Package 'halfmoon'

May 30, 2023

Title Techniques to Build Better Balance

Version 0.1.0

Description Build better balance in causal inference models. 'halfmoon' helps you assess propensity score models for balance between groups using metrics like standardized mean differences and visualization techniques like mirrored histograms. 'halfmoon' supports both weighting and matching techniques.

License MIT + file LICENSE

URL https://github.com/r-causal/halfmoon, https://r-causal.github.io/halfmoon/

BugReports https://github.com/r-causal/halfmoon/issues

Depends R (>= 2.10)

Imports cli, ggplot2, tidyselect, tidysmd (>= 0.2.0)

Suggests covr, testthat (>= 3.0.0), vdiffr

Config/testthat/edition 3

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

NeedsCompilation no

Author Malcolm Barrett [aut, cre, cph] (<https://orcid.org/0000-0003-0299-5825>)

Maintainer Malcolm Barrett <malcolmbarrett@gmail.com>

Repository CRAN

Date/Publication 2023-05-30 18:20:02 UTC

R topics documented:

`geom_ecdf` *Calculate weighted and unweighted empirical cumulative distributions*

Description

The empirical cumulative distribution function (ECDF) provides an alternative visualization of distribution. `geom_ecdf()` is similar to `ggplot2::stat_ecdf()` but it can also calculate weighted ECDFs.

Usage

```
geom_ecdf(
  mapping = NULL,
  data = NULL,
  geom = "step",
  position = "identity",
  ...,
  n = NULL,
  pad = TRUE,
  na.rm = FALSE,
  show.legend = NA,
  inherit.aes = TRUE
)
```

Arguments

`mapping`	Set of aesthetic mappings created by `aes()`. If specified and `inherit.aes = TRUE` (the default), it is combined with the default mapping at the top level of the plot. You must supply `mapping` if there is no plot mapping.
`data`	The data to be displayed in this layer. There are three options: If `NULL`, the default, the data is inherited from the plot data as specified in the call to `ggplot()`. A `data.frame`, or other object, will override the plot data. All objects will be fortified to produce a data frame. See `fortify()` for which variables will be created. A `function` will be called with a single argument, the plot data. The return value must be a `data.frame`, and will be used as the layer data. A `function` can be created from a `formula` (e.g. `~ head(.x, 10)`).
`geom`	The geometric object to use to display the data, either as a `ggproto` `Geom` subclass or as a string naming the geom stripped of the `geom_` prefix (e.g. `"point"` rather than `"geom_point"`)
`position`	Position adjustment, either as a string naming the adjustment (e.g. `"jitter"` to use `position_jitter`), or the result of a call to a position adjustment function. Use the latter if you need to change the settings of the adjustment.

...	Other arguments passed on to layer(). These are often aesthetics, used to set an aesthetic to a fixed value, like colour = "red" or size = 3. They may also be parameters to the paired geom/stat.
n	if NULL, do not interpolate. If not NULL, this is the number of points to interpolate with.
pad	If TRUE, pad the ecdf with additional points (-Inf, 0) and (Inf, 1)
na.rm	If FALSE (the default), removes missing values with a warning. If TRUE silently removes missing values.
show.legend	logical. Should this layer be included in the legends? NA, the default, includes if any aesthetics are mapped. FALSE never includes, and TRUE always includes. It can also be a named logical vector to finely select the aesthetics to display.
inherit.aes	If FALSE, overrides the default aesthetics, rather than combining with them. This is most useful for helper functions that define both data and aesthetics and shouldn't inherit behaviour from the default plot specification, e.g. borders().

Value

a geom

Aesthetics

In addition to the aesthetics for ggplot2::stat_ecdf(), geom_ecdf() also accepts:

- weights

Examples

```
library(ggplot2)

ggplot(
  nhefs_weights,
  aes(x = smokeyrs, color = qsmk)
) +
  geom_ecdf(aes(weights = w_ato)) +
  xlab("Smoking Years") +
  ylab("Proportion <= x")
```

---

geom_mirror_histogram *Create mirrored histograms*

Description

Create mirrored histograms

Usage

```
geom_mirror_histogram(
  mapping = NULL,
  data = NULL,
  position = "stack",
  ...,
  binwidth = NULL,
  bins = NULL,
  na.rm = FALSE,
  orientation = NA,
  show.legend = NA,
  inherit.aes = TRUE
)
```

Arguments

mapping	Set of aesthetic mappings created by aes(). If specified and inherit.aes = TRUE (the default), it is combined with the default mapping at the top level of the plot. You must supply mapping if there is no plot mapping.
data	The data to be displayed in this layer. There are three options: If NULL, the default, the data is inherited from the plot data as specified in the call to ggplot(). A data.frame, or other object, will override the plot data. All objects will be fortified to produce a data frame. See fortify() for which variables will be created. A function will be called with a single argument, the plot data. The return value must be a data.frame, and will be used as the layer data. A function can be created from a formula (e.g. ~ head(.x, 10)).
position	Position adjustment, either as a string naming the adjustment (e.g. "jitter" to use position_jitter), or the result of a call to a position adjustment function. Use the latter if you need to change the settings of the adjustment.
...	Other arguments passed on to layer(). These are often aesthetics, used to set an aesthetic to a fixed value, like colour = "red" or size = 3. They may also be parameters to the paired geom/stat.
binwidth	The width of the bins. Can be specified as a numeric value or as a function that calculates width from unscaled x. Here, "unscaled x" refers to the original x values in the data, before application of any scale transformation. When specifying a function along with a grouping structure, the function will be called once per group. The default is to use the number of bins in bins, covering the range of the data. You should always override this value, exploring multiple widths to find the best to illustrate the stories in your data. The bin width of a date variable is the number of days in each time; the bin width of a time variable is the number of seconds.
bins	Number of bins. Overridden by binwidth. Defaults to 30.
na.rm	If FALSE, the default, missing values are removed with a warning. If TRUE, missing values are silently removed.

`orientation`	The orientation of the layer. The default (`NA`) automatically determines the orientation from the aesthetic mapping. In the rare event that this fails it can be given explicitly by setting `orientation` to either `"x"` or `"y"`. See the *Orientation* section for more detail.
`show.legend`	logical. Should this layer be included in the legends? `NA`, the default, includes if any aesthetics are mapped. `FALSE` never includes, and `TRUE` always includes. It can also be a named logical vector to finely select the aesthetics to display.
`inherit.aes`	If `FALSE`, overrides the default aesthetics, rather than combining with them. This is most useful for helper functions that define both data and aesthetics and shouldn't inherit behaviour from the default plot specification, e.g. `borders()`.

Value

a geom

Examples

```
library(ggplot2)
ggplot(nhefs_weights, aes(.fitted)) +
  geom_mirror_histogram(
    aes(group = qsmk),
    bins = 50
  ) +
  geom_mirror_histogram(
    aes(fill = qsmk, weight = w_ate),
    bins = 50,
    alpha = 0.5
  ) +
  scale_y_continuous(labels = abs)
```

---

nhefs_weights *NHEFS with various propensity score weights*

Description

A dataset containing various propensity score weights for `causaldata::nhefs_complete`.

Usage

```
nhefs_weights
```

Format

A data frame with 1566 rows and 14 variables:

qsmk Quit smoking

race Race

age Age

sex Sex

education Education level

smokeintensity Smoking intensity

smokeyrs Number of smoke-years

exercise Exercise level

active Daily activity level

wt71 Participant weight in 1971 (baseline)

w_ate ATE weight

w_att ATT weight

w_atc ATC weight

w_atm ATM weight

w_ato ATO weight

.fitted Propensity score

Index