Package: lgrdata (via r-universe)

October 31, 2026

Type Package

Title Example Datasets for a Learning Guide to R

Version 0.1.2

Description A collection of example datasets, including several classics. Many of these datasets are well suited for regression, classification, and visualization.

Encoding UTF-8

Depends R (>= 2.10)

Suggests ggplot2

LazyData false

License CC0

Config/roxygen2/version 8.1.0

Repository https://remkoduursma.r-universe.dev

Date/Publication 2026-10-01 18:05:44 UTC

RemoteUrl https://github.com/remkoduursma/lgrdata

RemoteRef HEAD

RemoteSha 0d9b65fff1bc485f568d170f328e1dd2b18a4d05

Contents

allometry 2
anthropometry 3
automobiles 4
berkeley 4
brunhild 5
callitrishydraulic 6
cereal1 6
cereal2 7
cereal3 7
cereals 7
choat_precipp50 8

coweeta . . . 9
dutchelection . . . 10
eucface_gasexchange . . . 11
eucfacegc . . . 11
fluxtower . . . 12
germination_fire . . . 13
germination_water . . . 14
hfeifbytree . . . 15
hfeifplotmeans . . . 16
hfemet2008 . . . 16
howell . . . 17
hydro . . . 18
icecream . . . 18
masslost . . . 19
memory . . . 20
oil . . . 21
pulse . . . 21
pupae . . . 22
rain . . . 23
sydney_hobart_times . . . 24
titanic . . . 25
treecanopy . . . 25
vessel . . . 26
weightloss . . . 27
wildmousemetabolism . . . 27

Index **29**

`allometry` *Allometry*

Description

This dataset contains measurements of tree dimensions and biomass. Data kindly provided by John Marshall, University of Idaho.

Usage

```
data(allometry)
```

Format

A data frame with 63 rows and 5 variables:

`species` factor The tree species (PSME = Douglas fir, PIMO = Western white pine, PIPO = Ponderosa pine).

`diameter` double Tree diameter at 1.3m above ground (cm).

`height` double Tree height (m).

`leafarea` double Total leaf area (m2)

`branchmass` double Total (oven-dry) mass of branches (kg).

Examples

```
data(allometry)
with(allometry, plot(diameter, height, pch=19, col=species))
```

`anthropometry` *Child anthropometry*

Description

Data include measurements of age, foot length, and height for 3898 children. These data are a small subset of many dozens of measurements on the same children, described in detail by Snyder (1977).

Usage

```
data(anthropometry)
```

Format

A data frame with 3898 rows and 4 variables:

`age` double Age in years

`gender` integer "female" or "male"

`foot_length` integer Total foot length (mm)

`height` double Total height (cm)

Source

<http://mreed.umtri.umich.edu/mreed/downloads.html>.

Examples

```
data(anthropometry)
with(anthropometry, plot(age, foot_length, pch=16, cex=0.5, col=gender))
```

automobiles *Cars data*

Description

Fuel efficiency, weight, acceleration, and other measurements on 398 cars. The majority of the data come from American cars (n = 249), and some European (n = 70) and Japanese (n = 79). Not to be confused with cars data provided by base R, see `cars` and `mtcars`.

Usage

```
data(automobiles)
```

Format

A data frame with 398 rows and 9 variables:

`car_name` character Make and model

`origin` factor 'American', 'European' or 'Japanese'

`build_year` double Year car was built

`fuel_efficiency` double Liters / 100km

`cylinders` integer Nr. of cylinders

`engine_volume` double Engine volume ('displacement') in liters.

`horsepower` integer Engine power (hp)

`weight` double Car weight in kg

`acceleration` double Time to accelerate to 60mph

Source

Data originally hosted on <http://lib.stat.cmu.edu/datasets/>, also used in ISLR (as the 'Auto' dataset). Converted to metric units for use in this package.

berkeley *Berkeley admissions data, 1973*

Description

A well-known example dataset, used as an excellent example for Simpson's Paradox. The Wikipedia page (see source), describes: "The admission figures for the fall of 1973 showed that men applying were more likely than women to be admitted, and the difference was so large that it was unlikely to be due to chance. But when examining the individual departments, it appeared that six out of 85 departments were significantly biased against men, whereas only four were significantly biased against women. In fact, the pooled and corrected data showed a 'small but statistically significant bias in favor of women.'"

