# Package 'alr4'

April 20, 2018

**Version** 1.0.6

**Date** 2018-04-20

**Title** Data to Accompany Applied Linear Regression 4th Edition

**Author** Sanford Weisberg <sandy@umn.edu>

**Maintainer** Sanford Weisberg <sandy@umn.edu>

**Depends** R (>= 3.0), car, effects

**LazyLoad** yes

**LazyData** yes

**Description** Datasets to Accompany S. Weisberg (2014, ISBN: 978-1-118-38608-8),
``Applied Linear Regression,'' 4th edition. Many data files
in this package are included in the `alr3` package as well, so only one of them
should be used.

**License** GPL (>= 2)

**URL** http://www.z.umn.edu/alr4ed

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2018-04-20 14:25:58 UTC

## R topics documented:

ais . . . . . 3
allshoots . . . . . 4
alr4Web . . . . . 5
baeskel . . . . . 5
BGSall . . . . . 6
BigMac2003 . . . . . 7
BlowBS . . . . . 8
Blowdown . . . . . 8
brains . . . . . 9
cakes . . . . . 10
cathedral . . . . . 10

caution . . . 11
Challeng . . . 12
cloud . . . 13
domedata . . . 13
Donner . . . 14
Downer . . . 15
drugcost . . . 16
dwaste . . . 17
florida . . . 18
Forbes . . . 18
ftcollinssnow . . . 19
ftcollinstemp . . . 20
fuel2001 . . . 20
galapagos . . . 21
galtonpeas . . . 22
Heights . . . 23
Highway . . . 23
Hooker . . . 24
Htwt . . . 25
jevons . . . 26
lakemary . . . 26
lakes . . . 27
landrent . . . 28
lathe1 . . . 29
mantel . . . 29
mile . . . 30
MinnLand . . . 31
MinnWater . . . 32
Mitchell . . . 33
MWwords . . . 34
npdata . . . 35
oldfaith . . . 35
physics . . . 36
pipeline . . . 37
prodscore . . . 37
rat . . . 38
Rateprof . . . 39
Rpdata . . . 40
salary . . . 41
salarygov . . . 42
segreg . . . 42
shocks . . . 43
sleep1 . . . 44
snake . . . 45
sniffer . . . 45
Stevens . . . 46
stopping . . . 47
swan96 . . . 48

turk0 . . . . . 48
turkey . . . . . 49
twins . . . . . 50
UBSprices . . . . . 50
ufc . . . . . 51
UN1 . . . . . 52
UN11 . . . . . 53
walleye . . . . . 54
water . . . . . 54
wblake . . . . . 55
Whitestar . . . . . 56
wm1 . . . . . 57
wm2 . . . . . 57

**Index** **59**

---

`ais` *Australian institute of sport data*

---

## Description

Data on 102 male and 100 female athletes collected at the Australian Institute of Sport.

## Format

This data frame contains the following columns:

**Sex** (0 = male or 1 = female)

**Ht** height (cm)

**Wt** weight (kg)

**LBM** lean body mass

**RCC** red cell count

**WCC** white cell count

**Hc** Hematocrit

**Hg** Hemoglobin

**Ferr** plasma ferritin concentration

**BMI** body mass index, weight/(height)**2

**SSF** sum of skin folds

**Bfat** Percent body fat

**Label** Case Labels

**Sport** Sport

## Source

Ross Cunningham and Richard Telford

## References

S. Weisberg (2014). *Applied Linear Regression*, 4th edition. New York: Wiley.

## Examples

```
head(ais)
```

---

`allshoots` *Apple shoots data*

---

## Description

Bland's Apple Shoot data. allshoots includes all the data, shortshoots just the short shoot data, and longshoots includes long shoots only.

## Format

This data frame contains the following columns:

**Day** days from dormancy

**n** number of shoots sampled

**ybar** average number of stem units

**SD** within-day standard deviation

**Type** 1 if long shoots, 0 if shortshoots.

## Source

Bland, J. (1978). A comparisonof certain aspects of ontogeny in the long and short shoots of McIntosh apple during one annual growth cycle. Unpublished Ph. D. dissertation, University of Minnesota, St. Paul, Minnesota.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(longshoots)
```

## alr4Web *Access to the Applied Linear Regression website*

### Description

These function will access the website for Applied Linear Regression, 3rd and 4th editions.

### Usage

```
alr4Web(page = c("webpage", "errata", "primer", "solutions"))
```

### Arguments

| | |
|---|---|
| page | A character string indicating what page to open. The default "webpage" will open the main webpage, "errata" displays the Errata sheet for the thrid edition of the book, "primer" fetches and displays the primer for R, and "solutions" gives solutions to odd-numbered problems. |

### Value

Either a webpage or a pdf document is displayed. This function gives quick access to the website for the book and in particular to the R primer and solutions to odd-numbered problems. The pdf files are formatted for viewing on a computer screen. With Adobe Reader, view the pdf files with the bookmarks showning at the left, using signle page view which is selected by View -> Page Dispaly -> Single Page View.

### Author(s)

Sanford Weisberg, based on the function UsingR in the UsingR package by John Verzani

### Examples

```
## Not run: alr4Web("primer")
```

## baeskel *Surface tension*

### Description

The data in the file were collected in a study of the effect of dissolved sulfur on the surface tension of liquid copper (Baes and Kellogg, 1953)

## Format

This data frame contains the following columns:

**Sulfur** Weight percent sulfur

**Tension** Decrease in surface tension, dynes/cm

## Source

Baes, C. and Kellogg, H. (1953). Effect of dissolved sulphur on the surface tension of liquid copper. *J. Metals*, 5, 643-648.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(baeskel)
```

---

# BGSall *Berkeley guidance study*

---

## Description

Data from the Berkeley guidance study of children born in 1928-29 in Berkeley, CA. BGSall contains all the data, BGSboys the boys only, and BGSgirls the girls only.

