

# Package: MedDataSets (via r-universe)

October 25, 2024

**Type** Package

**Title** Comprehensive Medical, Disease, Treatment, and Drug Datasets

**Version** 0.1.0

**Maintainer** Renzo Caceres Rossi <arenzocaceresrossi@gmail.com>

**Description** Provides an extensive collection of datasets related to medicine, diseases, treatments, drugs, and public health. This package covers topics such as drug effectiveness, vaccine trials, survival rates, infectious disease outbreaks, and medical treatments. The included datasets span various health conditions, including AIDS, cancer, bacterial infections, and COVID-19, along with information on pharmaceuticals and vaccines. These datasets are sourced from the R ecosystem and other R packages, remaining unaltered to ensure data integrity. This package serves as a valuable resource for researchers, analysts, and healthcare professionals interested in conducting medical and public health data analysis in R.

**License** GPL-3

**URL** <https://github.com/lightbluetitan/meddatasets>

**Encoding** UTF-8

**LazyData** true

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

**RoxygenNote** 7.3.2

**Config/testthat/edition** 3

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Renzo Caceres Rossi [aut, cre]

**Depends** R (>= 3.5.0)

**Repository** CRAN

**Date/Publication** 2024-10-24 14:20:02 UTC

## Contents

|                                    |    |
|------------------------------------|----|
| Aids2_df . . . . .                 | 3  |
| anorexia_df . . . . .              | 4  |
| antibiotics_tbl_df . . . . .       | 4  |
| avandia_tbl_df . . . . .           | 5  |
| babies_crawl_tbl_df . . . . .      | 6  |
| babies_tbl_df . . . . .            | 6  |
| bacteria_df . . . . .              | 7  |
| bdims_tbl_df . . . . .             | 8  |
| biontech_teens_tbl_df . . . . .    | 9  |
| birthwt_df . . . . .               | 10 |
| colmozzie_tbl_df . . . . .         | 11 |
| covid19sf_hospital_df . . . . .    | 12 |
| covid19sf_tests_df . . . . .       | 13 |
| Cushings_df . . . . .              | 14 |
| diabetes2_tbl_df . . . . .         | 14 |
| drugbank_df . . . . .              | 15 |
| ebola_survey_tbl_df . . . . .      | 16 |
| edgar_df . . . . .                 | 16 |
| esoph_df . . . . .                 | 17 |
| fdeaths_ts . . . . .               | 18 |
| GAGurine_df . . . . .              | 18 |
| gehan_df . . . . .                 | 19 |
| genotype_df . . . . .              | 20 |
| heart_transplant_tbl_df . . . . .  | 21 |
| infert_df . . . . .                | 22 |
| ldeaths_ts . . . . .               | 23 |
| leuk_df . . . . .                  | 23 |
| malaria_tbl_df . . . . .           | 24 |
| mala_df . . . . .                  | 25 |
| mdeaths_ts . . . . .               | 25 |
| MedDataSets . . . . .              | 26 |
| Melanoma_df . . . . .              | 27 |
| Mixtures_Drug_tbl_df . . . . .     | 28 |
| muscle_df . . . . .                | 28 |
| Pima_te_df . . . . .               | 29 |
| Pima_tr2_df . . . . .              | 30 |
| Pima_tr_df . . . . .               | 31 |
| Puromycin_df . . . . .             | 32 |
| sinusitis_tbl_df . . . . .         | 32 |
| sleep_deprivation_tbl_df . . . . . | 33 |
| smallpox_tbl_df . . . . .          | 34 |
| smoking_tbl_df . . . . .           | 35 |
| stent30_tbl_df . . . . .           | 36 |
| ToothGrowth_df . . . . .           | 36 |
| transplant_tbl_df . . . . .        | 37 |
| VADeaths_matrix . . . . .          | 38 |

|                       |           |
|-----------------------|-----------|
| <i>Aids2_df</i>       | 3         |
| VA_df . . . . .       | 38        |
| wtloss_df . . . . .   | 39        |
| yawn_tbl_df . . . . . | 40        |
| <b>Index</b>          | <b>41</b> |

---

|          |                                      |
|----------|--------------------------------------|
| Aids2_df | <i>Australian AIDS Survival Data</i> |
|----------|--------------------------------------|

---

### Description

The dataset name has been changed to 'Aids2\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(Aids2_df)
```

### Format

A data frame with 2843 observations and 7 variables:

**state** A factor indicating the state of residence of the patient (4 levels).

**sex** A factor indicating the sex of the patient (2 levels).

**diag** An integer indicating the year of diagnosis.

**death** An integer indicating the year of death.

**status** A factor indicating the status of the patient (2 levels: alive or deceased).

**T.categ** A factor indicating the T-cell category of the patient (8 levels).

**age** An integer indicating the age of the patient at diagnosis.

### Details

This dataset provides information on the survival rates and characteristics of AIDS patients in Australia, including their demographic details and clinical data.

### Source

Australian Department of Health.

---

anorexia\_df                      *Anorexia Data on Weight Change*

---

**Description**

The dataset name has been changed to 'anorexia\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(anorexia_df)
```

**Format**

A data frame with 72 observations and 3 variables:

**Treat** A factor indicating the treatment group (with 3 possible levels).

**Prewt** A numeric value representing the weight of the patient before treatment (in pounds).

**Postwt** A numeric value representing the weight of the patient after treatment (in pounds).

**Details**

This dataset contains information on weight changes among patients diagnosed with anorexia, including their treatment and weight measurements before and after treatment.

**Source**

Data collected from clinical studies on anorexia treatment and weight change.

