Package 'PulmoDataSets'

June 3, 2025

Type Package

Title A Curated Collection of Pulmonary and Respiratory Disease Datasets

Version 0.1.0

Maintainer Renzo Caceres Rossi <arenzocaceresrossi@gmail.com>

Description Provides a comprehensive and curated collection of datasets related to the lungs, respiratory system, and associated diseases.
This package includes epidemiological, clinical, experimental, and simulated datasets on conditions such as lung cancer, asthma,
Chronic Obstructive Pulmonary Disease (COPD), tuberculosis, whooping cough, pneumonia, influenza, and other respiratory illnesses.
It is designed to support data exploration, statistical modeling, teaching, and research in pulmonary medicine, public health,
environmental epidemiology, and respiratory disease surveillance.

License GPL-3

Language en

URL https://github.com/lightbluetitan/pulmodatasets, https://lightbluetitan.github.io/pulmodatasets/

BugReports https://github.com/lightbluetitan/pulmodatasets/issues

Encoding UTF-8

LazyData true

Depends R (>= 4.1.0)

Imports utils

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

RoxygenNote 7.3.2

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Author Renzo Caceres Rossi [aut, cre]

Repository CRAN

Date/Publication 2025-06-03 13:00:09 UTC

Contents

`air_polution_mortality_df` *Air Pollution and Mortality*

Description

This dataset, air_polution_mortality_df, is a data frame containing information from an early study exploring the relationship between air pollution and mortality across 5 Standard Metropolitan Statistical Areas in the U.S. between 1959 and 1961.

Usage

```
data(air_polution_mortality_df)
```

Format

A data frame with 60 observations and 7 variables:

City Metropolitan area (factor with 60 levels)

Mort Mortality rate (numeric vector)

Precip Annual precipitation in inches (integer vector)

Educ Median years of education (numeric vector)

NonWhite Percentage of non-white population (numeric vector)

NOX Nitrogen oxide concentration (integer vector)

SO2 Sulfur dioxide concentration (integer vector)

Details

The dataset name has been kept as 'air_polution_mortality_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets 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 Sleuth3 package version 1.0-6

ai_ipn_performance_dt *AI Assessment of Pulmonary Nodules*

Description

This dataset, ai_ipn_performance_dt, is a data table containing performance metrics of an artificial intelligence tool for risk stratification of 200 indeterminate pulmonary nodules (IPNs) on chest CT scans.

Usage

```
data(ai_ipn_performance_dt)
```

Format

A data table with 200 observations and 2 variables:

cancer Malignancy status (0 = benign, 1 = malignant) (integer)

rating AI risk assessment rating (integer)

Details

The dataset name has been kept as 'ai_ipn_performance_dt' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'dt' indicates that this is a data table object. The original content has not been modified in any way.

Source

Data taken from the R4HCR package version 0.1

asthma_patients_tbl_df *COPD and Asthma Patients*

Description

This dataset, asthma_patients_tbl_df, is a tibble containing clinical information about 300 asthma (COPD) patients tracked over 3 years, including demographics, smoking status, diagnosis details, medications, and peak flow measurements.

Usage

```
data(asthma_patients_tbl_df)
```

Format

A tibble with 300 observations and 7 variables:

Patient_ID Unique patient identifier (numeric)

Age Patient age in years (numeric)

Gender Patient gender (character)

Smoking_Status Current/Former/Never smoker status (character)

Asthma_Diagnosis Specific asthma/COPD diagnosis (character)

Medication Prescribed treatment regimen (character)

Peak_Flow Peak expiratory flow rate (numeric)

Details

The dataset name has been kept as 'asthma_patients_tbl_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the Pulmo-DataSets package and assists users in identifying its specific characteristics. The suffix 'tbl_df' indicates that the dataset is a tibble object. The original content has not been modified in any way.

Source

Data taken from Kaggle: https://www.kaggle.com/datasets/jatinthakur706/copd-asthma-patient-dataset

---

`bronchitis_Cardiff_df` *Chronic Bronchitis in Cardiff Men*

Description

This dataset, bronchitis_Cardiff_df, is a data frame containing information from a study assessing the effects of smoking and pollution on bronchitis diagnosis in a sample of 212 men from Cardiff.

