# Package 'forestplotNMR'

August 1, 2018

**Title** Create Custom Forestplots for NMR Metabolomics Data

**Version** 0.1.2.9000

**Author** Maria Kalimeri (Nightingale Health Ltd.), Qin Wang (Computational Medicine, Faculty of Medicine, University of Oulu)

**Maintainer** Maria Kalimeri <maria.kalimeri@nightingalehealth.com>

**Description** Foreplot is a package for visualization of epidimiological analysis results. It's main function, forestplot, plots is a single or double column forest plots (a.k.a. blobbogram) associations of outcomes to NMR metabolomics biomarker data.

**Depends** R (>= 3.5.0),
dplyr,
tidyverse

**Suggests** testthat

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 6.0.1

## R topics documented:

## biomarkers *NMR-quantified Biomarker Names and Abbreviations*

### Description

A dataframe containing the abbreviations and spelled-out names of 228 NMR-quantified biomarkers.

### Usage

```
biomarkers
```

### Format

A data frame (tibble) with 228 rows and 4 columns:

**abbrev** Biomarker abbreviation

**fullname** Biomarker full name

**forest_plot_disp_name1** Biomarker display name option 1 (for forest plot purposes)

**forest_plot_categories** Suggestive biomarker category (for forest plot purposes)

**forest_plot_disp_name2** Biomarker display name option 2 (for forest plot purposes)

### Details

The main difference between forest_plot_disp_name1 and forest_plot_disp_name2 is how the names of the lipoprotein subclasses are displayed. For example, option 2 will display XXL-VLDL-TG

## bmr_selected_grouping *Generate Ordered, Grouped Subsets of Selected Biomarkers*

### Description

This function generates ordered, grouped subsets of selected biomarkers in the form of named lists.

### Usage

```
bmr_selected_grouping(bmr_grouping_choice = c("serum_all", "edta_plasma_all",
  "example_short"))
```

### Arguments

`bmr_grouping_choice`
: A string. Must be one of the following 3: "serum_all", "edta_plasma_all", "example_short".

## Details

The function currently generates 3 different groupings. The need for these specific groupings and orderings came from the need to generate different versions of forest plots. The categories "serum_all" and "edta_plasma_all" are expected to be the most useful as they group all the blood biomarkers for either serum (228 biomarkers) or plasma (225 biomarkers) samples. The other category "example_short" is meant as an example case.

## Value

A named list of selected biomarkers in a predefined order

## Author(s)

Maria Kalimeri

## Examples

```
# Get the grouped biomarkers
biomarker_groups <- bmr_selected_grouping("edta_plasma_all")
```

---

# demo_beta *Linear Associations of NMR-quantified Biomarkers to BMI*

---

## Description

A dataframe containing linear associations of NMR-quantified biomarkers to BMI as estimated using simulated data.

## Usage

```
demo_beta
```

## Format

A data frame (tibble) with 228 rows and 2 columns:

**abbrev** Biomarker abbreviation

**cohort1** Association values for simulated cohort 1

**cohort2** Association values for simulated cohort 2

## demo_pval — *P-values of Linear Associations of NMR-quantified Biomarkers to BMI*

### Description

A dataframe containing p-values for linear associations of NMR-quantified biomarkers to BMI as estimated using simulated data. p-values correspond to the demo_beta and demo_se dataframe values

### Usage

```
demo_pval
```

### Format

A data frame (tibble) with 228 rows and 2 columns:

**abbrev** Biomarker abbreviation

**cohort1** p-values for simulated cohort 1

**cohort2** p-values for simulated cohort 2

## demo_se — *Standard Error Values of Linear Associations of NMR-quantified Biomarkers to BMI*

### Description

A dataframe containing standard error values for linear associations of NMR-quantified biomarkers to BMI as estimated using simulated data. Std. error values correspond to the demo_beta dataframe values

### Usage

```
demo_se
```

### Format

A data frame (tibble) with 228 rows and 2 columns:

**abbrev** Biomarker abbreviation

**cohort1** Std. error values for simulated cohort 1

**cohort2** Std. error values for simulated cohort 2

forestplot_nmr *One- or Two- Column Foresplot of Biomarker Associations*

## Description

This function reads three dataframes, a biomarker association dataframe, its corresponding standard error dataframe and its corresponding p-value dataframe and plots a custom-size and layout, 1- or 2- column, forest plot.

## Usage

```
forestplot_nmr(beta, se, pval, biomarker_groups_as_list, indices = NULL,
  filename = "forest_plot.pdf", plot_title = NULL, is_log_odds_ratio = F,
  xlabel = "beta", signif_cutoff = 0.05, plotcolors = NULL,
  plotpointshape = 21, legend_vars = NULL, cex_text = NULL,
  bottom_margin = 2, left_margin = NULL, top_margin = 2,
  right_margin = 3, ylabelpos = NULL, biomarker_name_option = 1, ...)
```

