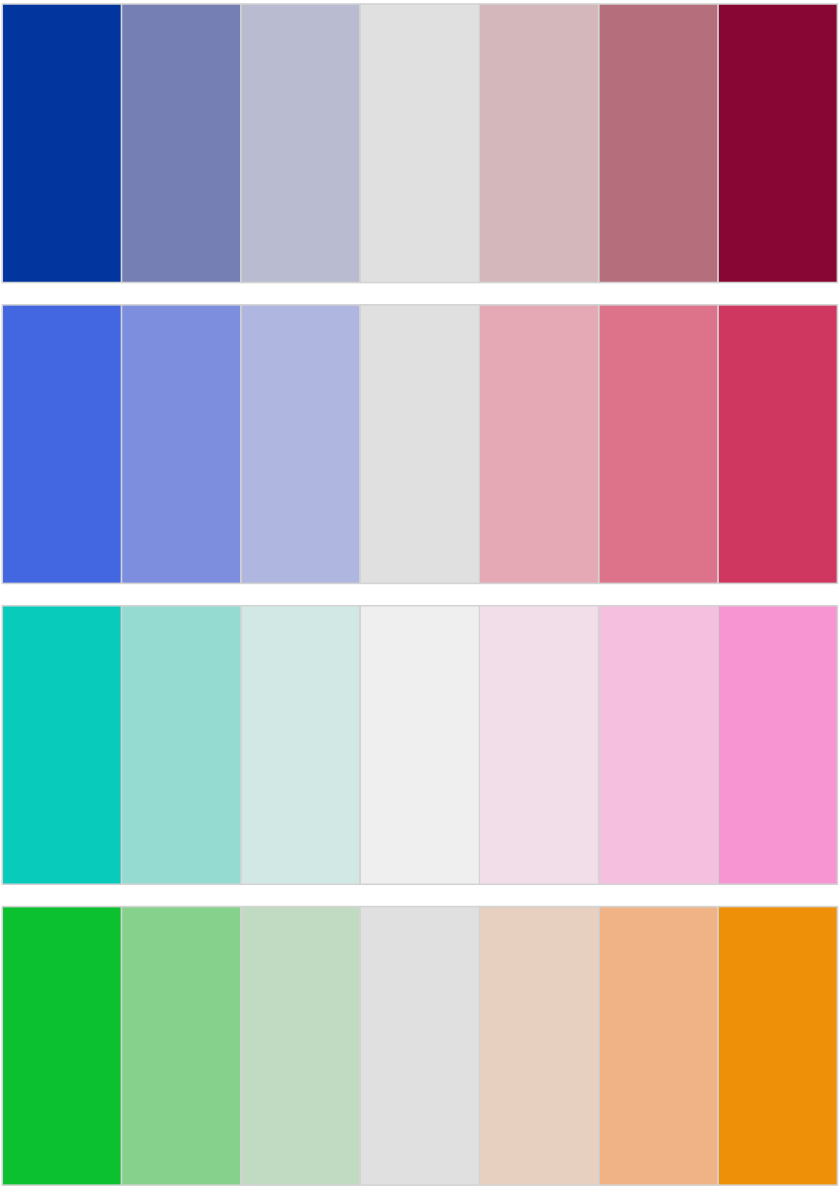

Figure 8: Examples for diverging palettes

Given sequential palettes, deriving diverging palettes is easy: two different hues are chosen for adding color to the same amount of 'gray' at a given intensity $|i|$. Figure 7 shows the chroma/luminance plane back to back for the hues $H = 0$ and 260 with two different paths—with slightly different emphasis on luminance or chroma contrasts—from a full red over a neutral grey to a full blue. These particular hues were chosen because they have rather similar chroma/luminance planes, allowing for many different combinations of maximal chroma and luminance. As Figure 5 illustrates, a bit of care is needed to choose two colors that are sufficiently similar in their chroma/luminance plane to allow for the same luminance and/or chroma contrasts.

In **vcd**, the following function is provided:

```
diverge_hcl(n, h = c(260, 0), c = 80, l = c(30, 90), power = 1.5, ...)
```

Figure 8 shows various examples of conceivable combinations of hue, chroma and luminance. The first palette uses a broader range on the luminance axis whereas the others use larger ranges on the chroma axis.

