tidyHeatmap: a modular R package for heatmap visualisation based on tidy principles

07 July 2020

Background

The heatmap is a powerful tool for visualising multi-dimensional data, where individual values can be organised in a two-dimensional matrix and their values expressed as colours. Row and columns can be ordered according to their reciprocal similarity using hierarchical clustering; and dendrograms can be added to the plot to facilitate the interpretation. Row- and column-wise visual annotations, such as coloured tiles, can also be included. Within the R environment, several packages have been produced to produce heatmaps. The simplest and most readily available tool, `heatmap`, is provided within the `stats` package (R Core Team 2013) and offers basic heatmaps with no annotations. The versatile package `ggplot2` can be used to produce basic heatmaps, but with fine aesthetical adjustable parameters (Wickham 2016). More powerful software exists for producing fully annotated, multi-panels heatmaps, such as `Pheatmap` (Kolde 2012) and `ComplexHeatmap` (Gu, Eils, and Schlesner 2016). The versatility of these packages comes at the cost of adding complexity in the user interface, characterised by many parameters and annotation functions that introduce a steep learning curve to produce complex, clear, and good-looking graphics.

Recently, efforts have been made toward the harmonisation of data frame structures and data analysis workflows using the concept of tidiness. Tidy data frames allow ease of manipulation, modelling, and visualisation; and are characterised by having a specific structure where each variable is a column and each observation is a row. The `tidyverse` is a suite of R libraries that defined the standard for tidy data and APIs (Wickham et al. 2019). The unique correspondence between quantities and annotations, characteristic of tidy data frames, allows complex operations to be performed from simple user inputs, such as a list of column names.

Tidy paradigm for visualisation

`tidyHeatmap` is a graphical R package that introduces tidy principles to the creation of information-rich heatmaps. It is available in the CRAN R repository. This package uses `ComplexHeatmap` as its graphical engine. The command-line user interface is organised in (i) the main plotting utility; (ii) the annotation layer utilities; and (iii) the file output utilities. The input data frame streams along the utility path using the pipe operator from `magrittr`, allowing a high-level of modularity. The main utility allows the user to plot a base heatmap with dendrograms. The annotation utilities allow the user to serially add tile, point, bar and/or line annotation boxes to the side on the heatmap. The orientation of the annotations (row- or column-wise) is inferred by the `tidyHeatmap` algorithms, based on the input data frame. The file output utility allows the user to write vector or bitmap images directly from the R object, in the style of `ggplot2`. User defined row- or column-wise clusters can be defined effortlessly by applying the `group_by` function from `dplyr` (Wickham et al. 2020) to the input data frame. Besides offering a modular and user-friendly interface, `tidyHeatmap` provides publication ready aesthetics such as `viridis` (Garnier 2018) and `brewer` (Neuwirth 2014) colour palettes and automatic sizing of row and column labels to avoid overlapping.