Usage

```
data(berkeley)
```

Format

A data frame with 6 rows and 5 variables:

`Department` integer University Department, A-F

`Admitted_Male` integer Nr. Admitted male applicants

`Denied_Male` integer Nr. Denied male applicants

`Admitted_Female` integer Nr. Addmitted female applicants

`Denied_Female` integer Nr. Denied female applicants.

Source

<https://en.wikipedia.org/wiki/Simpson

---

brunhild *A Baboon Named Brunhilda*

Description

The observed responses are Geiger counter counts (times 10-4) used to measure the amount of radioactively tagged sulfate drug in the blood of a baboon named Brunhilda after an injection of the drug.

Usage

```
data(brunhild)
```

Format

A data frame with 21 rows and 2 variables:

`Hours` integer Hours after drug injection

`Sulfate` double Tagged sulfate concentration in blood

Source

<http://www.statsci.org/data/general/brunhild.html>

callitrishydraulic *Cavitation resistance for Callitris branches*

Description

Measurements of so-called 'percent loss conductivity' (PLC) curves on terminal twigs of Callitris trees (a member of the Cupressaceae in Australia). Twigs are subjected to increasingly negative xylem pressure (Psi, included as a positive pressure in MPa), and the loss in conductivity (i.e. the conductivity of water transport in the xylem) is measured.

Usage

```
data(callitrishydraulic)
```

Format

A data frame with 31 rows and 3 variables:

`Rep` integer Replicate - four branches are included.

`Psi` double Positive-valued negative xylem water pressure (MPa)

`PLC` double Percent loss conductivity (sometimes < 0)

Examples

```
data(callitrishydraulic)
with(callitrishydraulic, plot(Psi, PLC, pch=Rep))
```

cereal1 *Cereal nutrition data - small subset nr1*

Description

Small subset nr1 of the Cereals data to practice merging, see `cereals` (available are cereal1, cereal2 and cereal3).

Usage

```
data(cereal1)
```

Format

An object of class `data.frame` with 10 rows and 2 columns.

cereal2 *Cereal nutrition data - small subset nr2*

Description

Small subset nr1 of the Cereals data to practice merging, see `cereals` (available are cereal1, cereal2 and cereal3).

Usage

```
data(cereal2)
```

Format

An object of class `data.frame` with 8 rows and 2 columns.

cereal3 *Cereal nutrition data - small subset nr3*

Description

Small subset nr1 of the Cereals data to practice merging, see `cereals` (available are cereal1, cereal2 and cereal3).

Usage

```
data(cereal3)
```

Format

An object of class `data.frame` with 6 rows and 2 columns.

cereals *Cereal nutrition data*

Description

This dataset summarizes 77 different brands of breakfast cereals, including calories, proteins, fats, and so on, and gives a 'rating' that indicates the overall nutritional value of the cereal.

Usage

```
data(cereals)
```

Format

A data frame with 77 rows and 13 variables:

`Cereal.name` character Cereal name

`Manufacturer` factor Cereal manufacturer (letter code)

`Cold.or.Hot` factor 'C' or 'H'

`calories` integer

`protein` integer

`fat` integer

`sodium` integer

`fiber` double

`carbo` double

`sugars` integer

`potass` integer

`vitamins` integer

`rating` double Health rating of the cereal (unknown calculation method).

Source

<https://dasl.datadescription.com/datafile/cereals/> (Originally at Statlib CMU).

---

choat_precipp50 *Choat's Plant Drought Tolerance*

Description

Data include a measure of plant drought tolerance (P50, more negative values indicate plant stems can tolerate lower water contents), and mean annual precipitation of the location where the sample was taken. Data are for 115 individual species (species name not included). Data are from original source were simplified for the purpose of this book.

Usage

```
data(choat_precipp50)
```

Format

A data frame with 115 rows and 2 variables:

`annualprecip` integer Annual rainfall (mm) where the plant was sampled.

`P50` double The negative water pressure in the xylem at which 50% of stem conductivity is lost. More negative indicates higher tolerance to drought.

