DEVELOPING R PACKAGES

Introduction to roxygen2

Aimée Gott
Education Practice Lead, Mango Solutions

Help files

sample_from_data {simutils} R Documentation

Sample from data

Description

Samples rows from a dataset.

Usage

```
sample_from_data(data, size, replace = TRUE)
```

Arguments

`data`	A data frame or matrix from which rows are to be sampled
`size`	Numeric. Number of rows to return
`replace`	Logical. Sample with replacement? TRUE by default.

Details

This function has been designed to sample from the rows of a two dimensional data set, returning all columns of sampled rows. Sampling is done with replacement by default.

Value

A data set of the same type as input with `size` rows.

Author(s)

Nic Crane

roxygen headers

```
#' Sample from data
#'
#' Samples rows from a dataset.
#'
#' This function has been designed to sample from the rows of a two dimensional
#' data set, returning all columns of sampled rows.  Sampling is done with replacement
#' by default.
#'
#' @param data A data frame or matrix from which rows are to be sampled
#' @param size Numeric. Number of rows to return
#' @param replace Logical. Sample with replacement? TRUE by default.
#'
#' @author Nic Crane
#'
#' @import dplyr
#'
#' @return A data set of the same type as input with \code{size} rows.
#' @export
#'
#' @examples
#' sample_from_data(airquality, size=10)
sample_from_data <- function(data, size, replace=TRUE) {

  if(!is.numeric(size)){
    stop("size must be a numeric value")
  }

  if(is.matrix(data)){
    data = as.data.frame(data)
  }
```

Title

```
#' Sample from data
#'
#' Samples rows from a dataset.
#'
#' This function has been designed to sample from the rows of a two dimensional
#' data set, returning all columns of sampled rows.  Sampling is done with replacement
#' by default.
#'
#' @param data A data frame or matrix from which rows are to be sampled
#' @param size Numeric. Number of rows to return
#' @param replace Logical. Sample with replacement? TRUE by default.
#'
#' @author Nic Crane
#'
#' @import dplyr
#'
#' @return A data set of the same type as input with \code{size} rows.
#' @export
#'
#' @examples
#' sample_from_data(airquality, size=10)
```

Description

```
#' Sample from data
#'
#' Samples rows from a dataset.
#'
#' This function has been designed to sample from the rows of a two dimensional
#' data set, returning all columns of sampled rows.  Sampling is done with replacement
#' by default.
#'
#' @param data A data frame or matrix from which rows are to be sampled
#' @param size Numeric. Number of rows to return
#' @param replace Logical. Sample with replacement? TRUE by default.
#'
#' @author Nic Crane
#'
#' @import dplyr
#'
#' @return A data set of the same type as input with \code{size} rows.
#' @export
#'
#' @examples
#' sample_from_data(airquality, size=10)
```

Details

```
#' Sample from data
#'
#' Samples rows from a dataset.
#'
#' This function has been designed to sample from the rows of a two dimensional
#' data set, returning all columns of sampled rows.  Sampling is done with replacement
#' by default.
#'
#' @param data A data frame or matrix from which rows are to be sampled
#' @param size Numeric. Number of rows to return
#' @param replace Logical. Sample with replacement? TRUE by default.
#'
#' @author Nic Crane
#'
#' @import dplyr
#'
#' @return A data set of the same type as input with \code{size} rows.
#' @export
#'
#' @examples
#' sample_from_data(airquality, size=10)
```

Arguments

```
#' Sample from data
#'
#' Samples rows from a dataset.
#'
#' This function has been designed to sample from the rows of a two dimensional
#' data set, returning all columns of sampled rows.  Sampling is done with replacement
#' by default.
#'
#' @param data A data frame or matrix from which rows are to be sampled
#' @param size Numeric. Number of rows to return
#' @param replace Logical. Sample with replacement? TRUE by default.
#'
#' @author Nic Crane
#'
#' @import dplyr
#'
#' @return A data set of the same type as input with \code{size} rows.
#' @export
#'
#' @examples
#' sample_from_data(airquality, size=10)
```

Imports

```
#' Sample from data
#'
#' Samples rows from a dataset.
#'
#' This function has been designed to sample from the rows of a two dimensional
#' data set, returning all columns of sampled rows.  Sampling is done with replacement
#' by default.
#'
#' @param data A data frame or matrix from which rows are to be sampled
#' @param size Numeric. Number of rows to return
#' @param replace Logical. Sample with replacement? TRUE by default.
#'
#' @author Nic Crane
#'
#' @import dplyr
#'
#' @return A data set of the same type as input with \code{size} rows.
#' @export
#'
#' @examples
#' sample_from_data(airquality, size=10)
```

DEVELOPING R PACKAGES

Let's practice!

DEVELOPING R PACKAGES

What Does Exporting a Function Mean and Why Do It?

