Package 'SRMData'

May 28, 2025

Type Package

Title Data Files Supporting ``Scientific Research and Methodology'' by Peter K. Dunn (2025)

Version 1.0.1

Author Peter K. Dunn [aut, cre]

Maintainer Peter K. Dunn <pdunn2@usc.edu.au>

Description Provides most of the data files used in the textbook ``Scientific Research and Methodology'' by Dunn (2025, ISBN:9781032496726; forthcoming).

License GPL (>= 2)

Language en-GB

Encoding UTF-8

LazyData false

Depends R (>= 3.5.0)

RoxygenNote 7.3.2

NeedsCompilation no

Repository CRAN

Date/Publication 2025-05-28 10:10:02 UTC

Contents

`AISsub` *AISsub*

Description

Body measurements from athletes at the Australian Institute of Sport.

Usage

```
data(AISsub)
```

Format

A data frame with 202 rows (each athlete) and 6 columns:

Sex The sex of the athlete; one of `F` or `M`

SSF The sum of skin folds

PBF The percentage body fat

Sport The sport played by the athlete; one of `BBall` (basketball), `Field`, `Gym` (gymnastics), `Netball`, `Rowing`, `Swim` (swimming), `T400m`, (track, further than 400m), `Tennis`, `TPSprnt` (track sprint events), `WPolo` (waterpolo)

Wt The weight of the athlete, in kg

Ht The height, in cm

Source

OzDASL, available on-line at http://www.statsci.org/data/.

References

Telford, R. D. and Cunningham, R. B. (1991). Sex, sport, and body-size dependency of hematology in highly trained athletes. *Medicine and Science in Sports and Exercise*, **23**(7):788–794.

`Anorexia` *Weight loss after treatment for anorexia*

Description

Weight changes in girls with anorexia: two treatments.

Usage

```
data(Anorexia)
```

Format

A data frame with 72 rows and 3 columns:

Treatment The treatment type; one of `CB` (cognitive behavioural treatment), `Control` (the control group) or `FT` (family therapy)

Before Weight (in kg) before the anorexia treatment

After Weight (in kg) after the anorexia treatment

Source

D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski (1994) A Handbook of Small Data Sets, London: Chapman and Hall. Dataset 285.

B12Diet *Vegetarianism and B12*

Description

B12 deficiency in vegetarian and non-vegetarian women.

Usage

```
data(B12Diet)
```

Format

A data frame with 124 rows (one for each person) and 2 columns:

B12 B12 deficiency; one of `1` (B12 deficient) or `2` (Not B12 deficient)

Diet The diet; one of `1` (Vegetarian) or `2` (non-vegetarian)

Source

Gammon, Cheryl S., Pamela R. von Hurst, Joan Coad, Rozanne Kruger, and Welma Stonehouse. 2012. Vegetarianism, Vitamin B12, and Insulin Resistance in a Group of Predominately Overweight/Obese South Asian Women. *Nutrition* **28**: 20–24.

BabyBoom *Baby births in one day at one hospital*

Description

Details of the births on one day from a Brisbane hospital.

Usage

```
data(BabyBoom)
```

Format

A data frame with 44 rows (one per birth) and 3 columns:

Gender The gender of the child; one of `Female` or `Male`

Weight The weight of the baby, in kg

Mins.Since.Midnight the time of birth, in minutes since midnight

Source

Steele, S. 1997. Babies by the Dozen for Christmas: 24-Hour Baby Boom. *The Sunday Mail*, 7.

Dunn, Peter K. 1999. A Simple Dataset for Demonstrating Common Distributions. *Journal of Statistics Education*, **7** (3).

Battery *Battery performance*

Description

Battery life for two brands of batteries.

Usage

```
data(Battery)
```

Format

A data frame with 108 rows (one per battery) and 4 columns:

Brand One of `Energizer` or `Ultracell` (ALDI home brand))

Voltage The voltages at which times were recorded

Time The time taken for 1.5V battery to reduce to the given voltage, in hours

Battery Which battery in the sequence

Source

Dunn, Peter K. 2013. Comparing the Lifetimes of Two Brands of Batteries. *Journal of Statistical Education*, **21** (1).

Bitumen *Bitumen content*

Description

Relationship between bitumen content and percentage air voids.

Usage

```
data(Bitumen)
```

Format

A data frame with 42 rows and 2 columns:

Bitumen The bitumen content (by percentage weight) in the bitumen sample

AirVoids The percentage of air voids, by volume

Source

Panda, R. P., Sudhanshu Sekhar Das, and P. K. Sahoo. 2018. Relation Between Bitumen Content and Percentage Air Voids in Semi Dense Bituminous Concrete. *Journal of The Institution of Engineers (India): Series A* **99** (2): 327–32.

BMI *BMI of Irish patients*

Description

The BMI and other health data number of Irish patients.

Usage

```
data(BMI)
```

Format

A data frame with 70 rows and 11 columns:

sex Sex of the person; one of `female` or `male`

age Age of person, in completed years

edu Level of education; one of `primary`, `secondary`, `postLeaving`, `complete3rd`

m_card whether the person has a medical card; one of `yes` or `no`

smoke smoking status; one of `daily`, `occasionally` or `not at all`

drink whether the person drinks alcohol weekly; one of `yes` or `no`

exercise The number of days per week the person walks or exercise for 30 minutes or more

diet whether the person thinks they have a healthy diet; one of `yes`, `no` or `dont know`

ob_weight_kg the observed (measured) weight, in kg

ob_height_m the observed (measured) height in metres

sr_weight_kg the weight reported by the person, in kg

sr_height_m the height reported by the person, in metres

bmi_perception the person perception of the BMI; one of `normalweight`, `overweight` or `obese`

Details

The data come from a survey.

Source

Johnson, E., Millar, S. R., & Shiely, F. (2021). The association between BMI self-selection, self-reported BMI and objectively measured BMI. *HRB Open Research*, **4**(37), 37.

BodyTemp *Body temperatures*

Description

Body temperature (in degrees C and F) for people.

Usage

```
data(BodyTemp)
```

Format

A data frame with 130 rows (each person) and 4 columns:

BodyTemp The measured body temperature, in degrees F, as given

Gender One of 1 (males) or 2 (females)

HeartRate Heart rate, in beats per minute

BodyTempC The measured body temperature in degrees C; converted from degrees F

Source

Allen, L. S. (1996). What's normal?–Temperature, gender, and heart rate. *Journal of Statistics Education*, **4**(2).

References

Wunderlich, C. 1868. Das Verhalten Der Eiaenwarme in Krankenheitem. Leipzig, Germany: Otto Wigard. Mackowiak, Philip A., Steven S.Wasserman, and Myron M. Levine. 1992. A Critical Appraisal of 98.6 degrees F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich. *Journal of the American Medical Association* **268** (12): 1578–80.

BoneQuality *Bone quality in South Koreans*

Description

Bone mass density of South Korean subjects, at three body locations.

Usage

```
data(BoneQuality)
```

Format

A data frame with 969 rows (one for each student) and 7 columns:

Sex The sex of the subject; one of `M` (male) or `F` (female)

Age The age of the subject, in years

Height The height of the subject, in cm

Weight The weight of the subject, in kg

LumbarBMD The bone mass density of the lumbar spine, in g/square-cm

HipBMD The bone mass density of the total hip, in g/square-cm

NeckBMD The bone mass density of the femoral neck, in g/square-cm

Details

Bone mass density and demographic information for 969 subjects in South Korea.

Source

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0260924#sec013

References

Kim, K. Y., & Kim, K. M. (2022). Similarities and differences between bone quality parameters, trabecular bone score and femur geometry. *PLOS One*, **17**(1), e0260924.

`Borers` *The impact of sugarcane borers*

Description

The impact of sugarcane borers on reducing sorghum fitness and grain production.

Usage

```
data(Borers)
```

Format

A data frame with 72 rows and 8 columns:

Hybrids The hybrid; one of `AG1090`, `BRS373` or `DKB590`

Insecticide Whether insecticide was used; one of `with` or `without`

Height The plant height, in cm

Tunnels The length of borers tunnels, in cm

PanicleLength The panicle (flower cluster) length, in cm

PanicleWeight The panicle (flower cluster) weight, in cm

Infestation The amount of infestation (the 'stem borer injury'), as a percentage

Yield The sorghum yield, in kg per hectare

Details

The data provide details of sorghum yield in the presence of borer infestation, from a study Brazil conducted over three years.

Source

Souza, Camila and Souza, Bruno and Fadini, Marcos and França, Joselia and Menezes, Cícero and Nascimento, Priscilla and Mendes, Simone (2025), "What is the potential of sugarcane borer in reducing sorghum fitness and grain production?", *Mendeley Data*, V2, doi: 10.17632/b6s9wnxgfm.2

References

Souza, C., de Souza, B. H. S., Fadini, M. A. M., França, J. C. O., de Menezes, C. B., Nascimento, P. T., and Mendes, S. M. (2024). What is the potential of sugarcane borer in reducing sorghum fitness and grain production?. *Journal of Applied Entomology*, **148**(7), 818–826.