Source

Choat B. et al., 2012, Global convergence in the vulnerability of forests to drought, Nature 491, pages 752–755 <https://www.nature.com/articles/nature11688>.

coweeta *Coweeta tree data*

Description

Tree measurements in the Coweeta LTER.

Usage

```
data(coweeta)
```

Format

A data frame with 87 rows and 9 variables:

`species` integer One of 10 tree species

`site` integer Site abbreviation

`elev` integer Elevation (m asl)

`age` integer Tree age (yr)

`DBH` double Diameter at breast height (cm)

`height` double Tree height (m)

`folmass` double Foliage mass (kg)

`SLA` double Specific leaf area (index of leaf thinness) (cm2 g-1)

`biomass` double Total tree biomass

Details

DETAILS

Source

Martin J.G., et al., 1998, Aboveground biomass and nitrogen allocation of ten deciduous southern Appalachian tree species, Canadian Journal of Forest Research 28, 1648-1659.

`dutchelection` *Dutch election data*

Description

Polls for the 12 leading political parties in the Netherlands, leading up to the general election on 12 Sept. 2012. Data are in 'wide' format, with a column for each party. Values are in percentages.

Usage

```
data(dutchelection)
```

Format

A data frame with 22 rows and 12 variables:

`Date` factor Date of poll (NOTE: has not been converted to Date class)

`VVD` double Vote for this part in percentage.

`PvdA` double Vote for this part in percentage.

`PVV` double Vote for this part in percentage.

`CDA` double Vote for this part in percentage.

`SP` double Vote for this part in percentage.

`D66` double Vote for this part in percentage.

`GL` double Vote for this part in percentage.

`CU` double Vote for this part in percentage.

`SGP` double Vote for this part in percentage.

`PvdD` double Vote for this part in percentage.

`FiftyPlus` double Vote for this part in percentage.

Source

<http://en.wikipedia.org/wiki/Dutch_general_election,_2012>

eucface_gasexchange *Leaf gas exchange at the EucFACE*

Description

Measurements of leaf net photosynthesis at the EucFACE experiment, on leaves of different trees growing in ambient and elevated CO$_2$ concentrations. Measurements were repeated four times during 2013 (labelled as Date=A,B,C,D).

Usage

```
data(eucface_gasexchange)
```

Format

A data frame with 84 rows and 7 variables:

`Date` factor Date label (A-D)

`CO2` integer CO2 treatment, Amb=ambient, Ele=elevated

`Ring` integer One of six plots ('rings') where treatment was applied

`Tree` integer Tree number

`Photo` double Rate of leaf photosynthesis (mu mol m-2 s-1)

`Trmmol` double Rate of leaf transpiration (mmol m-2 s-1)

`VpdL` double Vapour pressure deficit (kPa)

Source

Gimeno T.E., 2015, Conserved stomatal behaviour under elevated CO2 and varying water availability in a mature woodland. Functional Ecology <https://doi.org/10.1111/1365-2435.12532>

eucfacegc *EucFACE ground cover data*

Description

This file contains estimates of plant and litter cover within the rings of the EucFACE experiment, evaluating forest ecosystem responses to elevated CO$_2$, on two dates. Within each ring are four plots and within each plot are four 1m by 1m subplots. Values represent counts along a grid of 16 points within each subplot.

Usage

```
data(eucfacegc)
```

Format

A data frame with 192 rows and 8 variables:

`Date` integer Date of measurement (d/m/y, not yet converted to Date class)

`Ring` integer The identity of the EucFACE Ring, the level at which the experimental treatment is applied.

`Plot` integer A total of four plots, nested within each level of Ring.

`Sub` integer A total of four subplots, nested within each level of Plot.

`Forbes` integer Number of points where dicot plants are observed.

`Grass` integer Number of points where grass is observed.

`Litter` integer Number of points where leaf litter is observed.

`Trt` integer The experimental treatment: `ctrl` for ambient levels of atmospheric carbon dioxide, `elev` for ambient plus 150ppm.

Source

Jeff Powell

`fluxtower` *Fluxtower data*

Description

This dataset contains measurements of CO$_2$ and H$_2$O fluxes (and related variables) over a pine forest in Quintos de Mora, Spain. The site is a mixture of *Pinus pinaster* and *Pinus pinea*, and was planted in the 1960's.

