Package: descriptr (via r-universe)

August 21, 2024

Type Package

Title Generate Descriptive Statistics

Version 0.5.2.9000

Description Generate descriptive statistics such as measures of location, dispersion, frequency tables, cross tables, group summaries and multiple one/two way tables.

Depends R(>= 3.3.0)

Imports dplyr, ggplot2, magrittr, rlang, scales, stats, tibble, tidyr, utils

Suggests covr, gridExtra, knitr, rmarkdown, testthat, vdiffr, xplorerr

License MIT + file LICENSE

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

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

Encoding UTF-8

LazyData true

VignetteBuilder knitr

RoxygenNote 7.1.1

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

RemoteUrl https://github.com/rsquaredacademy/descriptr

RemoteRef HEAD

RemoteSha 99fc706e7b04db4c4f04a43d9e96c68d6318cb4c

Contents

`descriptr` `descriptr` *package*

Description

Generate descriptive statistics and explore statistical distributions

ds_auto_freq_table *Multiple One & Two Way Tables*

Description

ds_auto_freq_table creates multiple one way tables by creating a frequency table for each categorical variable in a data frame. ds_auto_cross_table creates multiple two way tables by creating a cross table for each unique pair of categorical variables in a data frame.

Usage

```
ds_auto_freq_table(data, ...)

ds_auto_cross_table(data, ...)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.

Details

ds_auto_freq_table is a extension of the ds_freq_table function. It creates a frequency table for each categorical variable in the dataframe. ds_auto_cross_table is a extension of the ds_cross_table function. It creates a two way table for each unique pair of categorical variables in the dataframe.

Deprecated Functions

ds_oway_tables() and ds_tway_tables() have been deprecated. Instead use ds_auto_freq_table() and ds_auto_cross_table().

See Also

link{ds_freq_table} link{ds_cross_table}

Examples

```
# frequency table for all columns
ds_auto_freq_table(mtcarz)

# frequency table for multiple columns
ds_auto_freq_table(mtcarz, cyl, gear)

# cross table for all columns
ds_auto_cross_table(mtcarz)

# cross table for multiple columns
```

```
ds_auto_cross_table(mtcarz, cyl, gear, am)
```

---

ds_auto_group_summary *Tabulation*

Description

Generate summary statistics for all continuous variables in data.

Usage

```
ds_auto_group_summary(data, ...)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.

Examples

```
# summary statistics of mpg & disp for each level of cyl & gear
ds_auto_group_summary(mtcarz, cyl, gear, mpg, disp)
```

---

ds_auto_summary_stats *Descriptive statistics and frquency tables*

Description

Generate summary statistics & frequency table for all continuous variables in data.

Usage

```
ds_auto_summary_stats(data, ...)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.

Examples

```
# all columns
ds_auto_summary_stats(mtcarz)

# multiple columns
ds_auto_summary_stats(mtcarz, disp, hp)
```

ds_cross_table *Two way table*

Description

Creates two way tables of categorical variables. The tables created can be visualized as bar plots and mosaic plots.

Usage

```
ds_cross_table(data, var_1, var_2)

## S3 method for class 'ds_cross_table'
plot(x, stacked = FALSE, proportional = FALSE, print_plot = TRUE, ...)

ds_twoway_table(data, var_1, var_2)
```

Arguments

data	A data.frame or a tibble.
var_1	First categorical variable.
var_2	Second categorical variable.
x	An object of class cross_table.
stacked	If FALSE, the columns of height are portrayed as stacked bars, and if TRUE the columns are portrayed as juxtaposed bars.
proportional	If TRUE, the height of the bars is proportional.
print_plot	logical; if TRUE, prints the plot else returns a plot object.
...	Further arguments to be passed to or from methods.

Examples

```
# cross table
k <- ds_cross_table(mtcarz, cyl, gear)
k

# bar plot
plot(k)

# stacked bar plot
plot(k, stacked = TRUE)

# proportional bar plot
plot(k, proportional = TRUE)

# returns tibble
ds_twoway_table(mtcarz, cyl, gear)
```

ds_css *Corrected Sum of Squares*

Description

Compute the corrected sum of squares

Usage

```
ds_css(data, x = NULL)
```

Arguments

data	A numeric vector or data.frame.
x	Column in data.

Examples

```
# vector
ds_css(mtcars$mpg)

# data.frame
ds_css(mtcars, mpg)
```

ds_cvar *Coefficient of Variation*

Description

Compute the coefficient of variation

Usage

```
ds_cvar(data, x = NULL)
```

Arguments

data	A numeric vector or data.frame.
x	Column in data.

Examples

```
# vector
ds_cvar(mtcars$mpg)