6. Illustrations

In this section, we show a few examples for the various types of palettes applied to statistical graphics. The first illustration is visualization of a bivariate density estimation for the Old Faithful geyser eruptions data. Figurer̃effig:bkde shows heatmaps of a bivariate kernel density estimate of waiting times between and duration of geyser eruptions in Yellowstone National Park. Both use a sequential palette as derived in Section 4 balanced towards the same gray levels with $L \in [30, 90]$ and $p_2 = 2$. The sequential palette in the left panel uses only gray colors (i.e., $C_{\max} = 0$) and the palette in the right panel additionally employs colors with $H \in [0, 90]$, $C \in [30, 80]$ and $p_1 = 1/5$.

To illustrate qualitative palettes, data from the 2005 election for the German parliament 'Bundestag' are employed. In that election, five parties were able to obtain enough votes to enter the Bundestag—the distribution of seats is depicted in a pie chart in Figure 10. The colors used are rough metaphors for the political parties, using a red hue $H = 0$ for the social democrats SPD, a blue hue $H = 240$ for the conservative CDU/CSU, a yellow hue $H = 60$ for the liberal FDP, a green hue $H = 120$ for the green party 'Die Gruenen' and a purple hue $H = 300$ for the leftist

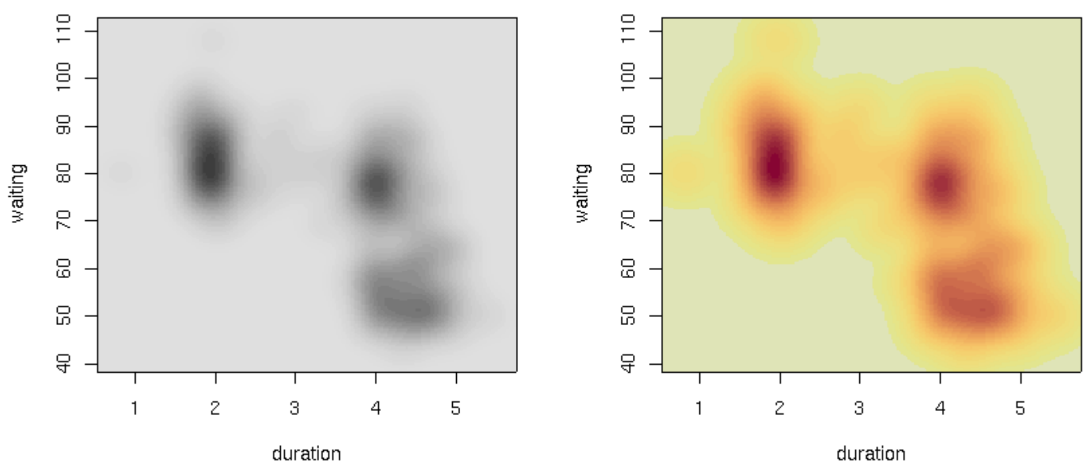

Figure 9: Bivariate density estimation for Old Faithful geyser eruptions

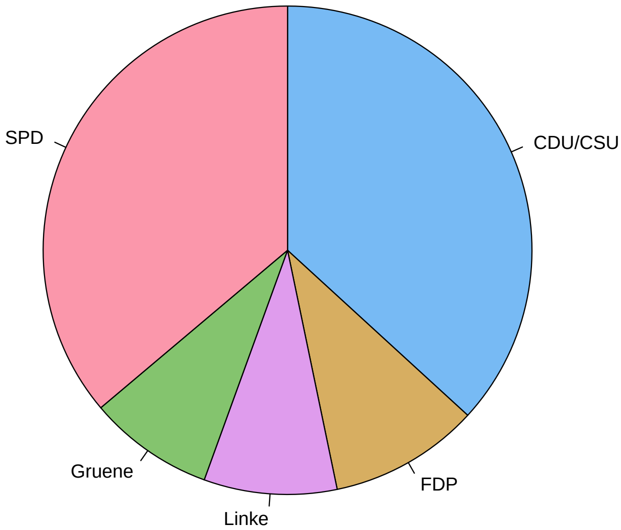

Figure 10: Seats in the German parliament

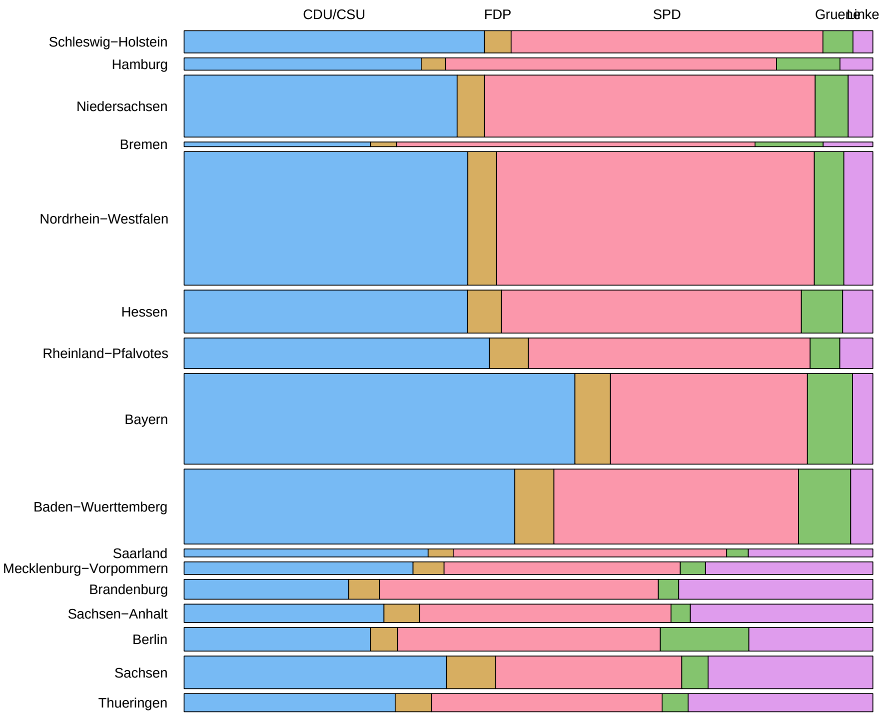

Figure 11: Votes in the German election 2005

party ‘Die Linke’. All colors use the same chroma $C = 60$ and luminance $L = 75$. The pie chart clearly shows that neither the governing coalition of SPD and Gruene nor the opposition of CDU and FDP could assemble a majority. Given that no party would enter a coalition with the leftists, this lead to a big coalition of CDU and SPD. Figure 11 shows the distribution of votes in this election stratified by province (Bundesland) in a highlighted mosaic display. The order of provinces is from north to south, first for the 10 Western provinces, then for the 6 Eastern provinces. Clearly, the SPD performed better in the North and the CDU better in the South; furthermore, the Die Linke performed particularly well in the Eastern provinces and the Saarland.

