

# Package ‘ipmisc’

June 19, 2020

**Type** Package

**Title** Miscellaneous Functions for Data Cleaning and Analysis

**Version** 3.1.0

**Maintainer** Indrajeet Patil <patilindrajeet.science@gmail.com>

**Description** Provides functions needed for data cleaning and formatting and forms data cleaning and wrangling backend for the following packages: 'broomExtra', 'ggstatsplot', 'groupedstats', 'pairwiseComparisons', 'statsExpressions', and 'tidyBF'.

**License** GPL-3 | file LICENSE

**URL** <https://indrajeetpatil.github.io/ipmisc/>,  
<https://github.com/IndrajeetPatil/ipmisc>

**BugReports** <https://github.com/IndrajeetPatil/ipmisc/issues>

**Depends** R (>= 3.6.0)

**Imports** crayon, dplyr, magrittr, rlang, rstudioapi, tibble, tidyr,  
zeallot

**Suggests** ggplot2, knitr, rmarkdown, spelling, testthat

**Encoding** UTF-8

**Language** en-US

**LazyData** true

**RoxygenNote** 7.1.0.9000

**NeedsCompilation** no

**Author** Indrajeet Patil [aut, cre, cph]  
(<<https://orcid.org/0000-0003-1995-6531>>)

**Repository** CRAN

**Date/Publication** 2020-06-19 08:40:03 UTC

## R topics documented:

bugs_long . . . . .	2
iris_long . . . . .	3
long_to_wide_converter . . . . .	4
outlier_df . . . . .	5
p_value_formatter . . . . .	6
set_cwd . . . . .	6
signif_column . . . . .	7
specify_decimal_p . . . . .	8
stats_type_switch . . . . .	8

<b>Index</b>	<b>10</b>
--------------	-----------

---

bugs_long	<i>Tidy version of the "Bugs" dataset.</i>
-----------	--

---

### Description

Tidy version of the "Bugs" dataset.

### Usage

```
bugs_long
```

### Format

A data frame with 372 rows and 6 variables

- subject. Dummy identity number for each participant.
- gender. Participant's gender (Female, Male).
- region. Region of the world the participant was from.
- education. Level of education.
- condition. Condition of the experiment the participant gave rating for (**LDLF**: low frighteningness and low disgustingness; **LFHD**: low frighteningness and high disgustingness; **HFHD**: high frighteningness and low disgustingness; **HFHD**: high frighteningness and high disgustingness).
- desire. The desire to kill an arthropod was indicated on a scale from 0 to 10.

### Details

This data set, "Bugs", provides the extent to which men and women want to kill arthropods that vary in frighteningness (low, high) and disgustingness (low, high). Each participant rates their attitudes towards all anthropods. Subset of the data reported by Ryan et al. (2013).

### Source

<https://www.sciencedirect.com/science/article/pii/S0747563213000277>

**Examples**

```
dim(iris_long)
head(iris_long)
dplyr::glimpse(iris_long)
```

---

iris_long	<i>Edgar Anderson's Iris Data in long format.</i>
-----------	---

---

**Description**

Edgar Anderson's Iris Data in long format.

**Usage**

```
iris_long
```

**Format**

A data frame with 600 rows and 5 variables

- id. Dummy identity number for each flower (150 flowers in total).
- Species. The species are *Iris setosa*, *versicolor*, and *virginica*.
- condition. Factor giving a detailed description of the attribute (Four levels: "Petal.Length", "Petal.Width", "Sepal.Length", "Sepal.Width").
- attribute. What attribute is being measured ("Sepal" or "Petal").
- measure. What aspect of the attribute is being measured ("Length" or "Width").
- value. Value of the measurement.

**Details**

This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica*.

This is a modified dataset from datasets package.

**Source**

<https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/iris.html>

**Examples**

```
dim(iris_long)
head(iris_long)
dplyr::glimpse(iris_long)
```

---

`long_to_wide_converter`*Converts dataframe from long/tidy to wide format with NAs removed*

---

### Description

This conversion is helpful mostly for repeated measures design.

### Usage

```
long_to_wide_converter(data, x, y, paired = TRUE, spread = TRUE, ...)
```

### Arguments

<code>data</code>	A dataframe (or a tibble) from which variables specified are to be taken. A matrix or tables will <b>not</b> be accepted.
<code>x</code>	The grouping variable from the dataframe <code>data</code> .
<code>y</code>	The response (a.k.a. outcome or dependent) variable from the dataframe <code>data</code> .
<code>paired</code>	Logical that decides whether the experimental design is repeated measures/within-subjects or between-subjects. The default is <code>FALSE</code> .
<code>spread</code>	Logical that decides whether the dataframe needs to be converted from long/tidy to wide (default: <code>TRUE</code> ), or the data needs to be returned as it is but with the NAs removed.
<code>...</code>	Currently ignored.

### Value

A dataframe in the wide (or Cartesian) format.

