Package 'DigestiveDataSets'

July 21, 2025

Type Package

Title A Curated Collection of Digestive System and Gastrointestinal Disease Datasets

Version 0.1.0

Maintainer Renzo Caceres Rossi <arenzocaceresrossi@gmail.com>

Description Provides an extensive and curated collection of datasets related to the digestive system, stomach, intestines, liver, pancreas, and associated diseases.
This package includes clinical trials, observational studies, experimental datasets, cohort data, and case series involving gastrointestinal disorders such as gastritis, ulcers, pancreatitis, liver cirrhosis, colon cancer, colorectal conditions, Helicobacter pylori infection, irritable bowel syndrome, intestinal infections, and post-surgical outcomes.
The datasets support educational, clinical, and research applications in gastroenterology, public health, epidemiology, and biomedical sciences.
Designed for researchers, clinicians, data scientists, students, and educators interested in digestive diseases, the package facilitates reproducible analysis, modeling, and hypothesis testing using real-world and historical data.

License GPL-3

Language en

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

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

Encoding UTF-8

LazyData true

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

Depends R (>= 4.1.0)

Imports utils

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 13:00:13 UTC

Contents

`anorexia_weight_change_df`

Anorexia Weight Change

Description

This dataset, anorexia_weight_change_df, is a data frame containing weight change data for young female anorexia patients. It includes pre- and post-treatment weights, along with the type of treatment administered.

Usage

```
data(anorexia_weight_change_df)
```

Format

A data frame with 72 observations and 3 variables:

Treat Factor indicating the treatment type (3 levels)

Prewt Numeric vector indicating the patient's weight before treatment (in kilograms)

Postwt Numeric vector indicating the patient's weight after treatment (in kilograms)

Details

The dataset name has been kept as 'anorexia_weight_change_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 MASS package version 7.3-65.

---

bleeding_ulcers_df *Recurrent Bleeding from Ulcers*

Description

This dataset, bleeding_ulcers_df, is a data frame containing data from 40 experiments designed to compare a new surgery for stomach ulcer with an older surgery.

Usage

```
data(bleeding_ulcers_df)
```

Format

A data frame with 80 observations and 9 variables:

author Factor indicating the author of the study (20 levels)

year Integer indicating the year of the study

quality Integer representing the quality score of the experiment

age Integer indicating the age of the patients

r Integer indicating the number of recurrent bleeds

m Integer indicating the total number of patients

bleed Integer indicating bleeding events

treat Factor indicating treatment type (6 levels)

table Factor representing the experiment table (40 levels)

Details

The dataset name has been kept as 'bleeding_ulcers_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 SMPracticals package version 1.4-3.1.

---

campylobacter_infections_ts

Campylobacter Infections Time Series

Description

This dataset, campylobacter_infections_ts, is a time series object containing the number of cases of campylobacter infections in northern Quebec (Canada), recorded in four-week intervals from January 1990 to October 2000. Campylobacterosis is an acute bacterial infectious disease attacking the digestive system.

Usage

```
data(campylobacter_infections_ts)
```

Format

A time series object (‘ts‘) with 140 observations:

Start c(1990, 1)

End c(2000, 10)

Frequency 13 (observations per year)

Details

The dataset name has been kept as 'campylobacter_infections_ts' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 **tscount** package version 1.4.3. Original source: Ferland, R., Latour, A. and Oraichi, D., "Integer-valued GARCH process". *Journal of Time Series Analysis*, 2006; 27(6): 923–942.

`cholera_deaths_1849_tbl_df`

Cholera Daily Deaths in England, 1849

Description

This dataset, cholera_deaths_1849_tbl_df, is a tibble containing daily deaths from Cholera and Diarrhaea in England for each day of the 12 months of 1849. It includes the month, cause of death, day of month, number of deaths, date, and day of week for each observation.