---

antibiotics\_tbl\_df                      *Pre-existing Conditions in Children*

---

**Description**

The dataset name has been changed to 'antibiotics\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(antibiotics_tbl_df)
```

**Format**

A tibble with 92 observations and 1 variable:

**condition** A factor indicating the pre-existing condition of the children (with 8 possible levels).

**Details**

This dataset contains information about pre-existing conditions in 92 children, providing insights into the prevalence of different conditions in the sample.

**Source**

Data collected from a clinical study assessing children's health conditions.

---

avandia\_tbl\_df

*Cardiovascular Problems for Two Types of Diabetes Medicines*

---

**Description**

The dataset name has been changed to 'avandia\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(avandia_tbl_df)
```

**Format**

A tibble with 227,571 observations and 2 variables:

**treatment** A factor indicating the type of diabetes medicine (with 2 possible levels).

**cardiovascular\_problems** A factor indicating the presence of cardiovascular problems (with 2 possible levels).

**Details**

This dataset contains information on cardiovascular problems associated with two types of diabetes medicines, providing insights into the safety and efficacy of these treatments.

**Source**

Data collected from clinical trials assessing the cardiovascular effects of diabetes medications.

---

`babies_crawl_tbl_df`     *Crawling Age*

---

### Description

The dataset name has been changed to 'babies\_crawl\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(babies_crawl_tbl_df)
```

### Format

A tibble with 12 observations and 5 variables:

**birth\_month** A factor indicating the month of birth (January to December).

**avg\_crawling\_age** A numeric value representing the average crawling age (in months).

**sd** A numeric value indicating the standard deviation of crawling age.

**n** An integer representing the number of infants included in the calculation.

**temperature** An integer indicating the average temperature (in degrees Celsius) during the month.

### Details

This dataset contains information on the average crawling age of infants based on the month of birth, as well as associated factors such as temperature.

### Source

Data collected on the crawling age of infants based on birth month and temperature.

---

`babies_tbl_df`     *The Child Health and Development Studies*

---

### Description

The dataset name has been changed to 'babies\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(babies_tbl_df)
```

**Format**

A tibble with 1,236 observations and 8 variables:

**case** An integer indicating the case number.

**bwt** An integer representing the birth weight of the infant (in grams).

**gestation** An integer indicating the gestation period (in weeks).

**parity** An integer representing the number of previous births.

**age** An integer indicating the age of the mother (in years).

**height** An integer indicating the height of the mother (in cm).

**weight** An integer indicating the weight of the mother (in kg).

**smoke** An integer indicating whether the mother smoked during pregnancy (1 = yes, 0 = no).

**Details**

This dataset contains information from the Child Health and Development Studies, focusing on various factors associated with infant health outcomes.

**Source**

Data collected from the Child Health and Development Studies.

---

bacteria\_df

*Presence of Bacteria after Drug Treatments*

---

**Description**

The dataset name has been changed to 'bacteria\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(bacteria_df)
```

**Format**

A data frame with 220 observations and 6 variables:

**y** A factor indicating the presence (1) or absence (0) of bacteria.

**ap** A factor indicating the result of an antibiotic susceptibility test (yes/no).

**hilo** A factor indicating a high or low bacterial load (high/low).

**week** An integer representing the week of treatment.

**ID** A factor representing the unique identifier for each patient (with 50 possible levels).

**trt** A factor indicating the treatment group (with 3 possible levels).

**Details**

This dataset contains information on the presence of bacteria in patients after receiving various drug treatments, including the treatment group and duration of treatment.

**Source**

Data collected from clinical trials on antibiotic treatments and their effects on bacterial presence.

---

bdims\_tbl\_df

*Body measurements of 507 physically active individuals.*

---

**Description**

The dataset name has been changed to 'bdims\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(bdims_tbl_df)
```

**Format**

A tibble with 507 observations and 25 variables:

**bia\_di** Numerical value of body measurement (in cm).

**bii\_di** Numerical value of body measurement (in cm).

**bit\_di** Numerical value of body measurement (in cm).

**che\_de** Numerical value of body measurement (in cm).

**che\_di** Numerical value of body measurement (in cm).

**elb\_di** Numerical value of body measurement (in cm).



**wri\_di** Numerical value of body measurement (in cm).  
**kne\_di** Numerical value of body measurement (in cm).  
**ank\_di** Numerical value of body measurement (in cm).  
**sho\_gi** Numerical value of body measurement (in cm).  
**che\_gi** Numerical value of body measurement (in cm).  
**wai\_gi** Numerical value of body measurement (in cm).  
**nav\_gi** Numerical value of body measurement (in cm).  
**hip\_gi** Numerical value of body measurement (in cm).  
**thi\_gi** Numerical value of body measurement (in cm).  
**bic\_gi** Numerical value of body measurement (in cm).  
**for\_gi** Numerical value of body measurement (in cm).  
**kne\_gi** Numerical value of body measurement (in cm).  
**cal\_gi** Numerical value of body measurement (in cm).  
**ank\_gi** Numerical value of body measurement (in cm).  
**wri\_gi** Numerical value of body measurement (in cm).  
**age** An integer indicating the age of the individual (in years).  
**wgt** A numeric value indicating the weight of the individual (in kg).  
**hgt** A numeric value indicating the height of the individual (in cm).  
**sex** An integer indicating the sex of the individual (1 = male, 2 = female).

### Details

This dataset contains body measurements of 507 physically active individuals, including various dimensions and physical attributes.

### Source

Data collected from physically active individuals to analyze body measurements.