### Examples

```
long_to_wide_converter(  
  data = iris_long,  
  x = condition,  
  y = value,  
  paired = TRUE  
)
```

---

`outlier_df`*Adding a column to dataframe describing outlier status*

---

**Description**

Adding a column to dataframe describing outlier status

**Usage**

```
outlier_df(data, x, y, outlier.label, outlier.coef = 1.5, ...)
```

**Arguments**

<code>data</code>	A dataframe (or a tibble) from which variables specified are to be taken. A matrix or tables will <b>not</b> be accepted.
<code>x</code>	The grouping variable from the dataframe data.
<code>y</code>	The response (a.k.a. outcome or dependent) variable from the dataframe data.
<code>outlier.label</code>	Label to put on the outliers that have been tagged. This can't be the same as <code>x</code> argument.
<code>outlier.coef</code>	Coefficient for outlier detection using Tukey's method. With Tukey's method, outliers are below (1st Quartile) or above (3rd Quartile) <code>coef</code> times the Inter-Quartile Range (IQR) (Default: 1.5).
<code>...</code>	Additional arguments.

**Value**

The dataframe entered as `data` argument is returned with two additional columns: `isanoutlier` and `outlier` denoting which observation are outliers and their corresponding labels.

**Examples**

```
# adding column for outlier and a label for that outlier
outlier_df(
  data = morley,
  x = Expt,
  y = Speed,
  outlier.label = Run,
  outlier.coef = 2
) %>%
  dplyr::arrange(outlier)
```

---

p\_value\_formatter      *Format p-values for creating labels from them for graphics*

---

### Description

Format  $p$ -values for creating labels from them for graphics

### Usage

```
p_value_formatter(data, k = 3L, ...)
```

### Arguments

data	Data frame from which variables specified are preferentially to be taken.
k	Number of digits after decimal point (should be an integer) (Default: k = 3L).
...	Currently ignored.

### Examples

```
# preparing a new dataframe
df <-
  cbind.data.frame(
    x = 1:5,
    y = 1,
    p.value = c(0.1, 0.5, 0.00001, 0.05, 0.01)
  )

# prepare label
p_value_formatter(data = df, k = 4L)
```

---

set\_cwd      *Setting Working Directory in RStudio to where the R Script is.*

---

### Description

This function will change the current working directory to whichever directory the R script you are currently working on is located. This preempts the trouble of setting the working directory manually.

### Usage

```
set_cwd()
```

### Value

Path to changed working directory.

**Note**

This function will work **only with RStudio IDE**. Reference: <https://eranraviv.com/r-tips-and-tricks-working-directory/>

---

`signif_column`*Creating a new column with significance labels*

---

**Description**

This function will add a new column with significance labels to a dataframe containing  $p$ -values.

**Usage**

```
signif_column(data, p, ...)
```

**Arguments**

<code>data</code>	Data frame from which variables specified are preferentially to be taken.
<code>p</code>	The column containing $p$ -values.
<code>...</code>	Currently ignored.

**Value**

Returns the dataframe in tibble format with an additional column corresponding to APA-format statistical significance labels.

**Examples**

```
# preparing a new dataframe
df <-
  cbind.data.frame(
    x = 1:5,
    y = 1,
    p.value = c(0.1, 0.5, 0.00001, 0.05, 0.01)
  )

# dataframe with significance column
signif_column(data = df, p = p.value)
```

---

specify\_decimal\_p      *Formatting numeric (p-)values*

---

### Description

Function to format an R object for pretty printing with a specified ( $k$ ) number of decimal places. The function also allows really small  $p$ -values to be denoted as " $p < 0.001$ " rather than " $p = 0.000$ ". Note that if `p.value` is set to `TRUE`, the minimum value of  $k$  allowed is 3. If  $k$  is set to less than 3, the function will ignore entered  $k$  value and use  $k = 3$  instead. **Important:** This function is not vectorized.

### Usage

```
specify_decimal_p(x, k = 3L, p.value = FALSE)
```

### Arguments

<code>x</code>	A numeric value.
<code>k</code>	Number of digits after decimal point (should be an integer) (Default: $k = 3L$ ).
<code>p.value</code>	Decides whether the number is a $p$ -value (Default: <code>FALSE</code> ).

### Value

Formatted numeric value.

### Examples

```
specify_decimal_p(x = 0.00001, k = 2, p.value = TRUE)
specify_decimal_p(x = 0.008, k = 2, p.value = TRUE)
specify_decimal_p(x = 0.008, k = 3, p.value = FALSE)
```

---

stats\_type\_switch      *Switch type of statistics.*

---

### Description

Relevant mostly for `ggstatsplot` and `statsExpressions` packages, where there are four types of statistics are supported: parametric, non-parametric, robust, and Bayesian. This switch function converts strings entered by users to a common pattern.

### Usage

```
stats_type_switch(type)
```



**Arguments**

type                    Character string describing the type of statistics.

**Examples**

```
stats_type_switch("p")  
stats_type_switch("bf")
```

# Index

## \*Topic **datasets**

bugs\_long, [2](#)

iris\_long, [3](#)

bugs\_long, [2](#)

iris\_long, [3](#)

long\_to\_wide\_converter, [4](#)

outlier\_df, [5](#)

p\_value\_formatter, [6](#)

set\_cwd, [6](#)

signif\_column, [7](#)

specify\_decimal\_p, [8](#)

stats\_type\_switch, [8](#)