Package 'thisutils'

July 20, 2025

Type Package

Title Collection of Utility Functions for Data Analysis and Computing

Version 0.0.7

Date 2025-07-19

Maintainer Meng Xu <mengxu98@qq.com>

Description Provides utility functions for data analysis and scientific computing. Includes functions for parallel processing, and other computational tasks to streamline workflows.

License MIT + file LICENSE

URL https://mengxu98.github.io/thisutils/

BugReports https://github.com/mengxu98/thisutils/issues

Depends R (>= 4.1.0)

Imports cli, doParallel, foreach, Matrix, methods, parallel, purrr, Rcpp, RcppArmadillo, RcppParallel, rlang, stats

Suggests httr2

LinkingTo Rcpp, RcppArmadillo, RcppParallel

Config/Needs/website mengxu98/thistemplate

Config/testthat/edition 3

Encoding UTF-8

RoxygenNote 7.3.2

Language en-US

NeedsCompilation yes

Author Meng Xu [aut, cre] (ORCID: <https://orcid.org/0000-0002-8300-1054>)

Repository CRAN

Date/Publication 2025-07-20 08:40:02 UTC

Contents

`thisutils-package` *Collection of Utility Functions for Data Analysis and Computing*

Description

Provides utility functions for data analysis and scientific computing. Includes functions for parallel processing, and other computational tasks to streamline workflows.

Author(s)

Meng Xu (Maintainer), <mengxu98@qq.com>

Source

https://mengxu98.github.io/thisutils/

See Also

Useful links:

- https://mengxu98.github.io/thisutils/
- Report bugs at https://github.com/mengxu98/thisutils/issues

---

`add_pkg_file` *Add package file*

Description

Automatically generate a file containing functions and related code for R package development.

Usage

```
add_pkg_file(
  desc_file,
  pkg_name = NULL,
  title = NULL,
  pkg_description = NULL,
  author_name = NULL,
  author_email = NULL,
  github_url = NULL,
  output_dir = NULL,
  use_figlet = TRUE,
  figlet_font = "Slant",
 colors = c("red", "yellow", "green", "magenta", "cyan", "yellow", "green", "white",
    "magenta", "cyan"),
  unicode = TRUE,
  verbose = TRUE
)
```

Arguments

`desc_file`	The DESCRIPTION file. Must be provided, it will be used to extract package information. Using `add_pkg_file("DESCRIPTION", output_dir = "R")`, will be created <pkg_name>-package.R based on the DESCRIPTION file in `R` dir. If you want to use some specific information, such as `author_name` or `author_email`, you can provide them manually.
`pkg_name`	Character string, the name of the package. Default is NULL, which will be read from DESCRIPTION file.
`title`	Character string, title of the package. Default is NULL, which will be read from DESCRIPTION file.
`pkg_description`	Character string, short description of the package. Default is NULL, which will be read from DESCRIPTION file.

author_name	Character string, name of the package author. Default is NULL, which will be read from DESCRIPTION file.
author_email	Character string, email of the package author. Default is NULL, which will be read from DESCRIPTION file.
github_url	Character string, GitHub URL of the package. Default is NULL, which will be read from DESCRIPTION file or constructed based on package name.
output_dir	Character string, directory where to save the package file. Default is NULL, you should specify it, such as 'R'.
use_figlet	Logical, whether to use figlet for ASCII art generation. Default is TRUE. Details see figlet.
figlet_font	Character string, figlet font to use. Default is "Slant".
colors	Character vector, colors to use for the logo elements.
unicode	Logical, whether to use Unicode symbols. Default is TRUE.
verbose	Logical, whether to print progress messages. Default is TRUE.

Value

Creates a file in specified output directory

---

as_matrix *Convert sparse matrix into dense matrix*

Description

Convert sparse matrix into dense matrix

Usage

```
as_matrix(x, parallel = FALSE, sparse = FALSE)
```

Arguments

x	A matrix.
parallel	Logical value, default is FALSE. Setting to parallelize the computation with RcppParallel::setThreadOptions.
sparse	Logical value, default is FALSE, whether to output a sparse matrix.