## Format

This data frame contains the following columns:

**Sex** 0 = males, 1 = females

**WT2** Age 2 weight (kg)

**HT2** Age 2 height (cm)

**WT9** Age 9 weight (kg)

**HT9** Age 9 height (cm)

**LG9** Age 9 leg circumference (cm)

**ST9** Age 9 strength (kg)

**WT18** Age 18 weight (kg)

**HT18** Age 18 height (cm)

**LG18** Age 18 leg circumference (cm)

**ST18** Age 18 strength (kg)

**BMI18** Body Mass Index, `WT18/(HT18/100)^2`, rounded to one decimal.

**Soma** Somatotype, a 1 to 7 scale of body type.

## Source

Tuddenham, R. D. and Snyder, M. M. (1954). Physical Growth of California Boys and Girls from Birth to Eighteen years. Univ. of Calif. Publications in Child Development, 1, 183-364.

## References

S. Weisberg (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(BGSall)
head(BGSboys)
head(BGSgirls)
```

---

# BigMac2003 *World cities data*

---

## Description

Prices in many world cities from a 2003 Union Bank of Switzerland report.

## Format

This data frame uses the name of the city as row names, and contains the following columns:

**BigMac** Minutes of labor to purchase a Big Mac

**Bread** Minutes of labor to purchase 1 kg of bread

**Rice** Minutes of labor to purchase 1 kg of rice

**FoodIndex** Food price index (Zurich=100)

**Bus** Cost in US dollars for a one-way 10 km ticket

**Apt** Normal rent (US dollars) of a 3 room apartment

**TeachGI** Primary teacher's gross income, 1000s of US dollars

**TeachNI** Primary teacher's net income, 1000s of US dollars

**TaxRate** Tax rate paid by a primary teacher

**TeachHours** Primary teacher's hours of work per week:

## Source

Union Bank of Switzerland report, *Prices and Earnings Around the Globe* (2003 version).

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(BigMac2003)
```

## BlowBS *Blowdown data, Black Spruce only*

### Description

Data from the Boundary Waters Canoe Area Wilderness Blowdown. The data frame `Blowdown` includes nine species of trees, but this file only includes black spruce, grouped by diameter.

### Format

This data frame contains the following columns:

**d** Tree diameter, in cm

**died** Number of trees of this value of `d` that died (blowdown)

**m** number of trees of this size class measured

### Source

Roy Rich

### References

S. Weisberg (2014). *Applied Linear Regression*, fourth edition. New York: Wiley.

### Examples

```
head(BlowBS)
```

## Blowdown *Blowdown data*

### Description

Data from the Boundary Waters Canoe Area Wilderness Blowdown. The data frame blowdown includes nine species of trees. The data for balsam fir, summarized by diameter class, are given in `BlowBF`.

### Format

This data frame contains the following columns:

**d** Tree diameter, in cm

**s** Proportion of basal area killed for the four species balsam fir, cedar, paper birch and blue spruse, a measure of local severity of the storm.

**spp** Tree species, a factor with 9 levels

**y** 1 if the tree died, 0 if it survived

## Source

Roy Rich

## References

S. Weisberg (2014). *Applied Linear Regression*, fourth edition. New York: Wiley.

## Examples

```
head(Blowdown)
```

---

`brains` *Mammal brain weights*

---

## Description

The data provided gives the average body weight in kilograms and the average brain weight in grams for sixty-two species of mammals.

## Format

This data frame uses species names as row labels and contains the following columns:

**BrainWt** Brain weight, grams

**BodyWt** Body weight, kg

## Source

Allison, T. and Cicchetti, D. (1976). Sleep in mammals: Ecology and constitutional correlates. *Science*, 194, 732-734.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(brains)
```

## cakes *Cakes data*

### Description

Oehlert (2000, Example 19.3) provides data from a small experiment on baking packaged cake mixes.

### Format

A data frame with 14 observations on the following 4 variables.

**block** a factor

**X1** Baking time, minutes

**X2** Baking temperature, degrees F

**Y** Palatability score

### Source

Oehlert, G. W. (2000). *A First Course in Design and Analysis of Experiments*. New York: Freeman.

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(cakes)
lm(Y~X1+X2+I(X1^2)+I(X2^2)+X1:X2, data=cakes)
```

## cathedral *Cathedrals*

### Description

Heights and lengths of Gothic and Romanesque cathedrals.

