Package 'rbin'

November 5, 2024

Type Package

Title Tools for Binning Data

Version 0.2.1

Description Manually bin data using weight of evidence and information value. Includes other binning methods such as equal length, quantile and winsorized. Options for combining levels of categorical
data are also available. Dummy variables can be generated based on the bins created using any of the available binning methods. References: Siddiqi, N. (2006) <doi:10.1002/9781119201731.biblio>.

License MIT + file LICENSE

URL https://github.com/rsquaredacademy/rbin,
https://rbin.rsquaredacademy.com

BugReports https://github.com/rsquaredacademy/rbin/issues

Depends R (>= 3.3)

Imports data.table, ggplot2, stats, utils

Suggests covr, graphics, knitr, miniUI, rmarkdown, rstudioapi, shiny,
testthat (>= 3.0.0), vdiffr

VignetteBuilder knitr

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Config/testthat/edition 3

NeedsCompilation no

Author Aravind Hebbali [aut, cre]

Maintainer Aravind Hebbali <hebbali.aravind@gmail.com>

Repository CRAN

Date/Publication 2024-11-05 11:50:03 UTC

Contents

---

mbank   *Bank marketing data set*

Description

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

Usage

    mbank

Format

A tibble with 4521 rows and 17 variables:

age age of the client

job type of job

marital marital status

education education level of the client

default has credit in default?

housing has housing loan?

loan has personal loan?

contact contact communication type

month last contact month of year

day_of_week last contact day of the week

duration last contact duration, in seconds

campaign number of contacts performed during this campaign and for this client

pdays number of days that passed by after the client was last contacted from a previous campaign

previous number of contacts performed before this campaign and for this clien

poutcome outcome of the previous marketing campaign

y has the client subscribed a term deposit?

Source

[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

---

rbinAddin *Bin continuous data*

Description

Manually bin continuous data using weight of evidence.

Usage

```
rbinAddin(data = NULL)
```

Arguments

data   A data.frame or tibble.

Examples

```
## Not run:
rbinAddin(data = mbank)

## End(Not run)
```

---

rbinFactorAddin   *Custom binning*

Description

Manually combine categorical variables using weight of evidence.

Usage

```
rbinFactorAddin(data = NULL)
```

Arguments

data   A data.frame or tibble.

Examples

```
## Not run:
rbinFactorAddin(data = mbank)

## End(Not run)
```

---

rbin_create   *Create dummy variables*

Description

Create dummy variables from bins.

Usage

```
rbin_create(data, predictor, bins)
```

Arguments

data   A data.frame or tibble.

predictor   Variable for which dummy variables must be created.

bins   An object of class rbin_manual or rbin_quantiles or rbin_equal_length
                or rbin_winsorized.

Value

data with dummy variables.

Examples

```
k <- rbin_manual(mbank, y, age, c(29, 39, 56))
rbin_create(mbank, age, k)
```

---

rbin_equal_freq *Equal frequency binning*

Description

Bin continuous data using the equal frequency binning method.

Usage

```
rbin_equal_freq(data = NULL, response = NULL, predictor = NULL, bins = 10)

## S3 method for class 'rbin_equal_freq'
plot(x, print_plot = TRUE, ...)
```

Arguments

data	A data.frame or tibble.
response	Response variable.
predictor	Predictor variable.
bins	Number of bins.
x	An object of class rbin_quantiles.
print_plot	logical; if TRUE, prints the plot else returns a plot object.
...	further arguments passed to or from other methods.

Value

A tibble.

Examples

```
bins <- rbin_equal_freq(mbank, y, age, 10)
bins

# plot
plot(bins)
```

rbin_equal_length	*Equal length binning*

Description

Bin continuous data using the equal length binning method.

Usage

```
rbin_equal_length(
  data = NULL,
  response = NULL,
  predictor = NULL,
  bins = 10,
  include_na = TRUE
)

## S3 method for class 'rbin_equal_length'
plot(x, print_plot = TRUE, ...)
```

Arguments

data	A data.frame or tibble.
response	Response variable.
predictor	Predictor variable.
bins	Number of bins.
include_na	logical; if TRUE, a separate bin is created for missing values.
x	An object of class rbin_equal_length.
print_plot	logical; if TRUE, prints the plot else returns a plot object.
...	further arguments passed to or from other methods.

Value

A tibble.

Examples

```
bins <- rbin_equal_length(mbank, y, age, 10)
bins

# plot
plot(bins)
```

---

rbin_factor   *Factor binning*

Description

Weight of evidence and information value for categorical data.

Usage

```
rbin_factor(data = NULL, response = NULL, predictor = NULL, include_na = TRUE)

## S3 method for class 'rbin_factor'
plot(x, print_plot = TRUE, ...)
```

Arguments

data	A data.frame or tibble.
response	Response variable.
predictor	Predictor variable.
include_na	logical; if TRUE, a separate bin is created for missing values.
x	An object of class rbin_factor.
print_plot	logical; if TRUE, prints the plot else returns a plot object.
...	further arguments passed to or from other methods.