Usage

```
data(bronchitis_Cardiff_df)
```

Format

A data frame with 212 observations and 4 variables:

cig Number of cigarettes smoked per day (numeric)

poll Pollution exposure level (numeric)

r Bronchitis diagnosis (0 = no, 1 = yes) (integer)

rfac Bronchitis diagnosis as a factor with 2 levels (factor)

Details

The dataset name has been kept as 'bronchitis_Cardiff_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets 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 gamclass package version 0.62.5

---

chicago_pollution_df *Chicago Mortality and Pollution*

Description

This dataset, chicago_pollution_df, is a data frame containing daily mortality, weather, and pollution data for Chicago from 1987 to 2000 from the National Morbidity, Mortality and Air Pollution Study (NMMAPS). It includes all-cause mortality, cardiovascular and respiratory deaths, temperature, humidity, and pollution levels (PM10 and ozone).

Usage

```
data(chicago_pollution_df)
```

Format

A data frame with 5114 observations and 14 variables:

date Date (Date object)

time Time index (integer vector)

year Year (numeric vector)

month Month (numeric vector)

doy Day of year (integer vector)

dow Day of week (factor with 7 levels)

death All-cause mortality count (integer vector)

cvd Cardiovascular mortality count (integer vector)

resp Respiratory mortality count (integer vector)

temp Temperature (numeric vector)

dptp Dew point temperature (numeric vector)

rhum Relative humidity (numeric vector)

pm10 PM10 pollution level (numeric vector)

o3 Ozone level (numeric vector)

Details

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

Source

Data taken from the dlnm package version 2.4.10

---

children_respiratory_rates_df

Children Respiratory Rates Data

Description

This dataset, children_respiratory_rates_df, is a data frame containing respiratory rate measurements from 618 Italian children aged between 15 days and 3 years, collected to establish normal respiratory rate distributions for clinical assessment.

Usage

```
data(children_respiratory_rates_df)
```

Format

A data frame with 618 observations and 2 variables:

Age Child's age in days (numeric vector)

Rate Respiratory rate in breaths per minute (integer vector)

Details

The dataset name has been kept as 'children_respiratory_rates_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets 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 Sleuth3 package version 1.0-6

`child_wheeze_pollution_df`

Child Wheeze and Pollution

Description

This dataset, child_wheeze_pollution_df, is a data frame containing longitudinal data on wheezing status for 16 children measured four times yearly at ages 9 through 12, with associated pollution exposure information.

Usage

```
data(child_wheeze_pollution_df)
```

Format

A data frame with 64 observations and 5 variables:

ID Child identifier (integer vector)

Wheeze Wheezing status (integer vector)

City City identifier (integer vector)

Age Child's age in years (integer vector)

Smoke Smoking exposure indicator (integer vector)

Details

The dataset name has been kept as 'child_wheeze_pollution_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets 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 geessbin package version 1.0.0

`danish_lung_incidence_df`

Lung cancer in 4 Danish cities 1968-71

Description

This dataset, danish_lung_incidence_df, is a data frame containing counts of incident lung cancer cases and population size in four neighbouring Danish cities by age group from 1968 to 1971.

Usage

```
data(danish_lung_incidence_df)
```

Format

A data frame with 24 observations and 4 variables:

city City of observation (factor with 4 levels)

age Age group (factor with 6 levels)

pop Population size (integer)

cases Number of incident lung cancer cases (integer)

Details

The dataset name has been kept as 'danish_lung_incidence_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a data frame object. The original content has not been modified in any way.

Source

Data taken from the ISwR package version 2.0-10

`engwales_cancer_mortality_df` *UK lung and nasal cancer deaths 1936–80*

Description

This dataset, engwales_cancer_mortality_df, is a data frame containing England and Wales mortality rates from lung cancer, nasal cancer, and all causes between 1936 and 1980. The 1936 rates are repeated as 1931 rates in order to accommodate follow-up for the nickel study.