Usage

```
data(cholera_deaths_1849_tbl_df)
```

Format

A tibble with 730 observations and 6 variables:

month Character indicating the month of observation

cause_of_death Factor with 2 levels indicating cause of death (Cholera or Diarrhaea)

day_of_month Character indicating the day of the month

deaths Numeric value indicating the number of deaths

date Date object indicating the exact date

day_of_week Ordered factor with 7 levels indicating the day of week

Details

The dataset name has been kept as 'cholera_deaths_1849_tbl_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets package and assists users in identifying its specific characteristics. The suffix 'tbl_df' indicates that the dataset is a tibble. The original content has not been modified in any way.

Source

Data taken from the **HistData** package version 0.9-3. Original source: Bingham P., Verlander, N. Q., Cheal M. J. (2004). "John Snow, William Farr and the 1849 outbreak of cholera that affected London: a reworking of the data highlights the importance of the water supply". *Public Health*, 118(6), 387–394, Table 2.

`colonoscopy_features_tbl_df` *Features from Colonoscopic Video*

Description

This dataset, colonoscopy_features_tbl_df, is a tibble containing features extracted from 76 colonoscopic videos. Each video was recorded using both White Light (WL) and Narrow Band Imaging (NBI). The dataset includes histology results (classification ground truth), the opinion of endoscopists (4 experts and 3 beginners), and 698 features derived from patients with gastrointestinal lesions.

Usage

```
data(colonoscopy_features_tbl_df)
```

Format

A tibble with 76 observations and 7 variables:

feature 294 Numeric feature extracted from colonoscopic videos

feature 441 Numeric feature extracted from colonoscopic videos

feature 472 Numeric feature extracted from colonoscopic videos

feature 486 Numeric feature extracted from colonoscopic videos

class_agreement Numeric score representing agreement among endoscopists

missinglabel_indicator Numeric indicator for missing labels

ground truth Character string representing the histology-based classification

Details

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

Source

Data taken from the gmmsslm package version 1.1.6.

`colon_stageBC_chemo_df`

Chemotherapy for Stage B/C Colon Cancer

Description

This dataset, colon_stageBC_chemo_df, is a data frame containing data from one of the first successful trials of adjuvant chemotherapy for stage B/C colon cancer. The dataset includes 1858 observations (with two records per patient: one for recurrence and one for death) and 16 clinical variables.

Usage

```
data(colon_stageBC_chemo_df)
```

Format

A data frame with 1858 observations and 16 variables:

id Numeric patient identifier

study Numeric study code

rx Factor with 3 levels indicating treatment group

sex Numeric gender code

age Numeric age in years

obstruct Numeric obstruction status

perfor Numeric perforation status

adhere Numeric adhesion status

nodes Numeric count of lymph nodes

status Numeric event status

differ Numeric differentiation grade

extent Numeric tumor extent

surg Numeric surgery code

node4 Numeric node4 status

time Numeric follow-up time

etype Numeric event type

Details

The dataset name has been kept as 'colon_stageBC_chemo_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 **OncoDataSets** package version 0.1.0.

`crc_mirnas_pubmed_tbl_df`

PubMed Data of miRNAs in Colorectal Cancer

Description

This dataset, crc_mirnas_pubmed_tbl_df, is a tibble containing information from PubMed abstracts related to microRNAs (miRNAs) in colorectal cancer. The data provides publication metadata, article abstracts, and associated miRNAs across 508 observations with 8 variables.

Usage

```
data(crc_mirnas_pubmed_tbl_df)
```

Format

A tibble with 508 observations and 8 variables:

PMID Numeric PubMed identifier

Year Numeric publication year

Title Character article title

Abstract Character full abstract text

Language Character publication language

Type Character article type

Topic Character research topic

miRNA Character microRNA identifiers

Details

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

Source

Data taken from the **OncoDataSets** package version 0.1.0.

`cystic_fibrosis_snps_df` *Cystic Fibrosis SNP*

Description

This dataset, cystic_fibrosis_snps_df, is a data frame containing genetic association data for cystic fibrosis, including a case-control indicator and 23 single nucleotide polymorphisms (SNPs) with specified inter-marker distances. The dataset contains 186 observations across 24 variables.

