

Package ‘diffdfs’

October 13, 2022

Title Compute the Difference Between Data Frames

Version 0.9.0

Description Shows you which rows have changed between two data frames with the same column structure. Useful for diffing slowly mutating data.

License MIT + file LICENSE

Imports arrow, dplyr, janitor, rlang

BugReports <https://github.com/riazarbi/diffdfs>

Encoding UTF-8

RoxygenNote 7.2.1

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Riaz Arbi [aut, cre]

Maintainer Riaz Arbi <diffdfs@arbidata.com>

Repository CRAN

Date/Publication 2022-08-09 13:50:06 UTC

R topics documented:

checkkey	2
diffdfs	2
Index	4

checkkey	<i>Check That A Dataframe Key Col Set Is Unique</i>
----------	---

Description

Checks that a provided vector of column names constitute a unique key (that is, no rows are duplicated) for a dataframe.

Usage

```
checkkey(df, key_cols, verbose = FALSE)
```

Arguments

df	a dataframe
key_cols	vector of column names
verbose	TRUE/FALSE should we print a message?

Value

TRUE if key cols have unique rows; FALSE if not

Examples

```
irisint = iris
irisint$rownum = 1:nrow(irisint)
key_cols = c("rownum")
checkkey(irisint, key_cols, TRUE)
checkkey(irisint, "Species", TRUE)
```

diffdfs	<i>Compute the Difference Between Dataframes</i>
---------	--

Description

Returns a dataframe describing the modifications required to transform `old_df` into `new_df`. The dataframes need `needBugReports`: <https://github.com/tidyverse/dplyr/issues> to have identical columns and column types and share unique index columns.

Usage

```
diffdfs(new_df, old_df = NA, key_cols = NA, verbose = FALSE)
```

Arguments

<code>new_df</code>	A dataframe of new data.
<code>old_df</code>	A dataframe of old data. <code>new_df</code> and <code>old_df</code> can (and usually do) have overlapping data.
<code>key_cols</code>	optional vector of column names that constitute a unique table key. If NA, <code>colnames(old_df)</code> will be used.
<code>verbose</code>	logical, default FALSE. Should the processing be chatty?

Value

a dataframe.

Examples

```
iris$key <- 1:nrow(iris)

old_df <- iris[1:100,]
old_df[75,1] <- 100
new_df <- iris[50:150,]
diffdfs(new_df, old_df, key_cols = "key")
```

Index

checkkey, [2](#)

diffdfs, [2](#)