Data need to be cleaned to some extent (the purpose of this example dataset).

Usage

```
data(fluxtower)
```

Format

A data frame with 244 rows and 8 variables:

`TIMESTAMP` factor Date and time

`FCO2` double Canopy CO2 flux (mu mol m-2 s-1)

`FH2O` double Canopy H2O flux (mmol m-2 s-1)

`ustar` double Roughness length (m s-1)

`Tair` double Air temperature (degrees C)

`RH` double Relative humidity (%)

`Tsoil` double Soil temperature (degrees C)

`Rain` integer Rainfall (mm half hour-1)

Source

Data kindly provided by Victor Resco de Dios (in 2011), and simplified somewhat.

`germination_fire` *Seed germination as affected by fire*

Description

Two datasets on the germination success of seeds of four *Melaleuca* species, when subjected to temperature, fire cue, and dehydration treatments. Seeds were collected from a number of sites and subjected to 6 temperature treatments and fire cues (in the fire germination data), or two a range of dehydration levels (in the water germination data).

This dataset contains the fire treatment data.

Usage

```
data(germination_fire)
```

Format

A data frame with 576 rows and 7 variables:

`species` factor One of four Melaleuca species

`temp` integer Temperature treatment (C)

`fire.cues` integer Fire cue treatment (yes or no)

`site` integer Coding for the site where the seed was collected

`cabinet` integer ID for the cabinet where seeds were treated

`germ` integer Number of germinated seeds

`n` integer Number of seeds tested (20 for all rows)

Source

Data are from Hewitt et al. 2015 (Austral Ecology 40(6):661-671), shared by Charles Morris, and simplified for the purpose of this book.

See Also

`germination_water`

germination_water *Seed germination as affected by water*

Description

Two datasets on the germination success of seeds of four *Melaleuca* species, when subjected to temperature, fire cue, and dehydration treatments. Seeds were collected from a number of sites and subjected to 6 temperature treatments and fire cues (in the fire germination data), or two a range of dehydration levels (in the water germination data).

This dataset contains the water treatment data.

Usage

```
data(germination_water)
```

Format

A data frame with 352 rows and 5 variables:

`species` factor One of four Melaleuca species

`site` integer Coding for the site where the seed was collected

`water.potential` double Water potential of the seed (Mpa) after incubation (low values is drier)

`germ` integer Number of germinated seeds

`n` integer Number of seeds tested (25 for all rows)

Source

Data are from Hewitt et al. 2015 (Austral Ecology 40(6):661-671), shared by Charles Morris, and simplified for the purpose of this package.

See Also

germination_fire

Examples

```
data(germination_water)
with(germination_water,
  plot(jitter(water.potential), germ/n,
    pch=21, bg=terrain.colors(4)[species])
)
```

hfeifbytree *I x F at the HFE - tree observations*

Description

Heights and stem diameters of trees growing in a fertilization x irrigation experiment in Richmond, New South Wales, Australia, as part pf the Hawkesbury Forest Experiment (HFE). A total of 16 plots, each with 72 Eucalyptus saligna trees, was remeasured 17 times between 2008 and 2012. Treatments to the plots were either control (C), applied with fertilizer (F), irrigation (I), or irrigation+fertilization (IF).

This dataset contains the tree-level observations, see `hfeifplotmeans` for averaged data.

Usage

```
data(hfeifbytree)
```

Format

A data frame with 9592 rows and 6 variables:

`ID` integer A unique identifier for each tree.

`plotnr` integer A total of sixteen plots (four treatments).

`treat` integer One of four treatments (I - irrigated, F - dry fertilized, IL - Liquid fertilizer plus irrigation, C - control)

`Date` factor The date of measurement (YYYY-MM-DD)

`height` double Mean height for the sample trees (m).

`diameter` double Mean diameter for the sample trees (cm).

Source

Data courtesy of Craig Barton and Burhan Amiji, from Western Sydney University.

Examples

```
# Variable sample sizes over time. On many occassions, subsamples were measured.
data(hfeifbytree)
ftable(xtabs(~Date+treat, data=hfeifbytree))
```

hfeifplotmeans *I x F at the HFE - plot-level observations*

Description

Heights and stem diameters of trees growing in a fertilization x irrigation experiment in Richmond, New South Wales, Australia, as part pf the Hawkesbury Forest Experiment (HFE). A total of 16 plots, each with 72 Eucalyptus saligna trees, was remeasured 17 times between 2008 and 2012. Treatments to the plots were either control (C), applied with fertilizer (F), irrigation (I), or irrigation+fertilization (IF).