Usage

```
data(engwales_cancer_mortality_df)
```

Format

A data frame with 150 observations and 5 variables:

year Year of observation (numeric)

age Age group (numeric)

lung Lung cancer mortality rate (numeric)

nasal Nasal cancer mortality rate (numeric)

other Mortality rate from all other causes (numeric)

Details

The dataset name has been kept as 'engwales_cancer_mortality_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a data frame object. The original content has not been modified in any way.

Source

Data taken from the ISwR package version 2.0-10

`influenza_us_1975_df` *US 1975-76 Influenza-Like Illness Data*

Description

This dataset, influenza_us_1975_df, is a data frame containing influenza-like illness (ILI) data for the lower 48 US states and District of Columbia during the 1975-76 season, which was dominated by the A H3N2 Victoria strain.

Usage

```
data(influenza_us_1975_df)
```

Format

A data frame with 49 observations (states + DC) and 7 variables:

State State identifier (integer)

Acronym State abbreviation (factor with 51 levels)

Pop State population (integer)

Latitude Geographic latitude (numeric)

Longitude Geographic longitude (numeric)

Start Week of season start (integer)

Peak Week of peak activity (integer)

Details

The dataset name has been kept as 'influenza_us_1975_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a standard data frame. The original content has not been modified in any way.

Source

Data taken from the epimdr package version 0.6-5

lungca_cancer_deaths_df

Male Lung Cancer by Smoking Duration

Description

This dataset, lungca_cancer_deaths_df, is a data frame containing data on man-years of smoking risk and observed lung cancer deaths among male smokers. It includes 63 observations across 4 variables measuring smoking exposure and mortality outcomes.

Usage

```
data(lungca_cancer_deaths_df)
```

Format

A data frame with 63 observations and 4 variables:

yrs_smk Years of smoking (factor with 9 levels)

pys Person-years of smoking exposure (numeric)

num_cigs Number of cigarettes smoked daily (factor with 7 levels)

deaths Number of lung cancer deaths (numeric)

Details

The dataset name has been kept as 'lungca_cancer_deaths_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a standard data frame. The original content has not been modified in any way.

Source

Data taken from the R4HCR package version 0.1

lung_cancer_survival_df

Lung Cancer Survival Data

Description

This dataset, lung_cancer_survival_df, is a data frame containing survival information for 228 lung cancer patients, with 10 clinical variables including survival time, patient status, age, gender, performance scores, and nutritional indicators.

Usage

```
data(lung_cancer_survival_df)
```

Format

A data frame with 228 observations (patients) and 10 variables:

inst Institution code where patient was treated (numeric)

time Survival time in days from diagnosis (numeric)

status Censoring status (1 = censored, 2 = died) (numeric)

age Patient age at diagnosis in years (numeric)

sex Gender (1 = male, 2 = female) (numeric)

ph.ecog ECOG performance score (0=asymptomatic to 4=fully disabled) (numeric)

ph.karno Karnofsky performance score (0-100) as rated by physician (numeric)

pat.karno Karnofsky performance score (0-100) as self-reported by patient (numeric)

meal.cal Daily calories consumed at meals (numeric)

wt.loss Weight loss in last six months (pounds) (numeric)

Details

The dataset name has been kept as 'lung_cancer_survival_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets 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 acro package version 0.1.4

---

lung_nodules_detection_dt

Incidental or Screen-Detected Lung Nodules

Description

This dataset, lung_nodules_detection_dt, is a data table containing clinical and radiological characteristics of 999 pulmonary nodules (up to 15mm in size) detected on routine chest CT scans from 3 UK academic centers.

Usage

```
data(lung_nodules_detection_dt)
```

Format

A data table with 999 observations and 8 variables:

sex Patient sex (factor with 2 levels)

age Patient age in years (numeric)

num.annotated Number of annotated nodules (numeric)

location Nodule location (factor with 6 levels)

spiculate Spiculation status (factor with 2 levels)

smoke.status Smoking history (factor with 5 levels)

diameter Nodule diameter in mm (numeric)

malignant Malignancy status (0=benign, 1=malignant) (numeric)

Details

The dataset name has been kept as 'lung_nodules_detection_dt' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'dt' indicates that this is a data table object. The original content has not been modified in any way.

