Package 'tabulator'

October 14, 2022

Title Efficient Tabulation with Stata-Like Output

Version 1.0.0

Description Efficient tabulation with Stata-like output.
For each unique value of the variable, it shows the number of observations with that value, proportion of observations with that value, and cumulative proportion, in descending order of frequency.
Accepts data.table, tibble, or data.frame as input.
Efficient with big data: if you give it a data.table, tab() uses data.table syntax.

Imports assertthat, dplyr, data.table, magrittr, purrr, rlang, stats, stringr, tibble, tidyr

Depends R (>= 3.4.0)

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

NeedsCompilation no

Author Sean Higgins [aut, cre]

Maintainer Sean Higgins <sean.higgins@kellogg.northwestern.edu>

Repository CRAN

Date/Publication 2021-01-08 13:20:02 UTC

R topics documented:

quantiles *Efficient quantiles*

Description

Produces quantiles of the variables. `quantiles` shows quantile values. Efficient with big data: if you give it a `data.table`, `quantiles` uses `data.table` syntax.

Usage

```
quantiles(df, ..., probs = seq(0, 1, 0.1), na.rm = FALSE)
```

Arguments

`df`	A data.table, tibble, or data.frame.
`...`	A column or set of columns (without quotation marks).
`probs`	numeric vector of probabilities with values in [0,1].
`na.rm`	logical; if true, any NA and NaN's are removed from x before the quantiles are computed.

Value

Quantile values.

Examples

```
# data.table
library(data.table)
library(magrittr)
a <- data.table(varname = sample.int(20, size = 1000000, replace = TRUE))
a %>% quantiles(varname)

# data.table: look at top 10% in more detail
a %>% quantiles(varname, probs = seq(0.9, 1, 0.01))

# tibble
library(dplyr)
b <- tibble(varname = sample.int(20, size = 1000000, replace = TRUE))
b %>% quantiles(varname, na.rm = TRUE)
```

tab *Efficient tabulation*

Description

Produces a tabulation: for each unique group from the variable(s), tab shows the number of observations with that value, proportion of observations with that value, and cumulative proportion, in descending order of frequency. Accepts data.table, tibble, or data.frame as input. Efficient with big data: if you give it a data.table, tab uses data.table syntax.

Usage

```
tab(df, ..., by, round)
```

Arguments

df	A data.table, tibble, or data.frame.
...	A column or set of columns (without quotation marks).
by	A variable by which you want to group observations before tabulating (without quotation marks).
round	An integer indicating the number of digits for proportion and cumulative proportion.

Value

Tabulation (frequencies, proportion, cumulative proportion) for each unique value of the variables given in ... from df.

Examples

```
# data.table
library(data.table)
library(magrittr)
a <- data.table(varname = sample.int(20, size = 1000000, replace = TRUE))
a %>% tab(varname)

# tibble
library(dplyr)
b <- tibble(varname = sample.int(20, size = 1000000, replace = TRUE))
b %>% tab(varname, round = 1)

# data.frame
c <- data.frame(varname = sample.int(20, size = 1000000, replace = TRUE))
c %>% tab(varname)
```

tabcount *Count distinct categories*

Description

Produces a count of unique categories, tabcount shows the number of unique categories for the selected variable. Accepts data.table, tibble, or data.frame as input. Efficient with big data: if you give it a data.table, tabcount uses data.table syntax.

Usage

```
tabcount(df, ...)
```

Arguments

df	A data.table, tibble, or data.frame
...	A column or set of columns (without quotation marks)

Value

Count of the number of unique groups formed by the variables given in ... from df.

Examples

```
# data.table
library(data.table)
library(magrittr)
a <- data.table(varname = sample.int(20, size = 1000000, replace = TRUE))
a %>% tabcount(varname)

# tibble
library(dplyr)
b <- tibble(varname = sample.int(20, size = 1000000, replace = TRUE))
b %>% tabcount(varname)
```

Index