Burros *The health of burros*

Description

The health of females burros in the Mojave Desert.

Usage

```
data(Burros)
```

Format

A data frame with 9 rows and 3 columns:

Status The reproductive status of the female burro; one of 1 (barren), 2 (pregnant (but not lactating)), or 3 (lactating)

Health The health of the burro; one of 1 (excellent), 2 (fair) or 3 (poor).

Counts The number of female burros in each cell

Details

The data provide the number of female burros of given health and reproductive status.

Source

Johnson, R. A., Carothers, S. W., & McGill, T. J. (1987). Demography of feral burros in the Mohave Desert. *The Journal of Wildlife Management*, **51**(4), 916–920.

Captopril *Captopril effectiveness*

Description

Blood pressure before and after treatment with Captopril.

Usage

```
data(Captopril)
```

Format

A data frame with 30 rows (one per person) and 3 columns:

Before The blood pressure before taking captopril, in mm Hg

After The blood pressure after taking captopril, in mm Hg

BP The type of blood pressure measured; `S` for systolic, and `D` for diastolic

Source

D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski (1994) A Handbook of Small Data Sets, London: Chapman and Hall. Dataset 72.

References

MacGregor, Graham A., N. D.Markandu, J. E. Roulston, and J. C. Jones. 1979. Essential Hypertension: Effect of an Oral Inhibitor of Angiotensin-Converting Enzyme. *British Medical Journal* **2**: 1106–1109.

CarCrashes *Car crashes*

Description

The number and type of of car crashes, in two different years.

Usage

```
data(CarCrashes)
```

Format

A data frame with 4 rows and 3 columns:

CrashType Whether or not the crash involved pedestrians (`1`) or other vehicle (`2`)

Year Either `2011` or `2015`

Counts The number of crashes in the combination defined by `CrashType` and `Year`

Details

The data provide the number of car crashes in a mountainous county in western China, some involving pedestrians and some involving other vehicles, in two years

Source

Wang, Liyang, Ruimin Li, Changjun Wang, and Zhiyong Liu (2020). "Driver Injury Severity Analysis of Crashes in a Western China's Rural Mountainous County: Taking Crash Compatibility Difference into Consideration.". *Journal of Traffic and Transportation Engineering* (English Edition).

`CherryRipe` *Cherry Ripe weights*

Description

The weight of 'Fun Size' Cherry Ripe chocolate bars.

Usage

```
data(CherryRipe)
```

Format

A data frame with 16 rows (each combination of the other variables) and 4 columns:

TotalWeight The weight of the wrapper bar, in g

WrapperWt The weight of the wrapper only, in g

BarWt The weight of the chocolate bar itself, in g, by subtraction

Year The year, from 2011, 2013 to 2015, 2017 to 2019

Details

The Cherry Ripe chocolate bars were weighted as an in-class activity, usually by weighing the bar+wrapper, and then the wrapper (for hygiene reasons) on a set of scales. The bars were in a Fun Size pack, of about 11 bars. Until 2015, the weights were listed in the nutrition panel as 18g. After 2015, this changed to 14g.

Source

Collected and weighed by Peter K. Dunn and students (who got to eat the chocolate bars).

Corollas *Price of second-hand Corollas*

Description

The price of second-hand Corollas advertised on Gum Tree (Australia).

Usage

```
data(Corollas)
```

Format

A data frame with 45 rows (one per vehicle) and 3 columns:

Year the year of manufacture of the vehicle

Price the advertised price, in AUD

Age the age of the vehicle, in years

Source

Collected by Peter K. Dunn, 2014, from www.gumtree.com.au

CrabShells2 *Crab shells and anemones (2x2)*

Description

The placement of anemones on their shells by hermit crabs.

Usage

```
data(CrabShells2)
```

Format

A data frame with 4 rows and 3 columns:

ShellColumn The column where anemone placed; one of 1 (Side) or 2 (Central)

ShellRow The row where anemone placed; one of 1 (Side) or 2 (Central)

Counts The number of anemones in the indicated sector on the shell

Details

The data provide the number of anemones placed on their shell by hermit crabs in indicated regions. Roughly, the shells are divided into a 3x3 grid of approximately equal areas (see `CrabShell3`) but here the 3x3 table has been collapsed to a 2x2 table.

Source

Brooks, W. R. (1989). Hermit crabs alter sea anemone placement patterns for shell balance and reduced predation. *Journal of Experimental Marine Biology and Ecology*, **132**(2), 109–121.

`CrabShells3` *Crab shells and anemones (3x3)*

Description

The placement of anemones on their shells by hermit crabs.

Usage

```
data(CrabShells3)
```

Format

A data frame with 9 rows and 3 columns:

ShellColumn The column where anemone placed; one of `1` (Side 1), `2` (Central) or `3` (Side 2)

ShellRow The row where anemone placed; one of `1` (Side 1), `2` (Central) or `3` (Side 2)

Counts The number of anemones in the indicated sector on the shell

Details

The data provide the number of anemones placed on their shell by hermit crabs in indicated regions. Roughly, the shells are divided into a 3 x 3 grid of approximately equal areas.

Source

Brooks, W. R. (1989). Hermit crabs alter sea anemone placement patterns for shell balance and reduced predation. *Journal of Experimental Marine Biology and Ecology*, **132**(2), 109–121.

`Cyclones` *Cyclones in the Australian region*

Description

The number of cyclones (severe; non-severe) and the ONI.

Usage

```
data(Cyclones)
```

Format

A data frame with 37 rows (one per person) and 8 columns:

Year The year

Severe The number of severe cyclones recorded in the Australian region

NonSevere The number of non-severe cyclones recorded in the Australian region

Total The total number of cyclones recorded in the Australian region

JFM the Ocean Nino Index, or oni, averaged over the months January to March; a numeric vector

AMJ the Ocean Nino Index, or oni, averaged over the months April to June; a numeric vector

JAS the Ocean Nino Index, or oni, averaged over the months July to September; a numeric vector

OND the Ocean Nino Index, or oni, averaged over the months October to December; a numeric vector

Source

Dunn, Peter K., and Gordon K. Smyth. 2018. *Generalized Linear Models with Examples in R*. Springer.

DanishLC *Danish lung cancer cases*

Description

The number of cases of lung cancer in four Danish cities.

Usage

```
data(DanishLC)
```

Format

A data frame with 24 rows (each combination) and 4 columns:

Cases The number of lung cancer cases for the given age group and city

Pop The population for the given age group and city

Age The age group; one of `40-54`, `55-59`, `60-64`, `65-69`, `70-74` or `>74`

City The city; one of `Fredericia`, `Horsens`, `Kolding` or `Vejle`

Source

James K. Lindsey (1995). *Modelling frequency and count data*. Clarendon Press, page 157.

References

E. B. Andersen (1977). Multiplicative Poisson models with unequal cell rates. *Scandinavian Journal of Statistics*, **4**, 153–158.

Deceleration *Deceleration of cars*

Description

The deceleration of cars after adding additional speed signage.

Usage

```
data(Deceleration)
```

Format

A data frame with 79 rows (one per car) and 2 columns:

When When the deceleration is measured: `Before` or `After` signage added

Deceleration The deceleration, in metres-per-second-squared

Source

Ma, Yongfeng, Wenbo Zhang, Xin Gu, and Jiguang Zhao. 2019. Impacts of Experimental Advisory Exit Speed Sign on Traffic Speeds for Freeway Exit Ramp. *PLoS One* **14** (11): e0225203.

Dental *Dental statistics*

Description

The data give the estimates of the mean number of decayed, missing and filled teeth (DMFT) at age 12 years, and the mean annual sugar consumption in the previous five years for 90 countries.

Usage

```
data(Dental)
```

Format

A data frame with 90 rows (one per person) and 4 columns:

Country the country; a factor

Indus whether the country is considered an industrialized country; a factor with levels `Yes` (industrialized) or `No` (not industrialized)

Sugar the mean annual sugar consumption in kilograms per person per year, computed over the five years (or as much as available) prior to the survey; a numeric vector

DMFT estimates of the mean number of decayed, missing and filled teeth at age 12; a numeric vector

Source

Woodward, M., and A. R. P.Walker. 1994. Sugar Consumption and Dental Caries: Evidence from 90 Countries. *British Dental Journal* **176**: 297–302

References

M. Woodward (2004) *Epidemiology: Study Design and Data Analysis*, second edition. Chapman and Hall.

`Diabetes` *Diabetes*

Description

Blood pressure on the first and second visits.

Usage

```
data(Diabetes)
```

Format

A data frame with 403 rows (one per person) and 4 columns; many values are missing

SBPfirst the systolic blood pressure from the first visit, in mm Hg

DBPfirst the diastolic blood pressure from the first visit, in mm Hg

SBPsecond the systolic blood pressure from the second visit, in mm Hg

DBPsecond the diastolic blood pressure from the second visit, in mm Hg

Source

Originally from <http://biostat.mc.vanderbilt.edu/DataSets>, though that URL no longer works. It seems to now appear at <https://hbiostat.org/data/repo/diabetes.html>

Dogs *Dog measurements*

Description

Measurements of Phu Quoc Ridgeback dogs.