### Format

This data frame uses cathedral names as row label andcontains the following columns:

**Type** Romanesque or Gothic

**Height** Total height, feet

**Length** Total length, feet

## Source

Stephen Jay Gould

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(cathedral)
```

---

`caution` *Caution data*

---

## Description

Artificial data to illustrate problems with residual plots.

## Format

This data frame contains the following columns:

**x1** Artificial data item.

**x2** Artificial data item.

**y** Artificial data item.

## Source

R. D. Cook and S. Weisberg (1999), Graphs in statistical analysis: Is the medium the message? *American Statistician,* 53, 29-37.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(caution)
```

Challeng *Challenger data*

## Description

Contains data from the performance of O-rings in 23 U.S. space shuttle flights prior to the Challenger disaster of January 20, 1986.

## Format

This data frame uses dates as row names and contains the following columns:

**temp** Air Temp at launch (degrees F)

**pres** Leak check pressure

**fail** Number of O-rings that failed

**n** 6, number of O-rings in launch

**erosion** Number of erosion incidents

**blowby** Number of blowby incidents

**damage** Total Damage Index

## Source

Dalal, S, Fowlkes, E. B. and Hoadley, B. (1989), Risk analysis of the space shuttle: Pre-challenger prediction of failure, *Journal of the American Statistical Association*, 84, 945-957. See also Tufte, E. R. (1997), *Visual and statistical Thinking: Displays of evidence for making decisions*, Cheshire, CT: Graphics Press.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(Challeng)
```

## cloud *Florida area cumulus experiment, FACE I.*

### Description

The data summarize the results of the first Florida Area Cumulus Experiment, or FACE-1, designed to study the effectiveness of cloud seeding to increase rainfall in a target area (Woodley, Simpson, Biondini, and Berkley, 1977).

### Format

This data frame contains the following columns:

**A** Action, 1=seed, 0=do not seed

**D** Day after June 16, 1975

**S** Suitability for seeding

**C** percent cloud cover in experimental area, measured using radar in Coral Gables, Florida

**P** $10^7 m^3$ prewetness

**E** echo motion category, either 1 or 2, a measure for type of cloud

**Rain** $10^7 m^3$ in target area

### Source

Woodley, W.L., Simpson, J., Biondini, R., and Berkley, J. (1977). Rainfall results 1970-75: Florida area cumulus experiment. *Science*, 195, 735-742.

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(cloud)
```

## domedata *Metrodome fan experiment*

### Description

These files give the results of two experiments to see if manipulating the air conditioning fans in the Minneapolis metrodome can effect the distance travelled by a baseball. The data in domedata were collected in April 2003. The experiment was repeated in May 2003 and domedata1 gives the combined data from the two experiments.

## Format

A data frame with 96 observations on the following 7 variables.

**Date** a factor with levels `March- May`

**Cond** a factor with levels `Headwind`, `Tailwind`

**Angle** the actual angle

**Velocity** in feet per second

**BallWt** weight of ball in grams

**BallDia** diameter of ball in inches

**Dist** distance in feet of the flight of the ball

## Source

Ivan Marusic

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(domedata1)
```

---

`Donner` *Donner party*

---

## Description

The Donner Party was the most famous tragedy in the history of the westward migration in the United States. In the winter of 1846-47, abount ninety wagon train emigrants were unable to cross the Sierra Nevada Mountains of California before winter, and almost one-half starved to death. Perhaps because they were ordinary people – farmers, merchants, parents, children. These data include some information about each of the members of the party from Johnson (1996).

## Format

This data frame uses the person's name as row labels and contains the following columns:

**age** Approximate age in 1846

**y** died or survived, a factor

**sex** Male or Female

**family.name** Either a family name, hired or single

**status** A factor with levels Family, Single or Hired

## Source

Johnson, K. (1996). *Unfortunate Emigrants: Narratives of the Donner Party.* Logan, UT: Utah State University Press, http://www.metrogourmet.com/crossroads/KJhome.htm.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(Donner)
```

---

`Downer` *Downer data*

---

## Description

For unknown reasons, some dairy cows become recumbant–they lay down. This condition can be serious, and may lead to death of the cow. These data are from a study of blood samples of over 500 cows studied at the Ruakura (N.Z.) Animal Health Laboratory during 1983-84. A variety of blood tests were performed, and for many of the animals the outcome (survived, died, or animal was killed) was determined. The goal is to see if survival can be predicted from the blood measurements. Case numbers 12607 and 11630 were noted as having exceptional care—and they survived.

## Format

This data frame contains the following columns:

