Package: tablecompare (via r-universe)

June 11, 2024

Type Package

Title Compare Data Frames

Version 0.1.1

Description A toolbox for comparing two data frames. This package is
defunct. I recommend you use the ``versus'' package instead.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports glue, magrittr, rlang (>= 0.4.3), tidyselect (>= 0.4.3), purrr

RoxygenNote 7.2.3

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

URL https://github.com/eutwt/tablecompare

BugReports https://github.com/eutwt/tablecompare/issues

Depends data.table (>= 1.14.2)

Repository https://eutwt.r-universe.dev

RemoteUrl https://github.com/eutwt/tablecompare

RemoteRef HEAD

RemoteSha fd1cd2f270e969147f87e85ea46bcfb046f05bcd

Contents

---

contents   *Show the contents of a data frame*

Description

Show the contents of a data frame

Usage

```
contents(.data)
```

Arguments

`.data`   A data frame or data table

Value

A `data.table` with one row per column in `.data` and columns "column": The name of the column
in `.data`, "class": the names of classes the column inherits from (as returned by `class()`), collapsed
into a single string.

Examples

```
contents(ToothGrowth)
```

---

count_dupes   *Check for duplicate rows*

Description

`count_dupes()` returns values of by variables for which the `.data` has multiple rows, along with
the number of rows for each combination of values.

`assert_unique()` throws an error if there are multiple rows for any combination of by variable
values

Usage

```
count_dupes(.data, by, setkey = FALSE)

assert_unique(.data, by, data_chr, by_chr)
```

Arguments

`.data`	A data frame or data table
`by`	tidy-select. Columns in `.data`
`setkey`	Logical. Should the output be keyed by by cols?
`data_chr`	optional. character. You can use this argument to manually specify the name of data shown in error messages. Useful when using these functions as checks inside other functions.
`by_chr`	optional. character. You can use this argument to manually specify the name of by shown in error messages. Useful when using these functions as checks inside other functions.

Value

count_dupes() A `data.table` with the (filtered) by columns and an additional column "n_rows" which shows the number of rows in `.data` having the combination of by values shown in the output row.

assert_unique() No return value. Called to throw an error depending on the input.

Examples

```
df <- read.table(text = "
x y z
1 6 1
2 6 2
3 7 3
3 7 4
4 3 5
4 3 6
", header = TRUE)

count_dupes(df, c(x, y))

## Not run:
assert_unique(df, c(x, y))

## End(Not run)
```

---

count_values	*Check for existence of multiple values per group*

Description

count_values() returns values of by variables for which the `.data` has multiple unique rows, along with the number of unique rows for each combination of values, only considering columns in col.

assert_single_value() throws an error if there are multiple unique rows for any combination of by variable values, only considering columns in col.

Usage

```
count_values(.data, col, by, setkey = FALSE)

assert_single_value(.data, col, by)
```

Arguments

`.data`	A data frame or data table
`col`	tidy-select. Columns in `.data`. When counting the number of unique rows, only the columns specified in `col` are considered.
`by`	tidy-select. Columns in `.data`.
`setkey`	Logical. Should the output be keyed by by cols?

Value

`count_values()` A `data.table` with the (filtered) by columns and an additional column "n_vals" which shows the number of unique rows in `.data` having the combination of by values shown in the output row.

`assert_single_value()` No return value. Called to throw an error depending on the input.

Examples

```
df <- read.table(text = "
x y z
a 1 3
a 1 3
a 2 4
a 2 4
a 2 2
b 1 1
b 1 2
", header = TRUE)

count_values(df, z, by = c(x, y))

## Not run:
assert_single_value(df, z, by = c(x, y))

## End(Not run)
```

---

tblcompare	*Compare two data frames. Using a key-column common to both tables, see which rows are common and highlight differing values by column.*

Description

Compare two data frames. Using a key-column common to both tables, see which rows are common and highlight differing values by column.

Usage

```
tblcompare(
  .data_a,
  .data_b,
  by,
  allow_bothNA = TRUE,
  ncol_by_out = 3,
  coerce = TRUE
)

value_diffs(comparison, col)

## S3 method for class 'tbcmp_compare'
value_diffs(comparison, col)

all_value_diffs(comparison)

## S3 method for class 'tbcmp_compare'
all_value_diffs(comparison)
```

Arguments

.data_a	A data frame or data table
.data_b	A data frame or data table
by	tidy-select. Selection of columns to use when matching rows between .data_a and .data_b. Both data frames must be unique on by.
allow_bothNA	Logical. If TRUE a missing value in both data frames is considered as equal
ncol_by_out	Number of by-columns to include in col_diffs and unmatched_rows output
coerce	Logical. If False only columns with the same class are compared.
comparison	An object of class "tbcmp_compare" (the output of a tablecompare::tablecompare() call)
col	tidy-select. A single column

Value

tblcompare() A "tbcmp_compare"-class object, which is a list of data.table's having the following elements:

tables A data.table with one row per input table showing the number of rows and columns in each.

by A data.table with one row per by column showing the class of the column in each of the input tables.

summ A data.table with one row per column common to .data_a and .data_b and columns "n_diffs" showing the number of values which are different between the two tables, "class_a"/"class_b" the class of the column in each table, and "value_diffs" a (nested) data.table showing the rows in each input table where values are unequal, the values in each table, and one column for each of the first ncol_by_out by columns for the identified rows in the input tables.

unmatched_cols A data.table with one row per column which is in one input table but not the other and columns "table": which table the column appears in, "column": the name of the column, and "class": the class of the column.

unmatched_rows A data.table which, for each row present in one input table but not the other, contains the columns "table": which table the row appears in, "i" the row number of the input row, and one column for each of the first ncol_by_out by columns for each row.

value_diffs() A data.table with one row for each element of col found to be unequal between the input tables (.data_a and .data_b from the original tblcompare() call) The output table has columns "i_a"/"i_b": the row number of the element in the input tables, "val_a"/"val_b": the value of col in the input tables, and one column for each of the first ncol_by_out by columns for the identified rows in the input tables.

all_value_diffs() A data.table of the value_diffs() output for all columns having at least one value difference, combined row-wise into a single table. To facilitate this combination into a single table, the "val_a" and "val_b" columns are coerced to character.

Index