

Package ‘PenguinR’

May 7, 2026

Type Package

Title A Comprehensive Collection of Penguin Datasets for Statistical Analysis and Experimental Design

Version 0.1.0

Maintainer Juan Pablo Vargas Perez <j.pablovargas340@gmail.com>

Description Offers a comprehensive collection of penguin-related datasets suitable for descriptive statistics, hypothesis testing, and experimental design. Derived from open ecological and biological sources such as Palmer Station studies, the package integrates datasets covering adult morphology, clutch size, blood isotope composition, and heart rate. It is designed for researchers, students, and educators to explore statistical methods including ANOVA, regression, multivariate analysis, and design of experiments in an accessible and reproducible context.

License GPL-3

URL <https://github.com/jpablovargas340/PenguinR>,
<https://jpablovargas340.github.io/PenguinR/>

BugReports <https://github.com/jpablovargas340/PenguinR/issues>

Encoding UTF-8

LazyData true

Suggests ggplot2, dplyr, testthat (>= 3.0.0), knitr, rmarkdown

RoxygenNote 7.3.3

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Author Juan Pablo Vargas Perez [aut, cre]

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2025-10-31 18:10:14 UTC

Contents

PenguinR	2
penguins_df	2
penguins_raw_df	3
peng_df	5
pinguinos_df	6
Index	7

PenguinR	<i>PenguinR: A Comprehensive Collection of Penguin Datasets for Statistical Analysis and Experimental Design</i>
----------	--

Description

Offers a comprehensive collection of penguin-related datasets suitable for descriptive statistics, hypothesis testing, and experimental design.

Details

PenguinR: A Comprehensive Collection of University and Engineering Datasets
 A Comprehensive Collection of University and Engineering Datasets.

Author(s)

Maintainer: Juan Pablo Vargas Perez <j.pablovargas340@gmail.com>

See Also

Useful links:

- <https://github.com/jpablovargas340/PenguinR>

penguins_df	<i>Size Measurements for Adult Foraging Penguins near Palmer Station, Antarctica</i>
-------------	--

Description

This dataset, penguins_df, is a data frame with measurements on adult penguins from three species and three islands in the Palmer Archipelago, Antarctica. Variables include flipper length, body mass, bill length and depth, sex, and year of observation. Some observations may contain missing values.

Usage

```
data(penguins_df)
```

Format

A data frame with 344 observations and 8 variables:

- species** Penguin species (factor with 3 levels).
- island** Island of observation (factor with 3 levels).
- bill_length_mm** Bill length in millimeters (numeric).
- bill_depth_mm** Bill depth in millimeters (numeric).
- flipper_length_mm** Flipper length in millimeters (integer).
- body_mass_g** Body mass in grams (integer).
- sex** Sex of the penguin (factor with 2 levels).
- year** Year of observation (integer).

Details

The columns in `penguins_df` are a subset of those in the more extensive `penguins_raw` data frame, which also includes nesting observations and blood isotope data. There are differences in column names and data types between the two datasets. See the *Format* section for details.

The dataset name has been kept as `'penguins_df'` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PenguinR package and assists users in identifying its specific characteristics. The suffix `'df'` indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the `palmerpenguins` package version 0.1.1

<code>penguins_raw_df</code>	<i>Penguin Size, Clutch, and Blood Isotope Data for Foraging Adults near Palmer Station, Antarctica</i>
------------------------------	---

Description

This dataset, `penguins_raw_df`, is a data frame containing comprehensive biological data for adult penguins observed near Palmer Station, Antarctica. It includes size measurements, clutch completion status, blood isotope values, and additional observational metadata. The dataset covers three penguin species from three islands in the Palmer Archipelago.

Usage

```
data(penguins_raw_df)
```

Format

A data frame with 344 observations and 17 variables:

studyName Name of the study or project.

Sample Number Unique numeric sample identifier.

Species Species name (character).

Region Region of the Palmer Archipelago (character).

Island Island of observation (character).

Stage Stage of breeding or observation period (character).

Individual ID Unique penguin identifier (character).

Clutch Completion Clutch completion status (character).

Date Egg Date the egg was recorded (Date).

Culmen Length (mm) Bill length in millimeters (numeric).

Culmen Depth (mm) Bill depth in millimeters (numeric).

Flipper Length (mm) Flipper length in millimeters (numeric).

Body Mass (g) Body mass in grams (numeric).

Sex Sex of the penguin (character).

Delta 15 N (o/oo) Nitrogen isotope ratio $\delta^{15}N$ (per mil).

Delta 13 C (o/oo) Carbon isotope ratio $\delta^{13}C$ (per mil).

Comments Additional notes and remarks (character).

Details

The columns of `penguins_raw_df` extend those of the `penguins_df` dataset, adding details such as study name, individual ID, clutch information, isotope composition ($\delta^{15}N$ and $\delta^{13}C$), and additional observational notes. Some variables include missing values.

The dataset name has been kept as `'penguins_raw_df'` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the `PenguinR` package and assists users in identifying its specific characteristics. The suffix `'df'` indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the `palmerpenguins` package version 0.1.1

peng_df

Size Measurements for Penguins near Palmer Station, Antarctica

Description

This dataset, `peng_df`, is a data frame containing morphometric measurements for penguins observed near Palmer Station in the Palmer Archipelago, Antarctica. It includes data on three species across multiple islands, with variables for bill dimensions, flipper length, body mass, sex, and year of observation. Some observations may contain missing values.

Usage

```
data(peng_df)
```

Format

A data frame with 333 observations and 8 variables:

species Penguin species (factor with 3 levels).

island Island of observation (factor with 3 levels).

bill_length Bill length in millimeters (numeric).

bill_depth Bill depth in millimeters (numeric).

flipper_length Flipper length in millimeters (integer).

body_mass Body mass in grams (integer).

sex Sex of the penguin (factor with 2 levels).

year Year of observation (integer).

Details

The data were originally sourced from the `palmerpenguins` dataset and are provided in the `heplots` package for illustrative and educational use in multivariate statistical analysis examples.

The dataset name has been kept as `'peng_df'` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the `PenguinR` package and assists users in identifying its specific characteristics. The suffix `'df'` indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the `heplots` package version 1.7.8

pinguinos_df

Penguin Heart Rate

Description

This dataset, `pinguinos_df`, is a data frame containing the slopes of regressions of mass-specific metabolic rate on heart rate for three groups of Macaroni Penguins. The data summarize physiological relationships between metabolic and cardiac activity under different breeding and sex conditions.

Usage

```
data(pinguinos_df)
```

Format

A data frame with 24 observations and 2 variables:

group Penguin group (factor with 3 levels: "breeding female", "breeding male", "nonbreeding female").

slope Slope of the regression of mass-specific metabolic rate on heart rate (numeric).

Details

The dataset includes one factor variable indicating the penguin group and one numeric variable representing the slope of the regression for each observation. Some observations may contain missing values.

The dataset name has been kept as 'pinguinos_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PenguinR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the `abd` package version 0.2-8

Index

[peng_df](#), [5](#)
[PenguinR](#), [2](#)
[PenguinR-package \(PenguinR\)](#), [2](#)
[penguins_df](#), [2](#)
[penguins_raw_df](#), [3](#)
[pinguinos_df](#), [6](#)