This dataset contains the plot-level means, see `hfeifbytree` for tree-level measurements.

Usage

```
data(hfeifplotmeans)
```

Format

A data frame with 320 rows and 5 variables:

`plotnr` integer A total of sixteen plots (four treatments).

`Date` factor The date of measurement (YYYY-MM-DD)

`diameter` double Mean diameter for the sample trees (cm).

`height` double Mean height for the sample trees (m).

`treat` integer One of four treatments (I - irrigated, F - dry fertilized, IL - Liquid fertilizer plus irrigation, C - control)

hfemet2008 *Weather data at the Hawkesbury Forest Experiment*

Description

Data for the weather station at the Hawkesbury Forest Experiment (HFE) for the year 2008. The HFE is in Richmond, New South Wales (in western Sydney), Australia.

Data are in 30min timestep.

Usage

```
data(hfemet2008)
```

Format

A data frame with 17568 rows and 9 variables:

`DateTime` integer Date Time (half-hourly steps)

`Tair` double Air temperature (degrees C)

`AirPress` double Air pressure (kPa)

`RH` double Relative humidity (%)

`VPD` double Vapour pressure deficit (kPa)

`PAR` double Photosynthetically active radiation (mu mol m-2 s-1)

`Rain` double Precipitation (mm)

`wind` double Wind speed (m s-1)

`winddirection` double Wind direction (degrees)

Source

Data courtesy of Craig Barton at Western Sydney University.

`howell` *Howell height, age and weight data*

Description

These data were also used by McElreath (2016, "Statistical Rethinking", CRC Press). Data include measurements of height, age and weight on Khosan people.

Usage

```
data(howell)
```

Format

A data frame with 783 rows and 4 variables:

`sex` factor male or female

`age` double Age (years)

`weight` double Body weight (kg)

`height` double Total height (cm)

Source

<https://tspace.library.utoronto.ca/handle/1807/17996>, subsetted for non-missing data and one outlier removed.

Examples

```
data(howell)
with(howell, plot(age, height, pch=19, col=sex))
```

hydro *Hydro dam storage data*

Description

This dataset describes the storage of the hydrodam on the Derwent river in Tasmania (Lake King William and Lake St. Clair), in equivalent of energy stored.

Usage

```
data(hydro)
```

Format

A data frame with 314 rows and 2 variables:

`Date` factor The date of the bi-weekly reading (d/m/yyyy)

`storage` integer Total water stored, in energy equivalent (GWh).

Details

DETAILS

icecream *Icecream sales and temperature*

Description

A synthetic dataset on weekly ice cream sales in two locations in Amsterdam, along with air temperature. The idea is that the ice cream salesman first sold icecream in 'Oosterpark', and decided to move shop to the 'Dappermarkt' the year after. Did sales improve? This dataset can be used to show that naive conclusions from simple linear model fits can be misleading, and that the use of covariates (here, air temperature) can change conclusions about effects.

Usage

```
data(icecream)
```

Format

A data frame with 40 rows and 3 variables:

`temperature` double Air temperature (C)

`sales` double Icecream sales per week (in local currency)

`location` factor Either 'Dappermarkt' or 'Oosterpark'

Examples

```
data(icecream)

# Linear model, temperature as covariate
fit_ice <- lm(sales ~ temperature*location, data=icecream)

# Try to guess from coefficients where the sales were higher:
summary(fit_ice)

# What about now?
with(icecream, plot(temperature, sales, pch=19, col=location))
legend("topleft", levels(icecream$location), fill=palette())
```

---

masslost *Genetically modified soybean litter decomposition*

Description

Soybean litter decomposition as a function of time (`date`), type of litter (`variety`), herbicides applied (`herbicide`), and where in the soil profile it is placed (`profile`). `masslost` refers to the proportion of the litter that was lost from the bag (decomposed) relative to the start of the experiment. Herbicide treatments were applied at the level of whole plots, with both treatments represented within each of four blocks. Both levels of variety and profile were each represented within each plot, with six replicates of each treatment added to each plot.

