

Package ‘tidyimpute’

March 5, 2018

Title Imputation the Tidyverse Way

Version 0.1.0

Date 2018-02-01

URL <https://github.com/decisionpatterns/tidyimpute>

Description Functions and methods for imputing missing values (NA) in tables and list patterned after the tidyverse approach of 'dplyr' and 'rlang'; works with data.tables as well.

BugReports <https://github.com/decisionpatterns/tidyimpute/issues>

Depends R (>= 3.1.0)

Imports methods, dplyr (>= 0.7.2), rlang (>= 0.1.2), na.tools (>= 0.1.0)

Suggests testthat (>= 1.0.2), data.table (>= 1.10), magrittr

License GPL-3 | file LICENSE

LazyData true

RoxygenNote 6.0.1.9000

Repository CRAN

Encoding UTF-8

Collate 'data.R' 'drop_cols.R' 'drop_rows.R' 'impute.R'
'make_impute.R' 'utils.R' 'impute_funs.R' 'na_predict.R'
'zzz.R'

NeedsCompilation no

Author Christopher Brown [aut, cre],
Decision Patterns [cph]

Maintainer Christopher Brown <chris.brown@decisionpatterns.com>

Date/Publication 2018-03-05 11:46:49 UTC

R topics documented:

| | |
|----------------------------|----------|
| drop_cols_all_na | 2 |
| drop_rows_all_na | 3 |
| impute | 4 |
| impute_functions | 6 |
| nacars | 7 |
| na_predict | 7 |
| Index | 8 |

| | |
|------------------|---|
| drop_cols_all_na | <i>Remove columns with missing values</i> |
|------------------|---|

Description

Remove columns of a table whose values are all NA or who have any NA

Usage

```
drop_cols_all_na(.tbl)
```

```
drop_cols_any_na(.tbl)
```

```
drop_na_cols(.tbl)
```

Arguments

.tbl table-like object

Details

drop_cols_all_na removes all cols whose only values are NA. drop_cols_any_na removes columns that have any NA. They work on all table-like objects.

Value

An object of the same class as data with cols containing all NA values removed

See Also

- [dplyr::select\(\)](#)

drop_rows_all_na *drop_rows_all_na, drop_rows_any_na*

Description

Drop rows of a table whose values are all NA

Usage

```
drop_rows_all_na(.tbl)
```

```
filter_all_na(.tbl)
```

```
drop_rows_any_na(.tbl)
```

```
filter_any_na(.tbl)
```

Arguments

`.tbl` data-like object

Details

`na_drop_rows` removes all rows whose only values are NA. It works for all table-like objects.

Value

An object of the same class as `.tbl` with rows containing all NA values removed

See Also

- [dplyr::filter\(\)](#)

Examples

```
data(iris)

.tbl <- iris[1:5,]
.tbl[1:2,] <- NA
.tbl[3,1] <- NA
.tbl

filter_all_na(.tbl)
filter_any_na(.tbl)

drop_rows_all_na(.tbl)
drop_rows_any_na(.tbl)
```

impute *Replace missing values in tables and lists*

Description

Replace missing values (NA) in a table and lists

Usage

```
impute(.tbl, .na, ...)
impute_at(.tbl, .na, .vars, ...)
impute_all(.tbl, .na, ...)
impute_if(.tbl, .na, .predicate, ...)
```

Arguments

| | |
|-------------------------|--|
| <code>.tbl</code> | list-like or table-like structure. |
| <code>.na</code> | scalar, vector or function as described in na.tools::na.replace() |
| <code>...</code> | additional args; either a unnamed list of columns (quoted or not) or name=function pairs. See Details. |
| <code>.vars</code> | character; names of columns to be imputed |
| <code>.predicate</code> | dplyr-type predicate functions |

Details

`impute` is similar to other *dplyr* verbs especially `dplyr::mutate()`. Like `dplyr::mutate()` it operates on columns. It changes only missing values (NA) to the value specified by `.na`.

