

Package ‘summarytools’

August 5, 2015

Type Package

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

Version 0.4

Date 2015-08-04

Author Dominic Comtois

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

Description Generates formatted summary tables for vectors and dataframes.

Three summarizing functions: 1) 'freq' generates frequency tables reporting counts and proportions (including cumulative); 2) 'descr' gives common central tendency statistics and measures of dispersion for numerical data; 3) 'dfSummary' gives as much information as possible about a dataframe and its components in a clear, legible table. All three functions can either display plain text tables in the console or write markdown, text or html files to disk. An additional misc function, 'what.is', extends base 'is' function, checking the object against all 'is' functions and gives a variety of other information about the object.

Depends htmltools

Imports methods, pryr, pander, xtable, rapporttools, rstudioapi, utils

License GPL-2

URL <https://github.com/dcomtois/summarytools>

NeedsCompilation no

Repository CRAN

Date/Publication 2015-08-05 09:30:57

R topics documented:

descr	2
dfSummary	3
examens	5

exams	5
freq	6
fumeurs	7
print.summarytools	8
smokers	9
view	9
what.is	10

Index	12
--------------	-----------

descr *Descriptive Univariate Statistics for Numerical Data*

Description

Calculates mean, standard deviation, min, max, median, MAD, IQR, CV, skewness, SE.skewness, and kurtosis.

Usage

```
descr(x, na.rm=TRUE, style="simple", round.digits=2,
      justify="right", plain.ascii=TRUE, file=NA,
      append=FALSE, transpose=FALSE,
      escape.pipe=FALSE, ...)
```

Arguments

x	Numerical vector.
na.rm	argument to be passed to statistical functions. Defaults to TRUE.
style	Style of pandor tables. Defaults to dQuotegrid.
round.digits	Number of sig. digits to keep. Defaults to 2.
justify	pandor argument. Defaults to “right”.
plain.ascii	pandor argument; when results are displayed in console, this avoids having markdown markup characters int the tables. Defaults to TRUE.
file	File name to write output to. Defaults to NA. To append output to an existing text file, use append=TRUE.
append	When “file” argument is supplied, this indicates whether to append output to existing file (TRUE) or to overwrite any existing file (FALSE, default). If TRUE and no file exists, a new file will be created.
transpose	Makes variables appears as columns, and stats as rows. Defaults to FALSE.
escape.pipe	Only useful when style='grid' and file argument is not NA, in which case it will escape the pipe character () to allow Pandoc to correctly convert multiline cells.
...	Additional arguments passed to pandor .

Value

A list with 2 elements; first one in a matrix (table) containing all the stats. Second, a table of statistics and table of counts (valid vs <NA> observations).

Author(s)

Dominic Comtois <dominic.comtois@gmail.com>

References

<https://github.com/dcomtois/summarytools>

Examples

```
data(exams)
descr(exams)
descr(exams, transpose=TRUE)
```

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.

Usage

```
dfSummary(x, style="multiline", justify="left",
          max.distinct.values=10, trim.strings=FALSE,
          max.string.width=15, round.digits=2,
          split.cells=40, display.labels=FALSE,
          file=NA, append=FALSE,
          escape.pipe=FALSE, ...)
```

Arguments

x	A dataframe.
style	The style to be used in pander table. Defaults to “multiline”. Alternatively, use “grid”; other styles (“simple” and “rmarkdown” are not supported.
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).

<code>trim.strings</code>	For character variables, remove any white space at the beginning or end of the string. This will impact the frequencies so interpret the frequencies cell accordingly. Defaults to FALSE.
<code>max.string.width</code>	Limits the number of characters to display in the frequency tables. Defaults to 15.
<code>round.digits</code>	Number of digits for rounding (used in numerical stats and in frequency tables).
<code>split.cells</code>	pander argument. Number of characters allowed on a line before splitting the cell. Defaults to 40.
<code>display.labels</code>	If TRUE, variable (as defined with rapportools or Hmisc 's label function) labels will be displayed. Defaults to FALSE.
<code>file</code>	The text file to be written to disk. Defaults to NA.
<code>append</code>	When "file" argument is supplied, this indicates whether to append output to existing file (TRUE) or to overwrite any existing file (FALSE, default). If TRUE and no file exists, a new file will be created.
<code>escape.pipe</code>	Only useful when <code>style='grid'</code> and <code>file</code> argument is not NA, in which case it will escape the pipe character (<code> </code>) to allow Pandoc to correctly convert multiline cells.
<code>...</code>	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>

References

<https://github.com/dcomtois/summarytools>

See Also

[summary.data.frame](#)

Examples

```
## Not run: view(dfSummary(iris))
data(smokers)
dfSummary(smokers)
```

examens	<i>examens sample data (French)</i>
---------	-------------------------------------

Description

This is merely a sample dataframe to allow users to experiment with summarytools. An English equivalent called 'exams' is also available.