Usage

```
data(masslost)
```

Format

A data frame with 246 rows and 8 variables:

`plot` integer A total of eight plots.

`block` integer A total of four blocks.

`variety` integer Soybean variety is genetically modified ('gm') or not ('nongm'); manipulated at the subplot level.

`herbicide` integer Herbicide applied is glyphosate ('gly') or conventional program ('conv'); manipulated at plot level.

`profile` integer Whether litter was 'buried' in the soil or placed at the soil 'surface'; manipulated at the subplot level.

`date` integer Date at which litter bags were recovered.

`sample` integer Factor representing timing of sampling ('incrop1', 'incrop2', 'postharvest').

`masslost` double The proportion of the initial mass that was lost from each litter bag during field incubation. Some values are lower than zero due to insufficient washing of dirt and biota from litter prior to weighing.

Source

Jeff Powell

`memory` *Memory of words dataset*

Description

A dataset on the number of words remembered from list, for various learning techniques, and in two age groups.

Usage

```
data(memory)
```

Format

A data frame with 100 rows and 3 variables:

`Age` integer Age of person tested (yr)

`Process` factor One of five methods used to memorize the words.

`Words` double Number of words recalled.

Details

Description taken from source: "Why do older people often seem not to remember things as well as younger people? Do they not pay attention? Do they just not process the material as thoroughly? One theory regarding memory is that verbal material is remembered as a function of the degree to which is was processed when it was initially presented. Eysenck (1974) randomly assigned 50 younger subjects and 50 older (between 55 and 65 years old) to one of five learning groups. The Counting group was asked to read through a list of words and count the number of letters in each word. This involved the lowest level of processing. The Rhyming group was asked to read each word and think of a word that rhymed with it. The Adjective group was asked to give an adjective that could reasonably be used to modify each word in the list. The Imagery group was instructed to form vivid images of each word, and this was assumed to require the deepest level of processing. None of these four groups was told they would later be asked to recall the items. Finally, the Intentional group was asked to memorize the words for later recall. After the subjects had gone through the list of 27 items three times they were asked to write down all the words they could remember."

Source

<http://www.statsci.org/data/general/eysenck.html>.

oil *Crude oil production*

Description

Crude oil production for the top 8 oil-producing countries (minus Russia, for which understandably no data were available pre-1990), for the period 1971-2017.

Usage

```
data(oil)
```

Format

A data frame with 376 rows and 3 variables:

`country` factor Country code

`year` integer 1971 - 2017

`production` double Annual crude oil production in TOE.

pulse *Pulse Rates before and after Exercise*

Description

Pulse rates measured on 110 participating students. Half of the students ran in place for one minute, before their pulse rate was measured again.

Usage

```
data(pulse)
```

Format

A data frame with 110 rows and 11 variables:

`Height` integer Height (cm)

`Weight` double Weight (kg)

`Age` integer Age (years)

`Gender` integer Sex (1 = male, 2 = female)

`Smokes` integer Regular smoker? (1 = yes, 2 = no)

`Alcohol` integer Regular drinker? (1 = yes, 2 = no)

`Exercise` integer Frequency of exercise (1 = high, 2 = moderate, 3 = low)

`Ran` integer Whether the student ran or sat between the first and second pulse measurements (1 = ran, 2 = sat)

`Pulse1` integer First pulse measurement (rate per minute)

`Pulse2` integer Second pulse measurement (rate per minute)

`Year` integer Year of class (93 - 98)

Details

Description taken from source: "Students in an introductory statistics class (MS212 taught by Professor John Eccleston and Dr Richard Wilson at The University of Queensland) participated in a simple experiment. The students took their own pulse rate. They were then asked to flip a coin. If the coin came up heads, they were to run in place for one minute. Otherwise they sat for one minute. Then everyone took their pulse again. The pulse rates and other physiological and lifestyle data are given in the data. Five class groups between 1993 and 1998 participated in the experiment. The lecturer, Richard Wilson, was concerned that some students would choose the less strenuous option of sitting rather than running even if their coin came up heads, so in the years 1995-1998 a different method of random assignment was used. In these years, data forms were handed out to the class before the experiment. The forms were pre-assigned to either running or non-running and there were an equal number of each. In 1995 and 1998 not all of the forms were returned so the numbers running and sitting was still not entirely controlled."

