

Package: vvfiller (via r-universe)

September 12, 2024

Title Fill Data Points

Version 0.6.7.9000

Description Provides numerous functions to fill data. These can be applied either to missing or skewed data. The functions are designed within the scope of Student Analytics.

URL <https://github.com/vusaverse/vvfiller>,
<https://vusaverse.github.io/vvfiller/>

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Imports dplyr, purrr, rlang

Repository <https://vusaverse.r-universe.dev>

RemoteUrl <https://github.com/vusaverse/vvfiller>

RemoteRef HEAD

RemoteSha b3ef9645fdce426b047c34ccfee0ac99b3c46b88

Contents

check_some_missing	2
fill_col_with_agg_by_group	2
fill_df_with_agg_by_group	3
fill_missing	4
fill_missing_interval	4
fill_missing_last	5
fill_missing_max	5
fill_missing_min	6
fill_missing_previous	6

fill_missing_rownumber	7
fill_missing_strict	7
fill_value	8
fill_vector_interval	9
fill_vector_last	9
fill_vector_max	9
fill_vector_min	10
fill_vector_previous	10
fill_vector_strict	10
na_impute_median	11

Index 12

check_some_missing	<i>Check if some missing values are present</i>
--------------------	---

Description

Check if some missing values are present, but not all are missing. returns a boolean. This check is done to save time for vectors where filling is not needed

Usage

```
check_some_missing(x)
```

Arguments

x	the vector to check
---	---------------------

Value

TRUE or FALSE

fill_col_with_agg_by_group	<i>Fill column with aggregate by group</i>
----------------------------	--

Description

Calculate a summary statistic (mean, median, `vvconverter::mode`, min, max etc.) by group and use it to fill missing values in a column. Primarily for use in `fill_with_agg_by_group()`.

Usage

```
fill_col_with_agg_by_group(df, group, col, statistic)
```

Arguments

df	tibble to use
group	string or vector of strings: columns to group by
col	string: column to impute
statistic	function: summary statistic to use (mean, median, min etc.). For now requires a function with na.rm argument

Value

a filled vector

```
fill_df_with_agg_by_group
  Fill with aggregate by group
```

Description

Function to calculate a summary statistic (mean, median, `vvconverter::mode`, min, max etc.) by group and use it to fill missing values. Note: this takes and produces a tibble rather than a vector.

Usage

```
fill_df_with_agg_by_group(
  df,
  group,
  columns,
  overwrite_col = FALSE,
  statistic = mean,
  fill_empty_group = FALSE
)
```

Arguments

df	tibble to use
group	string or vector of strings: columns to group by
columns	string or vector of strings: columns to impute
overwrite_col	boolean: whether to overwrite column. If FALSE, a new column with suffix <code>_imputed</code> will be created
statistic	function: summary statistic to use (mean, median, min etc.). For now requires a function with na.rm argument
fill_empty_group	boolean: If TRUE, fills groups that only contain NA with summary statistic of entire column

Value

a tibble with filled column(s)

fill_missing	<i>Fill missing</i>
--------------	---------------------

Description

wrapper function to do check and call all fill_vector functions

Usage

```
fill_missing(x, min_known_n = NULL, min_known_p = NULL, type)
```

Arguments

x	The vector to fill
min_known_n	numeric value: the minimum number of not-missing values
min_known_p	numeric value between 0 and 1: the minimum fraction of not-missing values
type	the type of fill missing function to be called

Value

filled vector

fill_missing_interval	<i>Fill missing interval</i>
-----------------------	------------------------------

Description

Fill all missing values for an interval observed in the vector

Usage

```
fill_missing_interval(x, min_known_n = NULL, min_known_p = NULL)
```

Arguments

x	The vector to fill
min_known_n	numeric value: the minimum number of not-missing values
min_known_p	numeric value between 0 and 1: the minimum fraction of not-missing values

Value

a filled vector

Examples

```
fill_missing_interval(c(NA, 1, 2, NA))
fill_missing_interval(c(NA, 10, 20, NA))
```

fill_missing_last	<i>Fill missing last</i>
-------------------	--------------------------

Description

Fill all missing values in a vector with the last value if it is known.

Usage

```
fill_missing_last(x, min_known_n = NULL, min_known_p = NULL)
```

Arguments

x	The vector to fill
min_known_n	numeric value: the minimum number of not-missing values
min_known_p	numeric value between 0 and 1: the minimum fraction of not-missing values

Value

a filled vector

Examples

```
fill_missing_last(c(1, 2, NA))  
fill_missing_last(c(NA, 1, 2, NA))
```

fill_missing_max	<i>Fill missing maximum</i>
------------------	-----------------------------

Description

Fill all missing values in a vector with the maximum value if it is known.