## Arguments

| | |
|---|---|
| beta | A data frame (either tibble or not) containing named columns in the following way: the first column must be the exact abbreviations of the NMR biomarkers (see built-in biomarkers dataset). This column must be named "abbrev". The rest of the columns, i.e. the study columns, must contain associations and they may be named with the name of the study, e.g. if the column contains the univariate associations of the biomarkers to BMI, the column may be named BMI. Use more than one study columns if you want to plot more than one studies in the same file. Avoid plotting more than 5 or 6 studies together, cause the result is not pretty and/or may be hard to read. It is important that the columns have all either linear associations or odds/hazard ratios. The reason is that odds/hazard ratios are plotted on a log axis, whereas linear associations not. |
| se | A data frame (either tibble or not) in the same format as beta parameter. Keep the same order of columns and preferably the same order of rows (although the latter is not necessary). |
| pval | A data frame (either tibble or not) in the same format as beta and se parameters. Keep the same order of columns and preferably the same order of rows (although the latter is not necessary). |
| biomarker_groups_as_list | A named list of character vactors containing the groups of biomarkers to plot (see examples). The category names, i.e. the names of the components of the list, can be anything. The actual character vectors must contain the exact biomarker abbreviations. See Results.tsv or Results.xlsx files or the built-in dataset biomarkers$abbrev. |
| indices | Either NULL or a list of numeric vectors, that has either 1, 2 or 4 components containing the rows from beta that will eventually be plotted. It basically allows to customize the layout of the forestplot. If NULL and biomarker_groups_as_list has all serum or plasma biomarkers then a 2-column, 2-page forestplot is printed, containing all biomarkers. If a list with 1 component, e.g. list(c(1:30)), then a 1-column, 1-page forestplot is printed, containing all the biomarkers from beta dataframe until row number tail(indices[[1]], 1), e.g. tail(list(c(1:30))[[1]], |

1)=30. If a list with 2 components, then a 2-column, 1 page forestplot is printed. If a list with 4 components, then a 2-column, 2-page forestplot is printed.

`filename` A character with the name of the pdf file that will contain the plot. Defaults to 'forest_plot.pdf'

`plot_title` A character (defaults to NULL) with a title for the plot. If NULL no title.

`is_log_odds_ratio` logical (defaults to F) specifying whether the associations are linear or not. If TRUE, provide the log odds ratio as the function will exponentiate the betas internally. For TRUE a log scale is be used.

`xlabel` A character with the xlab to display. Defaults to "beta".

`signif_cutoff` Numeric specifying the cutoff for statistical significance. E.g. often a cutoff of 0.05 is used. Associations with values larger than that will be plotted with an empty circular point.

`plotcolors` A vector of characters specifying the color of the plotted points. Defaults to NULL in which case, if only one study, black is used. If more than one studies, the script generates a default palette.

`plotpointshape` An integer or vector of integers (default 21) signifying the shape of points used for the plot. The values must be one of 21,22,23,24 or 25 in order for the insignificant cases to be displayed as empty shapes.

`legend_vars` A vector of characters specifying the legend names for when more than one studies are plotted.

`cex_text` The size of the y- and x-label. Legends and titles will be adjusted with respect to that.

`bottom_margin` The margin from the bottom of the plot (the forestplot will be plotted in a A4 paper).

`left_margin` The margin from the left edge of the plot.

`top_margin` The margin from the top of the plot.

`right_margin` The margin from the right edge of the plot.

`ylabelpos` The distance of the ylabels from the plot. This parameter will most likley need to be adjusted in conjuction with the margins and the cex_tex, especially when the non-default layout is used.

`biomarker_name_option` numeric (defaults to option 1), currently takes values 1 (for option 1) and 2 (for option 2). The main difference between the two options is how the names of the lipoprotein subclasses are displayed. For example, option 2 will display XXL-VLDL-TG % for the ratio of triglycerides in XXL VLDL particles, whereas option 1 assumes that plotting will be done according to lipid type, e.g. all triglycerides plotted in the same subgroup, therefore it would only display "Extremely large VLDL" (under the category "Triglycerides in lipoproteins").

`...` Arguments to be passed to the `pdf` device, like `paper`, `width`, `height` e.t.c.

## Details

The parameters `biomarker_groups_as_list` and `indices` allow for customization of the layout in 1- or 2- column, 1- or 2- page forestplots. Specifically, if the indices is specified (as opposed to the default NULL) it must be a list of numeric vectors, that has either 1, 2 or 4 components, which allow for a 1-column/1-page, 2-column/1-page or 2-column/2-page forestplot, respectively (see examples). The list essentially defines the rows from beta that will be plotted. By adding NA in selected positions of the biomarker_groups_as_list while increasing the number of indices, accordingly, one can add extra white space between biomarker of biomarker categories.

## Note

It is important that the columns have all either linear associations or odds/hazard ratios. The reason is that odds/hazard ratios are plotted on a log axis, whereas linear associations not.

## Author(s)

Qin Wang, Maria Kalimeri

## Examples

```
bmr_all_grouped <- bmr_selected_grouping(bmr_grouping_choice = "serum_all")

forestplot_nmr(beta=demo_beta,
           se=demo_se,
           pval=demo_pval,
           biomarker_groups_as_list=bmr_all_grouped,
           filename='plot_linear_comparison.pdf',
           plot_title="Linear associations to BMI",
           is_log_odds_ratio=FALSE,
           xlabel="SD difference (95% CI)",
           signif_cutoff=0.05,
           legend_vars=names(demo_beta)[2:3],
           paper="a4",
           height = 12,
           width = 9)
```

---

# fpcolors *A Palette of 5 Colors*

---

## Description

A named vector containing the hex numbers of 5 colors.

## Usage

```
fpcolors
```

## Format

An object of class `character` of length 5.

# Index