**calving** a factor with levels before and after

**daysrec** Days recumbent

**ck** Serum creatine phosphokinase (U/l at 30C)

**ast** serum asparate amino transferase (U/l at 30C)

**urea** serum urea (mmol/l)

**pcv** Packed Cell Volume (Haemactocrit),

**inflamat** inflamation 0=no, 1=yes

**myopathy** Muscle disorder, a factor with levels present, and absent

**outcome** a factor with levels died and survived

## Source

Clark, R. G., Henderson, H. V., Hoggard, G. K. Ellison, R. S. and Young, B. J. (1987). The abiltiy of biochemical and haematolgical tests to predict recovery in periparturient recumbent cows. *NZ Veterinary Journal*, 35, 126-133.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(Downer)
```

---

`drugcost` *Drug cost.*

---

## Description

These data are to try to understand the effect of health plan characteristics on drug costs. Health plans vary in size, given as member months. Some plans use generic drugs more than others. All differ on copayments. Some have strong restrictions on which drugs can be dispensed value of RI=0 means that all drugs are dispensed, RI=100 means that only one per category is avaiable. The goal is to determine the terms that are related to cost, and in particular to understand the role of GS and RI in determining cost.

## Format

This data frame uses a short code name for the drug plan as row labels and contains the following columns:

**COST** Ave. cost to plan for 1 prescription for 1 day

**RXPM** Number of prescriptions per member per year

**GS** Percent generic substitution, number between 0 (no substitution) to 100 (always use generic substitute)

**RI** Restrictiveness index (0=none, 100=total)

**COPAY** Average Rx copayment

**AGE** Average age of member

**F** Percent female members

**MM** Member months, a measure of the size of the plan

## Source

Mark Siracuse

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(drugcost)
```

dwaste *Crock data.*

## Description

An experiment was conducted to study the *O2UP*, oxygen uptake in milligrams of oxygen per minute, given five chemical measurements: biological oxygen demand (BOD), total Kjeldah nitrogen (TKN), total solids (TS), total vital solids (TVS), which is a component of TS, and chemical oxygen demand (COD), each measured in milligrams per liter (Moore, 1975).

## Format

This data frame contains the following columns:

**Day** Day number

**BOD** Biological oxygen demand

**TKN** Total Kjeldahl nitrogen

**TS** Total Solids

**TVS** Total volatile solids

**COD** Chemical oxygen demand

**O2UP** Oxygen uptake

## Source

Moore, J. (1975). Total Biomedical Oxygen Demand of Animal Manures. Unpublished Ph. D. disseration, University of Minnesota.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(dwaste)
```

## florida *Florida presidential election*

### Description

County-by-county vote for president in Florida in 2000 for Bush, Gore and Buchanan.

### Format

A data frame three vaiaables for each of Florida's 67 counties.

**Gore** Vote for Gore

**Bush** Vote for Bush

**Buchanan** Vote for Buchanan

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(florida)
## maybe str(florida) ; plot(florida) ...
```

## Forbes *Forbes' data*

### Description

The data consists of 17 pairs of numbers corresponding to observed boiling point and corrected barometric pressure, at locations in the Alps.

### Format

This data frame contains three columns. The first two columns are identical to the data set named forbes in the MASS package.

**bp** Adjusted boiling point of water in degrees F.

**pres** Atmospheric pressure, in inches of Mercury

**lpres** 100 times log10(pres), rounded to two decimals

### Source

Forbes, J. (1857). Further experiments and remarks on the measurement of heights and boiling point of water. *Transactions of the Royal Society of Edinburgh*, 21, 235-243.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(Forbes)
```

---

`ftcollinssnow` *Ft. Collins snowfall*

---

## Description

Monthly snowfall data for Fort Collins, CO, 1900-01 to 1992-93

## Format

This data frame contains the following columns:

**YR1** Year corresponding to the September to December data

**Early** September to December snowfall, inches

**Late** January to June snowfall, inches

## Source

http://ccc.atmos.colostate.edu/cgi-bin/monthlydata.pl

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(ftcollinssnow)
```

## ftcollinstemp *Ft. Collins Temperature*

### Description

Monthly average temperature data for Fort Collins, CO weather station 53005, 1900-01 to 2010-11

### Format

This data frame contains the following columns:

**year** Year corresponding to the September to November data

**fall** September to November mean temperature, degrees F

**winter** December to February mean temperature, degrees F

### Source

http://ccc.atmos.colostate.edu/cgi-bin/monthlydata.pl

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(ftcollinstemp)
```

## fuel2001 *Fuel consumption*

### Description

Data on motor fuel consumption and related variables, for the year 2001. The unit is a state in the United States or the District of Columbia. Data are for 2001, unless noted.

### Format

This data frame contains the following columns. Row labels are the two-digit US Postal abbreviations for the US states.

**Drivers** Number of Licensed drivers in the state

**FuelC** Gasoline sold for road use (1000s of gal.)

**Income** Per capita personal income (year 2000)

**Miles** Miles of Federal-aid highway miles in the state

**MPC** Estimated miles driven per capita

**Pop** Population age 16 and over

**Tax** Gasoline state tax rate, cents per gallon

## Source

http://www.fhwa.dot.gov/ohim/hs01/index.htm

## References

Weisberg, S. (2014). *Applied Linear Regression*, third edition. New York: Wiley.

## Examples

```
head(fuel2001)
# Most of the examples in ALR3 that use these data first
# transform several of the columns
fuel2001 <- transform(fuel2001,
     Dlic=1000 * Drivers/Pop,
     Fuel=1000 * FuelC/Pop,
     Income=Income/1000)
pairs(Fuel~Tax + Dlic + Income + log2(Miles), data=fuel2001)
```

---

# galapagos *Galapagos species data*

---

## Description

Johnson and Raven (1973) have presented data giving the number of species and related variables for 29 different islands in the Galapagos Archipelago.

## Format

This data frame uses the island name as row labels and contains the following columns:

**NS** Number of Species

**ES** Number of endemic species (orrur only on that island)

**Area** Surface area of island, hectares

**Anear** Area of closest island, hectares

**Dist** Distance to closest island, km

**DistSC** Distance from Santa Cruz Island, km

**Elevation** Elevation in m, missing values given as zero

**EM** 1 if elevation is observed, 0 if missing

## Source

Johnson, M.P., and Raven, P.H. (1973). Species number and endemism: The Galapagos Archipelago revisited. *Science*, 179, 893-895.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(galapagos)
```

---

`galtonpeas` *Galton's peas*

---

## Description

In a paper presented to the Royal Institute on February 9, 1877, Sir Francis Galton discussed his experiments on sweet peas in which he compared the sweet peas produced by parent plants to those produced by offspring plants. In these experiments he could observe inheritance from one generation to the next. Galton categorized the parent plants according to the typical diameter of the peas they produced.