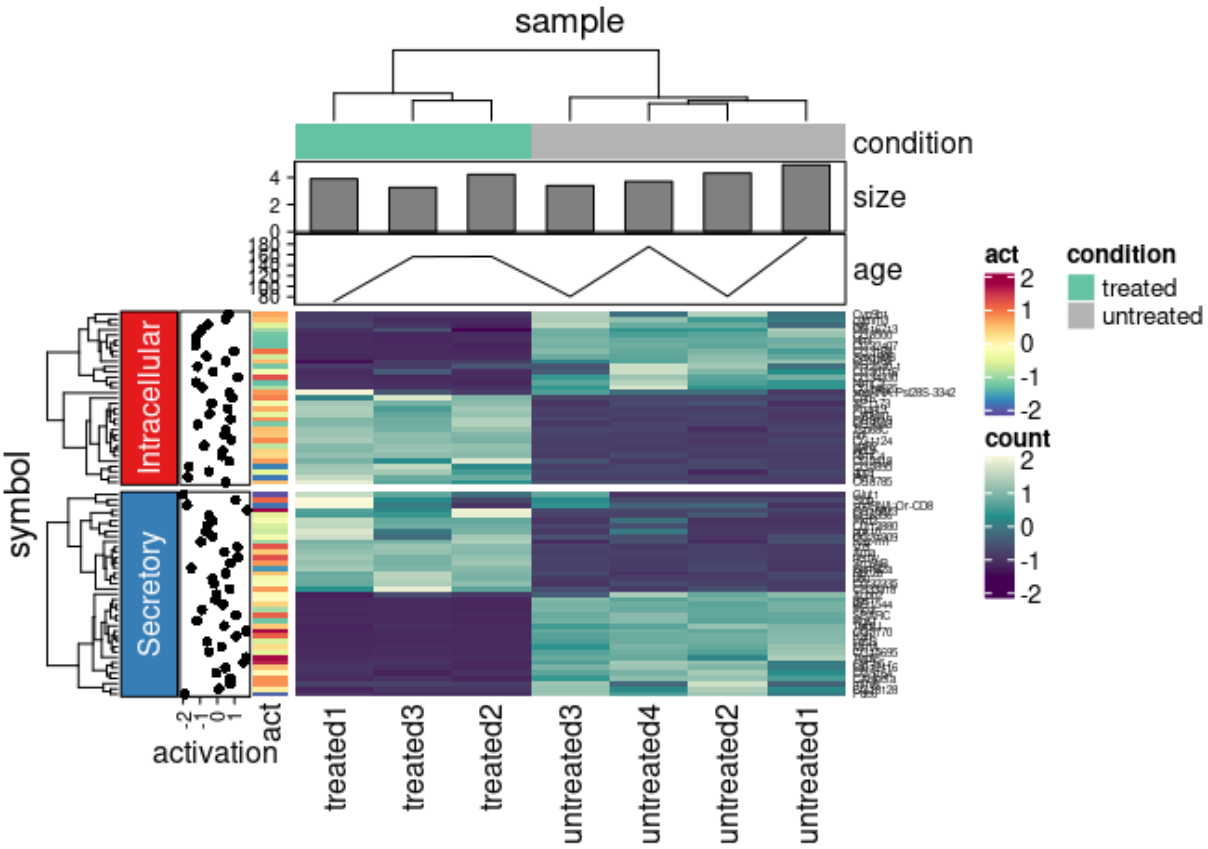

Figure 1: Heatmap of the pasilla dataset including grouping and multiple annotations. Some annotation data was simulated for visualisation purposes.

The input is a tidy data frame with the three basic columns including row and column element identifiers, and values that will be converted into colours. Additionally, further columns can include information about element grouping and annotation.

element	feature	value	annotation	group
`chr` or `fctr`	`chr` or `fctr`	`numeric`	...	...

The code interface consists of modular functions linked through the pipe operator (Figure 1). Custom colour palette can be used passing an array of colours of a colour function (e.g., circlize (Gu et al. 2014)) to the palette argument of the annotation utilities.

```
my_heatmap =

    # Grouping
    input_df %>%
    group_by(location) %>%

    # Plotting
    heatmap(feature, element, value) %>%

    # Annotation
    add_tile(condition) %>%
    add_tile(act) %>%
    add_point(activation) %>%
    add_bar(size) %>%
    add_line(age)

# Saving
my_heatmap %>% save_pdf("my_file.pdf")
```

Conclusions

The use of disjointed data structures demands time consuming and bug-prone information matching for performing complex tasks. Joint, tidy data frames allow to lower the cost/benefit ratio for the user, automatic a large part on data manipulation. `tidyHeatmap` introduces a modular paradigm for specifying information-rich heatmaps, just requiring column names as input. Due to its intuitive user interface and its advanced default aesthetic features, `tidyHeatmap` is ideal for the quick production of publication-ready heatmaps. This software is designed for modular expandability.

Acknowledgements

We acknowledge contributions all the Papenfuss Lab for feedback, and Maria Doyle for constant support.

References

Garnier, Simon. 2018. *Viridis: Default Color Maps from 'Matplotlib'*. https://CRAN.R-project.org/package=viridis.

Gu, Zuguang, Roland Eils, and Matthias Schlesner. 2016. "Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data." *Bioinformatics* 32 (18): 2847–9.

Gu, Zuguang, Lei Gu, Roland Eils, Matthias Schlesner, and Benedikt Brors. 2014. "Circlize Implements and Enhances Circular Visualization in R." *Bioinformatics* 30 (19): 2811–2.

Kolde, R. 2012. "Pheatmap: Pretty Heatmaps." *R Package Version*.

Neuwirth, Erich. 2014. *RColorBrewer: ColorBrewer Palettes*. https://CRAN.R-project.org/package=RColorBrewer.

R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.