Package 'ReportSubtotal'

November 15, 2024

Title Adds Subtotals to Data Reports

Version 0.1

Description
Adds subtotal rows / sections (a la the SAS ``Proc Tabulate'' ``All'' option) to a Group By output by running a series of Group By functions with partial sets of the same variables and combining the results with the original. Can be used to add comprehensive information to a data report or to quickly aggregate Group By outputs used to gain a greater understanding of data.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Imports dplyr, tidyselect, utils, magrittr

NeedsCompilation no

Author Yoni Aboody [aut, cre, cph]

Maintainer Yoni Aboody <yoniaboody@gmail.com>

Repository CRAN

Date/Publication 2024-11-15 09:00:13 UTC

Contents

---

subtotal_dupe_removal *Duplicate Subtotal Row Removal Function*

Description

Removes duplicate subtotal rows, which may be created by totalling a variable with only one level. In tables with many variables, some may have only one level within one section and many in other sections.

Usage

```
subtotal_dupe_removal(
  data,
  column,
  iterator = 2,
  skip = 0,
  remove = "All",
  suffix = "_Default_Suffix"
)
```

Arguments

data	Data frame or tibble to remove duplicate row labels from.
column	Column containing duplicate row labels.
iterator	Minimum number of rows meant to be between each section. Usually two.
skip	Number of rows to skip removing rows from. Usually zero. Can be used to dodge NA values.
remove	Label of subtotals to be removed. Usually "All".
suffix	Suffix to disambiguate required column name if already taken.

Details

Adds a leading version of the requested column, which places each observation in the same row as the next observation. Usually the observation 2 rows on - determined by the iterator. From here if both the original and leading column equal the value to be removed, then the row is a duplicate subtotal and is removed. Note: the last few rows will have NA values in the leading column, so they are covered separately. Note: If you already have columns named Lead or Index, they receive a temporary suffix.

Value

The data report without duplicate subtotal rows.

Examples

```
library(dplyr)

group_by(mtcars, cyl, vs) %>% summarise(sum(wt), .groups = "keep") %>%
subtotal_row(mtcars, "wt") %>%
subtotal_dupe_removal(2)
group_by(mtcars, cyl, vs, am) %>% summarise(mean(hp), .groups = "keep") %>%
subtotal_row(mtcars, "hp", "mean") %>%
subtotal_dupe_removal(3, skip = 1)
group_by(mtcars, cyl, vs, am) %>% summarise(mean(hp), .groups = "keep") %>%
subtotal_row(mtcars, "hp", "mean") %>%
subtotal_dupe_removal(3, skip = 1)
```

---

subtotal_row   *Subtotal Row Addition Function*

Description

Adds subtotal rows to a data report.

Usage

```
subtotal_row(
  report,
  frame,
  vars = "Population",
  aggregator = "sum",
  exclude = numeric(0),
  agg_parameter = character(0),
  subtotal_label = "All"
)
```

Arguments

report	A data report.
frame	Data frame summarised by the data report.
vars	Names of column(s) in the data frame aggregated in the data report.
aggregator	Function to aggregate the data with.
exclude	Vector of column indices determining which variables don't require subtotal rows.
agg_parameter	Optional parameter for the aggregation function to use.
subtotal_label	Label to be used for subtotal rows.

Details

The dataset and report are factorized, and a series of reports with the same variables are then generated, but with some variables replaced by a subtotal label, which effectively concentrates all levels of those variables into one subtotal row for those variables. The subtotal reports are all combined with the original report, and the combined report is sorted, sorting the subtotal label to the top for all variables.

Value

The data report with subtotal rows included.

Examples

```
library(dplyr)

group_by(iris, Species, Petal.Width) %>%
summarise(sum(Petal.Length), .groups = "keep") %>%
subtotal_row(iris, vars = "Petal.Length")

group_by(iris, Species, Petal.Width) %>%
summarise(mean(Sepal.Width), .groups = "keep") %>%
subtotal_row(iris, vars = "Sepal.Width", aggregator = "mean")

group_by(mtcars, cyl, gear, carb) %>%
summarise(median(wt), median(hp), .groups = "keep") %>%
subtotal_row(mtcars, vars = c("wt", "hp"), aggregator = "median")

group_by(mtcars, cyl, gear, carb) %>%
summarise(Med_Weight = median(wt), Med_Hrspw = median(hp), .groups = "keep") %>%
subtotal_row(mtcars, vars = c("wt", "hp"), aggregator = "median", exclude = 1)

group_by(mtcars, vs, am, drat, carb) %>%
summarise(min(mpg), min(disp), min(carb), .groups = "keep") %>%
subtotal_row(mtcars, vars = c("mpg", "disp", "carb"),
aggregator = "min", exclude = c(2, 4), subtotal_label = "Min_Cars_Total", agg_parameter = "na.rm")
```

---

subtotal_section   *Subtotal Section Addition Function*

Description

Adds subtotal sections to a data report.

Usage

```
subtotal_section(
  report,
  frame,
  vars = "Population",
```

```
    aggregator = "sum",
    exclude = numeric(0),
    agg_parameter = character(0),
    subtotal_label = "All"
)
```

Arguments

`report`	A data report.
`frame`	Data frame summarised by the data report.
`vars`	Names of column(s) in the data frame aggregated in the data report.
`aggregator`	Function to aggregate the data with.
`exclude`	Vector of column indices determining which variables only require subtotal rows (as opposed to sections).
`agg_parameter`	Optional parameter for the aggregation function to use.
`subtotal_label`	Label to be used for subtotal rows.

Details

The dataset and report are factorized, and a series of reports with the same variables are then generated, but with some variables replaced by a subtotal label, which effectively concentrates all levels of those variables into one subtotal row for those variables - ultimately making up entire subtotal sections for some variables. The subtotal reports are all combined with the original report, and the combined report is sorted, sorting the subtotal label to the top for all variables.

Value

The data report with subtotal sections included.

Examples

```
library(dplyr)

group_by(iris, Species) %>%
summarise(sum(Petal.Length), .groups = "keep") %>%
subtotal_section(iris, vars = "Petal.Length")

group_by(mtcars, cyl, gear) %>%
summarise(mean(mpg), .groups = "keep") %>%
subtotal_section(mtcars, vars = "mpg", aggregator = "mean")

group_by(mtcars, cyl, gear) %>%
summarise(mean(mpg), mean(wt), .groups = "keep") %>%
subtotal_section(mtcars, vars = c("mpg", "wt"),
aggregator = "mean", exclude = 1:2)

group_by(iris, Species, Petal.Width) %>%
summarise(max(Sepal.Width), max(Sepal.Length), .groups = "keep") %>%
subtotal_section(iris, vars = c("Sepal.Width", "Sepal.Length"),
```

```
    aggregator = "max", agg_parameter = "na.rm")

group_by(mtcars, qsec, carb, hp, gear) %>%
summarise(min(cyl), min(drat), min(wt), .groups = "keep") %>%
subtotal_section(mtcars, vars = c("cyl", "drat", "wt"),
    aggregator = "min", agg_parameter = "na.rm",
    subtotal_label = "Cars_Total", exclude = c(1, 4))
```

Index