## Format

This data frame contains the following columns:

**Parent** mean diameter of parent

**Progeny** mean diameter of offspring

**SD** offspring standard deviation

## Source

Pearson, K. (1930). *Life and Letters and Labours of Francis Galton*, Vol IIIa. Cambridge: Cambridge University Press.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(galtonpeas)
```

## Heights *Pearson-Lee data*

### Description

Karl Pearson organized the collection of data on over 1100 families in England in the period 1893 to 1898. This particular data set gives the Heights in inches of mothers and their daughters, with up to two daughters per mother. All daughters are at least age 18, and all mothers are younger than 65. Data were given in the source as a frequency table to the nearest inch. Rounding error has been added to remove discreteness from graph.

### Format

This data frame contains the following columns:

**mheight** Mother's ht, in.

**dheight** Daughter's ht, in.

### Source

K. Pearson and A. Lee (1903), On the laws of inheritance in man, *Biometrika*, 2, 357–463, Table 31.

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(Heights)
```

## Highway *Highway accidents*

### Description

The data comes from a unpublished master's paper by Carl Hoffstedt. They relate the automobile accident rate, in accidents per million vehicle miles to several potential terms. The data include 39 sections of large Highways in the state of Minnesota in 1973. The goal of this analysis was to understand the impact of design variables, acpts, slim, Sig, and shld that are under the control of the Highway department, on accidents.

## Format

This data frame contains the following columns:

**adt** average daily traffic count in thousands

**trks** truck volume as a percent of the total volume

**lane** total number of lanes of traffic

**acpt** number of access points per mile

**sigs** number of signalized interchanges per mile

**itg** number of freeway-type interchanges per mile

**slim** speed limit in 1973

**len** length of the Highway segment in miles

**lwid** lane width, in feet

**shld** width in feet of outer shoulder on the roadway

**htype** An indicator of the type of roadway or the source of funding for the road; "mc" for major collector, "fai" for Federal interstate highways, "pa" for principal arterial highway, and "ma" for major arterial highways

**rate** 1973 accident rate per million vehicle miles

## Source

Carl Hoffstedt

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(Highway)
```

---

`Hooker` *Hooker's data*

---

## Description

In his original paper, Forbes provided additional data collected by the botanist Dr. Joseph Hooker on temperatures and boiling points measured often at higher altitudes in the Himalaya Mountains.

## Format

This data frame contains the following columns:

**bp** Measured boiling temperature, degrees F.

**pres** Measured air pressure, inches of Mercury.

**lpres** 100 times pres rounded to two decimals.

## Source

Forbes, J. (1957). Further experiments and remarks on the measurement of heights by boiling point of water. *Transactions of the Royal Society of Edinburgh*, 21, 235-243.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(Hooker)
```

---

# Htwt *Artificial height and weight data*

---

## Description

The data for this table are a sample size of ten 18-year old girls taken from the study that was conducted by Tuddenham and Snyder (1954).

## Format

This data frame contains the following columns:

**ht** Height (cm) at age 18

**wt** Weight (kg) at age 18

## Source

Tuddenham, R., and Snyder, M. (1954). Physical growth of California boys and girls from birth to age 18. *California Publications on Child Development*, 1, 183-364.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(Htwt)
```

## jevons *Jevon's gold coin data*

### Description

In a study of coinage, W. Stanley Jevons weighed 274 gold sovereigns that he had collected from circulation in Manchester, England. For each coin, he recorded the weight, after cleaning, to the nearest .001 gram, and the date of issue. The age classes are coded 1 to 5, roughly corresponding to the age of the coin in decades. The standard weight of a gold sovereign was suppose to be 7.9876 grams; minimum legal weight was 7.9379 grams.

### Format

This data frame contains the following columns:

**Age** Age of coins, decades

**n** Number of coins

**Weight** Average weight, grams

**SD** Standard deviation.

**Min** Minimum weight

**Max** Maximum weight

### Source

Stephen Stigler

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(jevons)
```

## lakemary *Lake Mary bluegills*

### Description

78 bluegills were captured from Lake Mary, Minnesota. On each fish, a key scale was removed. The age of a fish is determined by counting the number of annular rings on the scale. The goal is to relate length at capture to the radius of the scale.

### Format

This data frame contains the following columns:

**Age** Years

**Length** mm

### Source

Collected by Richard Frie, and discussed in S. Weisberg (1986), A linear model approach to the backcalculation of fish length, *J. Amer. Statist. Assoc.*, 81, 922-929.