Usage

```
data(cystic_fibrosis_snps_df)
```

Format

A data frame with 186 observations and 24 variables:

y Integer case-control indicator

loc1 Integer SNP genotype at location 1

loc2 Integer SNP genotype at location 2

loc3 Integer SNP genotype at location 3

loc4 Integer SNP genotype at location 4

loc5 Integer SNP genotype at location 5

loc6 Integer SNP genotype at location 6

loc7 Integer SNP genotype at location 7

loc8 Integer SNP genotype at location 8

loc9 Integer SNP genotype at location 9

loc10 Integer SNP genotype at location 10

loc11 Integer SNP genotype at location 11

loc12 Integer SNP genotype at location 12

loc13 Integer SNP genotype at location 13

loc14 Integer SNP genotype at location 14

loc15 Integer SNP genotype at location 15

loc16 Integer SNP genotype at location 16

loc17 Integer SNP genotype at location 17

loc18 Integer SNP genotype at location 18

loc19 Integer SNP genotype at location 19

loc20 Integer SNP genotype at location 20

loc21 Integer SNP genotype at location 21

loc22 Integer SNP genotype at location 22

loc23 Integer SNP genotype at location 23

Details

The dataset name has been kept as 'cystic_fibrosis_snps_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 **gap.datasets** package version 0.0.6. Original source: Liu JS, Sabatti C, Teng J, Keats BJB, Risch N (2001). "Bayesian Analysis of Haplotypes for Linkage Disequilibrium Mapping". *Genome Research*, 11:1716–1724.

`DigestiveDataSets` *DigestiveDataSets: A Curated Collection of Digestive System and Gastrointestinal Disease Datasets*

Description

This package provides a wide variety of datasets focused on the digestive system, stomach, intestines, liver, pancreas, and associated diseases, including clinical trials, observational studies, experimental datasets, cohort data, and case series involving gastrointestinal disorders such as gastritis, ulcers, pancreatitis, liver cirrhosis, colon cancer, colorectal conditions, Helicobacter pylori infection, irritable bowel syndrome, intestinal infections, and post-surgical outcomes.

Details

DigestiveDataSets: A Curated Collection of Digestive System and Gastrointestinal Disease Datasets

A Curated Collection of Digestive System and Gastrointestinal Disease Datasets.

Author(s)

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

See Also

Useful links:

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

digestive_cancer_survival_df *Digestive Cancer Survival Times*

Description

This dataset, digestive_cancer_survival_df, is a data frame containing survival times (in days) of cancer patients with advanced cancer of the stomach, bronchus, colon, ovary, or breast. All patients included in this dataset received treatment that involved supplemental ascorbate.

Usage

```
data(digestive_cancer_survival_df)
```

Format

A data frame with 17 observations and 5 variables:

stomach Integer values indicating survival times (in days) for patients with stomach cancer

bronchus Integer values indicating survival times (in days) for patients with bronchial cancer

colon Integer values indicating survival times (in days) for patients with colon cancer

ovary Integer values indicating survival times (in days) for patients with ovarian cancer

breast Integer values indicating survival times (in days) for patients with breast cancer

Details

The dataset name has been kept as 'digestive_cancer_survival_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 RbyExample package version 0.0.100.

ecoli_infections_df *E. coli Infections Time Series*

Description

This dataset, ecoli_infections_df, is a data frame containing the weekly number of reported disease cases caused by Escherichia coli in the state of North Rhine-Westphalia (Germany) from January 2001 to May 2013, excluding cases of EHEC and HUS.

Usage

```
data(ecoli_infections_df)
```

Format

A data frame with 646 observations and 3 variables:

year Numeric value indicating the year of observation

week Numeric value indicating the week of observation

cases Numeric value indicating the number of reported E. coli cases

Details

The dataset name has been kept as 'ecoli_infections_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 **tscount** package version 1.4.3.

gastric_cancer_trial_df *Gastric Cancer Clinical Trial*

Description

This dataset, gastric_cancer_trial_df, is a data frame containing data from a randomized clinical trial conducted by the Gastrointestinal Tumor Study Group on patients with gastric cancer. It includes survival time, event occurrence, and group assignment.