Usage

```
data(Dogs)
```

Format

A data frame with 30 rows (one per dog) and 4 columns:

BL Body length, in cm

BH Body height, in cm

Chest Chest measurement, in cm

Waist Waist measurement, in cm

Source

Quan, Quoc-Dang, Hoang-Dung Tran, and Anh-Dung Chung. 2017. The Relation of Body Score (Body Height/Body Length) and Haplotype E on Phu Quoc Ridgeback Dogs (Canis Familiaris). *Journal of Entomology and Zoology Studies* **5**: 388–94

DogsLife *Lifespan of dogs*

Description

The average weight of dog breeds, and the average lifespan of dog breeds, using over 50 individuals for each breed.

Usage

```
data(DogsLife)
```

Format

A data frame with 73 rows and 5 columns:

Breed The breed name

Weight The average breed weight (in kg)

LitterSize The average breed litter size

BirthWeight The average breed birthweight (in kg)

Lifespan The average breed lifespan (in years)

Details

The original data list many more breeds, but these are (as best as I can determine) those based on at least 50 individuals, as noted in the original article.

Source

da Silva, Jack and Cross, Bethany (2022). Data and code for: Dog lifespans and the evolution of ageing [Dataset]. *Dryad* https://doi.org/10.5061/dryad.wwpzgmsn6

References

da Silva, J., & Cross, B. J. (2023). Dog life spans and the evolution of aging. *The American Naturalist*, **201**(6), E140–E152.

DogWalks *Dog walks*

Description

Dog walking in the city and country.

Usage

```
data(DogWalks)
```

Format

A data frame with 8 rows and 3 columns:

Location One of 1 (City) or 2 (Farm)

WalkLength One of 1 (Under 30 mins), 2 (30 to under 60 mins), 3 (60 to under 120 mins), or 4 (varies; mostly long walk but some shorter walks)

Counts The number of dogs in each cell

Details

The data provide the number of dogs being walked for given times, in the city and country.

Source

Naughton, Violetta, Teresa Grzelak, and Patrick J. Naughton. (2024). "Association Between Household Lo-cation (Urban Versus Rural) and Fundamental Care Provided to Domestic Dogs (Canis Familiaris) in Northern Ireland." In *Nutrition and Metabolism of Dogs and Cats*, 217–236. Springer.

EarInfection *Ear infections in Sydney*

Description

Ear infections for swimmers at a Sydney beach.

Usage

```
data(EarInfection)
```

Format

A data frame with 287 rows and 6 columns:

Swimmer The type of swimmer; one of `Occasional` or `Frequent`

Location The usul swimming location; one of `Non-beach` or `Beach`

Age The age group; one of `15 to 19`, `20 to 24`, or `25 to 29`

Sex The sex of the person; one of `Male` or `Female`

NumInfections The number of self-reported ear infections

Infections Whether the person had experienced an ear infection; one of `Yes` or `No`

Source

James K. Lindsey (1995). This data file was downloaded from OzDASL (http://www.statsci.org/data/oz/earinf.html) where it was prepared by Dr Gordon Smyth from Hand et al (1994) Dataset 328.

References

D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski (1994) A Handbook of Small Data Sets, London: Chapman and Hall. Dataset 328.

EDpatients *ED patients and welfare*

Description

Welfare distribution and emergency department (ED) patients.

Usage

```
data(EDpatients)
```

Format

A data frame with 30 rows (one per person) and 2 columns:

Days The number of days after welfare distribution

ED The mean number of emergency department (ED) patients

Source

Data read from the scatterplot in Brunette, Douglas D., John Kominsky, and Ernest Ruiz. 1991. Correlation of Emergency Health Care Use, 911 Volume, and Jail Activity with Welfare Check Distribution. *Annals of Emergency Medicine* **20** (7): 739–42.

`Elephants` *Elephant measurements*

Description

Physical measurements of elephants.

Usage

```
data(Elephants)
```

Format

A data frame with 1470 rows and 5 columns:

Sex Sex of the elephant; one of `A` or `B` (anonymised)

Age Age of elephant, in completed years

Chest Chest girth, in cm

Height Height to shoulder, in cm

Mass Body mass, in kg

Source

Lalande, Lucas; Lummaa, Virpi; Aung, Htoo Htoo; Htut, Win; Nyein, U. Kyaw; Berger, Verane; Briga, Michael (2022). Sex-specific body mass aging trajectories in adult Asian elephants. *Dryad*. https://doi.org/10.5061/dryad.5dv41ns59

References

Lalande, L. D., Lummaa, V., Aung, H. H., Htut, W., Nyein, U. K., Berger, V., & Briga, M. (2022). Sex-specific body mass ageing trajectories in adult Asian elephants. *Journal of Evolutionary Biology*, **35**(5), 752–762.

EmeraldAug *Emerald rainfall in Augusts*

Description

The total monthly rainfall in Emerald, Australia, and the average monthly SOI.

Usage

```
data(EmeraldAug)
```

Format

A data frame with 114 rows (one per August over 114 years) and 4 columns:

Year The year

Rain The rainfall in August of the given year; in mm

SOI The monthly average Southern Oscillation Index (SOI)

Phase the SOI phase (see Stone and Auliciems, 1992); a factor with these values: 1 (consistently negative), 2 (consistently positive), 3 (rapidly falling), 4 (rapidly rising), or 5 (consistently near zero)

Source

Data obtained from the Australian Bureau of Meteorology (<http://www.bom.gov.au>) and iri/ldeo Climate Data Library (<http://www.longpaddock.qld.gov.au/seasonalclimateoutlook/southernoscillationindex/soidatafiles/in on 21 December 2010, then compiled. The values of the SOI used here is that used by *LongPaddock*, which is slightly different than that used by the BoM (based on a different period of standardisation), because the SOI Phases are computed from these SOI values.

R. C. Stone and A. Auliciems (1992). SOI phase relationships with rainfall in eastern Australia, *International Journal of Climatology*, **12**, 625–636.

References

Dunn, Peter K., and Gordon K. Smyth. 2018. *Generalized Linear Models with Examples in R*. Springer.

EVpurchase *EV purchasing*

Description

Details of people regarding the purchase of an EV.

Usage

```
data(EVpurchase)
```

Format

A data frame with 4 rows (corresponding to the 4 cells in a 2 times 2 table) and three columns:

Education The level of education; one of 1 ('no post-graduate study') or 2 (post-graduate study')

PurchaseEV Whether respondent would purchase an electric vehicle in the next 10 years'; one of 1 (Yes) or 2 (No)

Counts The number of respondents in the given cell

Source

Egbue, Ona and Long, Suzanna (2012). Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. *Energy Policy*, **48**, 717–729.

Ferritin *Ferritin changes*

Description

Ferritin concentration changes.

Usage

```
data(Ferritin)
```

Format

A data frame with 20 rows (one per patient) and 3 columns:

September The patients' ferritin content (in micrograms/L) in September

March The patients' ferritin content (in micrograms/L) in March

Reduction The reduction in the patients' ferritin content (in micrograms/L) between September and the following March, during which time they had treatment

Source

Cressie, N. A. C., L. J. Sheffield, and H. J.Whitford. 1984. Use of the One Sample t-Test in the Real World. *Journal of Chronic Diseases* **37** (2): 107–14.

`Flowering` *Flowering shrubs*

Description

First-flowering dates for two shrubs.

Usage

```
data(Flowering)
```

Format

A data frame with 25 rows (one per person) and 4 columns:

Willow The (Julian) date on which flowering began for the encroaching Salix (willows)

Skypilot The (Julian) date on which flowering began for the native Polemonium viscosum (alpine skypilot)

MinTemp The minimum June temperature (in degrees C)

Altitude The altitude (in m)

Source

Kettenbach, Jessica A.; Miller-Struttmann, Nicole; Moffett, Zoë; Galen, Candace (2018). Data from: How shrub encroachment under climate change could threaten pollination services for alpine wildflowers: a case study using the alpine skypilot, Polemonium viscosum [Dataset]. Dryad. https://doi.org/10.5061/dryad.2p2bh

---

`Fluoro` *Fluoroscopic scanning*

Description

The data give the total procedure time during CT fluoroscopic scanning, and the radiation dose received.

Usage

```
data(Fluoro)
```

Format

A data frame with 19 rows and 2 columns:

Time The total procedure time, in minutes

Dose The total radiation dose received, in rads

Source

Kelly H. Zou, Kemal Tuncali, and Stuart G. Silverman (2003). Correlation and simple linear regression. *Radiology*, **227**, 617–628.

References

The data were originally used, but not given, in: S. G. Silverman, K. Tuncali, D. F. Adams, R. D. Nawfel, K. H. Zou, and P. F. Judy (1999). CT fluoroscopy-guided abdominal interventions: techniques, results, and radiation exposure. *Radiology*, **212**, 673–681.

ForwardFall *Forward-falling women*

Description

The forward-leaning angle before women fall over.

