Hierarchical qualitative color palettes

Category: Research

ABSTRACT

Color is an important attribute to display categorical data in statistical graphics, such as bar charts and mosaic plots. Categorical data are often hierarchically structured, for instance according a classification tree. Qualitative color palettes that represent tree structures would be very useful, both for hierarchically and non-hierarchically structured visualizations.

We present a method to map colors from the Hue-Chroma-Luminance (HCL) color model to tree structures. The HCL color space, which is a transformation of the CIELUV color space, is known for its well balanced perceptual properties.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—User-centered design

1 INTRODUCTION

Hierarchical data are of crucial importance in daily official statistics. Publications levels are often hierarchically structured, for instance geographic region or economic activity. A recent trend in current data analysis practice, is to follow a top-down approach rather than to analyze input data (from surveys or administrative sources) on record level. Several data visualization methods are useful to explore hierarchical statistical data, for instance treemaps [5]. Color palettes that resemble hierarchical structures would be very useful in this context.

Assigning colors to categorical data points is far from trivial. On the one hand, colors should be easy to distinguish, but on the other hand they should not introduce perceptual bias. The selection of color palettes for categorical data first depends on the type of data. For nominal data, such as gender or nationality, qualitative color palettes are used, while for ordinal data, such as level of urbanization, sequential or diverging palettes are used [1, 7]. However, for categorical data with a hierarchical structure, there are no specific guidelines for selecting color palettes, to the best of our knowledge.

Although many tree visualizations are proposed in literature [3], most of them use color only to a small extent. A visualization technique that uses color as a major attribute is the InterRing [6], a navigation tool with a radial layout. The leaf nodes are assigned to a different hue values. The color of a parent node is derived from averaging the colors of its children, where larger branches have more weight. An implicit effect is that colors of higher hierarchical levels are less saturated, except for one-child-per-parent branches.

2 METHOD

The Hue-Chroma-Luminance (HCL) space, which is a transformation of the CIELUV color space, is designed with the aim to control human color perception [2]. Colors with different hue values are perceptually uniform in colorfulness and brightness, which does not hold for the popular Hue-Saturation-Value (HSV) color space [7]. We use the HCL space to select colors for hierarchical structures. The hue H takes values from 0 to 360, and the chroma C and luminance L take values from 0 to 100.

We describe out method using the European classification system of economic activity NACE. Section F (Construction) is depicted in Figure 1. The nodes are colored with the resulting hierarchical color palette.

For selecting hue values we use the following breadth-first algorithm. Starting with the root node, which is assigned to a predefined hue range (by default $[0, 360]$), the procedure per node v is:

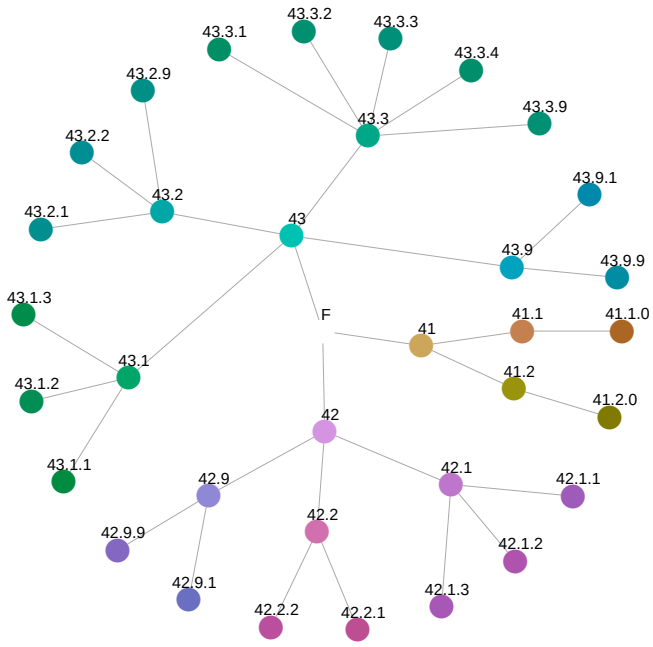


Figure 1: Tree of economic sector F of NACE.

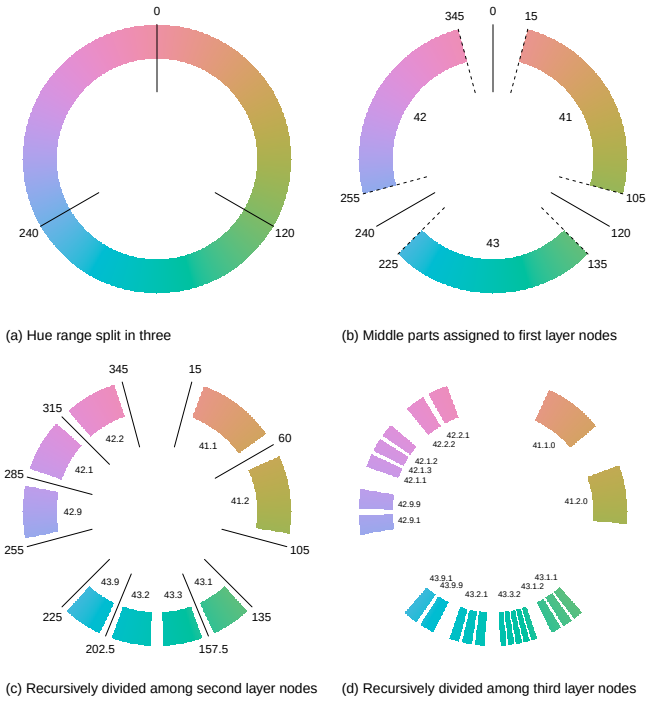
1. assign the middle hue value in the hue range to v (ignore coloring the root node itself);
2. if v has children:
   (a) divide the hue range equally among v's children;
   (b) keep the middle parts;
   (c) visit the v's children.