Source

<http://www.statsci.org/data/oz/ms212.html>

Examples

```
data(pulse)
with(pulse, plot(Weight, Pulse2-Pulse1,
  pch=19, col=c("red2", "dimgrey")[Ran]))
abline(h=0, lty=5)
```

---

pupae *Pupae data*

Description

This dataset is from an experiment where larvae were left to feed on *Eucalyptus* leaves, in a glasshouse that was controlled at two different levels of temperature and CO$_2$ concentration. After the larvae pupated (that is, turned into pupae), the body weight was measured, as well as the cumulative 'frass' (larvae excrement) over the entire time it took to pupate.

Usage

```
data(pupae)
```

Format

A data frame with 84 rows and 5 variables:

`T_treatment` integer Temperature treatments ('ambient' and 'elevated')

`CO2_treatment` integer CO$_2$ treatment (280 or 400 ppm).

`Gender` integer The gender of the pupae : 0 (male), 1 (female)

`PupalWeight` double Weight of the pupae (g)

`Frass` double Frass produced (g)

Source

Data courtesy of Tara Murray, and simplified for the purpose of this package.

`rain` *Rain data*

Description

This dataset contains ten years (1995-2006) of daily rainfall amounts as measured at the Richmond RAAF base.

Usage

```
data(rain)
```

Format

A data frame with 3653 rows and 3 variables:

`Year` integer Year

`DOY` integer Day of year (1-366)

`Rain` double Daily rainfall amount (mm)

Source

<http://www.bom.gov.au/climate/data/>, simplified and adjusted for this package.

`sydney_hobart_times` *Sydney to Hobart winning times*

Description

Winning times for the Sydney to Hobart Yacht Race. An annual sail yacht race over 1170km, from Sydney's harbour, to Hobart in Tasmania. The race is infamous for the rough conditions, long distance, and large number of dropouts in some years. The data include the winning time, and the number of starting yachts, and the number of yachts reaching the finish.

Usage

```
data(sydney_hobart_times)
```

Format

A data frame with 72 rows and 5 variables:

`Year` integer Year race was held

`Time` double Total time (days)

`fleet_start` integer Number yachts at start

`fleet_finish` integer Number yachts at finish

`Time_record` double Record race up to this year

Source

<https://en.wikipedia.org/wiki/Sydney_to_Hobart_Yacht_Race>

Examples

```
data(sydney_hobart_times)
with(sydney_hobart_times, {
    plot(Year, Time)
    lines(Year, Time_record, type='s', col="red")
})
```

titanic *Passengers on the Titanic*

Description

Survival status of passengers on the Titanic, together with their names, age, sex and passenger class. Not to be confused with the dataset `Titanic`, provided with R, which lists only tables of passengers. This dataset on the other hand provides one row per passenger.

Usage

```
data(titanic)
```

Format

A data frame with 1313 rows and 5 variables:

`Name` integer Recorded name of passenger

`PClass` integer Passenger class: 1st, 2nd or 3rd

`Age` double Age in years (many missing)

`Sex` integer male or female

`Survived` integer 1 = Yes, 0 = No

Details

DETAILS

Source

<http://www.statsci.org/data/general/titanic.html>

treecanopy *Tree canopy gradients in the Priest River Experimental Forest (PREF)*

Description

Leaves of two pine species (35 trees in total) were sampled throughout their canopy, usually 8 samples were taken at various heights. The height is expressed as the 'distance from top', i.e. the distance to the apex of the tree. Leaves (conifer needles) were analysed for nitrogen content (narea), and an index of leaf thickness, the 'leaf mass per area'. The data show the usual pattern of higher leaf thickness (higher LMA) toward the top of the trees, but individual trees show a lot of variation in LMA.