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(lakemary)
```

---

## lakes *Lake zooplankton diversity*

---

### Description

These data give the number of known crustacean zooplankton species for 69 world lakes. Also included are a number of characteristics of each lake. There are missing values.

### Format

This data frame uses lake name as row label and contains the following columns:

**Species** Number of zooplankton species

**MaxDepth** Maximum lake depth, m

**MeanDepth** Mean lake depth, m

**Cond** Specific conductance, micro Siemans

**Elev** Elevation, m

**Lat** N latitude, degrees

**Long** W longitude, degrees

**Dist** distance to nearest lake, km

**NLakes** number of lakes within 20 km

**Photo** Rate of photosynthesis, mostly by the 14C method

**Area** Lake area, in hectares

## Source

Dodson, S. (1992), Predicting curstacean zooplankton species richness, *Limnology and Oceanography*, 37, 848–856.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(lakes)
```

---

`landrent` *Land rent*

---

## Description

The data were collected by Douglas Tiffany to study the variation in rent paid in 1977 for agricultural land planted to alfalfa.

## Format

This data frame contains the following columns:

**X1** average rent for all tillable land

**X2** density of dairy cows (number per square mile)

**X3** proportion of farmland used for pasture

**X4** 1 if liming required to grow alfalfa; 0 otherwise

**Y** average rent per acre planted to alfalfa

## Source

Douglas Tiffany

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(landrent)
```

## lathe1 *Lathe data*

### Description

These data are the results of an experiment to study the performance of cutting-tool material in cutting steel on a lathe. The two factors are revolution speed and feed rate. The response is tool life in minutes.

### Format

This data frame contains the following columns:

**Feed** Coded feed rate, coded as (actual feed rate -13)/6. Feed is in thousandths of an inch per revolution.

**Speed** Coded speed, coded as (actual speed -900)/300. Speed is in feet per minute.

**Life** Life of tool until failure, minutes

### Source

M. R. Delozier

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(lathe1)
```

## mantel *Mantel's artifical data for stepwise regression*

### Description

An artificial data set suggested by N. Mantel to illustrate stepwise regression methods.

### Format

A data frame with 5 observations on the following 4 variables.

**Y** the response

**X1** predictor 1

**X2** predictor 2

**X3** predictor 3

## Source

Mantel, N. (1970). Why stepdown procedures in variable selection? *Technometrics*, 12, 621–625.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(mantel)
```

---

`mile` *World records for the mile run*

---

## Description

World record times for the mile run, 1861–2003.

## Format

A data frame with 46 observations:

**Year** Year in which the record was set

**Time** Running time, in seconds

**Name** Name of person setting the record

**Country** Country of residence of the record setter

**Place** Place the record was set

**Gender** Gender of the record holder

## Source

Data source: http://www.saunalahti.fi/~sut/eng/

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(mile)
```

MinnLand *Conservation restrictions and farm prices in Minnesota*

## Description

These data include nearly every farm sale in 6 economic regions in Minnesota from 2002-2011 that either has land enrolled in the federal Conservation Reserve Program, or CRP, or has no restrictions. A few sales with non-crp land easements were excluded. CRP enrollment is for a fixed period during which farmers agree not to grow crops for a fixed payment. This can effect sale price of land since buyers have fewer choices on use of land that could lower values, but also have guaranteed income for a fixed period that could raise values.

## Usage

```
data(MinnLand)
```

## Format

A data frame with 18700 observations on the following 10 variables.

`acrePrice` sale price in dollars per acre. Sale prices were adjusted to a common date within the year. No inflation adjustment is made between years.

`region` a factor with levels giving the geographic names of six economic regions of Minnesota. Excluded economic regions had few farm sales.

`improvements` percentage of property value due to improvements. Minnesota assessors estimate values separately for land and buildings. This variable is the ratio of the building value to the total value.

`year` year of sale, as a continuous variable, not as a factor. Most uses of this variable would require converting it to a factor.

`acres` size of the farm in acres

`tillable` percentage of farm acreaage that is rated arable by the assessor

`financing` a factor with levels `title transfer` and `seller finance`

`crpPct` the percentage of all farm acres enrolled in CRP

`productivity` average agronomic productivity scaled 1 to 100, with larger numbers for more productive land. This score is based on University of Minnesota soil studies. This value is frequently missing because some counties never had the study done, and some county assessors are inconsistent in including this value in the record of the sale.

## Details

Data is collected from Minnesota counties. Some counties do not include the `productivity` value in sales records, accounting for most of the missing values. The variable `tillable` is also frequently missing.

## Source

S. J. Taff

## References

Taff, S. J. and Weisberg, S. (2007). Compensated shrot-term conservation restrictions may reduce sale prices. *The Appraisal Journal*, 75(1), 45.

## Examples

```
head(MinnLand)
## Not run:
require(mice)
md.pattern(MinnLand)

## End(Not run)
```

---

# MinnWater *Total Water Usage in Minnesota*

---

## Description

Yearly water consumption in Minnesota from 1988-2011.

## Usage

```
data(MinnWater)
```

## Format

A data frame with 24 observations on the following variables.