Source

Data taken from the R4HCR package version 0.1

`neonatal_intubation_times_df`

Neonatal Intubation Simulation

Description

This dataset, neonatal_intubation_times_df, is a data frame containing execution times (in seconds) for specific actions performed by 37 midwife students during a high-fidelity neonatal resuscitation simulation. The simulation was video recorded, and each critical action in the intubation process was tagged for timing analysis.

Usage

```
data(neonatal_intubation_times_df)
```

Format

A data frame with 37 observations and 7 variables:

id Participant ID (integer)

deci_intub Time to decision to intubate (seconds) (integer)

stop_ventil Time to stop ventilation (seconds) (integer)

blade_in Time to insert laryngoscope blade (seconds) (integer)

insert_tube Time to insert endotracheal tube (seconds) (integer)

blade_out Time to remove laryngoscope blade (seconds) (integer)

restart_ventil Time to restart ventilation (seconds) (integer)

Details

The dataset name has been kept as 'neonatal_intubation_times_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a data frame object. The original content has not been modified in any way.

Source

Data taken from the ViSiElse package version 1.2.2

---

nicotine_gum_df *Nicotine Gum and Smoking Cessation*

Description

This dataset, nicotine_gum_df, is a data frame containing meta-analysis data on the effectiveness of nicotine gum for smoking cessation across 26 studies.

Usage

```
data(nicotine_gum_df)
```

Format

A data frame with 26 observations (studies) and 4 variables:

qt Number of successful quitters in treatment group (integer)

tt Total participants in treatment group (integer)

qc Number of successful quitters in control group (integer)

tc Total participants in control group (integer)

Details

The dataset name has been kept as 'nicotine_gum_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets 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 HSAUR3 package version 1.0-15

ohio_children_wheeze_df

Ohio Children Wheeze Status

Description

This dataset, ohio_children_wheeze_df, is a data frame containing wheeze status data from 2148 observations of children in Ohio. The data are part of a subset from the Six-City Study, a longitudinal study examining the health effects of air pollution on children.

Usage

```
data(ohio_children_wheeze_df)
```

Format

A data frame with 2148 observations and 4 variables:

resp Wheeze status (0 = no wheeze, 1 = wheeze) (integer)

id Child identifier (integer)

age Age of the child in years (integer)

smoke Parental smoking status (0 = no, 1 = yes) (integer)

Details

The dataset name has been kept as 'ohio_children_wheeze_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a data frame object. The original content has not been modified in any way.

Source

Data taken from the geepack package version 1.3.12

patients_lung_diseases_tbl_df

Lung Disease Patients

Description

This dataset, patients_lung_diseases_tbl_df, is a tibble containing detailed clinical information about 5,200 patients with various lung conditions, including demographics, smoking status, lung capacity measurements, disease types, treatments received, hospital visits, and recovery status.

Usage

```
data(patients_lung_diseases_tbl_df)
```

Format

A tibble with 5,200 observations and 8 variables:

Age Patient age in years (numeric)

Gender Patient gender (character)

Smoking Status Smoker or non-smoker status (character)

Lung Capacity Measured lung function (numeric)

Disease Type Specific lung condition (character)

Treatment Type Therapy, medication or surgery received (character)

Hospital Visits Number of hospital visits (numeric)

Recovered Recovery status (character)

Details

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

Source

Data taken from Kaggle: https://www.kaggle.com/datasets/samikshadalvi/lungs-diseases-dataset

---

pneumonia_influenza_ts

Monthly Pneumonia and Influenza Deaths in the U.S.

Description

This dataset, pneumonia_influenza_ts, is a time series containing monthly rates of pneumonia and influenza deaths in the United States from 1968 to 1978.