---

biontech\_teens\_tbl\_df *Pfizer-BioNTech COVID-19 Vaccine Efficacy in Adolescents*

---

### Description

The dataset name has been changed to 'biontech\_teens\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(biontech_teens_tbl_df)
```

**Format**

A tibble with 2,260 observations and 2 variables:

**group** A factor indicating the group (e.g., vaccinated vs. unvaccinated).

**outcome** A factor indicating the outcome (e.g., infection status).

**Details**

This dataset contains information on the efficacy of the Pfizer-BioNTech COVID-19 vaccine among adolescents, detailing the outcomes based on different groups.

**Source**

Data collected during clinical trials to evaluate vaccine efficacy in adolescents.

---

birthwt\_df

*Risk Factors Associated with Low Infant Birth Weight*

---

**Description**

The dataset name has been changed to 'birthwt\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(birthwt_df)
```

**Format**

A data frame with 189 observations and 10 variables:

**low** An integer indicating whether the infant's birth weight is low (1) or not (0).

**age** An integer representing the age of the mother (in years).

**lwt** An integer indicating the mother's weight at last menstrual period (in pounds).

**race** An integer indicating the race of the mother (coded as 1, 2, or 3).

**smoke** An integer indicating whether the mother smoked during pregnancy (1 for yes, 0 for no).

**ptl** An integer indicating the number of premature labors.

**ht** An integer indicating whether the mother had a history of hypertension (1 for yes, 0 for no).

**ui** An integer indicating whether the mother had a history of uterine irritability (1 for yes, 0 for no).

**ftv** An integer indicating the number of physician visits during the first trimester.

**bwt** An integer representing the infant's birth weight (in grams).

## Details

This dataset contains information on risk factors associated with low infant birth weight, including maternal characteristics and behaviors during pregnancy.

## Source

Data collected from maternal health studies related to infant birth weight.

---

|                  |  |
|------------------|--|
| colmozzie_tbl_df | <i>Weekly Notified Dengue Cases in Sri Lanka</i> |
|------------------|--|

---

## Description

The dataset name has been changed to 'colmozzie\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

## Usage

```
data(colmozzie_tbl_df)
```

## Format

A tibble with 279 observations and 12 variables:

**Cases** Number of dengue cases reported during the week (integer).

**Year** Year of the reported cases (integer).

**Week** Week of the year (integer).

**TEM** Average temperature (numeric).

**TMAX** Maximum temperature recorded (numeric).

**Tm** Minimum temperature recorded (numeric).

**SLP** Sea level pressure (character).

**H** Humidity levels (numeric).

**PP** Precipitation levels (numeric).

**VV** Wind velocity (numeric).

**V** Another wind variable (numeric).

**VM** Yet another wind variable (numeric).

## Details

This dataset contains weekly reported cases of dengue fever in Sri Lanka, along with various meteorological variables that may be associated with the incidence of the disease.

## Source

The dataset is based on public health records and meteorological data from Sri Lanka.

---

covid19sf\_hospital\_df *San Francisco COVID-19 Hospital Capacity*

---

## Description

The dataset name has been changed to 'covid19sf\_hospital\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

## Usage

```
data(covid19sf_hospital_df)
```

## Format

A data frame with 4,514 observations and 5 variables:

**hospital** The name of the hospital (character).

**date** The date of the reported data (Date).

**bed\_type** The type of bed (character), such as ICU, general, etc.

**status** The status of the beds (character), indicating if they are occupied, available, etc.

**count** The number of beds reported (integer).

## Details

This dataset provides information on hospital capacity in San Francisco during the COVID-19 pandemic. It details the number of available hospital beds categorized by type and status, along with the respective hospitals and dates. The dataset is crucial for understanding the hospital system's response and capacity to handle COVID-19 cases.

## Source

San Francisco Department of Public Health COVID-19 hospital capacity data.

---

|                    |                                     |
|--------------------|-------------------------------------|
| covid19sf_tests_df | <i>San Francisco COVID-19 Tests</i> |
|--------------------|-------------------------------------|

---

## Description

The dataset name has been changed to 'covid19sf\_tests\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

## Usage

```
data(covid19sf_tests_df)
```

## Format

A data frame with 652 observations and 6 variables:

**specimen\_collection\_date** The date when the specimen was collected (Date).

**tests** The total number of tests conducted (integer).

**pos** The number of positive test results (integer).

**pct** The percentage of positive tests (numeric).

**neg** The number of negative test results (integer).

**indeterminate** The number of indeterminate test results (integer).

## Details

This dataset contains information on COVID-19 tests conducted in San Francisco, detailing the number of tests performed, the number of positive and negative results, as well as other related metrics. It provides insights into the testing patterns and results during the COVID-19 pandemic.

## Source

San Francisco Department of Public Health COVID-19 testing data.

---

Cushings\_df

*Diagnostic Tests on Patients with Cushing's Syndrome*

---

### Description

The dataset name has been changed to 'Cushings\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(Cushings_df)
```

### Format

A data frame with 27 observations and 3 variables:

**Tetrahydrocortisone** A numeric vector representing the levels of Tetrahydrocortisone.

**Pregnanetriol** A numeric vector representing the levels of Pregnanetriol.

**Type** A factor indicating the type of test conducted (4 levels).

### Details

This dataset contains results from diagnostic tests conducted on patients suspected of having Cushing's Syndrome, focusing on the measurement of specific hormonal metabolites.

### Source

Data collected from clinical trials and patient records.

---

diabetes2\_tbl\_df

*Type 2 Diabetes Clinical Trial for Patients Aged 10-17*

---

### Description

The dataset name has been changed to 'diabetes2\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(diabetes2_tbl_df)
```

**Format**

A tibble with 699 observations and 2 variables:

**treatment** A factor indicating the type of treatment administered (e.g., different medication types).

**outcome** A factor indicating the outcome of the treatment (e.g., improvement or no improvement).

**Details**

This dataset contains information from a clinical trial focusing on Type 2 diabetes in patients aged 10 to 17 years, detailing the treatment approaches and their outcomes.

