Package ‘overviewR’

November 23, 2020

Type Package

Title Easily Extracting Information About Your Data

Version 0.0.7

Description Makes it easy to display descriptive information on a data set. Getting an easy overview of a data set by displaying and visualizing sample information in different tables (e.g., time and scope conditions). The package also provides publishable 'LaTeX' code to present the sample information.

License GPL-3

URL https://github.com/cosimameyer/overviewR

BugReports https://github.com/cosimameyer/overviewR/issues

Depends R (>= 3.5.0)

Imports dplyr (>= 1.0.0), ggplot2 (>= 3.3.2), tibble (>= 3.0.1)

Suggests covr, devtools, knitr, pkgdown, rmarkdown, spelling, testthat

VignetteBuilder knitr, rmarkdown

Encoding UTF-8

Language en-US

LazyData true

RoxygenNote 7.1.1

NeedsCompilation no

Author Cosima Meyer [cre, aut], Dennis Hammerschmidt [aut]

Maintainer Cosima Meyer <cosima.meyer@gmail.com>

Repository CRAN

Date/Publication 2020-11-23 09:20:03 UTC
Description

Sorts a data set conditionally in a cross table. This can be helpful to get a sense of the time and scope conditions of a data set. Note, if used with a data set that has multiple observations on the id-time unit, the function automatically aggregates this information using the mean.

Usage

overview_crosstab(dat, cond1, cond2, threshold1, threshold2, id, time)

Arguments

dat A data set object
cond1 Variable that describes the first condition
cond2 Variable that describes the second condition
threshold1 A threshold for cond1
threshold2 A threshold for cond2
id Scope (e.g., country codes or individual IDs)
time Time (e.g., time periods given by years, months, ...)

Value

A data frame object that contains a summary of the data set that can later be converted to a ‘LaTeX’ output using overview_print
overview_heat

Examples

```r
data(toydata)
overview_crosstab(
  dat = toydata,
  cond1 = gdp,
  cond2 = population,
  threshold1 = 25000,
  threshold2 = 27000,
  id = ccode,
  time = year
)
```

Description

This function plots a heat map to visualize the coverage of the time-scope-units of the data. Options include total number of cases per time-scope-unit or relative number in percentage.

Usage

```r
overview_heat(
  dat, 
  id, 
  time, 
  perc = FALSE, 
  exp_total, 
  xaxis = "Time frame", 
  yaxis = "Sample", 
  col_low = "#dceaf2", 
  col_high = "#2A5773", 
  label = TRUE
)
```

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dat</td>
<td>The data set</td>
</tr>
<tr>
<td>id</td>
<td>The scope (e.g., country codes or individual IDs). The axis is ordered in ascending order by default.</td>
</tr>
<tr>
<td>time</td>
<td>The time (e.g., time periods given by years, months, ...)</td>
</tr>
<tr>
<td>perc</td>
<td>If FALSE (default) plot returns the total number of observations per time-scope-unit. If TRUE, it returns the number of observations per time-scope-unit in percentage</td>
</tr>
<tr>
<td>exp_total</td>
<td>Expected total number of observations (i.e. maximum) for time unit.</td>
</tr>
<tr>
<td>xaxis</td>
<td>Label of your x axis (&quot;Time frame&quot; is default)</td>
</tr>
</tbody>
</table>
Overview

This function plots a ggplot to visualize the distribution of NAs across all variables in the data set.

Usage

```
overview_na(dat, yaxis = "Variables", perc = TRUE)
```

Arguments

- **dat**: Your data set
- **yaxis**: Label of your y axis ("Variables" is default)
- **perc**: If TRUE (default) plot returns the number of NAs in percentage

Value

A ggplot figure that presents the distribution of NAs in the data set

Examples

```
data(toydata)
overview_na(toydata, perc = FALSE)
```
Description

This function plots a ggplot to visualize the distribution of scope objects across the time frame.