Usage

```
data("examens")
```

Format

A data frame with 25 observations on the following 6 variables.

français a numeric vector

anglais a numeric vector

maths a numeric vector

geo a numeric vector

histoire a numeric vector

economie a numeric vector

Examples

```
data(examens)  
descr(examens)
```

exams	<i>exams sample data</i>
-------	--------------------------

Description

This is merely a sample dataframe to allow users to experiment with summarytools. A French equivalent called 'examens' is also available.

Usage

```
data("exams")
```

Format

A data frame with 25 observations on the following 6 variables.

french a numeric vector
 english a numeric vector
 math a numeric vector
 geo a numeric vector
 history a numeric vector
 economics a numeric vector

Examples

```
data(exams)
descr(exams)
```

freq	<i>Frequency Tables for Factors and other Discrete Data</i>
------	---

Description

Displays frequencies as well as valid/missing observation information.

Usage

```
freq(x, round.digits=2, style="simple", justify="right",
     plain.ascii=TRUE, file=NA, append=FALSE,
     escape.pipe=FALSE, ...)
```

Arguments

x	Vector of discrete (categorical) data.
round.digits	Number of sig. digits to keep in output. Defaults to 2.
style	Style of pandoc tables, one of “simple” (default), “multiline”, “grid” and “rmarkdown”.
justify	pander argument. Defaults to "right".
plain.ascii	pander argument; when results are displayed in console, this avoids having <i>markdown</i> markup characters in the tables. Defaults to TRUE.
file	File name to write output to. Defaults to NA. To append output to an existing text file, use append=TRUE.
append	When “file” argument is supplied, this indicates whether to append output to existing file (TRUE) or to overwrite any existing file (FALSE, default). If TRUE and no file exists, a new file will be created.
escape.pipe	Only useful when style='grid' and file argument is not NA, in which case it will escape the pipe character () to allow Pandoc to correctly convert multiline cells.
...	Additional arguments passed to pander .

Value

A frequency table of class matrix.

Author(s)

Dominic Comtois <dominic.comtois@gmail.com>

References

<https://github.com/dcomtois/summarytools>

Examples

```
data(smokers)
freq(smokers$gender)
freq(smokers$gender, style="rmarkdown")
```

fumeurs

fumeurs sample data (French)

Description

This is merely a sample dataframe to allow users to experiment with summarytools. An english equivalent called 'smokers' is also available.

Usage

```
data("fumeurs")
```

Format

A data frame with 300 observations on the following 5 variables.

sexe Factor with 2 levels, 'F' and 'H'

age Factor with 5 levels (age categories)

fumeur Factor with 2 levels (smoking status)

malade Factor with 2 levels (diseased)

imc Numerical vector (indice de masse corporelle / Body Mass Index)

Examples

```
data(fumeurs)
freq(fumeurs$sexe)
```

print.summarytools *Print method for objects of class summarytools.*

Description

Display summarytools objects in the console, in Web Browser or in *RStudio*'s Viewer.

Usage

```
## S3 method for class 'summarytools'  
print(x, method="pander", ...)
```

Arguments

x	summarytools object that was generated with freq, descr or dfSummary.
method	One of "pander", "viewer" or "browser". Defaults to "pander". If "viewer" is used outside RStudio, Web Browser will be used instead.
...	Additional arguments (not used for now).

Value

NULL when method="pander" ; a *file path* (returned invisibly) when method="viewer" or method="browser". In the latter case, the *html file* is passed to shell.exec which opens the file in default browser.

Author(s)

Dominic Comtois <dominic.comtois@gmail.com>

References

<https://github.com/dcomtois/summarytools>

Examples

```
data(iris)  
print(freq(iris$Species))
```

smokers	<i>smokers sample data</i>
---------	----------------------------

Description

This is merely a sample dataframe to allow users to experiment with summarytools. A French equivalent called 'fumeurs' is also available.