Examples

```
bins <- rbin_factor(mbank, y, education)
bins

# plot
plot(bins)
```

---

rbin_factor_combine   *Combine levels*

Description

Manually combine levels of categorical data.

Usage

```
rbin_factor_combine(data, var, new_var, new_name)
```

Arguments

`data`	A `data.frame` or `tibble`.
`var`	An object of class `factor`.
`new_var`	A character vector; it should include the names of the levels to be combined.
`new_name`	Name of the combined level.

Value

A `tibble`.

Examples

```
upper <- c("secondary", "tertiary")
out <- rbin_factor_combine(mbank, education, upper, "upper")
table(out$education)

out <- rbin_factor_combine(mbank, education, c("secondary", "tertiary"), "upper")
table(out$education)
```

---

rbin_factor_create   *Create dummy variables*

Description

Create dummy variables for categorical data.

Usage

```
rbin_factor_create(data, predictor)
```

Arguments

`data`	A `data.frame` or `tibble`.
`predictor`	Variable for which dummy variables must be created.

Value

A `tibble` with dummy variables.

Examples

```
upper <- c("secondary", "tertiary")
out <- rbin_factor_combine(mbank, education, upper, "upper")
rbin_factor_create(out, education)
```

`rbin_manual`	*Manual binning*

Description

Bin continuous data manually.

Usage

```
rbin_manual(
  data = NULL,
  response = NULL,
  predictor = NULL,
  cut_points = NULL,
  include_na = TRUE
)

## S3 method for class 'rbin_manual'
plot(x, print_plot = TRUE, ...)
```

Arguments

`data`	A `data.frame` or `tibble`.
`response`	Response variable.
`predictor`	Predictor variable.
`cut_points`	Cut points for binning.
`include_na`	logical; if TRUE, a separate bin is created for missing values.
`x`	An object of class `rbin_manual`.
`print_plot`	logical; if TRUE, prints the plot else returns a plot object.
`...`	further arguments passed to or from other methods.

Details

Specify the upper open interval for each bin. 'rbin' follows the left closed and right open interval. If you want to create_bins 10 bins, the app will show you only 9 input boxes. The interval for the 10th bin is automatically computed. For example, if you want the first bin to have all the values between the minimum and including 36, then you will enter the value 37.

Value

A `tibble`.

Examples

```
bins <- rbin_manual(mbank, y, age, c(29, 31, 34, 36, 39, 42, 46, 51, 56))
bins

# plot
plot(bins)
```

---

rbin_quantiles *Quantile binning*

Description

Bin continuous data using quantiles.

Usage

```
rbin_quantiles(
  data = NULL,
  response = NULL,
  predictor = NULL,
  bins = 10,
  include_na = TRUE
)

## S3 method for class 'rbin_quantiles'
plot(x, print_plot = TRUE, ...)
```

Arguments

data	A data.frame or tibble.
response	Response variable.
predictor	Predictor variable.
bins	Number of bins.
include_na	logical; if TRUE, a separate bin is created for missing values.
x	An object of class rbin_quantiles.
print_plot	logical; if TRUE, prints the plot else returns a plot object.
...	further arguments passed to or from other methods.

Value

A tibble.

Examples

```
bins <- rbin_quantiles(mbank, y, age, 10)
bins

# plot
plot(bins)
```

---

rbin_winsorize   *Winsorized binning*

Description

Bin continuous data using winsorized method.

Usage

```
rbin_winsorize(
  data = NULL,
  response = NULL,
  predictor = NULL,
  bins = 10,
  include_na = TRUE,
  winsor_rate = 0.05,
  min_val = NULL,
  max_val = NULL,
  type = 7,
  remove_na = TRUE
)

## S3 method for class 'rbin_winsorize'
plot(x, print_plot = TRUE, ...)
```

Arguments

data	A data.frame or tibble.
response	Response variable.
predictor	Predictor variable.
bins	Number of bins.
include_na	logical; if TRUE, a separate bin is created for missing values.
winsor_rate	A value from 0.0 to 0.5.
min_val	the low border, all values being lower than this will be replaced by this value. The default is set to the 5 percent quantile of predictor.
max_val	the high border, all values being larger than this will be replaced by this value. The default is set to the 95 percent quantile of predictor.

type	an integer between 1 and 9 selecting one of the nine quantile algorithms detailed in `quantile()` to be used.
remove_na	logical; if TRUE NAs will removed while calculating quantiles
x	An object of class `rbin_winsorize`.
print_plot	logical; if TRUE, prints the plot else returns a plot object.
...	further arguments passed to or from other methods.

Value

A `tibble`.

Examples

```
bins <- rbin_winsorize(mbank, y, age, 10, winsor_rate = 0.05)
bins

# plot
plot(bins)
```

Index