Usage

```
data(pneumonia_influenza_ts)
```

Format

A time series with 132 monthly observations from January 1968 to December 1978:

Value Mortality rate (numeric vector)

Time Monthly index from 1968 to 1978 (time series vector)

Details

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

Source

Data taken from the astsa package version 2.2

PulmoDataSets *PulmoDataSets: A Curated Collection of Pulmonary and Respiratory Disease Datasets*

Description

This package provides a wide variety of datasets focused on the lungs, respiratory system, tuberculosis, whooping cough, pneumonia, influenza and associated diseases.

Details

PulmoDataSets: A Curated Collection of Pulmonary and Respiratory Disease Datasets

A Curated Collection of Pulmonary and Respiratory Disease Datasets.

Author(s)

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

See Also

Useful links:

- https://github.com/lightbluetitan/pulmodatasets

---

respiratory_clinical_trial_df *Respiratory Clinical Trial*

Description

This dataset, respiratory_clinical_trial_df, is a data frame containing information from a clinical trial of patients with respiratory illness, where 111 patients from two different clinics were randomized to receive either placebo or an active treatment. Patients were examined at baseline and at four visits during treatment. The respiratory status was determined at each visit, with 1 representing good status and 0 representing poor status.

Usage

```
data(respiratory_clinical_trial_df)
```

Format

A data frame with 444 observations and 8 variables:

center Study identifier (integer vector)

id Patient identifier (integer vector)

treat Treatment group (factor with 2 levels)

sex Patient sex (factor with 2 levels)

age Patient age in years (integer vector)

baseline Baseline respiratory status (integer vector)

visit Visit number (integer vector)

outcome Respiratory status (integer vector)

Details

The dataset name has been kept as 'respiratory_clinical_trial_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets 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 geepack package version 1.3.12

---

respiratory_infections_df

Azithromycin for Respiratory Infections

Description

This dataset, respiratory_infections_df, is a data frame containing results from 15 clinical trials comparing the effectiveness of azithromycin versus amoxycillin or amoxycillin/clavulanic acid (amoxyclav) in the treatment of acute lower respiratory tract infections.

Usage

```
data(respiratory_infections_df)
```

Format

A data frame with 15 observations and 11 variables:

author Study author(s) (character vector)

year Year of publication (integer vector)

ai Number of successful treatments in azithromycin group (integer vector)

n1i Total number of participants in azithromycin group (integer vector)

ci Number of successful treatments in control group (integer vector)

n2i Total number of participants in control group (integer vector)

age Patient age characteristics (character vector)

diag.ab Number diagnosed with acute bronchitis (integer vector)

diag.cb Number diagnosed with chronic bronchitis (integer vector)

diag.pn Number diagnosed with pneumonia (integer vector)

ctrl Type of control treatment (character vector)

Details

The dataset name has been kept as 'respiratory_infections_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets 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 metadat package version 1.4-0

`respiratory_trial_df` *Respiratory Illness Clinical Trial*

Description

This dataset, respiratory_trial_df, is a data frame containing the respiratory status of patients recruited for a randomized clinical multicenter trial, with 555 observations across 111 subjects.

Usage

```
data(respiratory_trial_df)
```

Format

A data frame with 555 observations and 7 variables:

centre Study center (factor with 2 levels)

treatment Treatment group (factor with 2 levels)

gender Patient gender (factor with 2 levels)

age Patient age in years (numeric)

status Respiratory status (factor with 2 levels)

month Follow-up month (ordered factor with 5 levels)

subject Patient identifier (factor with 111 levels)

Details

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

Source

Data taken from the HSAUR3 package version 1.0-15

---

respiratory_trial_outcomes_df

Ordinal respiratory outcomes

Description

This dataset, respiratory_trial_outcomes_df, is a data frame containing outcome data from a randomized clinical trial described in Miller et al. (1993) evaluating a new treatment for respiratory disorder. The study includes 111 patients who were randomly assigned to one of two treatments (active or placebo). The patients were followed up at four visits, and their response status was classified on an ordinal scale at each visit.