Usage

```
data(gastric_cancer_trial_df)
```

Format

A data frame with 90 observations and 3 variables:

time Numeric vector representing survival time

event Numeric vector indicating event occurrence (e.g., death or relapse)

group Factor with 2 levels representing treatment groups

Details

The dataset name has been kept as 'gastric_cancer_trial_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 package coin version 1.4-3.

---

gi_damage_prevention_df

Gastrointestinal Damage Prevention

Description

This dataset, gi_damage_prevention_df, is a data frame containing results from four randomised clinical trials on the prevention of gastrointestinal damages by Misoprostol, reported by Lanza et al. (1987–1989).

Usage

```
data(gi_damage_prevention_df)
```

Format

A data frame with 198 observations and 3 variables:

study Factor indicating the clinical trial (4 levels)

treatment Factor indicating the treatment group (2 levels: control or Misoprostol)

classification Ordered factor indicating the degree of gastrointestinal damage (5 levels)

Details

The dataset name has been kept as 'gi_damage_prevention_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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.

`helicobacter_children_tbl_df`

Helicobacter pylori Infection in Preschoolers

Description

This dataset, helicobacter_children_tbl_df, is a tibble containing the prevalence of Helicobacter pylori infection in preschool children according to parental history of duodenal or gastric ulcer.

Usage

```
data(helicobacter_children_tbl_df)
```

Format

A tibble with 863 observations and 2 variables:

ulcer Factor with 2 levels indicating parental history of duodenal or gastric ulcer

infected Factor with 2 levels indicating Helicobacter pylori infection status

Details

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

Source

Data taken from the package pubh version 2.0.0.

`horse_colic_surgery_df`

Colic Horse Surgery

Description

This dataset, horse_colic_surgery_df, is a data frame containing clinical observations of horses with colic, where the primary task is to determine if the lesion requires surgery. The data consists of 300 cases with 31 clinical variables, modified from the original UCI repository version with adjusted factor levels.

Usage

```
data(horse_colic_surgery_df)
```

Format

A data frame with 300 observations and 31 variables:

surgery Factor with 2 levels indicating surgical requirement

age Factor with 1 level (age group)

hospitalID Integer hospital identifier

temp_rectal Numeric rectal temperature

pulse Numeric pulse rate

respiratory_rate Numeric respiratory rate

temp_extreme Factor with 4 levels (temperature extremes)

pulse_peripheral Factor with 4 levels (peripheral pulse)

capillayr_refill_time Factor with 3 levels (capillary refill time)

pain Numeric pain score

peristalsis Numeric peristalsis measure

abdominal_distension Numeric distension score

nasogastric_tube Numeric tube measure

nasogastric_reflux Numeric reflux quantity

nasogastric_reflux_PH Numeric reflux pH

rectal_examination Numeric exam result

abdomen Numeric abdomen assessment

cell_volume Numeric cell volume

protein Numeric protein level

abdominocentesis_appearance Numeric appearance score

abdomcentesis_protein Numeric protein measure

outcome Factor with 3 levels (outcome status)

surgical_lesion Factor with 2 levels (lesion type)

lesion_type1 Factor with 60 levels (primary lesion type)

lesion_type2 Integer secondary lesion code

lesion_type3 Integer tertiary lesion code

cp_data Factor with 2 levels (CP data)

temp_extreme_ordered Ordered factor with 4 levels (temperature)

temp_extreme_num Numeric temperature measure

mucous_membranes_col Factor with 6 levels (membrane color)

mucous_membranes_group Factor with 5 levels (membrane group)

Details

The dataset name has been kept as 'horse_colic_surgery_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 beyond factor level adjustments.

Source

Data taken from the **VIM** package version 6.2.2 (originally from UCI repository).

---

ibs_cam_trials_df *Studies on CAM for Irritable Bowel Syndrome*

Description

This dataset, ibs_cam_trials_df, is a data frame containing results from 19 clinical trials examining complementary and alternative medicine (CAM) interventions for irritable bowel syndrome (IBS). The dataset includes 12 variables characterizing each trial and its outcomes.

