

aheatmap: a Powerful Annotated Heatmap Engine

Package *NMF* - Version 0.21

Renaud Gaujoux

March 26, 2014

Abstract

This vignette showcases the main features of the annotated heatmap engine implemented by the function `aheatmap`. This engine is a highly enhanced modification of the function `pheatmap` from the *pheatmap* package¹, and provides convenient and quick ways of producing high quality and customizable annotated heatmaps. Currently this function is part of the package *NMF*, but will most probably eventually constitute a separate package on its own.

Contents

1 Overview	1	7 Colours	3
2 Preliminaries	2	8 Labels	3
2.1 Installation	2		
2.2 Sample data	2	9 Legends	3
3 Components	2	9.1 Colour scale	3
4 Annotation tracks	3	9.1.1 Colours and breaks	4
5 Dendrograms	3	9.1.2 Look and position	4
6 Column/row ordering	3	9.2 Annotations	4
		10 Session Info	4

1 Overview

The development of the function `aheatmap` started as modification of the function `pheatmap` from the *pheatmap* package². The initial objective was to improve and increase its capabilities, as well as defining a simplified interface that was more consistent with the R core function `heatmap`. It is evolving into a general, flexible, powerful and easy to use engine for drawing annotated heatmaps.

The function `aheatmap` has many advantages compared to other heatmap functions such as `heatmap`, `gplots::heatmap2`, `heatmap.plus::heatmap.plus`, or `pheatmap`:

- Annotations: unlimited number of annotation tracks can be added to *both* columns and rows, with automated colouring for categorical and numeric variables.
- Compatibility with both base and grid graphics: the function can be directly called in drawing contexts such as `grid`, `mflow` or `layout`. We believe that this is a feature many R users will enjoy, and that is strictly impossible with base heatmaps.
- Legends: default automatic legend and colouring;

¹<http://cran.r-project.org/package=pheatmap>

²<http://cran.r-project.org/package=pheatmap>

- Customisation: clustering methods, annotations, colours and legend can all be customised, even separately for rows and columns;
- Convenient interface: many arguments provide multiple ways of specifying their value(s), which speeds up developping/writing and reduce the amount of code required to generate customised plots (e.g. see ??).
- Aesthetics: the heatmaps look globally cleaner, the image and text components are by default well proportioned relatively to each other, and all fit within the graphic device – if not set to an unreasonably small size.

2 Preliminaries

2.1 Installation

The `aheatmap` function is currently part of the *NMF* package³, which can be installed from any CRAN mirror or from the GitHub repository⁴, for the development version, with the following commands:

```
# latest stable
install.packages("NMF")
# development version
devtools::install_github("NMF", "renozao", "devel")
```

2.2 Sample data

For the purpose of illustrating the capabilities of the function `aheatmap`, we first generate some random data that we will use throughout the vignette:

```
# data matrix
x <- rmatrix(20, 10)
```

3 Components

Annotated heatmaps essentially use `grid` graphics⁵, composing the global picture by putting together the following components (or viewports in `grid` language):

dendrograms clusters and order columns/rows;

annotations are additional *tracks* that provide extra information about each column/row according to some associated auxiliary data;

data matrix, i.e. the heatmap itself, shown as coloured cells;

labels associates each column/row with some textual information;

legends such as value scales or color code used for the data matrix or annotations;

other information like main title, sub-title, extra information pane.

Figure 1 shows a diagram of two possible grid layout that combined the above listed components into a complete annotated heatmap.

³<http://cran.r-project.org/package=NMF>

⁴<http://github.com/renozao/NMF>

⁵Except for drawing dendrograms, which are plotted using the proven and well optimised base function `plot.dendrogram`.



Figure 1: Grid layout diagram of annotated heatmaps: (left) default layout and (right) an alternative layout, with separate specification for rows and columns – passed as a single string.

4 Annotation tracks

5 Dendrograms

6 Column/row ordering

7 Colours

8 Labels

9 Legends

Annotated heatmaps have two types of legends, one showing the colour-value scale used to visualise the data matrix and another one for the annotation tracks.

9.1 Colour scale

The very principle of a heatmap is to bin the data values into a certain number of intervals or breaks, and associate each of these with a given colour. The colour scale is the legend that provides details about how to read the resulting colour coded data matrix. As such, it serves multiple purposes:

- provide the mapping between colours and value intervals;
- show the actual range of displayed values;
- optionnaly show the overall distribution of values.

9.1.1 Colours and breaks

9.1.2 Look and position

As for other components in annotated heatmaps, the position of the colour scale is controlled by the argument `layout`, which can also be used to specify if the scale should be expanded over the full height/width or have a fixed size.

By default the scale is placed on the top-right corner of the data matrix, with a fixed size. Figure 2 illustrates how to obtain some other commonly used positions/look. However, more options are available, as detailed in the manual page for `aheatmap.layout`.