Usage

```
data(respiratory_trial_outcomes_df)
```

Format

A data frame with 111 observations and 5 variables:

y1 Ordinal response at visit 1 (integer)

y2 Ordinal response at visit 2 (integer)

y3 Ordinal response at visit 3 (integer)

y4 Ordinal response at visit 4 (integer)

trt Treatment group (0 = placebo, 1 = active) (integer)

Details

The dataset name has been kept as 'respiratory_trial_outcomes_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a data frame object. The original content has not been modified in any way.

Source

Data taken from the geepack package version 1.3.12

---

smoking_doctors_df *Smoking Deaths Among Doctors (British)*

Description

This dataset, smoking_doctors_df, is a data frame containing data from a study on smoking habits and coronary artery disease mortality among British doctors. It includes 10 observations across 5 variables representing person-years of observation and deaths during the study period.

Usage

```
data(smoking_doctors_df)
```

Format

A data frame with 10 observations and 5 variables:

age Age group (factor with 5 levels)

smoke Smoking status (numeric)

n Number of person-years at risk (numeric)

y Number of deaths from coronary artery disease (numeric)

ns Standardized mortality ratio (numeric)

Details

The dataset name has been kept as 'smoking_doctors_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a standard data frame. The original content has not been modified in any way.

Source

Data taken from the boot package version 1.3-31

smoking_lung_cancer_df *Smoking and Lung Cancer*

Description

This dataset, smoking_lung_cancer_df, is a data frame containing data from a retrospective case-control study comparing smoking status between 86 lung cancer patients and 86 controls.

Usage

```
data(smoking_lung_cancer_df)
```

Format

A data frame with 2 observations and 3 variables:

Smoking Smoking status (factor with 2 levels: "NonSmokers", "Smokers")

Cancer Number of lung cancer cases (integer vector)

Control Number of control cases (integer vector)

Details

The dataset name has been kept as 'smoking_lung_cancer_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets 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 Sleuth3 package version 1.0-6

smoking_UK_tbl_df *UK Smoking Habits*

Description

This dataset, smoking_UK_tbl_df, is a tibble containing survey data on smoking habits from the UK, with demographic characteristics and tobacco consumption patterns from 1,691 respondents.

Usage

```
data(smoking_UK_tbl_df)
```

Format

A tibble with 1,691 observations and 12 variables:

gender Gender of respondent (factor with 2 levels)

age Age in years (integer)

marital_status Marital status (factor with 5 levels)

highest_qualification Highest education qualification (factor with 8 levels)

nationality Nationality (factor with 8 levels)

ethnicity Ethnic group (factor with 7 levels)

gross_income Income bracket (factor with 10 levels)

region UK region (factor with 7 levels)

smoke Smoking status (factor with 2 levels)

amt_weekends Cigarettes smoked on weekends (integer)

amt_weekdays Cigarettes smoked on weekdays (integer)

type Type of tobacco used (factor with 5 levels)

Details

The dataset name has been kept as 'smoking_UK_tbl_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the Pulmo-DataSets package. The suffix 'tbl_df' indicates that this is a tibble data frame. The original content has not been modified in any way.

Source

Data taken from the openintro package version 2.5.0

`smoking_youth_tbl_df` *Youth Smoking and Lung Function*

Description

This dataset, smoking_youth_tbl_df, is a tibble containing data from the Childhood Respiratory Disease Study collected in the late 1970s, examining the effects of smoking and second-hand smoke exposure on pulmonary function in 654 youths.

Usage

```
data(smoking_youth_tbl_df)
```

Format

A tibble with 654 observations and 5 variables:

age Age in years (integer)

FEV Forced Expiratory Volume in liters (numeric)

height Height in centimeters (numeric)

sex Sex of participant (character)

smoker Smoking status (character)

Details

The dataset name has been kept as 'smoking_youth_tbl_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'tbl_df' indicates that this is a tibble data frame. The original content has not been modified in any way.

Source

Data taken from the LSTbook package version 0.6

`tlc_lung_capacity_df` *Total Lung Capacity*

Description

This dataset, tlc_lung_capacity_df, is a data frame containing data on pretransplant total lung capacity (TLC) measured by whole-body plethysmography for recipients of heart-lung transplants.