Usage

```
data(ibs_cam_trials_df)
```

Format

A data frame with 19 observations and 12 variables:

id Integer trial identifier

study Character study name/location

year Integer publication year

country Character country where study was conducted

ibs.crit Character IBS diagnostic criteria used

days Integer study duration in days

visits Integer number of study visits

jadad Integer Jadad score for study quality

x.a Integer active treatment events

n.a Integer active treatment sample size

x.p Integer placebo group events

n.p Integer placebo group sample size

Details

The dataset name has been kept as 'ibs_cam_trials_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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.

`intestinal_smartpill_df` *SmartPill Intestinal Transit*

Description

This dataset, intestinal_smartpill_df, is a data frame from a prospective cohort study evaluating gastric emptying, small bowel transit time, and total intestinal transit time using a SmartPill motility capsule. The study involved 8 critically ill trauma patients and 87 healthy volunteers. The capsule wirelessly transmitted pH, pressure, and temperature to a recorder attached to each subject's abdomen.

Usage

```
data(intestinal_smartpill_df)
```

Format

A data frame with 95 observations and 22 variables:

Group Numeric indicator of group membership

Gender Numeric indicator of gender

Race Numeric code indicating racial background

Height Height in centimeters

Weight Weight in kilograms

Age Age in years

GE.Time Gastric emptying time (minutes)

SB.Time Small bowel transit time (minutes)

C.Time Colon transit time (minutes)

WG.Time Whole gut transit time (minutes)

S.Contractions Number of contractions in the stomach

S.Sum.of.Amplitudes Sum of contraction amplitudes in the stomach

S.Mean.Peak.Amplitude Mean peak amplitude in the stomach

S.Mean.pH Mean pH level in the stomach

SB.Contractions Number of contractions in the small bowel

SB.Sum.of.Amplitudes Sum of contraction amplitudes in the small bowel

SB.Mean.Peak.Amplitude Mean peak amplitude in the small bowel

SB.Mean.pH Mean pH level in the small bowel

Colon.Contractions Number of contractions in the colon

Colon.Sum.of.Amplitudes Sum of contraction amplitudes in the colon

C.Mean.Peak.Amplitude Mean peak amplitude in the colon

C.Mean.pH Mean pH level in the colon

Details

The dataset name has been kept as 'intestinal_smartpill_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 **medicaldata** package version 0.2.0. Original source: Rauch et al., "Use of Wireless Utility Capsule to Determine Gastric Emptying and Small Intestinal Transit Times in Critically Ill Trauma Patients". *Journal of Critical Care*, 2012; 27(5): 534.e7–534.e12.

`intestinal_surgery_df` *Satellite Tumors in GI Surgery*

Description

This dataset, intestinal_surgery_df, is a data frame containing intestinal surgery data from 844 cancer patients. The data consists of pairs (n_i, s_i) where n_i is the number of satellites removed and s_i is the number of satellites found to be malignant.

Usage

```
data(intestinal_surgery_df)
```

Format

A data frame with 844 observations and 2 variables:

n Numeric value representing the number of satellites removed

s Numeric value representing the number of malignant satellites found

Details

The dataset name has been kept as 'intestinal_surgery_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 **deconvolveR** package version 1.2-1. Original source: Efron, B. (2016). "Empirical Bayes deconvolution estimates". *Biometrika*, 103(1), 1–20.

---

liver_cirrhosis_prednisone_df

Prednisone vs Placebo in Liver Cirrhosis

Description

This dataset, liver_cirrhosis_prednisone_df, is a data frame containing data from a randomized control trial comparing prednisone (n=251) versus placebo (n=237) in 488 liver cirrhosis patients. The dataset includes both survival and longitudinal measurements of prothrombin index development over time, with 2968 total observations across 9 variables.

Usage

```
data(liver_cirrhosis_prednisone_df)
```

Format

A data frame with 2968 observations and 9 variables:

ID Integer patient identifier

Time Numeric time measurement

death Integer death indicator

obstime Numeric observation time

proth Integer prothrombin index value

Trt Factor with 2 levels indicating treatment group (prednisone/placebo)

start Numeric start time

stop Numeric stop time

event Numeric event indicator

Details

The dataset name has been kept as 'liver_cirrhosis_prednisone_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 **JSM** package version 1.0.1.