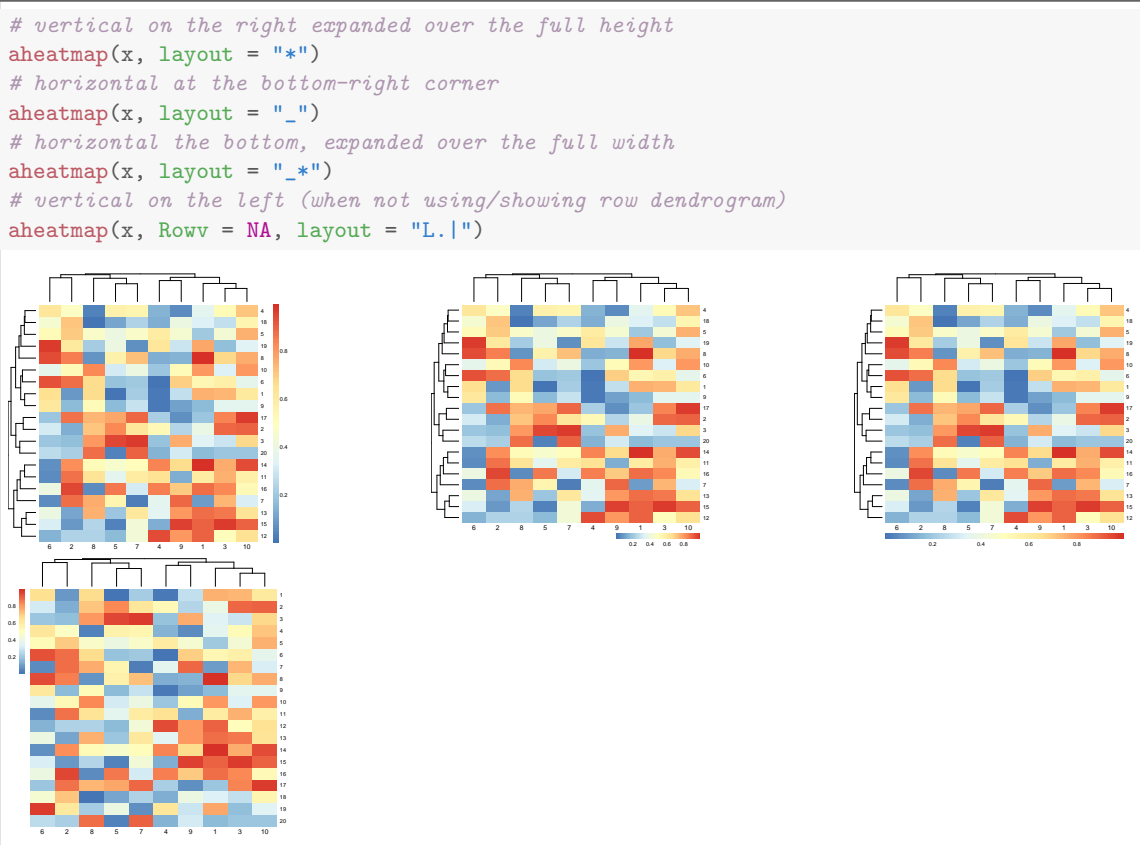


Figure 2: Colour scale alternative layouts: the scale can be placed in different areas around the data matrix and expanded to full height/width.

9.2 Annotations

10 Session Info

- R version 3.0.3 (2014-03-06), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=en_US.UTF-8, LC_NUMERIC=C, LC_TIME=en_US.UTF-8, LC_COLLATE=en_US.UTF-8, LC_MONETARY=en_US.UTF-8, LC_MESSAGES=en_US.UTF-8, LC_PAPER=en_US.UTF-8, LC_NAME=C, LC_ADDRESS=C, LC_TELEPHONE=C, LC_MEASUREMENT=en_US.UTF-8, LC_IDENTIFICATION=C
- Base packages: base, datasets, graphics, grDevices, methods, parallel, stats, utils

- Other packages: BH 1.51.0-4, bigmemory 4.4.6, bigmemory.sri 0.1.2, Biobase 2.22.0, BiocGenerics 0.8.0, cluster 1.15.1, knitr 1.5, NMF 0.21, pkgmaker 0.21, RColorBrewer 1.0-5, registry 0.2, rngtools 1.2.4, synchronicity 1.1.2
- Loaded via a namespace (and not attached): codetools 0.2-8, colorspace 1.2-4, dichromat 2.0-0, digest 0.6.4, doParallel 1.0.8, evaluate 0.5.1, foreach 1.4.1, formatR 0.10, ggplot2 0.9.3.1, grid 3.0.3, gridBase 0.4-7, gtable 0.1.2, highr 0.3, iterators 1.0.6, labeling 0.2, MASS 7.3-30, munsell 0.4.2, plyr 1.8.1, proto 0.3-10, Rcpp 0.11.1, reshape2 1.2.2, scales 0.2.3, stringr 0.6.2, tools 3.0.3, xtable 1.7-3