Usage

```
data(ForwardFall)
```

Format

A data frame with 15 rows (one per patient) and 2 columns:

LeanAngle The angle at which patients could lean forward and still recover

Group The age group; `1` means 'younger women' and `2` mean 'older women'

Source

Wojcik, Laura A., Darryl G. Thelen, Albert B. Schultz, James A. Ashton-Miller, and Neil B. Alexander. 1999. Age and Gender Differences in Single-Step Recovery from a Forward Fall. *Journal of Gerentology* **54A** (1): M44–50.

FriesWt *McDonald's fries*

Description

The weights of McDonald's large fries.

Usage

```
data(FriesWt)
```

Format

A data frame with 32 observations. The data give the weights of large fries bought from a McDonald (target: 171g).

FriesWt The weight of 32 large French fry order at McDonalds, in grams

Source

The data were extracted by reading Figure 2 in: Wetzel, Nathan (2005). "McDonald's french fries: Would you like small or large fries?" *STATS*, **43**, 12–14.

Fruit *Fruit statistics from farms*

Description

Details of fruit from different farms.

Usage

```
data(Fruit)
```

Format

A data frame with 37 rows (one per person) and 11 columns:

Farm The farm identifier

Flowers2014 The number of flowers in 2014

Flowers2015 The number of flowers in 2015

Fruit2014 The total number of fruits formed in 2014

Fruit2015 The total number of fruits formed in 2015

FLength2014 The fruit length (in cm) in 2014

FLength2015 The fruit length (in cm) in 2015

FBreadth2014 The fruit breadth (in cm) in 2014

FBreadth2015 The fruit breadth (in cm) in 2015

FWeight2014 The fruit weight (in g) in 2014

FWeight2015 The fruit weight (in g) in 2015

Source

Ronita Mukherjee, Rittik Deb and Soubadra Devy (2020). Diversity matters: effects of density compensation in pollination service during rainfall shift [Dataset]. Dryad. https://doi.org/10.5061/dryad.0n5v168

References

Mukherjee, Ronita; Deb, Rittik; Devy, Soubadra (2020). Diversity matters: Effects of density compensation in pollination service during rainfall shift *Ecology and Evolution*, **9**(17), 9701–9711.

`Gorillas` *Chest-beating rates in gorillas*

Description

Chest-beating rates in Gorillas.

Usage

```
data(Gorillas)
```

Format

A data frame with 25 rows (one per gorilla) and 7 columns:

Male An identifier

NoChestBeats The number of chest beats

FocalTime The focal time in hours (i.e., time spent watching gorilla)

ChestBeatRate The rate of chest beating, in beats per 10 hours

BackBreadth The breadth of the gorilla's back, in cm

Age Mean age during the study period, in years

Age20 Whether the gorillas is aged under 20 or not; one of `Younger` or `Older`

Source

Wright, Edward, Sven Grawunder, Eric Ndayishimiye, Jordi Galbany, Shannon C.McFarlin, Tara S. Stoinski, and Martha M. Robbins. 2021. Chest Beats as an Honest Signal of Body Size in Male Mountain Gorillas (Gorilla Beringei Beringei). *Scientific Reports* **11** (1): 6879.

HatSunglasses *Wearing hats and sunglasses*

Description

The wearing of hats and sunglasses in Brisbane.

Usage

```
data(HatSunglasses)
```

Format

A data frame with 16 rows (each combination of the other variables) and 5 columns:

Gender Gender of person; one of `Male` or `Female`

Hat Whether the person was wearing a hat; one of `Yes` or `No`

Sunglasses Whether the person was wearing sunglasses; one of `Yes` or `No`

Phone Whether the person had easy access to their phone; one of `Easy` or `Not easy`

Count The number if people meeting the given combination

Source

Dexter, Ben, Rachel King, Simone L. Harrison, Alfio V. Parisi, and Nathan J. Downs. 2019. A Pilot Observational Study of Environmental Summertime Health Risk Behavior in Central Brisbane, Queensland: Opportunities to Raise Sun Protection Awareness in Australia's Sunshine State. *Photochemistry and Photobiology* **95** (2): 650–55

HCrabs *Hermit crabs*

Description

The number of male crabs attached to female horseshoe crabs

Usage

```
data(HCrabs)
```

Format

A data frame with 173 rows (each crab) and 5 columns:

Col The female's carapace colour; one of `LM` (light medium), `M` (medium), `DM` (dark medium) or `D` (dark)

Spine The female's spine condition; one of `BothOK`, `OneOK` or `NoneOK`

Width The female's carapace width, in cm

Wt The weight of the female, in grams

Sat The number of male crabs attached ('satellites')

Source

H. J. Brockmann (1996) Satellite male groups in horseshoe crabs, *Limulus polyphemus*. *Ethology*, **102**(1), 1–21.

IgE *IgE concentrations*

Description

IgE concentration before and after intervention.

Usage

```
data(IgE)
```

Format

A data frame with 11 rows (one per child) and 3 columns:

Before IgE (before intervention), in micrograms/L

After IgE (after intervention), in micrograms/L

Reduction The reduction in IgE, in micrograms/L

Source

Lothian, James B. and Grey, Vijaylaxmi and Lands, Larry C. (2006). "Effect of whey protein to modulate immune response in children with atopic asthma", *International Journal of Food Science and Nutrition*, **57** (3/4), 204–211.

Insulation *Insulation and energy*

Description

Energy consumption before and after adding insulation.

Usage

```
data(Insulation)
```

Format

A data frame with 10 rows (each house) and 2 columns:

Before Energy consumption before adding insulation, in MWh

After Energy consumption after adding insulation, in MWh

Source

D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski (1994) A Handbook of Small Data Sets, London: Chapman and Hall. Dataset 86.

References

Originally from: The Open University. 1983. *MDST242 Statistics in Society, Unit A0: Introduction.* The Open University.

Jeans *Jeans' pockets*

Description

Measurements of pockets in men's and women's jeans.

Usage

```
data(Jeans)
```

Format

A data frame with 80 rows (each pair of jeans) and 14 columns:

Brand The brand of jeans; 22 brands are represented

Style The style of jeans; one of `boot-cut`, `regular`, `skinny`, `slim` or `straight`

Sex Whether the jeans are men's or women's jeans; one of `men` or `women`

Price The price, in US dollars

MaxHeightFront The height (in cm) of the front pocket from the top of the highest rivet to the lowest point of the pocket (along the left-hand side or zipper side)

MinHeightFront The height (in cm) of the front pocket from the top of the highest rivet to the lowest point of the pocket (along the right-hand side or non-zipper side)

MaxWidthFront The width (in cm) from the widest point of the front pocket

MinWidthFront The width (in cm) from the highest rivet to the right or non-zipper side of the pocket

MaxHeightBack The height (in cm) from the deepest point of the back pocket (usually in the pocket's center) to the top of the pocket

MinHeightBack The height (in cm) from the shallowest point of the back pocket to the top of the pocket

MaxWidthBack The width of the pocket at the very top (opening)

MinWidthBack The width of the pocket at its narrowest (just before the pocket tapers to a point)

Area The area of the pocket, from the pocket's measurements (in square cm)

Style2 The style, where `skinny` now means `Style == "skinny" | "slim"` and where `straight` means `Style == "straight" | "boot-cut"`

Note

The `githib` source contains a diagram explaining the pocket measurements more clearly. All jeans that were measured have a 32-inch waistband, as indicated by the brand.

Source

https://github.com/the-pudding/data/tree/master/pockets (used with permission).

References

Diehm, Jan & Thomas, Amber (August 2018). Women's pockets are inferior. *The Pudding*.

Jellyfish *Length and width of jellyfish*

Description

Width and length of jellyfish at two locations.

Usage

```
data(Jellyfish)
```

Format

A data frame with 46 rows (one per jellyfish) and 3 columns:

Location the location of the jellyfish; one of `Dangar` (Dangar Island) or `Salamander` (Salamander Bay)

Width the width (breadth) of the jellyfish, in mm

Length the length of the jellyfish, in mm

Source

D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski (1994) A Handbook of Small Data Sets, London: Chapman and Hall. Dataset 72.

References

Lunn, A. D. and McNeil, D. R. (1991). *Computer-Interactive Data Analysis*, Chichester: John Wiley and Sons.

Jumping *Jumping and footwear*

Description

Double-leg jumping distance, wearing shoes and barefoot.

Usage

```
data(Jumping)
```

Format

A data frame with 80 rows (each person) and 2 columns:

Shoes The jumping distance, while wearing shoes, in cm

Barefoot The jumping distance, while barefoot, in cm

Source

Hébert-Losier, K., Boswell-Smith, C., & Hanzlíková, I. (2023). Effect of Footwear Versus Barefoot on Double-Leg Jump-Landing and Jump Height Measures: A Randomized Cross-Over Study. *International Journal of Sports Physical Therapy*, **18**(4), 845.

KStones *Kidney stone treatments*

Description

Treatment of kidneys stones, and the result.

Usage

```
data(KStones)
```

Format

A data frame with 8 rows (each variable combination) and 4 columns:

Counts The number of people with the combination of the other variables

Size One of `Small` or `Large`, the kidney stone size

Method The method used; one of `Method A` or `Method B`

Result The result of the procedure; one of `Success` or `Failure`

Source

Charig, C. R.,D. R. Webb, S. R.Payne, and J. E. A. Wickham. 1986. Comparison of Treatment of Renal Calculi by Open Surgery, Percutaneous Nephrolithotomy, and Extracorporeal Shockwave Lithotripsy. *British Medical Journal* **292**: 879–82.