Usage

```
data(tlc_lung_capacity_df)
```

Format

A data frame with 32 observations and 4 variables:

age Age in years (integer)

sex Sex (0 = female, 1 = male) (integer)

height Height in centimeters (integer)

tlc Total lung capacity in liters (numeric)

Details

The dataset name has been kept as 'tlc_lung_capacity_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a data frame object. The original content has not been modified in any way.

Source

Data taken from the ISwR package version 2.0-10

`tuberculosis_vaccine_df`

BCG Vaccine Against Tuberculosis

Description

This dataset, tuberculosis_vaccine_df, is a data frame containing results from 13 clinical trials examining the effectiveness of the Bacillus Calmette-Guerin (BCG) vaccine against tuberculosis.

Usage

```
data(tuberculosis_vaccine_df)
```

Format

A data frame with 13 observations and 9 variables:

trial Trial identifier number (integer vector)

author Study author(s) (character vector)

year Year of publication (integer vector)

tpos Number of TB positive cases in vaccinated group (integer vector)

tneg Number of TB negative cases in vaccinated group (integer vector)

cpos Number of TB positive cases in control group (integer vector)

cneg Number of TB negative cases in control group (integer vector)

ablat Absolute latitude of study location (integer vector)

alloc Method of treatment allocation (character vector)

Details

The dataset name has been kept as 'tuberculosis_vaccine_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets 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 metadat package version 1.4-0

UK_female_lung_deaths_ts

UK Female Lung Disease Deaths

Description

This dataset, UK_female_lung_deaths_ts, is a time series object containing monthly deaths from bronchitis, emphysema and asthma in the UK from 1974 to 1979, for females.

Usage

```
data(UK_female_lung_deaths_ts)
```

Format

A time series (ts) object with 72 monthly observations from 1974 to 1979.

value Number of deaths (numeric vector)

time Time index (1974 to 1979)

Details

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

Source

Data taken from the datasets package (R version 4.5.0), fdeaths dataset

---

UK_male_lung_deaths_ts

UK Male Lung Disease Deaths

Description

This dataset, UK_male_lung_deaths_ts, is a time series object containing monthly deaths from bronchitis, emphysema and asthma in the UK from 1974 to 1979, for males.

Usage

```
data(UK_male_lung_deaths_ts)
```

Format

A time series (ts) object with 72 monthly observations from 1974 to 1979.

value Number of deaths (numeric vector)

time Time index (1974 to 1979)

Details

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

Source

Data taken from the datasets package (R version 4.5.0), mdeaths dataset

---

USMortality_df *US Mortality Rates by Cause and Gender*

Description

This dataset, USMortality_df, is a data frame containing mortality rates across all ages in the USA by cause of death, sex, rural and urban status from 2011 to 2013. The data represent national aggregate rates under the Department of Health and Human Services (HHS).

Usage

```
data(USMortality_df)
```

Format

A data frame with 40 observations and 5 variables:

Status Rural/Urban status (factor with 2 levels)

Sex Gender (factor with 2 levels)

Cause Cause of death (factor with 10 levels)

Rate Mortality rate (numeric vector)

SE Standard error of mortality rate (numeric vector)

Details

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

Source

Data taken from the lattice package version 0.22-6

`USRegionalMortality_df`
US Regional Mortality Rates by Cause and Gender

Description

This dataset, USRegionalMortality_df, is a data frame containing region-wise mortality rates across all ages in the USA by cause of death, sex, rural and urban status from 2011 to 2013. The data represent rates for each administrative region under the Department of Health and Human Services (HHS).

