

Package: simplanonym (via r-universe)

October 23, 2024

Type Package

Title Consistent Anonymisation Across Datasets

Version 0.1.0

Description A simple function that anonymises a list of variables in a consistent way: anonymised factors are not recycled and the same original levels receive the same anonymised factor even if located in different datasets.

License Apache License (≥ 2)

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

URL <https://github.com/dkgaraujo/simplanonym>

BugReports <https://github.com/dkgaraujo/simplanonym/issues>

Imports dplyr ($\geq 1.0.10$), forcats ($\geq 0.5.1$), tidyselect ($\geq 1.2.0$)

Suggests testthat ($\geq 3.0.0$), vroom ($\geq 1.5.7$)

Config/testthat/edition 3

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

RemoteUrl <https://github.com/dkgaraujo/simplanonym>

RemoteRef HEAD

RemoteSha 8e86242d05d95670a9f77d2937f49880baadf7ae

Contents

anonymise	2
Index	3

`anonymise`*Anonymise factor columns across datasets in a consistent way*

Description

`'anonymise()'` is a useful function for anonymising factor columns across different datasets using consistent anonymised levels. In other words, if the same factor level appears in more than one dataset, then `'anonymise()'` will use the same anonymous factor for that level.

Usage

```
anonymise(data_list, prefix = "", return_original_levels = FALSE)
```

Arguments

<code>data_list</code>	A list of data frames or tibbles.
<code>prefix</code>	A character prefix to insert in front of the random labels.
<code>return_original_levels</code>	Whether or not the resulting list should also include the original, non-anonymised levels. Default: FALSE.

Value

A list containing the original data, but with consistently anonymised factors

Examples

```
library(simplanonym)

rand_tbl_1 <- vroom::gen_tbl(10, 4, col_types = "fffd")
rand_tbl_2 <- vroom::gen_tbl(10, 2, col_types = "fd")
rand_tbl_2$X3 <- rand_tbl_1$X3

# note:
# * rand_tbl_1 and rand_tbl_2 share three column names,
#   of which X2 is a factor in one but not the other.
# * X1 factors do not overlap, but their anonymisation
#   should still be consistent (ie, different levels should
#   have their own unique anonymised factors).
# * For X3, the anonymised factors should consider the levels
#   at both `rand_tbl_1$X3` and `rand_tbl_2$X3`.

data_list <- list(rand_tbl_1, rand_tbl_2)
data_list

data_list |> anonymise(return_original_levels = TRUE)
```

Index

anonymise, [2](#)