Package: superheat (via r-universe)

September 12, 2024

Type Package

Title A Graphical Tool for Exploring Complex Datasets Using Heatmaps

Version 1.0.0

Description A system for generating extendable and customizable heatmaps for exploring complex datasets, including big data and data with multiple data types.

Depends R (>= 3.1)

License CC0

LazyData TRUE

Imports dplyr (>= 0.4.3), ggplot2 (>= 2.0.0), gtable (>= 0.1.2), magrittr (>= 1.5), plyr (>= 1.8.3), scales (>= 0.3.0), stringr (>= 1.2.0), ggdendro, forcats,

Suggests knitr (>= 1.11), RColorBrewer, rmarkdown, testthat (>= 0.11.0),

RoxygenNote 6.0.1

Repository https://rlbarter.r-universe.dev

RemoteUrl https://github.com/rlbarter/superheat

RemoteRef HEAD

RemoteSha 2c619578c2ae8bc46e1383183276f98b66b44a53

Contents

superheat *Generate supervised heatmaps.*

Description

Superheat is used to generate and customize heatmaps. Scatterplots, boxplots, barplots, line plots and boxplots can be plotted adjacent to the columns and rows of the heatmap, adding an additional layer of information. For usage, see the vignette at https://rlbarter.github.io/superheat/.

Usage

```
superheat(X, X.text = NULL, yt = NULL, yr = NULL,
  membership.rows = NULL, membership.cols = NULL, pretty.order.rows = F,
  pretty.order.cols = F, row.dendrogram = F, col.dendrogram = F,
  n.clusters.rows = NULL, n.clusters.cols = NULL,
  clustering.method = c("kmeans", "hierarchical"),
  dist.method = c("euclidean", "maximum", "manhattan", "canberra", "binary",
  "minkowski"), linkage.method = c("complete", "ward.D", "ward.D2", "single",
  "average", "mcquitty", "median", "centroid"), order.cols = NULL,
  order.rows = NULL, smooth.heat = FALSE, scale = FALSE,
  left.label = NULL, bottom.label = NULL, heat.col.scheme = c("viridis",
  "red", "purple", "blue", "grey", "green"), heat.pal = NULL,
  heat.pal.values = NULL, heat.na.col = "grey50", heat.lim = NULL,
  extreme.values.na = TRUE, X.text.size = 5, X.text.col = "black",
  X.text.angle = 0, yt.plot.type = c("scatter", "bar", "boxplot",
  "scattersmooth", "smooth", "scatterline", "line"),
  yr.plot.type = c("scatter", "bar", "boxplot", "scattersmooth", "smooth",
  "scatterline", "line"), legend = TRUE, legend.height = 0.1,
  legend.width = 1.5, legend.text.size = 12, legend.num.ticks = 4,
  legend.breaks = NULL, legend.vspace = 0.1, grid.hline = TRUE,
  grid.vline = TRUE, grid.hline.size = 0.5, grid.vline.size = 0.5,
  grid.hline.col = "black", grid.vline.col = "black",
  force.grid.hline = F, force.grid.vline = F,
  smoothing.method = c("loess", "lm"), smooth.se = TRUE, yt.axis = T,
  yr.axis = T, yt.num.ticks = 3, yr.num.ticks = 3, yt.plot.size = 0.3,
  yr.plot.size = 0.3, yt.axis.name = NULL, yr.axis.name = NULL,
  yr.axis.size = 10, yt.axis.size = 10, yr.axis.name.size = 10,
  yt.axis.name.size = 10, yr.axis.name.angle = NULL,
  yt.axis.name.angle = NULL, yt.obs.col = NULL, yr.obs.col = NULL,
  yt.cluster.col = NULL, yr.cluster.col = NULL, yt.bar.col = NULL,
  yr.bar.col = NULL, yt.point.size = 2, yt.point.alpha = 1,
  yr.point.size = 2, yr.point.alpha = 1, yr.line.col = NULL,
  yt.line.col = NULL, yr.line.size = NULL, yt.line.size = NULL,
  yr.lim = NULL, yt.lim = NULL, yr.breaks = NULL, yt.breaks = NULL,
  yr.break.labels = NULL, yt.break.labels = NULL,
  bottom.label.text.size = 5, left.label.text.size = 5,
  bottom.label.text.angle = NULL, left.label.text.angle = NULL,
```

```
bottom.label.size = NULL, left.label.size = NULL, left.label.col = NULL,
bottom.label.col = NULL, left.label.text.col = NULL,
bottom.label.text.col = NULL, left.label.text.alignment = NULL,
bottom.label.text.alignment = NULL, force.left.label = F,
force.bottom.label = F, column.title = NULL, row.title = NULL,
column.title.size = 5, row.title.size = 5, padding = 1, title = NULL,
title.size = 5, title.alignment = NULL, print.plot = TRUE)
```