**Source**

Data collected from clinical trials assessing treatment efficacy for Type 2 diabetes in adolescents.

---

drugbank\_df

*Relationship Between Gene and Disease*

---

**Description**

The dataset name has been changed to 'drugbank\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(drugbank_df)
```

**Format**

A data frame with 27,728 observations and 2 variables:

**gene** Gene associated with the disease (factor).

**disease** Disease associated with the gene (factor).

**Details**

This dataset contains information about the relationships between genes and diseases, providing insights into how specific genes may be associated with various diseases.

**Source**

The dataset is derived from drug interaction databases and gene-disease relationships.

---

ebola\_survey\_tbl\_df      *Survey on Ebola Quarantine*

---

### Description

The dataset name has been changed to 'ebola\_survey\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(ebola_survey_tbl_df)
```

### Format

A tibble with 1,042 observations and 1 variable:

**quarantine** A factor indicating the responses related to quarantine measures (e.g., yes or no).

### Details

This dataset contains survey responses regarding quarantine measures during the Ebola outbreak, focusing on public perceptions and behaviors.

### Source

Data collected from surveys conducted during the Ebola outbreak to assess public sentiment towards quarantine.

---

edgar\_df      *Relationship Between Gene and Disease in Edgar*

---

### Description

The dataset name has been changed to 'edgar\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(edgar_df)
```



**Format**

A data frame with 1,038,340 observations and 2 variables:

**Gene** Gene associated with the disease (factor).

**Disease** Disease associated with the gene (factor).

**Details**

This dataset contains information about the relationships between genes and diseases, specifically focusing on data sourced from the Edgar database, providing insights into how specific genes may be associated with various diseases.

**Source**

The dataset is derived from the Edgar database and focuses on gene-disease relationships.

---

|          |  |
|----------|--|
| esoph_df | <i>Smoking, Alcohol and (O)esophageal Cancer</i> |
|----------|--|

---

**Description**

The dataset name has been changed to 'esoph\_df' to avoid confusion with datasets from other packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(esoph_df)
```

**Format**

A data frame with 88 observations and 5 variables:

**agegp** Age group of the individuals (ordered factor).

**alcgp** Alcohol consumption group (ordered factor).

**tobgp** Tobacco consumption group (ordered factor).

**ncases** Number of cases (numeric).

**ncontrols** Number of controls (numeric).

**Details**

This dataset contains data from a case-control study investigating the association between smoking, alcohol consumption, and esophageal cancer. It includes the number of cancer cases and controls for various age, alcohol consumption, and smoking groups.

**Source**

Data from a case-control study on esophageal cancer.

---

|            |  |
|------------|--|
| fdeaths_ts | <i>Monthly Deaths from Lung Diseases in the UK (Females)</i> |
|------------|--|

---

**Description**

The dataset name has been changed to 'fdeaths\_ts' to avoid confusion with datasets from other packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_ts' indicates that this dataset is a time series object, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(fdeaths_ts)
```

**Format**

A time series object with 72 observations, from 1974 to 1980:

**fdeaths** A numeric vector representing the number of monthly deaths due to lung diseases in females.

**Details**

This dataset contains the number of monthly deaths from lung diseases in the UK, specifically for females, from 1974 to 1980.

**Source**

UK Health Authority data on lung disease deaths.

---

|             |  |
|-------------|--|
| GAGurine_df | <i>Level of GAG in Urine of Children</i> |
|-------------|--|

---

**Description**

The dataset name has been changed to 'GAGurine\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(GAGurine_df)
```

**Format**

A data frame with 314 observations and 2 variables:

**Age** A numeric vector representing the age of the children in years.

**GAG** A numeric vector representing the levels of glycosaminoglycans (GAG) in urine (in some appropriate unit).

**Details**

This dataset contains measurements of glycosaminoglycan (GAG) levels in urine samples collected from children, aimed at studying potential health implications associated with abnormal GAG levels.

**Source**

Data collected from pediatric health studies.

---

gehan\_df

*Remission Times of Leukaemia Patients*

---

**Description**

The dataset name has been changed to 'gehan\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(gehan_df)
```

**Format**

A data frame with 42 observations and 4 variables:

**pair** An integer representing the patient pair identifier.

**time** An integer indicating the time to remission (in days).

**cens** An integer indicating whether the observation was censored (1 for censored, 0 for not censored).

**treat** A factor indicating the treatment group (with 2 possible levels).

**Details**

This dataset contains information on the remission times of leukaemia patients, including treatment information and whether the remission time was censored.

**Source**

Data collected from clinical studies on leukaemia treatment and remission.

---

genotype\_df

*Rat Genotype Data*

---

**Description**

The dataset name has been changed to 'genotype\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(genotype_df)
```

**Format**

A data frame with 61 observations and 3 variables:

**Litter** A factor indicating the litter group (with 4 possible levels).

**Mother** A factor indicating the mother of the rats (with 4 possible levels).

**Wt** A numeric value representing the weight of the rats (in grams).

**Details**

This dataset contains genotype data from rats, including information on litter, maternal lineage, and weight measurements.

**Source**

Data collected from genetic studies involving rat populations.