Usage

```
data(treecanopy)
```

Format

A data frame with 249 rows and 7 variables:

`ID` integer ID of the individual tree

`species` integer Pinus ponderosa or Pinus monticola

`dfromtop` double Distance from top of tree (where leaf sample was taken) (m)

`totheight` double Total height of the tree (m)

`height` double Height from the ground (where sample was taken) (m)

`LMA` double Leaf mass per area (g m-2)

`narea` double Nitrogen per area (gN m-2)

Source

Marshall, J.D., Monserud, R.A. 2003. Foliage height influences specific leaf area of three conifer species. Can J For Res 33:164-170

Examples

```
data(treecanopy)
if(require(ggplot2)){
 ggplot(treecanopy, aes(dfromtop,LMA,group=ID,col=species)) +
   geom_point() +
   stat_smooth(method="lm",se=FALSE) +
   theme_minimal()
}
```

---

vessel *Xylem vessel diameters*

Description

Measurements of diameters of xylem (wood) vessels on a single *Eucalyptus saligna* tree grown at the Hawkesbury Forest Experiment.

Usage

```
data(vessel)
```

Format

A data frame with 550 rows and 3 variables:

`position` integer Either 'base' or 'apex' : the tree was sampled at stem base and near the top of the tree.

`imagenr` integer At the stem base, six images were analyzed (and all vessels measured in that image). At apex, three images.

`vesseldiam` double Diameter of individual water-conducting vessels (mu m).

Source

Sebastian Pfautsch

`weightloss` *Weight loss data*

Description

This dataset contains measurements of a Jeremy Zawodny over a period of about 3 months while he was trying to lose weight. This is an example of an irregular timeseries dataset (intervals between measurements vary).

Usage

```
data(weightloss)
```

Format

A data frame with 67 rows and 2 variables:

`Date` factor Date, d/m/yy

`Weight` double Weight, in pounds

Source

<http://jeremy.zawodny.com/blog/archives/006851.html>

`wildmousemetabolism` *Mouse metabolism*

Description

Wild mice were placed in a device where the metabolic rate (energy used by the animal) can be measured directly, and continuously. Measurements were made at varying temperature (15, 20 and 31C), mice were provided with food or not, and were able to exercise (with a treadmill) or not.

Usage

```
data(wildmousemetabolism)
```

Format

A data frame with 864 rows and 9 variables:

`id` integer Individual number

`run` integer The experiment was repeated three times (run = 1,2,3)

`day` integer Day of experiment (1-6)

`temp` integer Temperature (deg C)

`food` integer Whether food was provided ('Yes') or not ('No')

`bm` double Body mass (g)

`wheel` integer Whether the mouse could use an exercise wheel ('Yes') or not ('No')

`rmr` double Resting metabolic rate (minimum rate of a running average over 12min) (kC hour-1)

`sex` integer Male or Female

Source

Christopher Turbill

Index

∗ **datasets**
allometry, 2
anthropometry, 3
automobiles, 4
berkeley, 4
brunhild, 5
callitrishydraulic, 6
cereal1, 6
cereal2, 7
cereal3, 7
cereals, 7
choat_precipp50, 8
coweeta, 9
dutchelection, 10
eucface_gasexchange, 11
eucfacegc, 11
fluxtower, 12
germination_fire, 13
germination_water, 14
hfeifbytree, 15
hfeifplotmeans, 16
hfemet2008, 16
howell, 17
hydro, 18
icecream, 18
masslost, 19
memory, 20
oil, 21
pulse, 21
pupae, 22
rain, 23
sydney_hobart_times, 24
titanic, 25
treecanopy, 25
vessel, 26
weightloss, 27
wildmousemetabolism, 27

allometry, 2
anthropometry, 3
automobiles, 4

berkeley, 4
brunhild, 5

callitrishydraulic, 6
cars, *4*
cereal1, 6
cereal2, 7
cereal3, 7
cereals, *6*, 7, *7*
choat_precipp50, 8
coweeta, 9

dutchelection, 10

eucface_gasexchange, 11
eucfacegc, 11

fluxtower, 12

germination_fire, 13, *14*
germination_water, *13*, 14

hfeifbytree, 15, *16*
hfeifplotmeans, *15*, 16
hfemet2008, 16
howell, 17
hydro, 18

icecream, 18

masslost, 19
memory, 20
mtcars, *4*

oil, 21

pulse, 21
pupae, 22

rain, 23

sydney_hobart_times, 24

Titanic, *25*
titanic, 25
treecanopy, 25

vessel, 26

weightloss, 27
wildmousemetabolism, 27