Usage

```
data(USRegionalMortality_df)
```

Format

A data frame with 400 observations and 6 variables:

Region HHS administrative region (factor with 10 levels)

Status Rural/Urban status (factor with 2 levels)

Sex Gender (factor with 2 levels)

Cause Cause of death (factor with 10 levels)

Rate Mortality rate (numeric vector)

SE Standard error of mortality rate (numeric vector)

Details

The dataset name has been kept as 'USRegionalMortality_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the Pulmo-DataSets 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 lattice package version 0.22-6

`veterans_lung_cancer_df` *Veterans Administration Lung Cancer Study*

Description

This dataset, veterans_lung_cancer_df, is a data frame containing information from a randomized trial of two treatment regimens for lung cancer. This is a standard survival analysis data set.

Usage

```
data(veterans_lung_cancer_df)
```

Format

A data frame with 137 observations and 8 variables:

trt Treatment group (numeric)

celltype Cell type (factor with 4 levels)

time Survival time in days (numeric)

status Censoring status (numeric)

karno Karnofsky performance score (numeric)

diagtime Time from diagnosis to randomization (numeric)

age Age in years (numeric)

prior Number of prior therapies (numeric)

Details

The dataset name has been kept as 'veterans_lung_cancer_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a data frame object. The original content has not been modified in any way.

Source

Data taken from the survival package version 3.8-3

view_datasets_pulmo *View Available Datasets in PulmoDataSets*

Description

This function lists all datasets available in the 'PulmoDataSets' package. If the 'PulmoDataSets' package is not loaded, it stops and shows an error message. If no datasets are available, it returns a message and an empty vector.

Usage

```
view_datasets_pulmo()
```

Value

A character vector with the names of the available datasets. If no datasets are found, it returns an empty character vector.

Examples

```
if (requireNamespace("PulmoDataSets", quietly = TRUE)) {
  library(PulmoDataSets)
  view_datasets_pulmo()
}
```

whooping_cough_dk_df *Copenhagen Whooping Cough 1900-1937*

Description

This dataset, whooping_cough_dk_df, is a data frame containing weekly incidence data of whooping cough in Copenhagen, Denmark between January 1900 and December 1937. It includes 1,982 weekly observations across 8 demographic and epidemiological variables.

Usage

```
data(whooping_cough_dk_df)
```

Format

A data frame with 1,982 weekly observations and 8 variables:

date Date of observation (factor)

births Number of births (integer)

day Day of month (integer)

month Month (integer 1-12)

year Year (integer 1900-1937)

cases Number of whooping cough cases (integer)

deaths Number of whooping cough deaths (integer)

popsize Population size (numeric)

Details

The dataset name has been kept as 'whooping_cough_dk_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a standard data frame. The original content has not been modified in any way.

Source

Data taken from the epimdr package version 0.6-5

`whooping_cough_phila_df`

Philadelphia Whooping Cough 1925-1947

Description

This dataset, whooping_cough_phila_df, is a data frame containing weekly incidence data of whooping cough in Philadelphia between 1925 and 1947, with 1,200 weekly observations across 5 variables.

Usage

```
data(whooping_cough_phila_df)
```

Format

A data frame with 1,200 weekly observations and 5 variables:

YEAR Year of observation (integer)

WEEK Week number (integer)

PHILADELPHIA Weekly incidence count of whooping cough cases (integer)

TIME Time index (numeric)

TM Time marker (integer)

Details

The dataset name has been kept as 'whooping_cough_phila_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'df' indicates that this is a standard data frame. The original content has not been modified in any way.

Source

Data taken from the epimdr package version 0.6-5

`whooping_cough_ts` *Whooping Cough Deaths in London (1740-1881)*

Description

This dataset, whooping_cough_ts, is a time series object containing annual counts of deaths from whooping cough in London from 1740 to 1881, with three measurement variables recorded each year.

Usage

```
data(whooping_cough_ts)
```

Format

A multivariate time series with 142 annual observations from 1740 to 1881 and 3 variables:

wcough Number of whooping cough deaths (integer)

ratio Death ratio (numeric)

alldeaths Total deaths from all causes (integer)

Details

The dataset name has been kept as 'whooping_cough_ts' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the PulmoDataSets package. The suffix 'ts' indicates that this is a time series object. The original content has not been modified in any way.

Source

Data taken from the DAAG package version 1.25.6

Index