---

`heart_transplant_tbl_df`*Heart Transplant Data*

---

## Description

The dataset name has been changed to 'heart\_transplant\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'Med-DataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

## Usage

```
data(heart_transplant_tbl_df)
```

## Format

A tibble with 103 observations and 8 variables:

**id** An integer identifier for each patient.

**acceptyear** The year the patient was accepted for transplantation.

**age** The age of the patient at the time of transplantation.

**survived** A factor indicating whether the patient survived post-transplant (e.g., yes or no).

**survtime** The time (in months) the patient survived after the transplant.

**prior** A factor indicating whether the patient had prior heart conditions (e.g., yes or no).

**transplant** A factor indicating the type of transplant (e.g., heart only, heart-lung).

**wait** The wait time (in days) for the transplant.

## Details

This dataset contains information on heart transplant patients, including demographics, survival outcomes, and wait times.

## Source

Data collected from heart transplant records to study patient outcomes and factors influencing survival.

---

`infert_df`*Infertility after Spontaneous and Induced Abortion*

---

### Description

The dataset name has been changed to 'infert\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(infert_df)
```

### Format

A data frame with 248 observations and 8 variables:

**education** A factor representing the education level of the subjects, with 3 levels.

**age** A numeric vector indicating the age of the subjects.

**parity** A numeric vector representing the number of previous pregnancies.

**induced** A numeric vector indicating the number of induced abortions.

**case** A numeric vector indicating the case status (infertile or not).

**spontaneous** A numeric vector indicating the number of spontaneous abortions.

**stratum** An integer representing the stratum in the study.

**pooled.stratum** A numeric vector representing the pooled stratum values.

### Details

This dataset examines the relationship between various factors such as education level, age, parity, and the incidence of infertility after spontaneous and induced abortion.

### Source

Data collected from clinical studies on infertility.

---

`ldeaths_ts`*Monthly Deaths from Lung Diseases in the UK*

---

**Description**

The dataset name has been changed to 'Ideaths\_ts' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_ts' indicates that this dataset is a time series, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(ldeaths_ts)
```

**Format**

A time series object with 72 observations:

**Ideaths** A numeric vector containing the number of monthly deaths from lung diseases in the UK.

**Details**

This dataset provides information on the monthly deaths from lung diseases in the UK, recorded from 1974 to 1980.

**Source**

Office for National Statistics, UK.

---

`leuk_df`*Survival Times and White Blood Counts for Leukaemia Patients*

---

**Description**

The dataset name has been changed to 'leuk\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(leuk_df)
```

**Format**

A data frame with 33 observations and 3 variables:

**wbc** An integer representing the white blood cell count (in thousands per microliter).

**ag** A factor indicating the treatment group (with 2 possible levels).

**time** An integer indicating the survival time (in days).

**Details**

This dataset contains survival times and white blood cell counts for leukaemia patients, providing insights into the relationship between blood counts and survival outcomes.

**Source**

Data collected from clinical studies on leukaemia patients.

---

|                |                              |
|----------------|------------------------------|
| malaria_tbl_df | <i>Malaria Vaccine Trial</i> |
|----------------|------------------------------|

---

**Description**

The dataset name has been changed to 'malaria\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(malaria_tbl_df)
```

**Format**

A tibble with 20 observations and 2 variables:

**treatment** A factor indicating the type of treatment administered (e.g., vaccine or placebo).

**outcome** A factor indicating the outcome of the treatment (e.g., success or failure).

**Details**

This dataset contains information from a malaria vaccine trial, focusing on the treatment administered and the outcomes observed in the participants.

**Source**

Data collected from a clinical trial assessing the efficacy of a malaria vaccine.



---

`mala_df`*The Relationship Between Gene and Disease*

---

**Description**

The dataset name has been changed to 'mala\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(mala_df)
```

**Format**

A data frame with 241,306 observations and 2 variables:

**disease** Disease associated with the gene (factor).

**gene** Gene associated with the disease (factor).

**Details**

This dataset contains information about the relationships between genes and diseases, providing insights into how specific genes are associated with various diseases. It offers a comprehensive view of gene-disease associations.

**Source**

The dataset contains gene-disease relationship data from various scientific studies and databases.

---

`mdeaths_ts`*Monthly Deaths from Lung Diseases in the UK*

---

**Description**

The dataset name has been changed to 'mdeaths\_ts' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_ts' indicates that this dataset is a time series, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(mdeaths_ts)
```

**Format**

A time series object with 72 observations:

**mdeaths** A numeric vector containing the number of monthly deaths from lung diseases in the UK.

**Details**

This dataset provides information on the monthly deaths from lung diseases in the UK, recorded from 1974 to 1980.

**Source**

Office for National Statistics, UK.

---

|             |  |
|-------------|--|
| MedDataSets | <i>MedDataSets: Comprehensive Medical, Disease, Treatment, and Drug Datasets</i> |
|-------------|--|

---

**Description**

The MedDataSets package provides an extensive collection of datasets related to medicine, diseases, treatments, drugs, and public health. It covers topics such as drug effectiveness, vaccine trials, survival rates, infectious disease outbreaks, and medical treatments.