Value

A dense or sparse matrix

Examples

```
m <- simulate_sparse_matrix(
  1000, 1000,
  decimal = 3
)

system.time(
  a <- as.matrix(m)
)
system.time(
  b <- as_matrix(m)
)
system.time(
  c <- as_matrix(m, parallel = TRUE)
)
system.time(
  d <- as_matrix(m, sparse = TRUE)
)

m[1:5, 1:5]
a[1:5, 1:5]
b[1:5, 1:5]
c[1:5, 1:5]

identical(a, b)
identical(a, c)
identical(b, c)
identical(a, d)
identical(b, d)
```

`capitalize` *Capitalizes the characters Making the first letter uppercase*

Description

Capitalizes the characters Making the first letter uppercase

Usage

```
capitalize(x, force_tolower = FALSE)
```

Arguments

`x`	A vector of character strings to be capitalized.
`force_tolower`	Whether to force the remaining letters to be lowercase.

Examples

```
x <- c(
  "hello world",
  "Hello world",
  "hello World"
)
capitalize(x)
```

`check_sparsity` *Check sparsity of matrix*

Description

Check sparsity of matrix

Usage

```
check_sparsity(x)
```

Arguments

`x`	A matrix.

Value

Sparsity of matrix

`download` *Download file from the Internet*

Description

Download file from the Internet

Usage

```
download(
  url,
  destfile,
  methods = c("auto", "wget", "libcurl", "curl", "wininet", "internal"),
  quiet = FALSE,
  ...,
  max_tries = 2
)
```

Arguments

`url`	a `character` string (or longer vector for the `"libcurl"` method) naming the URL of a resource to be downloaded.
`destfile`	a character string (or vector, see the `url` argument) with the file path where the downloaded file is to be saved. Tilde-expansion is performed.
`methods`	Methods to be used for downloading files. The default is to try different download methods in turn until the download is successfully completed.
`quiet`	If `TRUE`, suppress status messages (if any), and the progress bar.
`...`	Other arguments passed to utils::download.file.
`max_tries`	Number of tries for each download method.

`figlet` *figlet*

Description

Create ASCII art text using figlet fonts.

Usage

```
figlet(
  text,
  font = "Slant",
  width = getOption("width", 80),
  justify = "left",
  absolute = FALSE,
  strip = TRUE
)
```

Arguments

`text`	Text to make bigger
`font`	Name of font, path to font, or ‘figlet_font‘ object
`width`	Width to use when justifying and breaking lines
`justify`	Text justification to use in rendering ("left", "centre", "right")
`absolute`	Logical, indicating if alignment is absolute
`strip`	Logical, indicating if whitespace should be removed

Value

An object of class ‘figlet_text‘ which is a character vector with a handy print method

References

http://www.figlet.org/ https://github.com/richfitz/rfiglet https://github.com/jbkunst/figletr

Examples

```
figlet("thisutils")
```

---

figlet_font *Get a figlet font*

Description

Get a figlet font

Usage

```
figlet_font(font)
```

Arguments

font	Path or name of the font to load

Value

A 'figlet_font' object for use with [figlet]

---

invoke_fun *Invoke a function with a list of arguments*

Description

Invoke a function with a list of arguments

Usage

```
invoke_fun(.fn, .args = list(), ..., .env = rlang::caller_env())
```

Arguments

.fn	A function, or function name as a string.
.args	A list of arguments.
...	Other arguments passed to the function.
.env	Environment in which to evaluate the call. This will be most useful if .fn is a string, or the function has side-effects.

Examples

```
f <- function(x, y) {
  x + y
}
invoke_fun(f, list(x = 1, y = 2))
invoke_fun("f", list(x = 1, y = 2))
invoke_fun("f", x = 1, y = 2)
```

---

list_figlet_fonts *List available figlet fonts*

Description

List all figlet font files available in the package or system.

Usage

```
list_figlet_fonts()
```

Value

Character vector of available font names.

Examples

```
list_figlet_fonts()
```

---

log_message *Print diagnostic message*

Description

Integrate the message printing function with the `cli` package, and the base::message function. The message could be suppressed by base::suppressMessages.

