

Package: GDadata (via r-universe)

September 3, 2024

Type Package

Title Datasets for the Book Graphical Data Analysis with R

Version 0.93

Date 2015-05-02

Author Antony Unwin

Maintainer Antony Unwin<unwin@math.uni-augsburg.de>

Description Datasets used in the book 'Graphical Data Analysis with R'
(Antony Unwin, CRC Press 2015).

Depends R (>= 2.10)

License GPL (>= 2)

Suggests ggplot2

LazyData yes

NeedsCompilation no

Date/Publication 2015-05-02 14:11:23

Repository <https://impaug.r-universe.dev>

RemoteUrl <https://github.com/cran/GDadata>

RemoteRef HEAD

RemoteSha faa2e0d46f3796bab8fb4fdb534157dc2e0dac63

Contents

Decathlon	2
EastIndiesTrade	3
foodnames	4
HRstars	5
MexLJ	5
SpeedSki	6
uniranks	7
Index	9

Decathlon

*Top performances in the Decathlon from 1985 to 2006.***Description**

The point scoring system for the Decathlon last changed in 1985. Best annual performances of 6800 points and over for a twenty-one year period after the new rules were introduced were downloaded from the excellent Estonian website Decathlon2000. Handtimed performances were not included. Names with accents have been simplified.

Usage

```
data(Decathlon)
```

Format

A data frame with 7968 observations on the following 24 variables.

Totalpoints the total points achieved over all 10 events

DecathleteName Decathlete's name

Nationality Decathlete's nationality

m100 Time for the 100 metres (secs)

Longjump Distance jumped (metres)

Shotput Distance putting the shot (metres)

Highjump Height jumped (metres)

m400 Time for the 400 metres (secs)

m110hurdles Time for the 110 metres hurdles (secs)

Discus Distance throwing the discus (metres)

Polevault Height achieved (metres)

Javelin Distance throwing the javelin (metres)

m1500 Time for the 1500 metres (secs)

yearEvent Year of performance

P100m Points for performance in 100 metres

P1j Points for performance in long jump

Psp Points for performance in putting the shot

Phj Points for performance in high jump

P400m Points for performance in 400 metres

P110h Points for performance in 110 metres hurdles

Ppv Points for performance in pole vault

Pdt Points for performance in discus

Pjt Points for performance in javelin

P1500 Points for performance in 1500 metres

Source

<http://www.decathlon2000.com>

Examples

```
data(Decathlon, package="GData")
summary(Decathlon[, grep("P.*", names(Decathlon))])
library(ggplot2)
ggplot(Decathlon, aes(P1j)) + geom_histogram()
ggplot(Decathlon, aes(P100m, P1j)) + geom_point()
```

EastIndiesTrade	<i>Figures for the trade between England and the East Indies in the 18th century.</i>
-----------------	---

Description

The data have been estimated from the graphic in the first edition of Playfair's *Commercial and Political Atlas* by the website 'Me, myself, and BI'.

Usage

```
data(EastIndiesTrade)
```

Format

A data frame with 81 observations on the following 3 variables.

Year the data go from 1700 to 1780

Exports Exports from England to the East Indies (millions of pounds)

Imports Imports to England from the East Indies (millions of pounds)

Source

<http://blog.bissantz.com/vis-a-vis>

Examples

```
data(EastIndiesTrade, package="GData")
library(ggplot2)
ggplot(EastIndiesTrade, aes(x=Year, y=Exports-Imports)) + geom_line()
```

foodnames	<i>Nutritional value of food.</i>
-----------	-----------------------------------

Description

Nutritional value of different foods based on standard serving sizes.

Usage

```
data(foodnames)
```

Format

A data frame with 961 observations on the following 9 variables.

Name name of food (not unique)

Measure serving description

Fat.grams grams of fat in a standard serving

Food.energy.calories calories per serving

Carbohydrates.grams grams of carbohydrates per serving

Protein.grams grams of protein per serving

Cholesterol.mg cholesterol in mg per serving

weight.grams weight in grams of a standard serving

Saturated.fat.grams grams of saturated fat per serving

Source

The data are used in A. Izenman (2008), *Modern Multivariate Statistical Techniques*, Springer and are available on the accompanying website <http://astro.temple.edu/~alan/MMST/>

Examples

```
data(foodnames, package="GDAdata")
summary(foodnames)
library(ggplot2)
ggplot(foodnames, aes(Fat.grams, Saturated.fat.grams)) + geom_point()
```

HRstars

Star data useful for drawing a Hertzsprung-Russell diagram.