**Details**

MedDataSets: Comprehensive Medical, Disease, Treatment, and Drug Datasets

Comprehensive Medical, Disease, Treatment, and Drug Datasets

**Author(s)**

**Maintainer:** Renzo Caceres Rossi <arenzocaceresrossi@gmail.com>

**See Also**

Useful links:

- <https://github.com/lightbluetitan/meddatasets>

---

Melanoma\_df

*Survival from Malignant Melanoma*

---

### Description

The dataset name has been changed to 'Melanoma\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(Melanoma_df)
```

### Format

A data frame with 205 observations and 7 variables:

**time** An integer representing the survival time of the patients (in months).

**status** An integer indicating the status of the patient at the end of the study; typically coded as 1 for dead and 0 for alive.

**sex** An integer representing the sex of the patient; usually coded as 1 for male and 0 for female.

**age** An integer indicating the age of the patient at diagnosis (in years).

**year** An integer representing the year of diagnosis.

**thickness** A numeric value indicating the thickness of the melanoma (in millimeters).

**ulcer** An integer indicating the presence of ulceration; usually coded as 1 for yes and 0 for no.

### Details

This dataset contains information on survival rates of patients diagnosed with malignant melanoma, including various clinical factors that may affect prognosis.

### Source

Data collected from clinical studies on malignant melanoma.

---

Mixtures\_Drug\_tbl\_df *Drug Mixture*

---

### Description

The dataset name has been changed to 'Mixtures\_Drug\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(Mixtures_Drug_tbl_df)
```

### Format

A tibble with 819 observations and 3 variables:

**name** Name of the drug mixture (character).

**ingredients** Ingredients that make up the drug mixture (character).

**parent\_key** Identifier linking the mixture to its parent compound or category (character).

### Details

This dataset provides information about various drug mixtures, including their names and ingredients.

### Source

The dataset is derived from publicly available pharmaceutical databases and research studies.

---

muscle\_df

*Effect of Calcium Chloride on Muscle Contraction in Rat Hearts*

---

### Description

The dataset name has been changed to 'muscle\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(muscle_df)
```

**Format**

A data frame with 60 observations and 3 variables:

**Strip** A factor indicating the specific muscle strip (with 21 possible levels).

**Conc** A numeric value representing the concentration of calcium chloride (in mM).

**Length** A numeric value indicating the length of the muscle strip (in millimeters).

**Details**

This dataset contains experimental data on the effect of calcium chloride on muscle contraction in rat hearts, including measurements of muscle strip length and calcium concentration.

**Source**

Data collected from experiments assessing the impact of calcium chloride on muscle contraction.

---

Pima\_te\_df

*Diabetes in Pima Indian Women*

---

**Description**

The dataset name has been changed to 'Pima\_te\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(Pima_te_df)
```

**Format**

A data frame with 332 observations and 8 variables:

**npreg** An integer representing the number of pregnancies.

**glu** An integer indicating the plasma glucose concentration (mg/dL) 2 hours after an oral glucose tolerance test.

**bp** An integer representing the diastolic blood pressure (mm Hg).

**skin** An integer indicating the skin thickness (mm).

**bmi** A numeric value indicating the body mass index (BMI).

**ped** A numeric value representing the diabetes pedigree function.

**age** An integer indicating the age of the individual (in years).

**type** A factor indicating whether the individual has diabetes (1) or not (0).

**Details**

This dataset contains medical examination data for Pima Indian women, including various health metrics that may be related to diabetes.

**Source**

Data collected from medical examinations of Pima Indian women.

---

Pima\_tr2\_df

*Diabetes in Pima Indian Women*

---

**Description**

The dataset name has been changed to 'Pima\_tr2\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(Pima_tr2_df)
```

**Format**

A data frame with 300 observations and 8 variables:

**npreg** An integer representing the number of pregnancies.

**glu** An integer indicating the plasma glucose concentration (mg/dL) 2 hours after an oral glucose tolerance test.

**bp** An integer representing the diastolic blood pressure (mm Hg).

**skin** An integer indicating the skin thickness (mm).

**bmi** A numeric value indicating the body mass index (BMI).

**ped** A numeric value representing the diabetes pedigree function.

**age** An integer indicating the age of the individual (in years).

**type** A factor indicating whether the individual has diabetes (1) or not (0).

**Details**

This dataset contains medical examination data for Pima Indian women, including various health metrics that may be related to diabetes.

**Source**

Data collected from medical examinations of Pima Indian women.

---

Pima\_tr\_df

*Diabetes in Pima Indian Women*

---

## Description

The dataset name has been changed to 'Pima\_tr\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

## Usage

```
data(Pima_tr_df)
```

## Format

A data frame with 200 observations and 8 variables:

**npreg** An integer representing the number of pregnancies.

**glu** An integer indicating the plasma glucose concentration (mg/dL) 2 hours after an oral glucose tolerance test.

**bp** An integer representing the diastolic blood pressure (mm Hg).

**skin** An integer indicating the skin thickness (mm).

**bmi** A numeric value indicating the body mass index (BMI).

**ped** A numeric value representing the diabetes pedigree function.

**age** An integer indicating the age of the individual (in years).

**type** A factor indicating whether the individual has diabetes (1) or not (0).

## Details

This dataset contains medical examination data for Pima Indian women, including various health metrics that may be related to diabetes.

## Source

Data collected from medical examinations of Pima Indian women.

---

Puromycin\_df

*Reaction Velocity of an Enzymatic Reaction*

---

### Description

The dataset name has been changed to 'Puromycin\_df' to avoid confusion with datasets from other packages in the R ecosystem and to follow the naming convention in the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, distinguishing it both within the package and from similar datasets in other R packages. The original content of the dataset has not been modified in any way.

### Usage

```
data(Puromycin_df)
```

### Format

A data frame with 23 observations and 3 variables:

**conc** Substrate concentration (numeric).

**rate** Reaction velocity (numeric).

**state** A factor with two levels: treated and untreated, indicating whether Puromycin was present.

The dataset contains additional metadata:

**reference** The reference for this dataset: "A1.3, p. 269".

### Details

This dataset contains data from an experiment on the reaction velocity of an enzymatic reaction with and without the use of Puromycin, an antibiotic that inhibits protein synthesis.

### Source

Experimental data on the effects of Puromycin on enzyme reaction rates.

---

sinusitis\_tbl\_df

*Sinusitis and Antibiotic Experiment*

---

### Description

The dataset name has been changed to 'sinusitis\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.



**Usage**