As pointed out in Section 5, diverging palettes are particularly useful when visualizing residuals or correlation (with natural neutral value 0) or probabilities in 2-class supervised learning (with neutral value 0.5). Examples for both situations are provided here. Figure 12 visualizes the outcome of a double-blind clinical trial investigating a new treatment for rheumatoid arthritis. The mosaic rectangles alone signal that the treatment lead to higher improvement compared to the placebo group which are shown to be significant by the shading based on the Pearson residuals. Positive residuals, corresponding to more observations in the corresponding cell than expected under independence, are depicted in blue, negative residuals in red. Light colors signal significance at 10% level, full colors significance at 1% level. Hence, it can be concluded that there are significantly more marked improvements in the treated group and significantly fewer in the placebo group than would be expected under independence between treatment and improvement. More details can be found in (Zeileis, Meyer, and Hornik 2005).

Figure 13 shows the fit of a support vector machine (SVM) to an artificial 2-class supervised learning example: a mixture of two bivariate normal distributions. The circles and triangles show the original observations, solid symbols correspond to the support vectors found. The shading underlying the plot visualizes the fitted decision values: values around 0 are on the decision boundary and are shaded in light gray, while regions that are firmly classified to one or the other class are shaded in full blue and red respectively.

7. Open questions

From our experience, the paths through HCL space described in the previous sections in principle work very well and can be used to effectively code qualitative and quantitative information. However, there are some open questions, formulated below. In particular, we are not sure how we can properly define trade-offs between increasing the range with respect to one dimension and decreasing the range with respect to another.