# data.frame
ds_cvar(mtcars, mpg)
```

`ds_extreme_obs` *Extreme observations*

Description

Returns the most extreme observations.

Usage

```
ds_extreme_obs(data, col, decimals = 2)
```

Arguments

`data`	A numeric vector or `data.frame` or `tibble`.
`col`	Column in `data`.
`decimals`	An option to specify the exact number of decimal places to use. The default number of decimal places is 2.

Examples

```
# data.frame
ds_extreme_obs(mtcarz, mpg)

# vector
ds_extreme_obs(mtcarz$mpg)

# decimal places
ds_extreme_obs(mtcarz$mpg, decimals = 3)
```

ds_freq_table *Frequency table*

Description

Frequency table for categorical and continuous data and returns the frequency, cumulative frequency, frequency percent and cumulative frequency percent. `plot.ds_freq_table()` creates bar plot for the categorical data and histogram for continuous data.

Usage

```
ds_freq_table(data, col, bins = 5)

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

Arguments

`data`	A `data.frame` or a `tibble`.
`col`	Column in `data`.
`bins`	Number of intervals into which the data must be split.
`x`	An object of class `ds_freq_table`.
`print_plot`	logical; if `TRUE`, prints the plot else returns a plot object.
`...`	Further arguments to be passed to or from methods.

See Also

`ds_cross_table`

Examples

```
# categorical data
ds_freq_table(mtcarz, cyl)

# barplot
k <- ds_freq_table(mtcarz, cyl)
plot(k)

# continuous data
ds_freq_table(mtcarz, mpg)

# barplot
k <- ds_freq_table(mtcarz, mpg)
plot(k)
```

ds_gmean *Geometric Mean*

Description

Computes the geometric mean

Usage

```
ds_gmean(data, x = NULL)
```

Arguments

data	A numeric vector or data.frame.
x	Column in data.

See Also

ds_hmean mean

Examples

```
# vector
ds_gmean(mtcars$mpg)

# data.frame
ds_gmean(mtcars, mpg)
```

ds_group_summary *Groupwise descriptive statistics*

Description

Descriptive statistics of a continuous variable for the different levels of a categorical variable. boxplot.group_summary() creates boxplots of the continuous variable for the different levels of the categorical variable.

Usage

```
ds_group_summary(data, group_by, cols)

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

Arguments

data	A data.frame or a tibble.
group_by	Column in data.
cols	Column in data.
x	An object of the class ds_group_summary.
print_plot	logical; if TRUE, prints the plot else returns a plot object.
...	Further arguments to be passed to or from methods.

Value

ds_group_summary() returns an object of class "ds_group_summary". An object of class "ds_group_summary" is a list containing the following components:

stats	A data frame containing descriptive statistics for the different levels of the factor variable.
tidy_stats	A tibble containing descriptive statistics for the different levels of the factor variable.
plotdata	Data for boxplot method.

See Also

ds_summary_stats

Examples

```
# ds_group summary
ds_group_summary(mtcarz, cyl, mpg)

# boxplot
k <- ds_group_summary(mtcarz, cyl, mpg)
plot(k)

# tibble
k$tidy_stats
```

---

ds_group_summary_interact

Category wise descriptive statistics

Description

Descriptive statistics of a continuous variable for the combination of levels of two or more categorical variables.

Usage

```
ds_group_summary_interact(data, col, ...)
```

Arguments

data	A data.frame or a tibble.
col	Column in data; continuous variable.
...	Columns in data; categorical variables.

See Also

ds_group_summary

Examples

```
ds_group_summary_interact(mtcarz, mpg, cyl, gear)
```

---

ds_hmean *Harmonic Mean*

Description

Computes the harmonic mean

Usage

```
ds_hmean(data, x = NULL)
```

Arguments

data	A numeric vector or data.frame.
x	Column in data.

See Also

ds_gmean mean

Examples

```
# vector
ds_hmean(mtcars$mpg)

# data.frame
ds_hmean(mtcars, mpg)
```

ds_kurtosis *Kurtosis*

Description

Compute the kurtosis of a probability distribution.

Usage

```
ds_kurtosis(data, x = NULL)
```

Arguments

data	A numeric vector or data.frame.
x	Column in data.

References

Sheskin, D.J. (2000) Handbook of Parametric and Nonparametric Statistical Procedures, Second Edition. Boca Raton, Florida: Chapman & Hall/CRC.

See Also

ds_skewness

Examples

```
# vector
ds_kurtosis(mtcars$mpg)

# data.frame
ds_kurtosis(mtcars, mpg)
```

ds_launch_shiny_app *Launch Shiny App*

Description

Launches shiny app

Usage