Usage

overview_plot(
  dat,
  id,
  time,
  xaxis = "Time frame",
  yaxis = "Sample",
  asc = TRUE
)

Arguments

dat         Your data set
id          Your scope (e.g., country codes or individual IDs). If the id variable contains NAs, they will not be included in the plot.
time        Your time (e.g., time periods given by years, months, ...)
xaxis       Label of the x axis ("Time frame" is default)
yaxis       Label of the y axis ("Sample" is default)
asc         Sorting the y axis in ascending order ("TRUE" is default)

Value

A ggplot figure that presents the sample information visually

Examples

data(toydata)
overview_plot(dat = toydata, id = ccode, time = year)
Overview

Description

Produces a 'LaTeX' output for output obtained via `overview_tab` and `overview_crosstab`

Usage

```r
overview_print(
  obj,
  title = "Time and scope of the sample",
  id = "Sample",
  time = "Time frame",
  crosstab = FALSE,
  cond1 = "Condition 1",
  cond2 = "Condition 2",
  save_out = FALSE,
  path,
  file
)
```

Arguments

- **obj**: Overview object produced by `overview_tab` or `overview_crosstab`
- **title**: Title of the table (default is "Time and scope of the sample")
- **id**: The name of the left column (default is "Sample"), will be ignored if `crosstab` is `TRUE`
- **time**: The name of the right column (default is "Time frame"), will be ignored if `crosstab` is `TRUE`
- **crosstab**: Logical argument, if `TRUE` produces a crosstab output, default is `FALSE`
- **cond1**: Description for the first condition (character), will be ignored if `crosstab` is `FALSE`. This should correspond to the input for `cond1` in `overview_crosstab`
- **cond2**: Description for the second condition (character), will be ignored if `crosstab` is `FALSE`. This should correspond to the input for `cond2` in `overview_crosstab`
- **save_out**: Optional argument, exports the output table as a .tex file, default is `FALSE`
- **path**: Specifies the path where the output should be saved
- **file**: Specifies name and file type (.tex)

Value

A 'LaTeX' output that can either be copy-pasted in a text document or exported directed as a .tex file
Overview Tab

Examples

data(toydata)

overview_object <- overview_tab(dat = toydata, id = ccode, time = year)
overview_print(
  obj = overview_object,
  title = "Some nice title",
  crosstab = FALSE
)

overview_ct_object <- overview_crosstab(
  dat = toydata,
  cond1 = gdp,
  cond2 = population,
  threshold1 = 25000,
  threshold2 = 27000,
  id = ccode,
  time = year
)
overview_print(
  obj = overview_ct_object,
  title = "Some nice title for a cross tab",
  crosstab = TRUE,
  cond1 = "Name of first condition",
  cond2 = "Name of second condition"
)

Description

Provides an overview table for the time and scope conditions of a data set

Usage

overview_tab(dat, id, time)

Arguments

dat A data set object
id Scope (e.g., country codes or individual IDs)
time Time (e.g., time periods given by years, months, ...)

Value

A data frame object that contains a summary of a sample that can later be converted to a 'LaTeX' output using overview_print
Examples

```r
data(toydata)
output_table <- overview_tab(dat = toydata, id = ccode, time = year)
```

---

### toydata

**Cross-sectional data for countries**

**Description**

Small, artificially generated toy data set that comes in a cross-sectional format where the unit of analysis is either country-year or country-year-month. It provides artificial information for five countries (Angola, Benin, France, Rwanda, and the UK) for a time span from 1990 to 1999 to illustrate the use of the package.

**Usage**

```r
data(toydata)
```

**Format**

An object of class "data.frame"

- **ccode**: ISO3 country code (as character) for the countries in the sample (Angola, Benin, France, Rwanda, and UK)
- **year**: A value between 1990 and 1999
- **month**: An abbreviation (MMM) for month (character)
- **gpd**: A fake value for GDP (randomly generated)
- **population**: A fake value for population (randomly generated)

**References**

This data set was artificially created for the overviewR package.

**Examples**

```r
data(toydata)
head(toydata)
```
Index

* datasets
  toydata, 8

overview_crosstab, 2
overview_heat, 3
overview_na, 4
overview_plot, 5
overview_print, 6
overview_tab, 7

toydata, 8