Usage

```
log_message(
  ...,
  verbose = TRUE,
  message_type = c("info", "success", "warning", "error"),
  cli_model = TRUE,
  timestamp = TRUE,
  timestamp_format = "%Y-%m-%d %H:%M:%S",
  level = 1,
  symbol = "  ",
```

```
  text_color = NULL,
  back_color = NULL,
  multiline_indent = TRUE,
  .envir = parent.frame(),
  .frame = .envir
)
```

Arguments

`...`	Text to print.
`verbose`	Logical value, default is `TRUE`. Whether to print the message.
`message_type`	Type of message, default is `info`. Could be choose one of `info`, `success`, `warning`, and `error`.
`cli_model`	Logical value, default is `TRUE`. Whether to use the `cli` package to print the message.
`timestamp`	Logical value, default is `TRUE`. Whether to show the current time in the message.
`timestamp_format`	Character value, default is `"%Y-%m-%d %H:%M:%S"`. Format string for timestamp display.
`level`	Integer value, default is `1`. The level of the message, which affects the indentation. Level 1 has no indentation, higher levels add more indentation.
`symbol`	Character value, default is `" "` (two spaces). The symbol used for indentation. When specified, it ignores the level parameter and uses the symbol directly.
`text_color`	Character value, default is `NULL`. Color for the message text. Available colors: "red", "green", "blue", "yellow", "magenta", "cyan", "white", "black".
`back_color`	Character value, default is `NULL`. Background color for the message text. Available colors: "red", "green", "blue", "yellow", "magenta", "cyan", "white", "black".
`multiline_indent`	Logical value, default is `TRUE`. Whether to apply consistent formatting (timestamp and indentation) to each line in multiline messages. When TRUE, each line gets the full formatting; when FALSE, only the first line gets the timestamp.
`.envir`	The environment to evaluate calls in. Default is `parent.frame()`.
`.frame`	The frame to use for error reporting. Default is `.envir`.

Value

`Formated message` printed to the console.

References

https://cli.r-lib.org/articles/index.html

Examples

```
# basic usage
log_message("Hello, ", "world!")
log_message("hello, world!")
log_message("Hello, world!", timestamp = FALSE)
log_message("Hello, ", "world!", message_type = "success")
log_message("Hello, world!", message_type = "warning")
log_message("Hello, ", "world!", cli_model = FALSE)

# suppress messages
suppressMessages(log_message("Hello, world!"))
log_message("Hello, world!", verbose = FALSE)
options(log_message.verbose = FALSE)
log_message("Hello, world!")
options(log_message.verbose = TRUE)

# cli formatting
log_message("hello, {.arg world}!")
message("hello, {.arg world}!")
log_message("hello, {.code world}!")
message("hello, {.code world}!")
log_message("{.emph R}")
log_message("Hello, {.file world}!")
log_message("press {.kbd ENTER}")
log_message("address: {.email example@example.com}")
log_message("URL: {.url https://example.com}")
log_message("Some {.field field}")
log_message("Hello, {.pkg world}!")
log_message("Hello, {.val world}!")

# set indentation
log_message("Hello, world!", level = 2)
log_message("Hello, world!", symbol = "->")

# multiline message
log_message("Line 1\nLine 2\nLine 3", multiline_indent = TRUE)
log_message(
  "Multi-line\ncolored\nmessage",
  text_color = "blue",
  multiline_indent = FALSE
)
log_message(
  "Multi-line\ncolored\nmessage",
  text_color = "blue",
  multiline_indent = FALSE,
  timestamp = FALSE
)

# color formatting
log_message(
  "This is a red message",
  text_color = "red"
```

```
)
log_message(
  "This is a message with background",
  back_color = "cyan"
)
log_message(
  "This is a message with both text and background",
  text_color = "red",
  back_color = "cyan"
)
# color formatting also works with `cli_model = FALSE`
log_message(
  "This is a red message",
  text_color = "red",
  cli_model = FALSE
)
log_message(
  "This is a message with background",
  back_color = "cyan",
  cli_model = FALSE
)
log_message(
  "This is a message with both text and background",
  text_color = "red",
  back_color = "cyan",
  cli_model = FALSE
)

# cli variables
fun <- function(x) {
  log_message("{.val x}")
  log_message("{.val {x}}")
  log_message("{.val {x + 1}}")
}
fun(1)
```