`lynch_ontario_families_df`
Ontario Lynch Syndrome families

Description

This dataset, lynch_ontario_families_df, is a data frame containing data from 32 Lynch Syndrome families segregating mismatch repair mutations selected from the Ontario Familial Colorectal Cancer Registry. The dataset includes 765 individuals (both probands and relatives) with 11 variables per observation.

Usage

```
data(lynch_ontario_families_df)
```

Format

A data frame with 765 observations and 11 variables:

famID Integer family identifier

indID Integer individual identifier

fatherID Integer father's identifier

motherID Integer mother's identifier

gender Integer gender code

status Integer disease status

time Integer time variable

currentage Integer current age

mgene Integer mutation gene status

proband Integer proband indicator

relation Integer relationship code

Details

The dataset name has been kept as 'lynch_ontario_families_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 **FamEvent** package version 3.2.

norovirus_derbyshire_df *Norovirus Outbreak in Derbyshire*

Description

This dataset, norovirus_derbyshire_df, is a data frame describing an outbreak of norovirus in the summer of 2001 in a primary school and nursery in Derbyshire, England. It contains 492 observations across 5 variables tracking illness patterns among students.

Usage

```
data(norovirus_derbyshire_df)
```

Format

A data frame with 492 observations and 5 variables:

class Factor with 15 levels representing school classes

day_absent Integer day of absence

start_illness Integer day when illness started

end_illness Integer day when illness ended

day_vomiting Integer day when vomiting occurred

Details

The dataset name has been kept as 'norovirus_derbyshire_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the Digestive-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 **outbreaks** package version 1.9.0. Original source: O'Neill and Marks (2005).

pancreatic_cancer_df *Pancreatic Cancer Clinical Trial*

Description

This dataset, pancreatic_cancer_df, is a data frame containing data from a Phase II clinical trial of patients with locally advanced or metastatic pancreatic cancer. It includes time-to-event data for disease progression and death, as well as staging information.

Usage

```
data(pancreatic_cancer_df)
```

Format

A data frame with 41 observations and 4 variables:

stage Factor indicating disease stage (locally advanced or metastatic)

onstudy Factor indicating time (in days) from enrollment

progression Factor indicating time (in days) to disease progression

death Factor indicating time (in days) to death

Details

The dataset name has been kept as 'pancreatic_cancer_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 asaur package version 0.50.

---

pbc_mayo_survival_df *Mayo Clinic Primary Biliary Cirrhosis*

Description

This dataset, pbc_mayo_survival_df, is a data frame containing data from a randomized control trial conducted at Mayo Clinic from 1974 to 1984, studying the progression of primary biliary cirrhosis. The dataset includes both survival and longitudinal measurements with 1945 observations across 16 clinical variables.

Usage

```
data(pbc_mayo_survival_df)
```

Format

A data frame with 1945 observations and 16 variables:

ID Integer patient identifier

Time Numeric time measurement

death Numeric death indicator

obstime Numeric observation time

serBilir Numeric serum bilirubin measurement

albumin Numeric serum albumin measurement

alkaline Integer alkaline phosphatase level

platelets Integer platelet count

drug Factor with 2 levels indicating treatment group

age Numeric age in years

gender Factor with 2 levels indicating patient sex

ascites Factor with 2 levels indicating presence of ascites

hepatom Factor with 2 levels indicating presence of hepatomegaly

start Numeric start time for interval

stop Numeric stop time for interval

event Numeric event indicator

Details

The dataset name has been kept as 'pbc_mayo_survival_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 **JSM** package version 1.0.1.

`post_ercp_pancreatitis_tbl_df`
Indomethacin for Post-ERCP Pancreatitis

Description

This dataset, post_ercp_pancreatitis_tbl_df, is a tibble containing results from a randomized, placebo-controlled, prospective 2-arm trial of rectal indomethacin (100 mg) versus placebo to prevent post-ERCP pancreatitis in 602 participants, as reported by Elmunzer, Higgins, et al. (2012) in the New England Journal of Medicine.