- In sequential/diverging palettes, there is a trade-off between high chroma and high luminance. How should this be chosen in practice? Our impression is that when a small set of colors (such as 3 or 4) are used, large differnces in chroma work well and large differences in luminance are not necessary. However, when a larger set of colors is used (e.g., for heatmaps where extreme values should be identifiable) it is much more important to have a big difference in luminance.
- How should the intensity $|i|$ be increased from 0 to 1? Our experience is that for a small set of colors linear increase is sufficient ($p = 1$) whereas in heatmaps where only very extreme regions are interesting a $p > 1$ should be used.
- For diverging palettes, how should pairs of colors be chosen? The hues $H = 0$ and $H = 260$ where chosen because they are on opposite sides of the color wheel but have a very similar chroma/luminance plane.
- To allow more convenient navigation in HCL space, some more R infrastructure could be helpful. We are not sure what would the best or most natural interface for this. A good start might be a function that computes the admissible value on one axis given a particular value on the other two axes. Maybe you have already experimented with something like this?

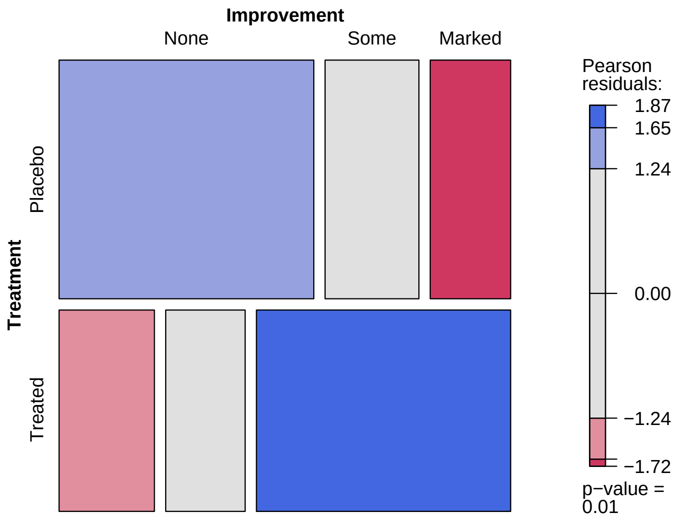

Figure 12: Extended mosaic display for arthritis data

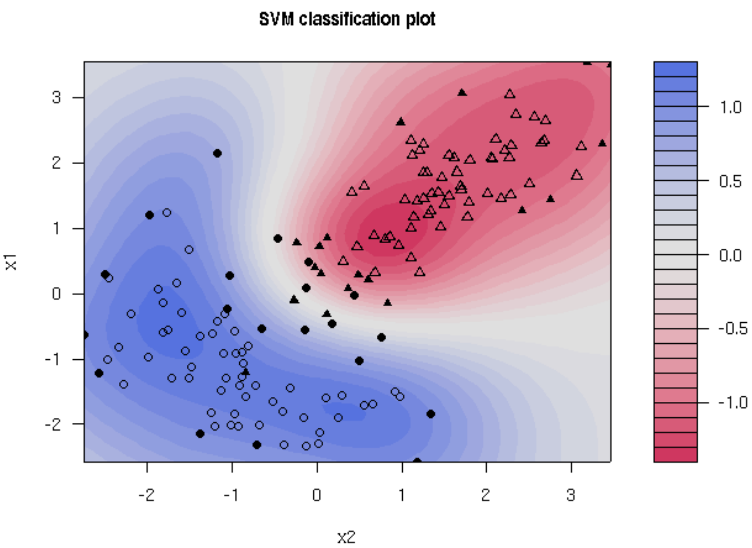

Figure 13: SVM classification plot

References

Brewer CA (1999). "Color Use Guidelines for Data Representation." In "Proceedings of the Section on Statistical Graphics, American Statistical Association," pp. 55–60. Alexandria, VA.

Commission Internationale de l'Éclairage (2004). *Colorimetry*. Publication CIE 15:2004, Vienna, Austria, 3rd edition. ISBN 3-901-90633-9.

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/.

Ihaka R (2004). **colorspace***: Colorspace Manipulation*. R package version 0.9.

Meyer D, Zeileis A, Hornik K (2006). **vcd***: Visualizing Categorical Data*. R package version 0.9-8.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org/.

Zeileis A, Meyer D, Hornik K (2005). "Residual-based Shadings for Visualizing (Conditional) Independence." *Report 20*, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Research Report Series. URL http://epub.wu-wien.ac.at/dyn/openURL?id=oai:epub.wu-wien.ac.at:epub-wu-01_871.