```
ds_launch_shiny_app()
```

Deprecated Function

launch_descriptr() has been deprecated. Instead use ds_launch_shiny_app().

Examples

```
## Not run:
ds_launch_shiny_app()

## End(Not run)
```

---

ds_mdev *Mean Absolute Deviation*

Description

Compute the mean absolute deviation about the mean

Usage

```
ds_mdev(data, x = NULL)
```

Arguments

data	A numeric vector or data.frame.
x	Column in data.

Details

The ds_mdev function computes the mean absolute deviation about the mean. It is different from mad in stats package as the statistic used to compute the deviations is not median but mean. Any NA values are stripped from x before computation takes place

See Also

mad

Examples

```
# vector
ds_mdev(mtcars$mpg)

# data.frame
ds_mdev(mtcars, mpg)
```

`ds_measures_location` *Measures of location*

Description

Returns the measures of location such as mean, median & mode.

Usage

```
ds_measures_location(data, ..., trim = 0.05, decimals = 2)
```

Arguments

`data`	A `data.frame` or `tibble` or numeric vector.
`...`	Column(s) in `data` or numeric vectors.
`trim`	The fraction of values to be trimmed before computing the mean.
`decimals`	An option to specify the exact number of decimal places to use. The default number of decimal places is 2.

Examples

```
# single column
ds_measures_location(mtcarz, mpg)

# multiple columns
ds_measures_location(mtcarz, mpg, disp)

# all columns
ds_measures_location(mtcarz)

# vector
ds_measures_location(mtcarz$mpg)

# vectors of different length
disp <- mtcarz$disp[1:10]
ds_measures_location(mtcarz$mpg, disp)

# decimal places
ds_measures_location(mtcarz, disp, hp, decimals = 3)
```

ds_measures_symmetry *Measures of symmetry*

Description

Returns the measures of symmetry such as skewness and kurtosis.

Usage

```
ds_measures_symmetry(data, ..., decimals = 2)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
decimals	An option to specify the exact number of decimal places to use. The default number of decimal places is 2.

Examples

```
# single column
ds_measures_symmetry(mtcarz, mpg)

# multiple columns
ds_measures_symmetry(mtcarz, mpg, disp)

# all columns
ds_measures_symmetry(mtcarz)

# vector
ds_measures_symmetry(mtcarz$mpg)

# vectors of different length
disp <- mtcarz$disp[1:10]
ds_measures_symmetry(mtcarz$mpg, disp)

# decimal places
ds_measures_symmetry(mtcarz, disp, hp, decimals = 3)
```

ds_measures_variation *Measures of variation*

Description

Returns the measures of location such as range, variance and standard deviation.

Usage

```
ds_measures_variation(data, ..., decimals = 2)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
decimals	An option to specify the exact number of decimal places to use. The default number of decimal places is 2.

Examples

```
# single column
ds_measures_variation(mtcarz, mpg)

# multiple columns
ds_measures_variation(mtcarz, mpg, disp)

# all columns
ds_measures_variation(mtcarz)

# vector
ds_measures_variation(mtcarz$mpg)

# vectors of different length
disp <- mtcarz$disp[1:10]
ds_measures_variation(mtcarz$mpg, disp)

# decimal places
ds_measures_variation(mtcarz, disp, hp, decimals = 3)
```

ds_mode *Mode*

Description

Compute the sample mode

Usage

```
ds_mode(data, x = NULL)
```

Arguments

data	A numeric vector or data.frame.
x	Column in data.

Details

Any NA values are stripped from x before computation takes place.

Value

Mode of x

See Also

mean median

Examples

```
# vector
ds_mode(mtcars$mpg)

# data.frame
ds_mode(mtcars, mpg)
```

ds_percentiles *Percentiles*

Description

Returns the percentiles

Usage

```
ds_percentiles(data, ..., decimals = 2)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
decimals	An option to specify the exact number of decimal places to use. The default number of decimal places is 2.

Examples

```
# single column
ds_percentiles(mtcarz, mpg)

# multiple columns
ds_percentiles(mtcarz, mpg, disp)

# all columns
ds_percentiles(mtcarz)

# vector
ds_percentiles(mtcarz$mpg)

# vectors of different length
disp <- mtcarz$disp[1:10]
ds_percentiles(mtcarz$mpg, disp)

# decimal places
ds_percentiles(mtcarz, disp, hp, decimals = 3)
```

ds_plot_bar *Generate bar plots*

Description

Creates bar plots if the data has categorical variables.