Arguments

`X`	a matrix whose values are to be plotted in the heatmap.
`X.text`	a matrix containing text entries to be plotted on top of the heatmap cells. The number of rows/columns must match either the number of rows/columns of `X` or the number of row/column clusters of `X`.
`yt`	a data frame whose columns consist of values to plot above the heatmap (the "top plot"). The length of `yt` must be equal to the number of columns of `X`.
`yr`	a data frame whose columns consist of values to plot to the right of the heatmap (the "right plot"). The length of `yr` must be equal to the number of rows of `X`.
`membership.rows`	a vector specifying the cluster membership of the rows in X.
`membership.cols`	a vector specifying the cluster membership of the columns in X.
`pretty.order.rows`	a logical specifying whether the cols should be reordered based on hierarchical clustering. Default is FALSE.
`pretty.order.cols`	a logical specifying whether the rows should be reordered based on hierarchical clustering. Default is FALSE.
`row.dendrogram`	a logical specifying whether a dendrogram should be placed next to the rows. Can only be used when `yr` is not specified and clustering is not performed.
`col.dendrogram`	a logical specifying whether a dendrogram should be placed next to the columns. Can only be used when `yt` is not specified and clustering is not performed.
`n.clusters.rows`	a number specifying the number of row clusters to generate. The default is 0 (indicating no clustering of the rows). This argument is ignored if `membership.rows` is provided.
`n.clusters.cols`	a number specifying the number of column clusters to generate. The default is 0 (indicating no clustering of the columns). This argument is ignored if `membership.columns` is provided.
`clustering.method`	the clustering method to use whenever `n.clusters.cols` or `n.clusters.rows` is specified. The default ("kmeans") is to use K-means clustering, the alternative option ("hierarchical") performs hierarchical clustering. Another (suggested) alternative is to provide a row and/or column membership vector.

`dist.method` the distance method to use for hierarchical clustering. This must be one of "euclidean", "maximum", "manhattan", "canberra", "binary" or "minkowski".

`linkage.method` the linkage method to use for hierarchical clustering. This must be one of "ward.D", "ward.D2", "single", "complete", "average" (= UPGMA), "mcquitty" (= WPGMA), "median" (= WPGMC) or "centroid" (= UPGMC).

`order.cols` a vector of specifying the ordering of the columns of `X`. If the columns are clustered, then this vector specifies the order within the clusters. Note that this vector must be a rearranged `1:ncol(X)` vector which specifies the new location of each column.

`order.rows` a vector of specifying the ordering of the rows of `X`. If the rows are clustered, then this vector specifies the order within the clusters. Note that this vector must be a rearranged `1:nrow(X)` vector which specifies the new location of each row.

`smooth.heat` a logical specifying whether or not to smooth the colour of the heatmap within clusters (by taking the median value).

`scale` a logical specifying whether or not to center and scale the columns of X.

`left.label` a character specifying the type of the label provided to the left of the heatmap. If clustering was performed on the rows, then the default type is "cluster" (which provides the cluster names). Otherwise, the default is "variable" (which provides the variable names). The final option, "none", removes the left labels all together.

`bottom.label` a character specifying the type of the label provided to the left of the heatmap. If clustering was performed on the columns, then the default type is "cluster" (which provides the cluster names). Otherwise, the default is "variable" (which provides the variable names). The final option, "none", removes the label all together.

`heat.col.scheme`
A character specifying the heatmap colour scheme. The default is "viridis", and other options include "red", purple", "blue", "grey" and "green". If you wish to supply your own colour scheme, use the `heat.pal` argument.

`heat.pal` a vector of colour names specifying a manual heatmap colour palette. This corresponds to the `colour` argument for the ggplot2 `scale_colour_gradientn` function.