`normalization` *Normalize numeric vector*

Description

Normalize numeric vector

Usage

```
normalization(x, method = "max_min", na_rm = TRUE, ...)
```

Arguments

`x`	Input numeric vector.

method	Method used for normalization.
na_rm	Whether to remove NA values, and if setting TRUE, using 0 instead.
...	Parameters for other methods.

Value

Normalized numeric vector

Examples

```
x <- c(runif(2), NA, -runif(2))
x
normalization(x, method = "max_min")
normalization(x, method = "maximum")
normalization(x, method = "sum")
normalization(x, method = "softmax")
normalization(x, method = "z_score")
normalization(x, method = "mad")
normalization(x, method = "unit_vector")
normalization(x, method = "unit_vector", na_rm = FALSE)
```

---

parallelize_fun *Parallelize a function*

Description

Parallelize a function

Usage

```
parallelize_fun(x, fun, cores = 1, export_fun = NULL, verbose = TRUE)
```

Arguments

x	A vector or list to apply over.
fun	The function to be applied to each element.
cores	The number of cores to use for parallelization with foreach::foreach. Default is 1.
export_fun	The functions to export the function to workers.
verbose	Logical value, default is TRUE. Whether to print progress messages.

Value

A list of computed results

Examples

```
parallelize_fun(1:3, function(x) {
  Sys.sleep(0.2)
  x^2
})

parallelize_fun(list(1, 2, 3), function(x) {
  Sys.sleep(0.2)
  x^2
}, cores = 2)
```

`pearson_correlation` *Correlation and covariance calculation for sparse matrix*

Description

Correlation and covariance calculation for sparse matrix

Usage

```
pearson_correlation(x, y = NULL)
```

Arguments

x	Sparse matrix or character vector.
y	Sparse matrix or character vector.

Value

A list with covariance and correlation matrices.

Examples

```
m1 <- simulate_sparse_matrix(
  100, 100
)
m2 <- simulate_sparse_matrix(
  100, 100,
  sparsity = 0.05
)
a <- pearson_correlation(m1, m2)
a$cov[1:5, 1:5]
a$cor[1:5, 1:5]
```

print.thisutils_logo *Print logo*

Description

Print logo

Usage

```
## S3 method for class 'thisutils_logo'
print(x, ...)
```

Arguments

x	Input information.
...	Other parameters.

Value

Print the ASCII logo

remove_space *Remove and normalize spaces*

Description

Remove leading/trailing spaces and normalize multiple spaces between words in character strings.

Usage

```
remove_space(
  x,
  trim_start = TRUE,
  trim_end = FALSE,
  collapse_multiple = TRUE,
  preserve_newlines = TRUE
)
```

Arguments

x	A vector of character strings.
trim_start	Logical value, default is 'TRUE'. Whether to remove leading spaces before the first word.
trim_end	Logical value, default is 'FALSE'. Whether to remove trailing spaces after the last word.

collapse_multiple
: Logical value, default is 'TRUE'. Whether to collapse multiple consecutive spaces between words into a single space.

preserve_newlines
: Logical value, default is 'TRUE'. Whether to preserve newline characters when collapsing spaces.

Value

A character vector with spaces normalized according to the specified parameters.

Examples

```
x <- c(
  " hello  world ",
  "  test   case  ",
  "no space",
  "   multiple   spaces   "
)
remove_space(x)
remove_space(x, trim_start = FALSE)
remove_space(x, trim_end = TRUE)
remove_space(x, collapse_multiple = FALSE)
remove_space(
  x,
  trim_start = FALSE,
  trim_end = FALSE,
  collapse_multiple = FALSE
)

# with newlines
multiline <- c(
  "hello\n\n  world  ",
  "  first  \n  second  "
)
remove_space(multiline)
remove_space(multiline, preserve_newlines = FALSE)
```

---

rescale *Rescale numeric vector*

Description

Rescale numeric vector

Usage

```
rescale(x, from = range(x, na.rm = TRUE, finite = TRUE), to = c(0, 1))
```

Arguments

x	A numeric vector.
from	The range of the original data.
to	The range of the rescaled data.