Usage

```
ds_plot_bar(data, ..., fill = "blue", print_plot = TRUE)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
fill	Color of the bars.
print_plot	logical; if TRUE, prints the plot else returns a plot object.

Examples

```
# plot single variable
ds_plot_bar(mtcarz, cyl)

# plot multiple variables
ds_plot_bar(mtcarz, cyl, gear)

# plot all variables
ds_plot_bar(mtcarz)
```

ds_plot_bar_grouped *Generate grouped bar plots*

Description

Creates grouped bar plots if the data has categorical variables.

Usage

```
ds_plot_bar_grouped(data, ..., print_plot = TRUE)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
print_plot	logical; if TRUE, prints the plot else returns a plot object.

Examples

```
# subset data
mt <- dplyr::select(mtcarz, cyl, gear, am)

# grouped bar plot
ds_plot_bar_grouped(mtcarz, cyl, gear)

# plot all variables
ds_plot_bar_grouped(mt)
```

---

ds_plot_bar_stacked *Generate stacked bar plots*

Description

Creates stacked bar plots if the data has categorical variables.

Usage

```
ds_plot_bar_stacked(data, ..., print_plot = TRUE)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
print_plot	logical; if TRUE, prints the plot else returns a plot object.

Examples

```
# subset data
mt <- dplyr::select(mtcarz, cyl, gear, am)

# stacked bar plot
ds_plot_bar_stacked(mtcarz, cyl, gear)

# plot all variables
ds_plot_bar_stacked(mt)
```

ds_plot_box_group *Compare distributions*

Description

Creates box plots if the data has both categorical & continuous variables.

Usage

```
ds_plot_box_group(data, ..., print_plot = TRUE)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
print_plot	logical; if TRUE, prints the plot else returns a plot object.

Examples

```
# subset data
mt <- dplyr::select(mtcarz, cyl, disp, mpg)

# plot select variables
ds_plot_box_group(mtcarz, cyl, gear, mpg)

# plot all variables
ds_plot_box_group(mt)
```

ds_plot_box_single *Generate box plots*

Description

Creates box plots if the data has continuous variables.

Usage

```
ds_plot_box_single(data, ..., print_plot = TRUE)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
print_plot	logical; if TRUE, prints the plot else returns a plot object.

Examples

```
# plot single variable
ds_plot_box_single(mtcarz, mpg)

# plot multiple variables
ds_plot_box_single(mtcarz, mpg, disp, hp)

# plot all variables
ds_plot_box_single(mtcarz)
```

---

ds_plot_density *Generate density plots*

Description

Creates density plots if the data has continuous variables.

Usage

```
ds_plot_density(data, ..., color = "blue", print_plot = TRUE)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
color	Color of the plot.
print_plot	logical; if TRUE, prints the plot else returns a plot object.

Examples

```
# plot single variable
ds_plot_density(mtcarz, mpg)

# plot multiple variables
ds_plot_density(mtcarz, mpg, disp, hp)

# plot all variables
ds_plot_density(mtcarz)
```

ds_plot_histogram *Generate histograms*

Description

Creates histograms if the data has continuous variables.

Usage

```
ds_plot_histogram(data, ..., bins = 5, fill = "blue", print_plot = TRUE)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
bins	Number of bins in the histogram.
fill	Color of the histogram.
print_plot	logical; if TRUE, prints the plot else returns a plot object.

Examples

```
# plot single variable
ds_plot_histogram(mtcarz, mpg)

# plot multiple variables
ds_plot_histogram(mtcarz, mpg, disp, hp)

# plot all variables
ds_plot_histogram(mtcarz)
```

ds_plot_scatter *Generate scatter plots*

Description

Creates scatter plots if the data has continuous variables.

Usage

```
ds_plot_scatter(data, ..., print_plot = TRUE)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.
print_plot	logical; if TRUE, prints the plot else returns a plot object.

Examples

```
# plot select variables
ds_plot_scatter(mtcarz, mpg, disp)

# plot all variables
ds_plot_scatter(mtcarz)
```

---

ds_range *Range*

Description

Compute the range of a numeric vector

Usage

```
ds_range(data, x = NULL)
```

Arguments

data	A numeric vector or data.frame.
x	Column in data.

Value

Range of x

See Also

range

Examples

```
# vector
ds_range(mtcars$mpg)

# data.frame
ds_range(mtcars, mpg)
```

ds_rindex *Index Values*

Description

Returns index of values.

Usage

```
ds_rindex(data, values)
```

Arguments

data	a numeric vector
values	a numeric vector containing the values whose index is returned

Value

Index of the values in data. In case, data does not contain index, NULL is returned.