```
data(sinusitis_tbl_df)
```

**Format**

A tibble with 166 observations and 2 variables:

**group** A factor indicating the treatment group (e.g., antibiotic vs. placebo).

**self\_reported\_improvement** A factor indicating the participants' self-reported improvement (e.g., yes or no).

**Details**

This dataset contains information from an experiment assessing the effects of antibiotics on patients with sinusitis, focusing on the group assignments and the self-reported improvement outcomes observed in the participants.

**Source**

Data collected from a clinical trial investigating the efficacy of antibiotics in treating sinusitis.

---

sleep\_deprivation\_tbl\_df

*Survey on Sleep Deprivation and Transportation Workers*

---

**Description**

The dataset name has been changed to 'sleep\_deprivation\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'Med-DataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(sleep_deprivation_tbl_df)
```

**Format**

A tibble with 1,087 observations and 2 variables:

**sleep** A factor indicating the reported sleep deprivation level (e.g., low, moderate, high).

**profession** A factor indicating the profession of the participants (e.g., truck driver, pilot, etc.).

**Details**

This dataset contains information from a survey conducted on transportation workers, focusing on the relationship between sleep deprivation and their professional roles. It includes variables on the amount of sleep reported and the professions of the respondents.

**Source**

Data collected from a survey targeting transportation workers to assess the impact of sleep deprivation on their performance and well-being.

---

|                 |                                 |
|-----------------|---------------------------------|
| smallpox_tbl_df | <i>Smallpox Vaccine Results</i> |
|-----------------|---------------------------------|

---

**Description**

The dataset name has been changed to 'smallpox\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(smallpox_tbl_df)
```

**Format**

A tibble with 6,224 observations and 2 variables:

**result** A factor indicating the outcome of the vaccination (e.g., successful, unsuccessful).

**inoculated** A factor indicating whether the individual was inoculated with the smallpox vaccine.

**Details**

This dataset contains the results of a study on the efficacy of the smallpox vaccine. It includes information on the vaccination outcomes of individuals who were inoculated, providing insight into the effectiveness of the vaccine in preventing the disease.

**Source**

Data collected from studies on smallpox vaccination and its outcomes.

---

`smoking_tbl_df`*UK Smoking Data*

---

### Description

The dataset name has been changed to 'smoking\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(smoking_tbl_df)
```

### Format

A tibble with 1,691 observations and 12 variables:

**gender** A factor indicating the gender of the respondent (e.g., male, female).

**age** An integer representing the age of the respondent.

**marital\_status** A factor indicating the marital status of the respondent (e.g., single, married).

**highest\_qualification** A factor indicating the highest qualification attained by the respondent.

**nationality** A factor indicating the nationality of the respondent.

**ethnicity** A factor indicating the ethnicity of the respondent.

**gross\_income** A factor indicating the gross income level of the respondent.

**region** A factor indicating the geographical region of the respondent.

**smoke** A factor indicating whether the respondent is a smoker (e.g., yes, no).

**amt\_weekends** An integer representing the amount smoked on weekends.

**amt\_weekdays** An integer representing the amount smoked on weekdays.

**type** A factor indicating the type of smoking behavior (e.g., occasional, regular).

### Details

This dataset contains information on smoking habits in the UK, focusing on various demographic factors and smoking behaviors. It provides insights into smoking patterns among different groups of people, helping to inform public health strategies.

### Source

Data collected from UK health surveys focusing on smoking behavior.

---

|                |   |
|----------------|---|
| stent30_tbl_df | <i>Stents for the Treatment of Stroke</i> |
|----------------|---|

---

**Description**

The dataset name has been changed to 'stent30\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(stent30_tbl_df)
```

**Format**

A tibble with 451 observations and 2 variables:

**group** A factor indicating the treatment group (e.g., stent vs. control).

**outcome** A factor indicating the outcome of the treatment (e.g., success or failure).

**Details**

This dataset contains information regarding the use of stents for the treatment of stroke, focusing on the group assignments and the outcomes observed in the participants.

**Source**

Data collected from a clinical trial assessing the efficacy of stents in stroke treatment.

---

|                |   |
|----------------|---|
| ToothGrowth_df | <i>The Effect of Vitamin C on Tooth Growth in Guinea Pigs</i> |
|----------------|---|

---

**Description**

The dataset name has been changed to 'ToothGrowth\_df' to avoid confusion with datasets from other packages in the R ecosystem and to align with the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from datasets in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(ToothGrowth_df)
```

**Format**

A data frame with 60 observations and 3 variables:

**len** Tooth length (numeric).

**supp** Type of supplement: either "VC" (Vitamin C) or "OJ" (Orange Juice) (factor).

**dose** Dose of Vitamin C administered in milligrams per day (numeric).

**Details**

This dataset explores the effect of Vitamin C on tooth growth in guinea pigs. It includes data on tooth length as a response to different doses of Vitamin C, administered through two delivery methods.

**Source**

Experimental data on the effect of Vitamin C on tooth growth in guinea pigs.