**year** year

**allUse** total ground water consumption, statewide, in billions of gallons

**muniUse** total municipal water consumption, statewide, in billions of gallons

**irrUse** consumption for irrigation in 13 counties, in billions of gallons

**agPrecip** average growing season June to August precipiciation (inches) for the 13 Minnesota counties that use the most irrigation

**muniPrecip** average May to September precipiciation (inches) for the 10 Minnesota counties with highest municipal water pumping

**statePop** estimated state population

**muniPop** estimated 10 county urban population

## Details

Is water usage increasing? How fast?

## Source

These data were provided by the Freshwater Society. They collected the data from the Minnesota Department of Natural Resources and from the Minnesota Climatology Working Group. Thanks to Tom Burk.

## Examples

```
data(MinnWater)
## maybe str(MinnWater) ; plot(MinnWater) ...
```

---

# Mitchell *Mitchell soil temperature*

---

## Description

Data collected by Kenneth G. Hubbard on soil temperature at 20 cm depth in Mitchell, Nebraska for 17 years (1976-1992) The variable month is the month number.

## Format

This data frame contains the following columns:

**Month** Months beginning Jan, 1976

**Temp** Average soil temperature, degrees C

## Source

Kenneth G. Hubbard

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(Mitchell)
```

`MWwords` *Word frequencies from Mosteller and Wallace*

## Description

The data give the frequencies of words in works from four different sources: the political writings of eighteenth century American political figures Alexander Hamilton, James Madison, and John Jay, and the book *Ulysses* by twentieth century Irish writer James Joyce.

## Format

This data frame uses the word as row labels and contains the following columns:

**Hamilton** Hamilton frequency

**HamiltonRank** Hamilton rank

**Madison** Madison frequency

**MadisonRank** Madison rank

**Jay** Jay frequency

**JayRank** Jay rank

**Ulysses** Word frequency in *Ulysses*

**UlyssesRank** Word rank in *Ulysses*

## Source

Mosteller, F. and Wallace, D. (1964). *Inference and Disputed Authorship: The Federalist.* Reading, MA: Addison-Wesley.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(MWwords)
```

## npdata *Northern pike catch per unit effort*

### Description

Catch per unit effort data for 16 Minnesota lakes

### Format

A data frame with 16 observations on the following 4 variables.

**CPUE** Estimated catch per unit effect

**SECPUE** Estimated standard error of CPUE

**Density** Estimated fish density

**SEdens** Estimated standard error of Density

### Source

R. Pierce, Minnesota Dept. of Natural Resources

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(npdata)
```

## oldfaith *Old Faithful Geyser data*

### Description

Data on eruptions of Old Faithful Geyser, October 1980. Variables are the duration in seconds of the current eruption, and the time in minutes to the next eruption. Collected by volunteers, and supplied by the Yellowstone National Park Geologist. Data was not collected between approximately midnight and 6 AM.

### Format

This data frame contains the following columns:

**Duration** Duration in seconds

**Interval** Time to next eruption

## Source

R. Hutchinson

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(oldfaith)
```

---

# physics *Physics data*

---

## Description

The file physics constains results for $\pi^+$ meson as input and $\pi^+$ meson as output. physics1 is for $\pi^-$ to $\pi^-$.

## Format

This data frame contains the following columns:

**x** Inverse total energy

**y** Scattering cross-section/sec

**SD** Standard deviation

## Source

Weisberg, H., Beier, H., Brody, H., Patton, R., Raychaudhari, K., Takeda, H., Thern, R. and Van Berg, R. (1978). s-dependence of proton fragmentation by hadrons. II. Incident laboratory momenta, 30–250 GeV/c. *Physics Review D*, 17, 2875–2887.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(physics1)
```

## pipeline *Alaska pipeline*

### Description

The Alaska pipeline data consists of in-field ultrasonic measurements of the depths of defects in the Alaska pipeline. The depth of the defects were then re-measured in the laboratory. These measurements were performed in six different batches. The data were analyzed to calibrate the bias of the field measurements relative to the laboratory measurements. In this analysis, the field measurement is the response variable and the laboratory measurement is the predictor variable.

These data were originally provided by Harry Berger, who was at the time a scientist for the Office of the Director of the Institute of Materials Research (now the Materials Science and Engineering Laboratory) of NIST. These data were used for a study conducted for the Materials Transportation Bureau of the U.S. Department of Transportation.

### Format

This data frame contains the following columns:

**Field** Number of defects measured in the field.

**Lab** Number of defects measured in the field.

**Batch** Batch number

### Source

http://www.itl.nist.gov/div898/handbook/pmd/section6/pmd621.htm

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(pipeline)
```

## prodscore *Soil productivity*

### Description

Soil productivity scores for farms in townships in four counties in the Minneapolis St. Paul metropolitan area, 1981-82. The goal is to see if the productivity score is a good predictor of the assessed value of the farmland. If so, then productivity score could be used to set assesed value for farms enrolled in the "green acres" program that requires that urban farmland be taxed at its agricultural value only without regard to development potential.