`heat.pal.values`
a vector specifying the location of each colour in the colour palette specified by `heat.pal`. Each entry should be a number between 0 and 1. This corresponds to the `values` argument for the ggplot2 `scale_colour_gradientn` function. The default values are the corresponding quantiles.

`heat.na.col` the color for NA values in the heatmap.

`heat.lim` a vector of length two consisting of the maximum and minimum value for the heatmap palette.

`extreme.values.na`
a logical describing whether values outside the range of heat.lim are presented as missing (TRUE, default) or as the max/min value of the range.

`X.text.size` a single number or a matrix of numbers (whose dimension matches that of `X.text`) that specifies the size of each text entry in `X.text`.

`X.text.col` a single character string or a matrix of character strings (whose dimension matches that of `X.text`) that specifies the colours of each text entry in `X.text`.

`X.text.angle` a single number or a matrix of numbers (whose dimension matches that of `X.text`) that specifies the angle of each text entry in `X.text`.

`yt.plot.type` a character specifying the `yt` plot type. The default is "scatter", and other options include "bar", "scattersmooth", "smooth", "boxplot", "scatterline" and "line".

`yr.plot.type` character specifying the `yr` plot type. The default is "scatter", and other options include "bar", "scattersmooth", "smooth", "boxplot", "scatterline", and "line".

`legend` logical. If set to `FALSE`, then no legend is provided.

`legend.height` a number specifying the height of the legend. The default is 0.1.

`legend.width` a number specifying the width of the legend. The default is 1.5.

`legend.text.size` a number specifying the size of the numbers on the legend axis. The default is 12.

`legend.num.ticks` a vector specifying the desired number of legend breaks (superheat may or may not actually give this number of breaks).

`legend.breaks` a vector specifying the actual legend breaks.

`legend.vspace` a number specifying the vertical gap between the heatmap and the legend

`grid.hline` a logical specifying whether horizontal grid lines are plotted in the heatmap.

`grid.vline` a logical specifying whether vertical grid lines are plotted in the heatmap.

`grid.hline.size` the thickness of the horizontal grid lines. The default is 0.5.

`grid.vline.size` the thickness of the vertical grid lines. The default is 0.5.

`grid.hline.col` the colour of the horizontal grid lines.

`grid.vline.col` the colour of the vertical grid lines.

`force.grid.hline` a logical describing whether or not to force the horizontal grid lines to appear (relevant only when X has more than 50 rows). Note that by default there are no horizontal grid lines when there are more than 50 rows.

`force.grid.vline` a logical describing whether or not to force the vertical grid lines to appear (relevant only when X has more than 50 columns). Note that by default there are no vertical grid lines when there are more than 50 columns.

`smoothing.method` if `plot.type = "scattersmooth"` or `"smooth"`, this argument specifies the smoothing method to use. The default is "loess" for a curve. The alternative option is "lm" for a line.

`smooth.se` a logical specifying whether the smoothed `yt` and `yr` curves have standard error curves.

`yt.axis` a logical specifying the presence of an axis for the `yt` plot.

`yr.axis` a logical specifying the presence of an axis for the `yr` plot.

`yt.num.ticks` the number of ticks on the `yt` axis. This does not always work perfectly as it is coerced into looking pretty.

`yr.num.ticks` the number of ticks on the `yr` axis. This does not always work perfectly as it is coerced into looking pretty.

`yt.plot.size` a number specifying the size of the `yt` plot.

`yr.plot.size` a number specifying the size of the `yr` plot.

`yt.axis.name` a character specifying the `yt` axis name.

`yr.axis.name` a character specifying the `yr` axis name.

`yr.axis.size` a number specifying the size of the numbers on the axis.

`yt.axis.size` a number specifying the size of the numbers on the axis.

`yr.axis.name.size`
a number specifying the size of the axis name.

`yt.axis.name.size`
a number specifying the size of the axis name.

`yr.axis.name.angle`
a number specifying the angle of the axis name.

`yt.axis.name.angle`
a number specifying the angle of the axis name.

`yt.obs.col` a vector specifying the colour of individual points in the `yt` plot.

`yr.obs.col` a vector specifying the colour of individual points in the `yr` plot.

`yt.cluster.col` a vector the same length as the number of clusters which specifies the colour of each cluster in `yt`.

`yr.cluster.col` a vector the same length as the number of clusters which specifies the colour of each cluster in `yr`.