LHconc *Accuracy of scientific instruments*

Description

Measurements of LH concentrations at different concentrations, for two instruments.

Usage

```
data(LHconc)
```

Format

A data frame with 36 rows and 4 columns:

High1 Instrument 1 measurement of luteotropichormone (LH) concentrations at a high level, in mIU/ml

Mid1 Instrument 1 measurement of LH concentrations at a middle level, in mIU/ml

High2 Instrument 2 measurement of LH concentrations at a high level, in mIU/ml

Mid2 Instrument 2 measurement of LH concentrations at a middle level, in mIU/ml

Note

The known values are, respectively, 64.31, 19.24, 64.97 and 19.40 mIU/ml.

Source

Feng, Yang-chun and Huang, Yan-chun and Ma, Xiu-min. 2017. The application of Student's t-test in internal quality control of clinical laboratory. *Frontiers in Laboratory Medicine* **1** (3): 125–128.

Lime *Lime tree foliage*

Description

The foliage biomass of small-leaved lime trees of different origins.

Usage

```
data(Lime)
```

Format

A data frame with 385 rows (each tree) and 4 columns:

Foliage The oven-dried foliage biomass, in kg

DBH The diameter at breast height, in cm

Age The age of the tree, in years

Origin The origin of the tree; one of `Coppice`, `Natural` or `Planted`

Source

Schepaschenko, Dmitry; Shvidenko, Anatoly; Usoltsev, Vladimir A; Lakyda, Petro; Luo, Yunjian; Vasylyshyn, Roman; Lakyda, Ivan; Myklush, Yuriy; See, Linda; McCallum, Ian; Fritz, Steffen; Kraxner, Florian; Obersteiner, Michael (2017). Biomass tree data base. doi:10.1594/PANGAEA.871491

In supplement to: Schepaschenko, D et al. (2017): A dataset of forest biomass structure for Eurasia. *Scientific Data*, 4, 170070, doi:10.1038/sdata.2017.70

Extracted from <https://doi.pangaea.de/10.1594/PANGAEA.871491>

References

The source (Schepaschenko et al.) obtains the data from various sources, which are given there.

`LungCap` *Lung capacity*

Description

The lung capacity of children.

Usage

```
data(LungCap)
```

Format

A data frame with 654 rows (each child) and 5 columns:

Age The age of the child, in years

FEV The forced expiratory volume, in litres

Ht The height, in inches

Gender The gender of the child; one of `F` or `M`

Smoke Whether the child is a smoker; one of `0` (non-smoker) or `1` (smoker)

Source

Kahn, M. (2003) Data Sleuth, *STATS*, **37**, 24.

Ira B. Tager, Scott T. Weiss, Alvaro Munoz, Bernard Rosner, and Frank E. Speizer (1983). Longitudinal study of the effects of maternal smoking on pulmonary function in children. *New England Journal of Medicine*, **309**(12):699–703.

References

Kahn, Michael (2005). An exhalent problem for teaching statistics. *The Journal of Statistical Education*, **13**(2). Available on-line.

Mandible *Mandible lengths*

Description

The mandible length and gestational age for 167 foetuses from the 12th week of gestation onwards

Usage

```
data(Mandible)
```

Format

A data frame with 167 rows (each foetus) and 2 columns:

Age The foetus age, in weeks

Length The foetus length, in mm

Source

Patrick Royston and Douglas G. Altman (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Applied Statistics*, **43**(3), 429–467.

MaryRiver *Mary River stream flow*

Description

The mean daily stream flow from the Mary River.

Usage

```
data(MaryRiver)
```

Format

A data frame with 21,659 rows and 3 columns:

Month The month (where 1 means January, etc.

Year The year

Mean The mean stream flow recording for given date, in ML

Source

Originally sourced from: <http://watermonitoring.dnrm.qld.gov.au/cgi/webhyd.pl?rsdf_org=138110A&cat=rs&lvl=1&0>, but the actual website address keeps changing...

Last time I checked it was: <https://water-monitoring.information.qld.gov.au>; then select "Streamflow data", "Mary Basin" and "Mary River at Bellbird Creek" (i.e., station 138110A).

Mumps *Mumps and isolating*

Description

Whether students complied with isolation orders duration a mumps outbreak.

Usage

```
data(Mumps)
```

Format

A data frame with 8 rows and 3 columns:

AgeGroup One of 1 (18 to 19), 2 (20 to 21) or 3 (Older than 22)

Compliance One of 1 (complied with isolation order) or 2 (did not comply

Counts The number of students in each cell

Details

The data provide the number of students complying and not complying with an isolation order during a mumps outbreak in Kansas in 2006.

Source

Soud, F. A., M. M. Cortese, A. T. Curns, P. J. Edelson, R. H. Bitsko, H. T. Jordan, A. S. Huang, J. M.Villalon-Gomez, and G. H. Dayan. (2009). "Isolation Compliance Among University Students During a Mumps Outbreak, Kansas 2006". *Epidemiology & Infection*, **137**(1): 30–37.

NMiner *Noisy miner (birds)*

Description

The number of noisy miners detected in various 2 hectare transects in buloke woodland patches within the Wimmera Plains of western Victoria, Australia

Usage

```
data(NMiner)
```

Format

A data frame with 31 rows (each transect) and 2 columns:

Eucs The number of eucalypt trees in the transect

Minerab The number of noisy miners ('abundance') in three 20 minute surveys in each transect

Source

Personal communication from Martine Maron.

References

Martine Maron (2007). Threshold effect of eucalypt density on an aggressive avian competitor. *Biological Conservation*, **136**, 100–107.

`Orthoses` *Orthoses for children*

Description

Details of children fitted with orthoses.

Usage

```
data(Orthoses)
```

Format

A data frame with 15 rows and 5 columns:

Gender The gender of the child; one of `M` (male) or `F` (female)

Age The age of the child, in years

Height The height of the child, in cm

Weight The weight of the child, in kg

GMFCS The value of the ordinal Gross Motor Function Classification System describing the impact of cerebral palsy on their motor function; lower levels mean better functionality; one of `1` or `2`

Source

Swinnen, Eva, Jean-Pierre Baeyens, Benjamin Van Mulders, Julian Verspecht, and Marc Degelaen (2017). "The Influence of the Use of Ankle-Foot Orthoses on Thorax, Spine, and Pelvis Kinematics During Walking in Children with Cerebral Palsy". *Prosthetics and Orthotics International*. **42**(2), 208–213.

OSA *Obstructive sleep apnea (OSA)*

Description

Sleeping information for adults with Down Syndrome.

Usage

```
data(OSA)
```

Format

A data frame with 60 rows (each patient) and 7 columns:

ID An identifier

Age The age of the patient, in years

Gender The gender of the patient; one of `1` (male) or `2` (female)

BMI The BMI of the patient

Neck The neck circumference of the patient, in cm

REI The Respiratory Event Index for the patient

SAOS The SAOS score; one of `Severe`, `Moderate` or `Low`

Source

de Carvalho, Anderson Albuquerque, Fabio Ferreira Amorim, Levy Aniceto Santana, Karlo Jozefo Quadros de Almeida, Alfredo Nicodemos Cruz Santana, and Francisco de Assis Rocha Neves. 2020. STOP-Bang Questionnaire Should Be Used in All Adults with Down Syndrome to Screen for Moderate to Severe Obstructive Sleep Apnea. *PloS ONE* **15** (5): e0232596.

The data are given at: <https://figshare.com/articles/dataset/Raw_database_and_statistical_analysis_results-STOP-Bang_questionnaire_should_be_used_in_all_adults_with_Down_Syndrome_to_screen_for_moderate_to_severe_obs

PainRelief *Pain relief for mothers*

Description

Pain relief for birthing mothers.

Usage

```
data(PainRelief)
```

Format

A data frame with 912 rows (228 mothers with four rows (`Time`) for each) and 8 columns:

ID The patient ID; a digit from `1` to `228`

Time The time point of the measurement; one of `1` (0 minutes), `2` (after 20 mins), `3` (after 40 mins) or `4` (after 60 mins)

Score Pain score

Group The type of pain-relief used; one of `palacetamol` or `coldpack`

Age The age of the mother, in years

Parity Which number child is this (e.g., 1 means this is the mother's first child)

ChildSex The sex of the baby; one of `female` or `male`

Birthweight The birthweight of the baby, in kg, to the nearest 0.5kg

Source

Augustino, J., Moshi, F., Joho, A., & Mageda, J. F. K. (2023). Dataset comparing the effectiveness of perineal cold pack application over oral paracetamol 1000mg on postpartum perineal pain among women after spontaneous vaginal delivery in Dodoma region. "Data in Brief", 109766.

`Peas` *Pea nutrition*

Description

Nutritional content of peas.