Examples

```
# returns index of 21
ds_rindex(mtcars$mpg, 21)

# returns NULL
ds_rindex(mtcars$mpg, 22)
```

ds_screener *Screen data*

Description

Screen data and return details such as variable names, class, levels and missing values. plot.ds_screener() creates bar plots to visualize of missing observations for each variable in a data set.

Usage

```
ds_screener(data)

## S3 method for class 'ds_screener'
plot(x, ...)
```

Arguments

data	A tibble or a data.frame.
x	An object of class ds_screener.
...	Further arguments to be passed to or from methods.

Value

ds_screener() returns an object of class "ds_screener". An object of class "ds_screener" is a list containing the following components:

Rows	Number of rows in the data frame.
Columns	Number of columns in the data frame.
Variables	Names of the variables in the data frame.
Types	Class of the variables in the data frame.
Count	Length of the variables in the data frame.
nlevels	Number of levels of a factor variable.
levels	Levels of factor variables in the data frame.
Missing	Number of missing observations in each variable.
MissingPer	Percent of missing observations in each variable.
MissingTotal	Total number of missing observations in the data frame.
MissingTotPer	Total percent of missing observations in the data frame.
MissingRows	Total number of rows with missing observations in the data frame.
MissingCols	Total number of columns with missing observations in the data frame.

Examples

```
# screen data
ds_screener(mtcarz)
ds_screener(airquality)

# plot
x <- ds_screener(airquality)
plot(x)
```

ds_skewness *Skewness*

Description

Compute the skewness of a probability distribution.

Usage

```
ds_skewness(data, x = NULL)
```

Arguments

data	A numeric vector or data.frame.
x	Column in data.

References

Sheskin, D.J. (2000) Handbook of Parametric and Nonparametric Statistical Procedures, Second Edition. Boca Raton, Florida: Chapman & Hall/CRC.

See Also

kurtosis

Examples

```
# vector
ds_skewness(mtcars$mpg)

# data.frame
ds_skewness(mtcars, mpg)
```

ds_std_error *Standard error of mean*

Description

Returns the standard error of mean.

Usage

```
ds_std_error(x)
```

Arguments

x	A numeric vector.

Examples

```
ds_std_error(mtcars$mpg)
```

---

ds_summary_stats *Descriptive statistics*

Description

Range of descriptive statistics for continuous data.

Usage

```
ds_summary_stats(data, ...)
```

Arguments

data	A data.frame or tibble.
...	Column(s) in data.

See Also

summary ds_freq_table ds_cross_table

Examples

```
# single variable
ds_summary_stats(mtcarz, mpg)

# multiple variables
ds_summary_stats(mtcarz, mpg, disp, hp)

# all variables
ds_summary_stats(mtcarz)
```

ds_tailobs *Tail Observations*

Description

Returns the n highest/lowest observations from a numeric vector.

Usage

```
ds_tailobs(data, n, type = c("low", "high"), decimals = 2)
```

Arguments

data	a numeric vector
n	number of observations to be returned
type	if low, the n lowest observations are returned, else the highest n observations are returned.
decimals	An option to specify the exact number of decimal places to use. The default number of decimal places is 2.

Details

Any NA values are stripped from data before computation takes place.

Value

n highest/lowest observations from data

See Also

top_n

Examples

```
# 5 lowest observations
ds_tailobs(mtcarz$mpg, 5)

# 5 highest observations
ds_tailobs(mtcarz$mpg, 5, type = "high")

# specify decimal places to display
ds_tailobs(mtcarz$mpg, 5, decimals = 3)
```

ds_tidy_stats *Tidy descriptive statistics*

Description

Descriptive statistics for multiple variables.

Usage

```
ds_tidy_stats(data, ...)
```

Arguments

data	A tibble or a data.frame.
...	Columns in x.

Value

A tibble.

Deprecated Functions

ds_multi_stats() have been deprecated. Instead use ds_tidy_stats().

Examples

```
# all columns
ds_tidy_stats(mtcarz)

# multiple columns
ds_tidy_stats(mtcarz, mpg, disp, hp)
```

hsb *High School and Beyond Data Set*

Description

A dataset containing demographic information and standardized test scores of high school students.

Usage

```
hsb
```

Format

A data frame with 200 rows and 10 variables:

id id of the student

female gender of the student

race ethnic background of the student

ses socio-economic status of the student

schtyp school type

prog program type

read scores from test of reading

write scores from test of writing

math scores from test of math

science scores from test of science

socst scores from test of social studies

Source

https://nces.ed.gov/surveys/hsb/

`mtcarz` *mtcarz*

Description

Copy of mtcars data set with modified variable types

Usage

```
mtcarz
```

Format

An object of class `data.frame` with 32 rows and 11 columns.

Index