Behavior:

Behavior depends on the values of `.na` and `...`

`impute` can be used for three replacement operations:

1. `impute(.tbl, .na) : (missing ...)` Replace missing values in **ALL COLS** by `.na`. This is analogous to `impute_all`.
2. `impute(.tbl, .na, ...)` : (`...` is an unnamed list) Replace column(s) specified in `...` by `.na`. Columns are specified as an unnamed list of quoted or unquoted column names. This is analogous to `impute_at` where `...` specifies `.vars`
3. `impute(.tbl, col1=na.*, col2=na.*) : (missing .na)` : Replace by column-specific `.na`

Additional arguments are to `.na` are not used; Use `impute_at` for this or create your own lambda functions.

`impute_all` is like `impute` without specifying `...` `...` is used for additional arguments to `.na`

Value

Returns a object as the same type as `.tbl`. Columns are mutated to replace missing values (NA) with value specied by `.na` and `...`

Note

`...` is used to specify columns in `impute` but is used as additional arguments to `.na` in the other `impute_*` functions.

See Also

- The `na.tools` package.
- `impute_functions`

Examples

```

data(nacars)

## Not run:
nacars %>% impute(0, mpg, cyl)
nacars %>% impute(1:6, mpg, cyl)

nacars %>% impute( na.mean )
nacars %>% impute( mean )      # unsafe
nacars %>% impute( length, mpg, disp )
nacars %>% impute( mean, mpg, disp )
nacars %>% impute( mpg=na.mean, cyl=na.max )
nacars %>% impute( na.mean, c('mpg','disp') )

## End(Not run)

## Not run:
nacars %>% impute_at( -99, .vars=1:3 )
nacars %>% impute_at( .na=na.mean, .vars=1:6 )

# Same, uses `...` for additional args
nacars %>%
  impute_at( .na=mean , .vars=1:6, na.rm = TRUE )

nacars %>% impute_at( .na=na.mean, .vars = c('mpg','cyl', 'disp') )

## End(Not run)

## Not run:
nacars %>% impute_all( -99 )
nacars %>% impute_all( na.min )

## End(Not run)

```

| | |
|------------------|---------------------------------|
| impute_functions | <i>Table imputation methods</i> |
|------------------|---------------------------------|

Description

Replace missing value methods with a variety of methods

Usage

```
impute_functions(.tbl, .na, .vars, .predicate)
```

Arguments

| | |
|-------------------------|--|
| <code>.tbl</code> | table-like or list-like structure |
| <code>.na</code> | value/function to be used for replacement |
| <code>.vars</code> | list of columns generated by <code>vars()</code> , or a character vector of column names, or a numeric vector of column positions. |
| <code>.predicate</code> | A predicate function to be applied to the columns or a logical vector. |
| <code>...</code> | addition passed to the imputation method |

Details

These methods are modelled closely after `dplyr::mutate()` and the select style verbs. Most of the functions depend on the **na.tools** package.

Function List:

- explicit: `impute_explicit`, `impute_explicit_at`, `impute_explicit_all`, `impute_explicit_if`
- zero: `impute_zero`, `impute_zero_at`, `impute_zero_all`, `impute_zero_if`
- inf: `impute_inf`, `impute_inf_at`, `impute_inf_all`, `impute_inf_if`
- neginf: `impute_neginf`, `impute_neginf_at`, `impute_neginf_all`, `impute_neginf_if`
- constant: `impute_constant`, `impute_constant_at`, `impute_constant_all`, `impute_constant_if`
- max: `impute_max`, `impute_max_at`, `impute_max_all`, `impute_max_if`
- min: `impute_min`, `impute_min_at`, `impute_min_all`, `impute_min_if`
- median: `impute_median`, `impute_median_at`, `impute_median_all`, `impute_median_if`
- mean: `impute_mean`, `impute_mean_at`, `impute_mean_all`, `impute_mean_if`
- most_freq: `impute_most_freq`, `impute_most_freq_at`, `impute_most_freq_all`, `impute_most_freq_if`
- quantile: `impute_quantile`, `impute_quantile_at`, `impute_quantile_all`, `impute_quantile_if`
- sample: `impute_sample`, `impute_sample_at`, `impute_sample_all`, `impute_sample_if`
- random: `impute_random`, `impute_random_at`, `impute_random_all`, `impute_random_if`
- replace: `impute_replace`, `impute_replace_at`, `impute_replace_all`, `impute_replace_if`

Examples

```
## Not run:
nacars %>% impute_zero()
nacars %>% impute_zero( mpg, cyl )
nacars %>% impute_zero( "mpg", "cyl" )
nacars %>% impute_zero( c("mpg","cyl") )
nacars %>% impute_zero( 1:2 )

## End(Not run)
```

| | |
|--------|---------------------------------|
| nacars | <i>data with missing values</i> |
|--------|---------------------------------|

Description

data with missing values

Usage

nacars

Format

An object of class `data.frame` with 6 rows and 11 columns.