Usage

```
fill_missing_max(x, min_known_n = NULL, min_known_p = NULL)
```

Arguments

x	The vector to fill
min_known_n	numeric value: the minimum number of not-missing values
min_known_p	numeric value between 0 and 1: the minimum fraction of not-missing values

Value

a filled vector

Examples

```
fill_missing_max(c(1, 2, NA))
fill_missing_max(c(NA, 1, 2, NA))
```

fill_missing_min	<i>Fill missing minimum</i>
------------------	-----------------------------

Description

Fill all missing values in a vector with the minimum value if it is known.

Usage

```
fill_missing_min(x, min_known_n = NULL, min_known_p = NULL)
```

Arguments

x	The vector to fill
min_known_n	numeric value: the minimum number of not-missing values
min_known_p	numeric value between 0 and 1: the minimum fraction of not-missing values

Value

a filled vector

Examples

```
fill_missing_min(c(1, 2, NA))
fill_missing_min(c(NA, 1, 2, NA))
```

fill_missing_previous	<i>Fill missing previous</i>
-----------------------	------------------------------

Description

Fill all missing values in a vector with the previous value if it is known.

Usage

```
fill_missing_previous(x, min_known_n = NULL, min_known_p = NULL)
```

Arguments

x	The vector to fill
min_known_n	numeric value: the minimum number of not-missing values
min_known_p	numeric value between 0 and 1: the minimum fraction of not-missing values

Value

a filled vector

Examples

```
fill_missing_previous(c(1, 2, NA))  
fill_missing_previous(c(NA, 1, 2, NA))
```

`fill_missing_rownumber` *Fill missing rownumber*

Description

Impute missing values of a count variable. Imputation is done by counting from the last known value. Example: `c(NA,4,NA,NA)` then becomes `c(NA,4,NA,NA)`.

Usage

```
fill_missing_rownumber(x)
```

Arguments

`x` Integer vector.

Value

Integer vector with filled values.

Examples

```
fill_missing_rownumber(c(NA, 4, NA, NA))
```

`fill_missing_strict` *Fill missing strict*

Description

Fill all missing values in a vector with the same value if it is known. Only fills the value when all known values are the same

Usage

```
fill_missing_strict(x, min_known_n = NULL, min_known_p = NULL)
```

Arguments

x	The vector to fill
min_known_n	numeric value: the minimum number of not-missing values
min_known_p	numeric value between 0 and 1: the minimum fraction of not-missing values

Value

a filled vector

Examples

```
fill_missing_strict(c(NA, 1))
```

fill_value	<i>fill missing value</i>
------------	---------------------------

Description

Returns a vector with all missing values filled with another value

Usage

```
fill_value(x, value)
```

Arguments

x	vectors. All inputs should have the same length
value	a value with the same class as x

Value

vector with the same length as the first vector

Examples

```
fill_value(c(NA,1), 2)
```

fill_vector_interval *fill_vector_interval*

Description

fill_vector_interval

Usage

fill_vector_interval(x)

Arguments

x the vector to be filled

fill_vector_last *fill_vector_last*

Description

fill_vector_last

Usage

fill_vector_last(x, x_na_omit)

Arguments

x the vector to be filled
x_na_omit the x vector without NA values

fill_vector_max *fill_vector_max*

Description

fill_vector_max

Usage

fill_vector_max(x, x_na_omit)

Arguments

x the vector to be filled
x_na_omit the x vector without NA values

fill_vector_min *fill_vector_min*

Description

fill_vector_min

Usage

```
fill_vector_min(x, x_na_omit)
```

Arguments

x	the vector to be filled
x_na_omit	the x vector without NA values

fill_vector_previous *fill_vector_previous*

Description

fill_vector_previous

Usage

```
fill_vector_previous(x)
```

Arguments

x	the vector to be filled
---	-------------------------

fill_vector_strict *fill_vector_strict*

Description

fill_vector_strict

Usage

```
fill_vector_strict(x, x_na_omit)
```

Arguments

x	the vector to be filled
x_na_omit	the x vector without NA values

na_impute_median	<i>NA impute median</i>
------------------	-------------------------

Description

Is a specialized function which takes a variable and turns it into two new variables to be used in a prediction model.

1. the variable for which missing values are imputed by the median for the given year.
2. an indicator when the variable is missing

Usage

```
na_impute_median(data, var, year = 2014, year_column)
```

Arguments

data	The data frame.
var	The variable used to create new variables.
year	Year used for the median for imputation.
year_column	Column with year to use median on.

Value

New data frame in which missing values are filled.

Index

* missing data functions

fill_missing_rownumber, 7

* vector calculations

fill_missing_rownumber, 7

check_some_missing, 2

fill_col_with_agg_by_group, 2

fill_df_with_agg_by_group, 3

fill_missing, 4

fill_missing_interval, 4

fill_missing_last, 5

fill_missing_max, 5

fill_missing_min, 6

fill_missing_previous, 6

fill_missing_rownumber, 7

fill_missing_strict, 7

fill_value, 8

fill_vector_interval, 9

fill_vector_last, 9

fill_vector_max, 9

fill_vector_min, 10

fill_vector_previous, 10

fill_vector_strict, 10

na_impute_median, 11