---

|                   |   |
|-------------------|---|
| transplant_tbl_df | <i>Transplant consultant success rate (fake data)</i> |
|-------------------|---|

---

**Description**

The dataset name has been changed to 'transplant\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(transplant_tbl_df)
```

**Format**

A tibble with 62 observations and 1 variable:

**outcome** A factor indicating the outcome of the transplant procedure (e.g., success, failure).

**Details**

This dataset contains fake data representing the success rates of transplant consultants. It provides insights into the outcomes of transplant procedures performed by different consultants, useful for evaluating consultant performance and patient outcomes.

**Source**

Synthetic data created for educational purposes.

---

|                 |                                       |
|-----------------|---------------------------------------|
| VADeaths_matrix | <i>Death Rates in Virginia (1940)</i> |
|-----------------|---------------------------------------|

---

### Description

The dataset name has been changed to 'VADeaths\_matrix' to avoid confusion with datasets from other packages in the R ecosystem and to align with the naming conventions of the 'MedDataSets' package. The suffix '\_matrix' indicates that this dataset is a matrix, helping to distinguish it from other datasets within the package and from datasets in the broader R ecosystem. The original content of the dataset has not been modified in any way.

### Usage

```
data(VADeaths_matrix)
```

### Format

A matrix with 5 rows and 4 columns:

[,1 ] Death rates for Rural Male (numeric).

[,2 ] Death rates for Rural Female (numeric).

[,3 ] Death rates for Urban Male (numeric).

[,4 ] Death rates for Urban Female (numeric).

**Row labels** Age groups: 50-54, 55-59, 60-64, 65-69, 70-74.

### Details

This dataset contains death rates per 1,000 individuals in various population groups in Virginia in 1940, classified by age group and rural/urban residency.

### Source

U.S. Census Bureau, Virginia (1940) Death Records.

---

|       |   |
|-------|---|
| VA_df | <i>Veteran's Administration Lung Cancer Trial</i> |
|-------|---|

---

### Description

The dataset name has been changed to 'VA\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(VA_df)
```

**Format**

A data frame with 137 observations and 8 variables:

**stime** A numeric value representing the survival time (in days).

**status** A numeric value indicating the status of the patient (1 if the patient died, 0 otherwise).

**treat** A factor indicating the treatment group (e.g., treatment A or B).

**age** A numeric value representing the age of the patient (in years).

**Karn** A numeric value representing the Karnofsky performance status score.

**diag.time** A numeric value indicating the time since diagnosis (in days).

**cell** A factor indicating the cell type of the lung cancer (with 4 possible levels).

**prior** A factor indicating prior treatment (yes/no).

**Details**

This dataset contains data from the Veteran's Administration Lung Cancer Trial, which includes information on patients diagnosed with lung cancer, their treatment, and other relevant variables.

**Source**

Data collected from the Veteran's Administration Lung Cancer Trial.

---

wtloss\_df

*Weight Loss Data from an Obese Patient*

---

**Description**

The dataset name has been changed to 'wtloss\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_df' indicates that this dataset is a data frame, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(wtloss_df)
```

**Format**

A data frame with 52 observations and 2 variables:

**Days** An integer representing the number of days in the weight loss program.

**Weight** A numeric value indicating the weight of the patient (in kilograms).

**Details**

This dataset contains weight loss data from an obese patient, detailing the weight changes over a specified number of days during a weight loss program.

**Source**

Data collected from a clinical study on weight loss in obese patients.

---

yawn\_tbl\_df

*Contagiousness of Yawning*

---

**Description**

The dataset name has been changed to 'yawn\_tbl\_df' to avoid confusion with other datasets from packages in the R ecosystem and to follow the naming conventions of the 'MedDataSets' package. The suffix '\_tbl\_df' indicates that this dataset is a tibble, helping to distinguish it from other datasets within the package and from those in the broader R ecosystem. The original content of the dataset has not been modified in any way.

**Usage**

```
data(yawn_tbl_df)
```

**Format**

A tibble with 50 observations and 2 variables:

**result** A factor indicating the result of the yawning observation (e.g., yawned, did not yawn).

**group** A factor representing the group to which the participants belong (e.g., control, experimental).

**Details**

This dataset investigates the contagiousness of yawning. It includes results from an experiment that examines whether individuals yawn more when they are in the presence of someone else who is yawning, providing insights into social behaviors and contagion phenomena.

**Source**

Data collected from a study on yawning contagion.



# Index

Aids2\_df, 3  
anorexia\_df, 4  
antibiotics\_tbl\_df, 4  
avandia\_tbl\_df, 5

babies\_crawl\_tbl\_df, 6  
babies\_tbl\_df, 6  
bacteria\_df, 7  
bdims\_tbl\_df, 8  
biontech\_teens\_tbl\_df, 9  
birthwt\_df, 10

colmozzie\_tbl\_df, 11  
covid19sf\_hospital\_df, 12  
covid19sf\_tests\_df, 13  
Cushings\_df, 14

diabetes2\_tbl\_df, 14  
drugbank\_df, 15

ebola\_survey\_tbl\_df, 16  
edgar\_df, 16  
esoph\_df, 17

fdeaths\_ts, 18

GAGurine\_df, 18  
gehan\_df, 19  
genotype\_df, 20

heart\_transplant\_tbl\_df, 21

infert\_df, 22

ldeaths\_ts, 23  
leuk\_df, 23

mala\_df, 25  
malaria\_tbl\_df, 24  
mdeaths\_ts, 25  
MedDataSets, 26

MedDataSets-package (MedDataSets), 26  
Melanoma\_df, 27  
Mixtures\_Drug\_tbl\_df, 28  
muscle\_df, 28

Pima\_te\_df, 29  
Pima\_tr2\_df, 30  
Pima\_tr\_df, 31  
Puromycin\_df, 32

sinusitis\_tbl\_df, 32  
sleep\_deprivation\_tbl\_df, 33  
smallpox\_tbl\_df, 34  
smoking\_tbl\_df, 35  
stent30\_tbl\_df, 36

ToothGrowth\_df, 36  
transplant\_tbl\_df, 37

VA\_df, 38  
VADeaths\_matrix, 38

wtloss\_df, 39

yawn\_tbl\_df, 40