Usage

```
data("smokers")
```

Format

A data frame with 300 observations on the following 5 variables.

gender Factor with 2 levels, 'F' and 'M'

age Factor with 5 levels (age group categories)

smoker Factor with 2 levels (smoking status)

diseased Factor with 2 levels, 'diseased' and 'healthy'

bmi Numerical vector (Body Mass Index)

Examples

```
data(smokers)
freq(smokers$gender)
```

view	<i>Wrapper function around print.summarytools to facilitate generation of html files.</i>
------	---

Description

Takes a summarytools object as argument and calls its print method using method="viewer" by default. This generates an *html* file in the session's temporary directory and displays this file in *RStudio*'s Viewer pane, falling back to default browser when used outside *RStudio*.

Usage

```
view(x, method = "viewer", ...)
```

Arguments

x	A summarytools object created with one of freq, descr and dfSummary.
method	One of the 3 following: “viewer” (default), “browser”, “pander”. Note: To display x without any formatting at all, simply use as .table(x)
...	Not used for now.

Details

This function makes it more practical to generate html files with a single line of code, such as in `view(freq(iris$Species))`.

Value

For `method=="pander"`, nothing (NULL) is returned. With the two other methods (“viewer” and “browser”), returns the *html* file’s absolute path.

Author(s)

Dominic Comtois <dominic.comtois@gmail.com>

Examples

```
data(iris)
## Not run: view(dfSummary(iris))
## Not run: view(descr(iris,transpose=TRUE), "browser")
view(freq(iris$Species), "pander")
```

what.is

Obtain most common information on data structures.

Description

Takes any object as argument and returns (1) a dataframe with the class(es), type, mode and storage mode of the object as well as the dim, length and object.size; (2) a character vector of all the identifier functions (starting with "is.") that yield TRUE with the provided object; (3) a named character vector giving all attributes (for instance, "names", "row.names", "class", "dim", and so on) along with their lengths.

Usage

```
what.is(x, show.all = FALSE, ignore.size.warn = FALSE)
```

Arguments

<code>x</code>	Any object.
<code>show.all</code>	When TRUE, all logical results from the “is.” identifier functions will be displayed, accompanied by a warning when only the result applies only to the first element in the structure. FALSE by default.
<code>ignore.size.warn</code>	Set to TRUE to force execution of the function for large (>20 K bytes) objects. Defaults to FALSE.

Details

The function is an attempt to centralize the most common “macro-level” functions that describe the data at hand. Instead of calling, in turn, `class`, `typeof`, `dim`, and so on, a single call to `what.is` will readily give all that information.

Value

A list with 3 named items: “obj.properties” containing `class`, `typeof`, `mode`, `storage.mode`, `dim` and `length`; “identifiers” containing the “is.” identifier functions that yield TRUE; “attributes” containing the names and lengths of attributes; optionally, a 4th item with all the identifier test results (when argument `show.all=TRUE`)

Author(s)

Dominic Comtois <dominic.comtois@gmail.com>

Examples

```
what.is(1)
what.is(NaN)
what.is(iris3)
```

Index

- *Topic **RStudio viewer**
 - `print.summarytools`, 8
- *Topic **attributes**
 - `what.is`, 10
- *Topic **class**
 - `what.is`, 10
- *Topic **data.frame**
 - `dfSummary`, 3
- *Topic **datasets**
 - `examens`, 5
 - `exams`, 5
 - `fumeurs`, 7
 - `smokers`, 9
- *Topic **descriptives**
 - `descr`, 2
- *Topic **dim**
 - `what.is`, 10
- *Topic **formatted print**
 - `view`, 9
- *Topic **frequencies**
 - `freq`, 6
- *Topic **frequency**
 - `freq`, 6
- *Topic **html**
 - `print.summarytools`, 8
 - `view`, 9
- *Topic **is**
 - `what.is`, 10
- *Topic **length**
 - `what.is`, 10
- *Topic **mode**
 - `what.is`, 10
- *Topic **numerical summary**
 - `dfSummary`, 3
- *Topic **print**
 - `print.summarytools`, 8
- *Topic **storage.mode**
 - `what.is`, 10
- *Topic **stylized output**
 - `view`, 9
- *Topic **summary**
 - `dfSummary`, 3
- *Topic **type**
 - `what.is`, 10
- *Topic **univariate**
 - `descr`, 2
- `descr`, 2
- `dfSummary`, 3
- `examens`, 5
- `exams`, 5
- `freq`, 6
- `fumeurs`, 7
- `print (print.summarytools)`, 8
- `print.summarytools`, 8
- `smokers`, 9
- `summary.data.frame`, 4
- `view`, 9
- `what.is`, 10