This division of the hue range is illustrated in Figure 2: in (a) the full hue range (for a constant $C = 60$ and $L = 70$) is divided among the three children of the root, in (b) the middle parts are kept, in (c) and (d) these steps are recursively taken for the deepest two hierarchical layers.

Ad 2(a) In most hierarchical structures, there is no ordinal order among siblings. When the nodes in such structure are plotted in a linear or radial layout, the colors of the siblings should not introduce a perceptual order. Therefore, the assigned hue ranges are permuted among the siblings. The used permutation order is based on the five-elements-permutation $[1, 3, 5, 2, 4]$. Furthermore, the permutation within even numbered branches is reversed to differentiate between branches.

Ad 2(b) The fraction of each hue range that is kept is by default set to 0.75. This choice is a trade-off between discriminating different main branches and discriminating different leaf nodes.

In order to show depth, we let C and L values only depend on the depth of the corresponding nodes. We let the L decrease linearly with depth and C increase, analog to ocean water colors at various depths.



(a) Hue range split in three

(b) Middle parts assigned to first layer nodes

(c) Recursively divided among second layer nodes

(d) Recursively divided among third layer nodes

Figure 2: Assignment of hue values.

3 Application

A treemap of a random variable assigned to the NACE tree of sector F is depicted in Figure 3. As a real application in official statistics, this variable could be turnover that is available for each business enterprise in a business register, and aggregated according to the NACE tree. The colored rectangles correspond to the deepest NACE layer, while the color of higher NACE layers is used to for the text label backgrounds. This treemap are created with the free and open source R package treemap [4].

As an example of a non-hierarchically-structured plot, a bar chart of random data is depicted in 4. Such graphics could be useful when the hierarchical structure will not be the main focus in the conducted analyses.

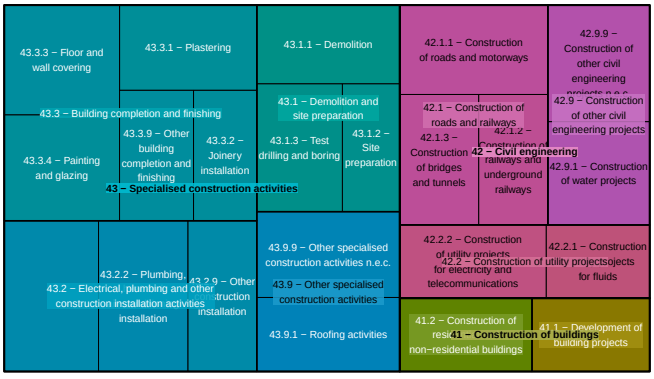


Figure 3: Treemap with hierarchical colors

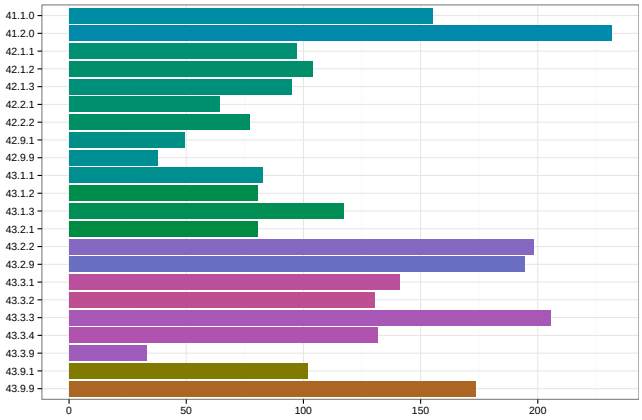


Figure 4: Bar chart with hierarchical colors

4 Further research

We recommend to further investigate the proposed hierarchical color selection method, and to evaluate the obtained color palettes in various statistical graphics.

References

[1] C. A. Brewer, G. W. Hatchard, and M. A. Harrower. Colorbrewer in print: A catalog of color schemes for maps. *Cartography and Geographic Information Science*, 30(1):5–32, 2003.

[2] R. Ihaka. Colour for presentation graphics. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna Austria*, 2003.

[3] H.-J. Schulz. Treevis.net: A tree visualization reference. *IEEE Comput. Graph. Appl.*, 31(6):11–15, Nov. 2011.

[4] M. Tennekes. *treemap: Treemap visualization*, 2013. R package version 2.0.1.

[5] M. Tennekes and E. de Jonge. Top-down data analysis with treemaps. In *Proceedings of the International Conference on Information Visualization Theory and Applications, IVAPP 2011*, 2011.

[6] J. Yang, M. O. Ward, and E. A. Rundensteiner. Interring: An interactive tool for visually navigating and manipulating hierarchical structures. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis'02)*, INFOVIS '02, pages 77–, 2002.

[7] A. Zeileis, K. Hornik, and P. Murrell. Escaping rgbland: Selecting colors for statistical graphics. *Comput. Stat. Data Anal.*, 53(9):3259–3270, July 2009.