Usage

```
data(Peas)
```

Format

A data frame with 96 rows (each seed) and 11 columns:

Origin The seed origin; a vector of strings listing locations

P The phosphorus content, in mg/g

K The potassium content, in mg/g

Ca The calcium content, in mg/g

Mg The magnesium content, in mg/g

S The sulphur content, in mg/g

Zn The zinc content, in mg/g

Fe The iron content, in mg/g

Cu The copper content, in mg/g

B The boron content, in mg/g

Mn The manganese content, in mg/g

Source

Hacisalihoglu, Gokhan, Nicole S. Beisel, and A.Mark Settles. 2021. Characterization of Pea Seed Nutritional Value Within a Diverse Population of Pisum Sativum. *PLoS One* **16** (11): e0259565.

`Perm` *Permeability of building materials*

Description

The permeability of building materials.

Usage

```
data(Perm)
```

Format

A data frame with 81 rows (each sample) and 3 columns:

Day The day of the data collection; `1` to `9`

Mach The machine; one of `A`, `B` or `C`

Perm The permeability of the sample, in seconds

Source

Bent Joergensen (1992) Exponential dispersion models and extensions: A review. *International Statistical Review*, **60**(1), 5–20

References

A. Hald (1952). *Statistical theory with engineering applications*. New York: Wiley.

`PetBirds` *Pet birds*

Description

Lung cancer and owning pet birds.

Usage

```
data(PetBirds)
```

Format

A data frame with 4 rows (each combination) and 3 columns:

LC Whether the adult had lung cancer; one of `Adults with lung cancer` or `Adults without lung cancer`

Pets Whether the adult kept pet birds; one of `Kept pet birds` or `Did not keep pet birds`

Counts The number of adults with the given combination

Source

Kohlmeier, L., G. Arminger, S. Bartolomeycik, B. Bellach, J. Rehm, and M. Thamm. 1992. Pet birds as an independent risk factor for lung cancer: case-control study. *British Medical Journal* **305** (6860): 986–89.

`PizzaSize` *Diameters of pizzas*

Description

The diameter of 12-inch pizzas from two companies.

Usage

```
data(PizzaSize)
```

Format

A data frame with 250 rows (one per pizza) and 5 columns:

Store the pizza chain; one of `Dominos` (Domino's Pizza) or `EagleBoys` (Eagle Boy's Pizza)

CrustDescription the type of crust for the pizza; one of `ClassicCrust`, `DeepPan`, `MidCrust`, `ThinCrust` or `ThinNCrispy` (some unique to one pizza company)

Topping the type of pizza topping; one of `BBQMeatlovers`, `Hawaiian`, `SuperSupremo` or `Supreme` (some unique to one pizza company)

Diameter the pizza diameter, in cm)

DiameterInches the pizza diameter, in inches (converted from cm)

Source

P. K. Dunn. Assessing claims made by a pizza chain. *Journal of Statistical Education*, **20**(1), 2012.

Placebos *Placebos and pain relief*

Description

Pain relief from analgesics and placebos.

Usage

```
data(Placebos)
```

Format

A data frame with 7 rows (each time point) and 6 columns:

Time The time after taking the treatment, in hours

Placebo The mean pain relief score for 22 patients given placebos

Distr The mean pain relief score for 22 patients given distalgesics

Asp The mean pain relief score for 22 patients given aspirin

Codis The mean pain relief score for 22 patients given codis

PlaceboRed The mean pain relief score for 22 patients given red placebos

Source

Read from Figures 3 and 4 of Huskisson, E. C. 1974. Simple Analgesics for Arthritis. *British Medical Journal* **4**: 196–200.

Possums *Possum weights*

Description

Sex and weight of possums at various elevations.

Usage

```
data(Possums)
```

Format

A data frame with 135 rows (each possum) and 3 columns:

Sex The sex of the possum; one of `Female` or `Male`

Wgt The weight of the possum, in g

DEM The elevation, in m, where the possum is found

Source

Williams, Jessica L., Dan Harley, Darcy Watchorn, Lachlan McBurney, and David B. Lindenmayer. 2022. Relationship Between Body Weight and Elevation in Leadbeater's Possum (Gymnobelideus Leadbeateri). *Australian Journal of Zoology* **69** (5): 167–74

`PremierL` *Premier league results*

Description

Premier League football (soccer) results from 2019 to 2020.

Usage

```
data(PremierL)
```

Format

A data frame with 208 rows (games) and 6 columns:

Date The data of the game

HomeTeam The name of the home team; for example `Liverpool` or `Man United`

AwayTeam The name of the away team; for example `Wolves` or `West Ham`

HomeGoals The number of goals scored by the home team

AwayGoals The number of goals scored by the away team

Result The result, one of `H` for the home team, `A` for the away team, or `D` for a draw

Source

The website https://sports-statistics.com/sports-data/soccer-datasets/

`QSchools` *Queensland school children*

Description

The number of four-year-old students enrolled at school in Queensland (Australia), classified by sex, school type and whether the students are First Nations students (in 2019).

Usage

```
data(QSchools)
```

Format

A data frame with 8 rows and 4 columns:

Sex Sex of the student; one of `F` (female) or `M` (male)

FNations The first-nations status; one of `Yes` (First Nations students) or `No` (non-First Nations students)

School The school type; one of `Government` or `Non-government`

Counts The number of four-year-old students meeting the designated criteria

Source

Collated by Peter K. Dunn, obtained from data at the *Australian Bureau of Statistics*, web page (https://www.abs.gov.au) in 2023.

References

Peter K. Dunn. Generalized linear models. In R. J. Tierney, F. Rizvi, and K. Erkican, editors, *International Encyclopedia of Education*, pages 583–589. Elsevier, 2023.

`ReactionTime` *Reaction times when driving*

Description

Reaction times when driving, when using and not using a mobile phone.

Usage

```
data(ReactionTime)
```

Format

A data frame with 64 rows (each student) and 2 columns:

Reaction The reaction time, in milliseconds

Group Which group the student was in; one of `Phone` or `Control`

Source

Reported by: Agresti, Alan, and Christine A. Franklin. 2007. *Statistics: The Art and Science of Learning from Data*.

Agresti & Franklin claim the data comes from: Strayer, David L., and William A. Johnston. 2001. Driven to Distraction: Dual-Task Studies of Simulated Driving and Conversing on a Cellular Telephone. *Psychological Science* **12** (6):462–66

RedDeer *Molar weights of red deer*

Description

The age and weight of molars in male red deer.

Usage

```
data(RedDeer)
```

Format

A data frame with 78 rows (each deer) and 2 columns:

Age The age of the deer, in years

Weight The weight of the first molar tooth, in g

Source

D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski (1994) A Handbook of Small Data Sets, London: Chapman and Hall. Dataset 170.

References

The data originally come from: Holgate, P. 1965. Fitting a Straight Line to Data from a Truncated Population. *Biometrics* **21**(3): 715–20

Removal *Biofiltration removal efficiency*

Description

The removal efficiency in biofiltration.

Usage

```
data(Removal)
```

Format

A data frame with 32 rows (each experiment) and 2 columns:

Removal The removal efficiency, in percent

Temp The inlet temperature, in degrees C

Source

Exercise 12.109 in Devore, Jay L., and Kenneth N. Berk. 2007. *Modern Mathematical Statistics with Applications*. Thomson Higher Education

References

The data originally come from: Chitwood, Derek E., and Joseph S. Devinny. 2001. Treatment of Mixed Hydrogen Sulfide and Organic Vapors in a Rock Medium Biofilter. *Water Environment Research* **73** (4): 426–35.

RipsID *Rip identification*

Description

Whether people of given age groups can correctly identify ocean rips.

Usage

```
data(RipsID)
```

Format

A data frame with 8 rows and 3 columns:

AgeGroup The age group of the person; one of 1 (18 to 24), 2 (25 to 34), 3 (25 to 50) or 4 (51 to 65)

Identification Whether the person correctly identified a rip from a picture; one of 1 (correctly) or 2 (incorrectly)

Counts The number of people in each cell

Details

The data provide the number of people correctly identifying a rip from a photo, by age group.

Source

Diez-Fern\'andez, P., Ruibal-Lista, B., Lobato-Alejano, F., & L\'opez-Garc\'ia, S. (2023). 'Rip current knowledge: do people really know its danger? Do lifeguards know more than the general public?'. *Heliyon*, **9**(7).

Running *Running data*

Description

The reliability vertical oscillation measurements in wearable devices for running.

Usage

```
data(Running)
```

Format

A data frame with 150 rows (15 participants by 10 reps each) and 8 columns:

ID The participant ID

Trial Which trial; one of 1 to 5

Speed The average running speed, in km.h

HRM The vertical oscillation (VO) as measured by the Garmin Heart Rate Monitor-Pro (HRM), in cm

NOVA The VO as measured by the the INCUS NOVA device, in cm

RDP The VO as measured by the Garmin Running Dynamics Pod (RDP), in cm

Footpod The VO as measured by the Stryd Running Power Meter Footpod (Footpod), in cm

Video The VO as measured by video analysis, in cm

Source

From Tables 1 and 5 of:

Smith, Craig P. and Fullerton, Elliott and Walton, Liam and Funnell, Emelia and Pantazis, Dimitrios and Lugo, Heinz (2022). The validity and reliability of wearable devices for the measurement of vertical oscillation for running. *Plos One*, 17 (11), p. e0277810.