Value

A numeric vector with rescaled values.

Examples

```
x <- rnorm(10)
rescale(x)
rescale(x, from = c(0, 1))
rescale(x, to = c(0, 2))
```

---

r_square *coefficient of determination (R^2)*

Description

coefficient of determination (R^2)

Usage

```
r_square(y_true, y_pred)
```

Arguments

y_true	A numeric vector with ground truth values.
y_pred	A numeric vector with predicted values.

Value

R^2 value

Examples

```
y <- rnorm(100)
y_pred <- y + rnorm(100, sd = 0.5)
r_square(y, y_pred)
```

`simulate_sparse_matrix` *Generate a simulated sparse matrix*

Description

This function generates a sparse matrix with a specified number of rows and columns, a given sparsity level, and a distribution function for the non-zero values.

Usage

```
simulate_sparse_matrix(
  nrow,
  ncol,
  sparsity = 0.95,
  distribution_fun = function(n) stats::rpois(n, lambda = 0.5) + 1,
  decimal = 0,
  seed = 1
)
```

Arguments

`nrow`	Number of rows in the matrix.
`ncol`	Number of columns in the matrix.
`sparsity`	Proportion of zero elements (sparsity level). Default is 0.95, meaning 95% of elements are zero (5% are non-zero).
`distribution_fun`	Function to generate non-zero values.
`decimal`	Numeric value, default is `0`. Controls the number of decimal places in the generated values. If set to `0`, values will be integers. When decimal > 0, random decimal parts are uniformly distributed across the full range.
`seed`	Random seed for reproducibility.

Value

A sparse matrix of class "dgCMatrix"

Examples

```
simulate_sparse_matrix(1000, 500) |>
  check_sparsity()

simulate_sparse_matrix(10, 10, decimal = 1)
simulate_sparse_matrix(10, 10, decimal = 5)
```

`sparse_cor` *A sparse correlation function*

Description

Safe correlation function which returns a sparse matrix without missing values

Usage

```
sparse_cor(
  x,
  y = NULL,
  method = "pearson",
  allow_neg = TRUE,
  remove_na = TRUE,
  remove_inf = TRUE,
  ...
)
```

Arguments

`x`	Sparse matrix or character vector.
`y`	Sparse matrix or character vector.
`method`	Method to use for calculating the correlation coefficient.
`allow_neg`	Logical. Whether to allow negative values or set them to 0.
`remove_na`	Logical. Whether to replace NA values with 0.
`remove_inf`	Logical. Whether to replace infinite values with 1.
`...`	Other arguments passed to stats::cor function.

Value

A correlation matrix.

Examples

```
m1 <- simulate_sparse_matrix(
  500, 100
)
m2 <- simulate_sparse_matrix(
  500, 100,
  seed = 2025
)
a <- sparse_cor(m1)
b <- sparse_cor(m1, m2)
c <- as_matrix(
  cor(as_matrix(m1)),
```

```
  sparse = TRUE
)
d <- as_matrix(
  cor(as_matrix(m1), as_matrix(m2)),
  sparse = TRUE
)

a[1:5, 1:5]
c[1:5, 1:5]
all.equal(a, c)

b[1:5, 1:5]
d[1:5, 1:5]
all.equal(b, d)

m1[sample(1:500, 10)] <- NA
m2[sample(1:500, 10)] <- NA

sparse_cor(m1, m2)[1:5, 1:5]

system.time(
  sparse_cor(m1)
)
system.time(
  cor(as_matrix(m1))
)

system.time(
  sparse_cor(m1, m2)
)
system.time(
  cor(as_matrix(m1), as_matrix(m2))
)
```

`split_indices` *Split indices.*

Description

An optimised version of split for the special case of splitting row indices into groups.