Description

Hertzsprung-Russell diagrams plot star luminosity (brightness) against temperature (colour). The first one was drawn just over 100 years ago. The dataset is the Yale Trigonometric Parallax Dataset and this version can be found on the webpage of the Astronomy Department of Case Western Reserve University.

Usage

```
data(HRstars)
```

Format

A data frame with 6220 observations on the following 5 variables.

ID star ID number

V apparent V magnitude

BV observed B-V color

Para observed parallax (in arcsec)

Uncert uncertainty in parallax (in milliarcsec)

Source

<http://burro.astr.cwru.edu/Academics/Astr221/HW/HW5/HW5.html>

Examples

```
data(HRstars, package="GDadata")
with(HRstars, hist(BV))
with(HRstars, hist(V))
```

MexLJ

Data from the longjump final in the 1968 Mexico Olympics.

Description

The best longjumps by the 16 finalists in the 1968 Mexico Olympics. Each athlete jumped up to six times, though the winner of the Gold Medal, Bob Beamon, only jumped twice.

Usage

```
data(MexLJ)
```

Format

A data frame with 16 observations on the following variable.

Jump Distance jumped measured in metres

Source

http://en.wikipedia.org/wiki/Athletics_at_the_1968_Summer_Olympics_-_Men's_long_jump

Examples

```
data(MexLJ, package="GData")
with(MexLJ, summary(Jump))
with(MexLJ, hist(Jump,breaks=seq(7.25,9,0.25)))
```

SpeedSki

World Speed Skiing Competition, Verbier 21st April, 2011

Description

There were separate Speed Skiing competitions for men (79 participants) and women (12 participants).

Usage

```
data(SpeedSki)
```

Format

A data frame with 91 observations on the following 8 variables.

Rank Finishing position by sex

Bib Start number

FIS.Code Skier's international skiing ID number

Name Skier's name

Year Skier's year of birth

Nation Skier's nationality

Speed Speed achieved in km/hr

Sex Female or Male

Event Speed Downhill, Speed Downhill Junior or Speed One

no.of.runs No of runs

Source

<http://www.fis-ski.com/de/606/612.html?sector=SS&raceid=262> (men)
<http://www.fis-ski.com/de/606/612.html?sector=SS&raceid=263> (women)

Examples

```
data(SpeedSki, package="GDadata")
with(SpeedSki, summary(Speed))
library(ggplot2)
ggplot(SpeedSki, aes(Speed)) + geom_histogram(binwidth=5)
```

uniranks

*The Guardian University League Table 2013***Description**

The Guardian newspaper in the UK publishes a ranking of British universities each year and it reported these data in May, 2012 as a guide for 2013.

Usage

```
data(uniranks)
```

Format

A data frame with 120 observations on the following 13 variables.

Rank Rank of the University

Institution University name

UniGroup Universities can be a member of one of five groups, 1994 Group, Guild HE, Million+, Russell, University Alliance, or none

HesaCode University's Higher Education Statistics Agency code

AvTeachScore Average Teaching Score

NSSTeaching University's National Student Survey teaching score

NSSOverall University's NSS overall score

SpendPerStudent University expenditure per student (depends on subject)

StudentStaffRatio Student to Staff ratio

CareerProspects Proportion of graduates in appropriate level employment or full-time study within six months of graduation

ValueAddScore "Based upon a sophisticated indexing methodology that tracks students from enrolment to graduation, qualifications upon entry are compared with the award that a student receives at the end of their studies." (Guardian)

EntryTariff Value dependent on the average points needed to get on the university's courses

NSSFeedback University's NSS feedback score

Source

<http://www.theguardian.com/news/datablog/2012/may/22/university-guide-2013-guardian-data>

Examples

```
data(uniranks, package="GDAdat")
with(uniranks, table(UniGroup))
library(ggplot2)
ggplot(uniranks, aes(x=NSSTeaching, y=NSSFeedback)) + geom_point()
ggplot(uniranks, aes(x=UniGroup, y=SpendPerStudent)) + geom_boxplot()
```


Index

* datasets

- Decathlon, [2](#)
- EastIndiesTrade, [3](#)
- foodnames, [4](#)
- HRstars, [5](#)
- MexLJ, [5](#)
- SpeedSki, [6](#)
- uniranks, [7](#)

Decathlon, [2](#)

EastIndiesTrade, [3](#)

foodnames, [4](#)

HRstars, [5](#)

MexLJ, [5](#)

SpeedSki, [6](#)

uniranks, [7](#)