Sanddollars *Sand dollars*

Description

Details about reproduction of sand dollars

Usage

```
data(Sanddollars)
```

Format

A data frame with 36 rows (each experiments) and 4 columns:

SD.temperatures The temperature, in degrees C, where the sand dollar is located

SD.fertilization Sand dollar fertilization rates, in percent

SD.speeds Sperm swimming velocities, in micrometres per second

SD.motility Sperm motility

Source

Leuchtenberger, Sara Grace, Maris Daleo, Peter Gullickson, Andi Delgado, Carly Lo, and Michael T. Nishizaki. 2022. The Effects of Temperature and pH on the Reproductive Ecology of Sand Dollars and Sea Urchins: Impacts on Sperm Swimming and Fertilization. *PLoS One* **17** (12): e0276134

The data are available directly from: Nishizaki, Michael T., Sara Grace Leuchtenberger, Maris Daleo, Peter Gullickson, Andi Delgado, and Carly Lo. 2022. Echinoderm Sperm Swimming and Fertilization. *Dryad*. <https://doi.org/10.5061/dryad.jwstqjqbz>

ScarHeight *Scar heights*

Description

Scar heights for men and women.

Usage

```
data(ScarHeight)
```

Format

A data frame with 4 rows (each combination) and 3 columns:

Counts The number of people with the given combination

Gender The gender of the person; one of `Women` or `Men`

ScarHt The scar height; one of `0mm` (i.e., smooth) or `1mm` (i.e., 0mm to 1mm)

Source

Wallace, Hilary J., Mark W. Fear, Margaret M. Crowe, Lisa J. Martin, and Fiona M. Wood.2017. Identification of Factors Predicting Scar Outcome After Burn in Adults: A Prospective Case-Control Study. *Burns* **43**: 1271–83

SDrink *Soft drink delivery*

Description

The time taken to deliver soft drinks to vending machines.

Usage

```
data(SDrink)
```

Format

A data frame with 25 rows (each delivery) and 3 columns:

Time The time taken to service the vending machine, in minutes

Cases The number of cases of soft drink stocked

Distance The distance walked by the driver to service the vending machine, in feet

Source

The data were obtained electronically from OzDASL <http://www.statsci.org/data/>.

References

D. C. Montgomery and E. A. Peck (1992). *Introduction to Regression Analysis*. Wiley, New York. Example 4.1

ShoppingBags *Shopping bags*

Description

Age of people, and whether they bring their own shopping bags.

Usage

```
data(ShoppingBags)
```

Format

A data frame with 6 rows and 3 columns:

AgeGroup The age group: `1` means '30 and under'; `2` means '31 to 40'; `3` means 'Over 40'

BringBags Whether people bring their own shopping bags or not; `y` means they do; `n` means they do not

Counts The number of people in each designated category

Source

From Tables 1 and 5 of: Choon, S. W., Tan, S. H., & Chong, L. L. (2017). The perception of households about solid waste management issues in Malaysia. *Environment, Development and Sustainability*, **19**, 1685–1700.

`SixMWT` *Six-minute walk time tests*

Description

Six-minute walk time data for two different walkway lengths.

Usage

```
data(SixMWT)
```

Format

A data frame with 50 rows (one per subject) and 3 columns:

Dist20 The 6MWT distance in a 20m corridor, in m

Dist30 The 6MWT distance in a 30m corridor, in m

Age The age of the subject, in completed years

Source

Saiphoklang, N., Pugongchai, A., & Leelasittikul, K. (2022). Comparison between 20 and 30 meters in walkway length affecting the 6-minute walk test in patients with chronic obstructive pulmonary disease: A randomized crossover study. *Plos One*, **17**(1), e0262238.

`Snakes` *Snakes*

Description

Measurements of snakes, some of which eat crayfish, and some of which do not.

Usage

```
data(Snakes)
```

Format

A data frame with 28 rows (each plot) and 4 columns:

Crayfish Whether the snake lives in a crayfish region or not; one of `Cfish` or `NoCfish`

Sex The snake sex; one of `male` or `female`

SVL The snout-to-length length, in cm

Teeth The number of number of maxillary teeth

Source

Javier Manjarrez, Constantino Macías Garcia, Hugh Drummond (2018). Data from: Morphological convergence in a Mexican garter snake associated with the ingestion of a novel prey [Dataset]. Dryad. https://doi.org/10.5061/dryad.mg152

References

Manjarrez, J., Macias Garcia, C., & Drummond, H. (2017). Morphological convergence in a Mexican garter snake associated with the ingestion of a novel prey. *Ecology and Evolution*, **7**(18), 7178–7186.

`SoilCN` *Soil carbon and nitrogen*

Description

Percentage of carbon and nitrogen in irrigated and non-irrigated plots.

Usage

```
data(SoilCN)
```

Format

A data frame with 28 rows (each plot) and 4 columns:

IrrigatedC The percentage carbon, in a paired irrigated plot

NonirrigatedC The percentage carbon, in a paired non-irrigated plot

IrrigatedN The percentage nitrogen, in a paired irrigated plot

NonirrigatedN The percentage nitrogen, in a paired non-irrigated plot

Source

Lambie, S. M., Mudge, P. L., & Stevenson, B. A. (2021). Microbial community composition and activity in paired irrigated and non-irrigated pastures in New Zealand. *Soil Research*, **60**(4), 337–348.

Soils *Soil properties*

Description

Properties of soil and the California Bearing Ratio.

Usage

```
data(Soils)
```

Format

A data frame with 16 rows (each sample) and 12 columns:

Sample An identifier

Gravel The percentage of gravel in the sample

Sand The percentage of sand in the sample

Clay The percentage of clay in the sample

PI Plasticity index (PI, a measure of the plasticity of the soil

CBR The California Bearing Ratio, a measure of flexibility, as a percentage

Source

Talukdar, Dilip Kumar. 2014. A Study of Correlation Between California Bearing Ratio (CBR) Value with Other Properties of Soil. *International Journal of Emerging Technology and Advanced Engineering* **4** (1): 559–62

Speed *Speed of vehicles*

Description

Speeds of vehicles before and after adding additional signage.

Usage

```
data(Speed)
```

Format

A data frame with 79 rows (each vehicle) and 2 columns:

When When the speed is measured; one of `Before` or `After` new signage added

Speed The measured speed, in km/h

Source

Ma, Yongfeng, Wenbo Zhang, Xin Gu, and Jiguang Zhao. 2019. Impacts of Experimental Advisory Exit Speed Sign on Traffic Speeds for Freeway Exit Ramp. *PLoS One* **14** (11):e0225203

Stress *Stress before surgery*

Description

Stress at two time-points before surgery.

Usage

```
data(Stress)
```

Format

A data frame with 19 rows and 2 columns:

BeforeHours beta-endorphin concentrations measured 12–14 hours before surgery, in fmol/ml

BeforeMins beta-endorphin concentrations measured 10 minutes before surgery, in fmol/ml

Source

D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski (1994) A Handbook of Small Data Sets, London: Chapman and Hall. Dataset 232.

References

The original source is given as Hoaglin, D. C., Mosteller, F. and Tukey. J. W. 1985. *Exploring data tables, trends and shapes*. New York: John Wiley & Sons.

StudentsEat *Students eating habits*

Description

Where students live and where they eat most of their meals.

Usage

```
data(StudentsEat)
```

Format

A data frame with 183 rows (each student) and 2 columns:

Meals Where the student eats most of their meals; one of `Most off-campus` or `Most on-campus`

Live Where the student lives; one of `Living with parents` or `Not living with parents`

Source

Mann, Linda, and Karen Blotnicky. 2017. Influences of Physical Environments on University Student Eating Behaviors. *International Journal of Health Sciences* **5** (2): 42–52

`StudentWt` *Students' weight changes*

Description

Weights of students from Week 1 to Week 12 of semester.

Usage

```
data(StudentWt)
```

Format

A data frame with 68 rows (each student) and 4 columns:

Student An identifier

WtWk1 The student's weight in Week 1, in kg

WtWk12 The student's weight in Week 12, in kg

GainWt The student's weight gain, in kg

Source

David. n.d. DASL: Data and Story Library. <https://dasl.datadescription.com/datafile/freshman-15/>

References

Levitsky, D. A., Halbmaier, C. A., & Mrdjenovic, G. (2004). The freshman weight gain: a model for the study of the epidemic of obesity. *International Journal of Obesity*, **28**(11), 1435–1442.

Tape *Kinesio tape use*

Description

The use of tapes to reduce pain.