Details

cars and **iris** data sets with missing data for demonstration purposes.

| | |
|------------|-------------------|
| na_predict | <i>na_predict</i> |
|------------|-------------------|

Description

replace NA values by predictions of a model

Usage

```
na_predict(x, object, data = x)
```

Arguments

| | |
|--------|----------------------------|
| x | data |
| object | object with predict method |
| data | data object |

Index

*Topic **datasets**

 nacars, 7

dplyr::filter(), 3
dplyr::mutate(), 4, 6
dplyr::select(), 2
drop_cols_all_na, 2
drop_cols_any_na (drop_cols_all_na), 2
drop_na_cols (drop_cols_all_na), 2
drop_rows_all_na, 3
drop_rows_any_na (drop_rows_all_na), 3

filter_all_na (drop_rows_all_na), 3
filter_any_na (drop_rows_all_na), 3

impute, 4
impute_all (impute), 4
impute_at (impute), 4
impute_constant (impute_functions), 6
impute_constant_all (impute_functions),
 6
impute_constant_at (impute_functions), 6
impute_constant_if (impute_functions), 6
impute_explicit (impute_functions), 6
impute_explicit_all (impute_functions),
 6
impute_explicit_at (impute_functions), 6
impute_explicit_if (impute_functions), 6
impute_functions, 6
impute_if (impute), 4
impute_inf (impute_functions), 6
impute_inf_all (impute_functions), 6
impute_inf_at (impute_functions), 6
impute_inf_if (impute_functions), 6
impute_max (impute_functions), 6
impute_max_all (impute_functions), 6
impute_max_at (impute_functions), 6
impute_max_if (impute_functions), 6
impute_mean (impute_functions), 6
impute_mean_all (impute_functions), 6
impute_mean_at (impute_functions), 6
impute_mean_if (impute_functions), 6
impute_median (impute_functions), 6
impute_median_all (impute_functions), 6
impute_median_at (impute_functions), 6
impute_median_if (impute_functions), 6
impute_min (impute_functions), 6
impute_min_all (impute_functions), 6
impute_min_at (impute_functions), 6
impute_min_if (impute_functions), 6
impute_most_freq (impute_functions), 6
impute_most_freq_all
 (impute_functions), 6
impute_most_freq_at (impute_functions),
 6
impute_most_freq_if (impute_functions),
 6
impute_neginf (impute_functions), 6
impute_neginf_all (impute_functions), 6
impute_neginf_at (impute_functions), 6
impute_neginf_if (impute_functions), 6
impute_quantile (impute_functions), 6
impute_quantile_all (impute_functions),
 6
impute_quantile_at (impute_functions), 6
impute_quantile_if (impute_functions), 6
impute_random (impute_functions), 6
impute_random_all (impute_functions), 6
impute_random_at (impute_functions), 6
impute_random_if (impute_functions), 6
impute_replace (impute_functions), 6
impute_replace_all (impute_functions), 6
impute_replace_at (impute_functions), 6
impute_replace_if (impute_functions), 6
impute_sample (impute_functions), 6
impute_sample_all (impute_functions), 6
impute_sample_at (impute_functions), 6
impute_sample_if (impute_functions), 6
impute_zero (impute_functions), 6

`impute_zero_all` (`impute_functions`), [6](#)
`impute_zero_at` (`impute_functions`), [6](#)
`impute_zero_if` (`impute_functions`), [6](#)

`na.tools::na.replace()`, [4](#)
`na_predict`, [7](#)
`nacars`, [7](#)
`nacars_dt` (`nacars`), [7](#)
`nairis` (`nacars`), [7](#)
`nairis_dt` (`nacars`), [7](#)