Usage

```
data(post_ercp_pancreatitis_tbl_df)
```

Format

A tibble with 602 observations and 33 variables:

id Numeric subject identifier

site Factor indicating study site (4 levels)

age Numeric age of the participant

risk Numeric risk score

gender Factor indicating gender (2 levels)

outcome Factor indicating study outcome (2 levels)

sod Factor indicating presence of sphincter of Oddi dysfunction (2 levels)

pep Factor indicating presence of post-ERCP pancreatitis (2 levels)

recpanc Factor indicating recurrent pancreatitis (2 levels)

psphinc Factor indicating pancreatic sphincterotomy (2 levels)

precut Factor indicating precut sphincterotomy (2 levels)

difcan Factor indicating difficult cannulation (2 levels)

pneudil Factor indicating pneumatic dilation (2 levels)

amp Factor indicating ampullary interventions (2 levels)

paninj Factor indicating pancreatic injury (2 levels)

acinar Factor indicating acinarization (2 levels)

brush Factor indicating brushing procedures (2 levels)

asa81 Factor indicating ASA 81 mg use (3 levels)

asa325 Factor indicating ASA 325 mg use (3 levels)

asa Factor indicating ASA status (3 levels)

prophystent Factor indicating prophylactic stent placement (2 levels)

therastent Factor indicating therapeutic stent use (2 levels)

pdstent Factor indicating pancreatic duct stent (2 levels)

sodsom Factor indicating somatostatin use for SOD (2 levels)

bsphinc Factor indicating biliary sphincterotomy (2 levels)

bstent Factor indicating biliary stent (2 levels)

chole Factor indicating cholecystectomy (2 levels)

pbmal Factor indicating presence of pancreaticobiliary malignancy (2 levels)

train Factor indicating if performed by trainee (2 levels)

status Factor indicating trial status (2 levels)

type Factor indicating procedure type (4 levels)

rx Factor indicating treatment group: placebo or indomethacin (2 levels)

bleed Numeric bleeding indicator

Details

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

Source

Data taken from the medicaldata package version 0.2.0.

---

ugi_bleeding_df *H2 Antagonists in UGIB*

Description

This dataset, ugi_bleeding_df, is a data frame containing results from 27 studies examining the effectiveness of histamine H2 antagonists (cimetidine or ranitidine) in treating acute upper gastrointestinal hemorrhage, with 14 variables per study.

Usage

```
data(ugi_bleeding_df)
```

Format

A data frame with 27 observations and 14 variables:

id Integer study identifier

trial Character trial name/location

year Integer publication year

ref Integer reference number

trt Character treatment description

ctrl Character control description

nti Integer treatment group sample size

b.xti Integer treatment group bleeding events

o.xti Integer treatment group other events

d.xti Integer treatment group deaths

nci Integer control group sample size

b.xci Integer control group bleeding events

o.xci Integer control group other events

d.xci Integer control group deaths

Details

The dataset name has been kept as 'ugi_bleeding_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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.

`view_datasets_digestive`

View Available Datasets in DigestiveDataSets

Description

This function lists all datasets available in the 'DigestiveDataSets' package. If the 'DigestiveDataSets' 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_digestive()
```

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("DigestiveDataSets", quietly = TRUE)) {
  library(DigestiveDataSets)
  view_datasets_digestive()
}
```

`weight_loss_df` *Obese Patient Weight Loss Data*

Description

This dataset, weight_loss_df, is a data frame containing the weight, in kilograms, of an obese patient measured at 52 time points over an 8-month period as part of a weight rehabilitation programme.

Usage

```
data(weight_loss_df)
```

Format

A data frame with 52 observations and 2 variables:

Days Integer vector indicating the number of days since the beginning of the programme

Weight Numeric vector indicating the weight (in kilograms) of the patient at each time point

Details

The dataset name has been kept as 'weight_loss_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the DigestiveDataSets 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 MASS package version 7.3-65.

Index