Usage

```
split_indices(group, n = 0L)
```

Arguments

`group`	Integer indices
`n`	The largest integer (may not appear in index). This is hint: if the largest value of `group` is bigger than `n`, the output will silently expand.

Value

A list of vectors of indices.

References

https://github.com/hadley/plyr/blob/d57f9377eb5d56107ba3136775f2f0f005f33aa3/src/split-numeric.cpp#L20

Examples

```
split_indices(sample(10, 100, rep = TRUE))
split_indices(sample(10, 100, rep = TRUE), 10)
```

---

str_wrap *Wrap text*

Description

Wrap text

Usage

```
str_wrap(x, width = 80)
```

Arguments

x	A character vector to wrap.
width	The maximum width of the lines.

Value

A character vector with wrapped text.

Examples

```
str_wrap(rep("Hello, world!", 10))
str_wrap(rep("Hello, world!", 10), width = 10)
```

thisutils_logo *The logo of thisutils*

Description

The thisutils logo, using ASCII or Unicode characters Use cli::ansi_strip to get rid of the colors.

Usage

```
thisutils_logo(unicode = cli::is_utf8_output())
```

Arguments

unicode	Unicode symbols on UTF-8 platforms. Default is cli::is_utf8_output.

References

https://github.com/tidyverse/tidyverse/blob/main/R/logo.R

Examples

```
thisutils_logo()
```

try_get *Try to evaluate an expression a set number of times before failing*

Description

The function is used as a fail-safe if your R code sometimes works and sometimes doesn't, usually because it depends on a resource that may be temporarily unavailable. It tries to evaluate the expression 'max_tries' times. If all the attempts fail, it throws an error; if not, the evaluated expression is returned.

Usage

```
try_get(expr, max_tries = 5, error_message = "", retry_message = "Retrying...")
```

Arguments

expr	The expression to be evaluated.
max_tries	The maximum number of attempts to evaluate the expression before giving up. Default is set to 5.
error_message	a string, additional custom error message you would like to be displayed when an error occurs.
retry_message	a string, a message displayed when a new try to evaluate the expression would be attempted.

Value

This function returns the evaluated expression if successful, otherwise it throws an error if all attempts are unsuccessful.

Examples

```
f <- function() {
  value <- runif(1, min = 0, max = 1)
  if (value > 0.5) {
    log_message("value is larger than 0.5")
    return(value)
  } else {
    log_message(
      "value is smaller than 0.5",
      message_type = "error"
    )
  }
}
f_evaluated <- try_get(expr = f())
print(f_evaluated)
```

---

unnest_fun *Unnest a list-column*

Description

Implement similar functions to the tidyr::unnest function.

Usage

```
unnest_fun(data, cols, keep_empty = FALSE)
```

Arguments

data	A data frame.
cols	Columns to unnest.
keep_empty	By default, you get one row of output for each element of the list your unchopping/unnesting. This means that if there's a size-0 element (like NULL or an empty data frame), that entire row will be dropped from the output. If you want to preserve all rows, use keep_empty = TRUE to replace size-0 elements with a single row of missing values.

Examples

```
data <- data.frame(
  id = 1:3,
  x = c("a", "b", "c"),
  stringsAsFactors = FALSE
)
data$data <- list(
  c(1, 2),
  c(3, 4, 5),
  c(6)
)
unnest_fun(data, cols = "data")

data2 <- data.frame(
  id = 1:3,
  x = c("a", "b", "c"),
  stringsAsFactors = FALSE
)
data2$data <- list(
  c(1, 2),
  numeric(0),
  c(6)
)
unnest_fun(data2, cols = "data")
unnest_fun(data2, cols = "data", keep_empty = TRUE)
```

%ss% *Value selection operator*

Description

This operator returns the left side if it's not NULL, otherwise it returns the right side.

Usage

```
a %ss% b
```

Arguments

a	The left side value to check
b	The right side value to use if a is NULL

Value

a if it is not NULL, otherwise b

Examples

```
NULL %ss% 10
5 %ss% 10
```

Index