Package 'summarytools'

August 11, 2014

Type Package

Title Dataframe Summaries, Frequency Tables and Numerical Summaries with Customizable Output

Version 0.1

Date 2014-08-06

Author Dominic Comtois

Maintainer Dominic Comtois <dominic.comtois@gmail.com>

Description

Includes 4 summarizing functions; two are to be used with uni-dimensional data, and the two others with dataframes. For uni-dimensional data, 'unistats' will generate common central tendency statistics and measures of dispersion for numerical data and 'frequencies' will generate a table of frequencies with counts and percentages (including cumulative). Both functions report number and proportion of valid values vs NA values. For dataframes, 'dfSummary' will generate a table containing as many rows as there are columns in the dataframe, each row giving variable information (class and type), labels if any, common statistics for numerical data and frequency tables for any type of data, along with number and proportion of valid (non-missing) values. Finally, 'properties' will conveniently report attributes of a dataframe object itself and also of all its individual components (columns).

Imports Hmisc, pander, timeDate

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2014-08-11 00:05:36

R topics documented:

`dfSummary` *Dataframe Summary*

Description

Summary of a dataframe consisting of: variable names and labels, factor levels, frequencies or numerical summary statistics, and valid/missing observations information. Optionally uses package 'pander' to display pretty tables.

Usage

```
dfSummary(x, echo=TRUE, style="grid", justify="left",
          max.distinct.values=10, str.distinct.values="distinct values",
          trim.strings=FALSE, max.string.width=15, round.digits=2,
          file=NA, display.labels=FALSE, ...)
```

Arguments

`x`	A dataframe.
`echo`	If TRUE (default), pander tables will be displayed and resulting dataframe will be returned silently.
`style`	The style to be used in pander table. Defaults to "grid".
`justify`	Pander argument. Defaults to "left".
`max.distinct.values`	The maximum number of items to be displayed in the frequency cell. If variable has more distinct values, no frequency will be shown (only a message stating the number of distinct values).
`str.distinct.values`	For internationalization; defaults to "distinct values".
`trim.strings`	For character variables, remove any white space at the beginning or end of the string. This will impact the frequencies so interpret the frequency tables accordingly. Defaults to FALSE.
`max.string.width`	Limits the number of characters to display in the frequency tables. Defaults to 15.
`round.digits`	Number of digits for rounding (used in numerical stats and in freq tables).
`file`	The text file to be written to disk. Defaults to NA.
`display.labels`	If TRUE,a column containing Hmisc-type labels will be displayed. Defaults to FALSE.
`...`	Additional arguments passed to pander().

Value

A dataframe containing as many rows as there are columns in x.

Author(s)

Dominic Comtois <dominic.comtois@gmail.com>

See Also

summary.data.frame

Examples

```
dfSummary(iris)
```

---

frequencies *Frequency Tables for Discrete Data and Factors.*

Description

Displays frequencies as well as valid/missing obs. information. Optionally uses package 'pander' to display pretty tables.

Usage

```
frequencies(x, round.digits=2, echo=TRUE, style="grid", justify="right",
            plain.ascii=TRUE, display.label=FALSE, ...)
```

Arguments

x	Atomic object of discrete (categorical) data.
round.digits	Number of sig. digits to keep. Defaults to 2.
echo	If TRUE, pander tables will be displayed. Defaults to TRUE.
style	Style of pander tables. Defaults to "grid".
justify	Pander argument. Defaults to "right".
plain.ascii	Pander argument. Defaults to TRUE.
display.label	If TRUE, Hmisc-type labels will be displayed. Defaults to FALSE.
...	Additional arguments passed to pander().

Value

A list containing the variable's name, label if applicable, frequency table and information on valid vs <NA> observations.

Author(s)

Dominic Comtois <dominic.comtois@gmail.com>

Examples

```
colors <- c("blue","red","blue","yellow",NA)
frequencies(colors)
frequencies(colors, echo=TRUE)
```

`properties` *Dataframe Attributes Along With Individual Column Attributes*

Description

Generates a table containing variable information: type, class and other attributes, label and first observation. Optionally uses package 'pander' to display pretty tables.

Usage

```
properties(x, echo=TRUE, style="grid", justify="left",
           plain.ascii=TRUE, display.labels=FALSE, ...)
```

Arguments

`x`	A dataframe
`echo`	If TRUE (default), pander tables will be displayed and resulting dataframe will be returned silently.
`style`	Pander argument. Defaults to "grid".
`justify`	Pander argument. Defaults to "left".
`plain.ascii`	Pander argument. Defaults to TRUE.
`display.labels`	If TRUE, a column containing Hmisc-type labels will be added. Defaults to FALSE.
`...`	Additional arguments to be passed to pander.

Value

A dataframe containing as many rows as there are columns in x.

Author(s)

Dominic Comtois <dominic.comtois@gmail.com>

Examples

```
properties(iris,echo=TRUE)
```

unistats *Univariate Statistics for Numerical Data*

Description

Calculates mean, standard deviation, min, max, median, IQR, skewness, str.err skewness, and kurtosis. Optionally uses package 'pander' to display pretty tables.

Usage

```
unistats(x, na.rm=TRUE, round.digits=2, echo=TRUE, style="grid",
         justify="right", plain.ascii=TRUE, display.label=FALSE, ...)
```

Arguments

x	Numerical vector.
na.rm	argument to be passed to statistical functions. Defaults to TRUE.
round.digits	Number of sig. digits to keep. Defaults to 2.
echo	If TRUE, pander tables will be displayed. Defaults to TRUE.
style	Style of pander tables. Defaults to "grid".
justify	Pander argument. Defaults to "right".
plain.ascii	Pander argument. Defaults to TRUE.
display.label	If TRUE, Hmisc-type labels will be displayed. Defaults to FALSE.
...	Additional arguments passed to pander().

Value

A list containing the variable's name, label if applicable, table of statistics and table of counts (valid vs <NA> observations).

Author(s)

Dominic Comtois <dominic.comtois@gmail.com>

Examples

```
data(cars)
unistats(cars$speed)
unistats(cars$speed, echo=FALSE)
```

Index