`yt.bar.col` a character which specifies the colour of the boundary of the bars in the barplot of `yt`.

`yr.bar.col` a character which specifies the colour of the boundary of the bars in the barplot of `yr`.

`yt.point.size` the size of the points in the `yt` scatterplot. The default is 2.

`yt.point.alpha` the transparency of the points in the `yt` scatterplot. The default is 1, which corresponds to no transparency.

`yr.point.size` the size of the points in the `yr` scatterplot. The default is 2.

`yr.point.alpha` the transparency of the points in the `yr` scatterplot. The default is 1, which corresponds to no transparency.

`yr.line.col` the color of the (smoothing) line in the `yr` plot.

`yt.line.col` the color of the (smoothing) line in the `yt` plot.

`yr.line.size` the thickness of the (smoothing) line in the `yr` plot.

`yt.line.size` the thickness of the (smoothing) line in the `yt` plot.

`yr.lim` a vector of length two describing the y-axis limits.

`yt.lim` a vector of length two describing the y-axis limits.

`yr.breaks` a vector describing the position of the y-axis breaks.

`yt.breaks` a vector describing the position of the y-axis breaks.

`yr.break.labels`
: a vector describing the printed labels of the y-axis breaks.

`yt.break.labels`
: a vector describing the printed labels of the y-axis breaks.

`bottom.label.text.size`
: the size of the bottom heatmap label text. The default is 5.

`left.label.text.size`
: the size of the left heatmap label text. The default is 5.

`bottom.label.text.angle`
: number of degrees to rotate the text on the bottom cluster/variable labels. The default is 90.

`left.label.text.angle`
: number of degrees to rotate the text on the left cluster/variable labels. The default is 0.

`bottom.label.size`
: a number specifying the size of the bottom cluster/variable label panel.

`left.label.size`
: a number specifying the size of the left cluster/variable label panel.

`left.label.col` a vector specifying the left cluster/variable label colour palette.

`bottom.label.col`
: a vector specifying the bottom cluster/variable label colour palette.

`left.label.text.col`
: a character or character vector specifying the left cluster/variable label text colour.

`bottom.label.text.col`
: a character or character vector specifying the bottom cluster/variable label text colour.

`left.label.text.alignment`
: the text alignment of the label text. The default is "center". Alternate options are "left" and "right".

`bottom.label.text.alignment`
: the text alignment of the label text. The default is "center". Alternate options are "left" and "right".

`force.left.label`
: a logical describing whether or not to force the left labels to appear (relevant only when X has more than 50 rows). Note that by default there are no labels when there are more than 50 rows.

`force.bottom.label`
: a logical describing whether or not to force the bottom labels to appear (relevant only when X has more than 50 columns). Note that by default there are no labels when there are more than 50 columns.

`column.title` a string specifying the overall column name (located below the bottom.labels).

`row.title` a string specifying the overall row name (located to the left of the left.labels).

`column.title.size`
: a number specifying the size of the column name. The default is 5.

`row.title.size`
: a number specifying the size of the row name. The default is 5.

`padding`
: the amount (in cm) of white space (padding) around the plot. The default is 1 cm.

`title`
: a character string specifying a main heading.

`title.size`
: the size of the title. The default is 5.

`title.alignment`
: the alignment of the title. The default is "center". Options are "left", "right", "center".

`print.plot`
: a logical specifying whether or not to output the plot.

Value

`plot` a plot with the properties specified by the above arguments.

`membership.cols` the column cluster membership vector

`membership.rows` the row cluster membership vector

`order.rows` the order of the rows in the heatmap

`order.cols` the order of the columns in the heatmap

`heat.pal.values` the heat.pal.values vector used

References

Barter and Yu (2018), Superheat: An R package for creating beautiful and extendable heatmaps for visualizing complex data, https://www.tandfonline.com/doi/full/10.1080/10618600.2018.1473780, Journal of Computational and Graphical Statistics

Examples

```
# plot a heatmap of the numerical iris variables
# cluster by species and plot Sepal.Length on the right
# save the superheat object to access the membership vectors
sh <- superheat(X = iris[,-c(1, 5)],
                yr = iris[,1],
                yr.axis.name = "Sepal.Length",
                membership.rows = iris$Species)
sh$membership.rows
```

Index