Nic Crane
Data Science Consultant, Mango Solutions

Exported Functions

Exported functions:

- visible to the end user
- key package functionality

Non-exported functions:

- not visible to end user
- utility functions

Exported and Non Exported Functions

```
#' Count NAs in a vector
#'
#' @param x A vector
#'
#' @return Number of NAs in x
#'
#' @examples
#' sumNa(airquality$Ozone)
sum_na <- function(x) {
  sum(is.na(x))
}
```

Exported and Non Exported Functions

```
#' Count all NAs in a data set
#'
#' @param data A data frame or matrix
#'
#' @import purrr
#'
#' @return Vector of NA counts
#' @export
#'
#' @examples
#' na_counter(airquality)
#'
na_counter <- function(data) {

  stopifnot(is.matrix(data) | is.data.frame(data))

  if(is.matrix(data)){
    data = as.data.frame(data)
  }

  map_int(data, sum_na)
}
```

Exported and Non-Exported Functions

```
library(simutils)
na_counter(airquality)
```

```
Ozone Solar.R    Wind    Temp   Month     Day
   37       7       0       0       0       0
```

Calling Non-Exported Functions

```
library(simutils)
sum_na(airquality$Ozone)
```

```
Error: could not find function "sum_na"
```

Calling Non-Exported Functions

```
simutils:::sum_na(airquality$Ozone)
```

Exporting Functions with roxygen Headers

```
#' Count all NAs in a data set
#'
#' @param data A data frame or matrix
#'
#' @import purrr
#'
#' @return Vector of NA counts
#' @export
#'
#' @examples
#' na_counter(airquality)
#'
na_counter <- function(data) {

  stopifnot(is.matrix(data) | is.data.frame(data))

  if(is.matrix(data)){
    data = as.data.frame(data)
  }

  map_int(data, sum_na)
}
```

DEVELOPING R PACKAGES

Let's practice!

DEVELOPING R PACKAGES

What Other Elements Can We Document with roxygen Headers?

Aimée Gott
Education Practice Lead, Mango Solutions

Documenting Examples

```
#' Sample from data
#'
#' Samples rows from a dataset.
#'
#' This function has been designed to sample from the rows of a two dimensional
#' data set, returning all columns of sampled rows.  Sampling is done with replacement
#' by default.
#'
#' @param data A data frame or matrix from which rows are to be sampled
#' @param size Numeric. Number of rows to return
#' @param replace Logical. Sample with replacement? TRUE by default.
#'
#' @author Nic Crane
#'
#' @import dplyr
#'
#' @return A data set of the same type as input with \code{size} rows.
#' @export
#'
#' @examples
#' sample_from_data(airquality, size=10)
```

Non-Running Examples

```
#' Count NAs in a vector
#'
#' @param x A vector
#'
#' @return Number of NAs in x
#'
#' @examples
#' \dontrun{
#'   sum_na(airquality$Ozone)
#' }
sum_na <- function(x) {
  sum(is.na(x))
}
```

Documenting Function Return Values

```
#' Sample from data
#'
#' Samples rows from a dataset.
#'
#' This function has been designed to sample from the rows of a two dimensional
#' data set, returning all columns of sampled rows.  Sampling is done with replacement
#' by default.
#'
#' @param data A data frame or matrix from which rows are to be sampled
#' @param size Numeric. Number of rows to return
#' @param replace Logical. Sample with replacement? TRUE by default.
#'
#' @author Nic Crane
#'
#' @import dplyr
#'
#' @return A data set of the same type as input with \code{size} rows.
#' @export
#'
#' @examples
#' sample_from_data(airquality, size=10)
```

Additional Documentation

```
#' Sample from data
#'
#' Samples rows from a dataset.
#'
#' This function has been designed to sample from the rows of a two dimensional
#' data set, returning all columns of sampled rows.  Sampling is done with replacement
#' by default.
#'
#' @param data A data frame or matrix from which rows are to be sampled
#' @param size Numeric. Number of rows to return
#' @param replace Logical. Sample with replacement? TRUE by default.
#'
#' @author Nic Crane
#'
#' @import dplyr
#'
#' @return A data set of the same type as input with \code{size} rows.
#' @export
#'
#' @examples
#' sample_from_data(airquality, size=10)
```

DEVELOPING R PACKAGES

Let's practice!

DEVELOPING R PACKAGES

Documenting a Package

Nic Crane
Data Science Consultant, Mango Solutions

Package Level Documentation

```
#' simutils: A package for performing common simulation tasks
#'
#' This package provides functionality for a variety of simulation tasks,
#' and plotting tools for viewing the results.
#'
#' @author Nic Crane \email{ncrane@mango-solutions.com}
#' @docType package
#' @name simutils
"_PACKAGE"
```

Minimum Level of Documentation

For each function, document:

- Title
- Description
- Arguments
- Exported (for exported functions only)

Documenting Data Objects

```
use_data(sim_dat, pkg = "simutils")
```

Documenting Data Objects

```
#' sim_dat data set
#'
#' We made some data for the package
#'
#' @format A data.frame with 3 columns
#' \describe{
#' \item{ID}{ID value}
#' \item{Value}{Measured value in pounds}
#' \item{Apples}{Logical. Do they like apples}
#' }
#' @source Simulated Data
#'
"sim_dat"
```

Creating man Files

```
document("simutils")
```

DEVELOPING R PACKAGES

Let's practice!