Usage

```
data(Tape)
```

Format

A data frame with 16 individuals having 18 observations:

Age The age of the participant, in years

Sex The sex of the participant; one of 1 or 2, but what they refer to is unknown

Pre.Left.KT.NoTension The pressure pain threshold (PPT) in the left arm, using Kinesio tape (KT), applied without tension: The level of pressure where pain was felt, in kPa

Pre.Right.KT.NoTension The PPT, in the right arm, using KT, 5 mins before application of KT, applied without tension: The level of pressure where pain was felt, in kPa

Post1.Left.KT.NoTension The PPT, in the left arm, using KT, 5 mins after application of KT, applied without tension: The level of pressure where pain was felt, in kPa

Post1.Right.KT.NoTension The PPT, in the right arm, using KT, 5 mins after application of KT, applied without tension: The level of pressure where pain was felt, in kPa

Post2.Left.KT.NoTension The PPT, in the left arm, using KT, 15–20 mins after application of KT, applied without tension: The level of pressure where pain was felt, in kPa

Post2.Right.KT.NoTension The PPT, in the right arm, using KT, 15–20 mins after application of KT, applied without tension: The level of pressure where pain was felt, in kPa

Post1.Left.75KT.Tension The PPT, in the left arm, using KT, 5 mins after application of KT, applied with 75% tension: The level of pressure where pain was felt, in kPa

Post1.Right.75KT.Tension The PPT, in the right arm, using KT, 5 mins after application of KT, applied with 75% tension: The level of pressure where pain was felt, in kPa

Post2.Left.75KT.Tension The PPT, in the left arm, using KT, 15–20 mins after application of KT, applied with 75% tension: The level of pressure where pain was felt, in kPa

Post2.Right.75KT.Tension The PPT, in the right arm, using KT, 15–20 mins after application of KT, applied with 75% tension: The level of pressure where pain was felt, in kPa

Pre.Left.NoTape The PPT, in the left arm, using no tape: The level of pressure where pain was felt, in kPa

Pre.Right.NoTape The PPT, in the right arm, using no tape: The level of pressure where pain was felt, in kPa

Post1.Left.NoTape The PPT, in the left arm, using no tape, 10 minutes after first test: The level of pressure where pain was felt, in kPa

Post1.Right.NoTape The PPT, in the right arm, using no tape, 10 minutes after first test: The level of pressure where pain was felt, in kPa

Post2.Left.NoTape The PPT, in the left arm, using no tape, 20–35 minutes after first test: The level of pressure where pain was felt, in kPa

Post2.Right.NoTape The PPT, in the right arm, using no tape, 20–35 minutes after first test: The level of pressure where pain was felt, in kPa

Source

Naugle, K. E., Hackett, J., Aqeel, D., & Naugle, K. M. (2021). "Effect of different Kinesio tape tensions on experimentally-induced thermal and muscle pain in healthy adults." *PloS One*, **16**(11), e0259433.

`Throttle` *Throttles*

Description

Throttle and manifold air pressure.

Usage

```
data(Throttle)
```

Format

A data frame with 68 rows (each student) and 2 columns:

ThrottleAngle The throttle angle, in degrees

MAPvalue The manifold air pressure, as a fraction of the maximum value

Source

Amin, Arslan Ahmed, and Khalid Mahmood-ul-Hasan. 2019. Robust Active Fault-Tolerant Control for Internal Combustion Gas Engine for Air-Fuel Ratio Control with Statistical Regression-Based Observer Model. *Measurement and Control*, 0020294018823031

Turbines *Turbine fissures*

Description

Fissure cracks appearing in turbines.

Usage

```
data(Turbines)
```

Format

A data frame with 4 rows and 3 columns:

Hours The approximate number of hours run by these turbines

Turbines The number of turbines run for the indicated number of hours

Fissures The number of fissure cracks in the turbines

Details

The data provide the number of turbines, and those with fissure cracks, for an approximate given hours of run-time. A two-way table of the data as given in not appropriate; `Turbines` includes all turbines, including those given in `Fissures`.

Source

Raymond H. Myers, Douglas C. Montgomery, and G. Geoffrey Vining (2002). *Generalized linear models with applications in engineering and the sciences*, Wiley.

TurtleNests *Turtle nests*

Description

Infected and non-infected turtle nests, and whether the nests were relocated.

Usage

```
data(TurtleNests)
```

Format

A data frame with 4 rows and 3 columns:

Infected Whether the nest was infected with fungi or bacteria; one of `0` (not infected) or `1`

Nest Whether the nest was relocated; one of `0` (Natural (not relocated)) or `1` (relocated)

Counts The number of nests in the combination defined by `Infected` and `Nest`

Details

The data provide the number of nests from Mediterranean loggerhead turtles that had fungal or bacterial infections. Some nests are relocated due to the risk if tidal inundation; researchers were interested to see if the relocation was related to the probability of infection.

Source

Candan, Ahmet Yavuz, Katilmis, Yusuf and Ergin, Cagri (2021). "First report of *Fusarium* species occurrence in loggerhead sea turtle (*Caretta caretta*) nests and hatchling success in Iztuzu Beach, Turkey". *Biologia*, **76**, 565–573.

Typing *Typing speeds*

Description

Typing speeds and accuracy.

Usage

```
data(Typing)
```

Format

A data frame with 1301 rows (one for each student) and 5 columns:

Subject The subject number

mTS The mean typing speed (wpm) for each subject

mAcc The mean typing accuracy for each subject

Age The age, in completed years .

Sex The sex of the subject; one of `female` or `male`

Details

Typing speeds measured online for students.

Source

https://osf.io/v92fy/files/osfstorage?view_only=87885752038b4be190d532143fdedb07

References

Pinet, Svetlana, Christelle Zielinski, F.-Xavier Alario, and Marieke Longcamp. Typing Expertise in a Large Student Population. *Cognitive Research: Principles and Implications* **7**, no. 1 (August 5, 2022): 77. https://doi.org/10.1186/s41235-022-00424-3.

`WaterAccess` *Water access*

Description

Water access for households in West Cameroon.

Usage

```
data(WaterAccess)
```

Format

A data frame with 150 rows (15 participants by 10 reps each) and 12 columns:

Region The region; one of `Mbeng`, `Mbih` or `Ntsingbeu`

Age The age of the woman in the household, in years

Education The level of education of the woman; one of `Primary or less` or `Secondary or higher`

SourceDistance The distance to the water source; one of `Under 100m`, `100m to 1000m` or `Over 1000m`

SourceQueueTime The queuing time at the water source; one of `Under 5 min`, `5 to 15 min` or `Over 15 min`

HasGarden Whether the household has a farming garden; one of `Y` or `N`

HasLivestock Whether the household keeps livestock; one of `Y` or `N`

HouseholdPeople The number of people in the household

HouseholdUnder5s The number of people under 5 in the household

WaterSource The water source; one of `Tap`, `Bore`, `Well` or `River`

WashContainer How often the water container is washed; one of `Before each fill`, `Once per week` or `Once per month`

Diarrhea Whether a child has had diarrhoea in the last two weeks; one of `Y` or `N`

Source

Nounkeu, C. D., Metapi, Y. D., Ouabo, F. K., Kamguem, A. S. T., Nono, B., Azza, N., Leumeni, P., Nguefack-Tsague, G., Todem, D., Dharod, J. M., & Kuate, D. (2022). "Assessment of drinking water access and household water insecurity: A cross sectional study in three rural communities of the Menoua division, West Cameroon". *PLOS Water*, **1**(8), e0000029.

WCTennis *Wheelchair tennis*

Description

The push time for wheelchair tennis players, with and without holding a racquet.

Usage

```
data(WCTennis)
```

Format

A data frame with 13 rows (each player) and 3 columns:

Person The person

PTwith The push time, when holding a racquet; in seconds

PTwithout The push time, without holding a racquet; in seconds

Source

I. Alberca, 2016, Kinetic and temporal parameters calculated from raw data collected via wireless instrumented wheel for measuring 3D pushrim kinetics of a racing wheelchair, https://doi.org/10.17026/dans-xjf-bs8v, *DANS Data Station Life Sciences*, V1.

References

Alberca, I., Chénier, F., Astier, M., Watelain, E., Vallier, J. M., Pradon, D., & Faupin, A. (2022). Sprint performance and force application of tennis players during manual wheelchair propulsion with and without holding a tennis racket. *PLoS ONE*, **17**(2), e0263392.

Windmill *Windmill and current*

Description

The amount of direct current (DC) output from windmills for varying wind velocities.

Usage

```
data(Windmill)
```

Format

A data frame with 25 rows (each windmill) and 2 columns:

Wind The wind velocity, in miles per hour

DC The DC output

Source

G. Joglekar, J. H. Schuenemeyer and V. LaRicca (1989) Lack-of-fit testing when replicates are not available. *American Statistician*, **43**, 135–143.

References

D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski (1994). *A Handbook of Small Data Sets*, London: Chapman and Hall. Dataset 271.

D. C. Montgomery and E. A. Peck (1982). *Introduction to Linear Regression Analysis*. New York: John Wiley.

`YieldDen` *Yield of onions*

Description

The mean yields per plant for three onion varieties.

Usage

```
data(YieldDen)
```

Format

A data frame with 30 rows (each plants) and 3 columns:

Yield The yield per plant, in grams

Dens The planting density, in plants per square foot

Var The variety; one of `1`. `2` or `3`

Source

R. Mead (1970). Plant density and crop yield. *Applied Statistics*, **19**(1), 64–81.

Index