## Format

This data frame contains the following columns:

**County** Name of the county

**Value** Assessed value in dollars per acre.

**P** Productivity score, a number between 1 and 100.

**Year** Tax year, either 1981 or 1982.

## Source

Douglas Tiffany

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(prodscore)
```

---

# rat *Rat data*

---

## Description

Data collected in an experiment in which rats were injected with a dose of a drug approximately proportional to body weight. At the end of the experiment, the animal's liver was weighed, and the fraction of the drug recoved in the liver was recorded. The experimenter expected the response to be independent of the predictors.

## Format

This data frame contains the following columns:

**BodyWt** BodyWt of the rat

**LiverWt** LiverWt measured after sacrifice

**Dose** Dose, roughly proportional to body weight

**y** dose of drug recovered after sacrifice of the animal

## Source

Dennis Cook

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(rat)
```

---

Rateprof *Professor ratings from RateMyProfessors.com.*

---

## Description

These data includes the summaries of the ratings of 364 instructors at one large campus in the Midwest from Bleske-Rechek and Fritsch (2011). Each instructor included in the data had at least 10 ratings over a several year period. Students provided ratings on 5 point scales. The data file provides the averages ratings and additional characteristics of the instructors

## Format

A data frame with 364 observations on the following 17 variables.

`gender` instructor gender, a factor with levels `female` `male`

`numYears` a numeric vector, number of years in which this instructor had ratings between 1999 and 2009.

`numRaters` number of ratings

`numCourses` number of different course titles included in the rating for this instructor

`pepper` a factor with levels `no` and `yes`. In addition to rating for quality, instructors are rated as attractive or not. A value of `yes` means that the consensus is that the instructor is attractive.

`discipline` a factor with levels `Hum` for humanities, `SocSci` for social sciences, `STEM` for science, technology, engineering and mathematics and `Pre-prof` for professional training

`dept` a factor with department names `Accounting`, `Anthropology`, `Art`, `Art and design`, `Art History`, `Astronomy/Physics`, `Biology`, `Business`, `Chemistry`, `Communication`, `Communication Disorders`, `Computer Science`, `Criminal Justice`, `Curriculum and Instruction`, `Dance`, `Economics`, `English`, `Environmental Public Health`, `Finance`, `FLTR`, `French`, `Geography`, `Geology`, `German`, `History`, `Information Systems`, `Japanese`, `Kins`, `Library Science`, `Management`, `Managerial Science`, `Marketing`, `Math`, `Music`, `Nursing`, `Philosophy`, `Physics`, `Physics & Astronomy`, `Physics and Astronomy`, `Political Science`, `Psychology`, `Religious Studies`, `Social Work`, `Sociology`, `Spanish`, `Special Education`, `Theater`, `Womens Studies`,

`quality` Average quality rating, between 1, worst, to 5, best

`helpfulness` Average helfpulness rating, between 1, worst, to 5, best

`clarity` Average clarity rating, between 1, worst, to 5, best

`easiness` Average easiness rating, between 1, worst, to 5, best

`raterInterest` Average rater interest, between 1, lowest, to 5, highest

`sdQuality` SD of quality rating

`sdHelpfulness` SD of helpfulness rating

`sdClarity` SD of clarity rating

`sdEasiness` SD of easiness rating

`sdRaterInterest` SD of rater interest

## Source

Provided by April Bleske-Rechek.

## References

Bleske-Rechek, A. and Fritsch, A. (2011). Student Consensus on RateMyProfessors.com. Practical Assessment, Research \& Evaluation, 16(18), http://pareonline.net/getvn.asp?v=16&n=18

## Examples

```
data(Rateprof)
```

---

# Rpdata *Always plot residuals.*

---

## Description

This example with aritifical data is designed to demonstrate the importance of plotting residuals.

## Usage

```
data(Rpdata)
```

## Format

A data frame with 990 observations on the following 7 variables.

`y` a numeric vector

`x1` a numeric vector

`x2` a numeric vector

`x3` a numeric vector

`x4` a numeric vector

`x5` a numeric vector

`x6` a numeric vector

## Source

Data generated using programs from http://www4.stat.ncsu.edu/~stefanski/NSF_Supported/Hidden_Images/stat_res_plots.html

## References

Stefanski, L. A. (2007). Residual (sur)Realism. The American Statistician, 61, 163-177. url=https://www.amstat.org/about/pc

## Examples

```
data(Rpdata)
## Not run:
require(car)
residualPlot(lm(Rpdata))

## End(Not run)
```

---

`salary` *Salary data*

---

## Description

Salary of faculty in a small Midwestern college in the early 1980s.

## Format

This data frame contains the following columns:

**degree** Factor with levels "PhD" or "Masters"

**rank** Factor, "Asst", "Assoc" or "Prof"

**sex** Factor, "Male" or "Female"

**year** Years in current rank

**ysdeg** Years since highest degree earned

**salary** dollars per year

## Source

Sanford Weisberg

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(salary)
```

## salarygov *Government salary study*

### Description

Data on non-unionized job classes in a US county in 1986. Included are the job class difficulty score, the number of employees in the class, number of female employees, and the name of the class.

### Format

This data frame contains the following columns:

**JobClass** Name of job class

**NW** Number of women employees

**NE** Total number of employees in a job class

**Score** Difficulty score for job class

**MaxSalary** Maximum salary for job class

### Source

Sanford Weisberg

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(salarygov)
```

## segreg *Energy consumption*

### Description

Data on electricity consumption (KWH) and mean temperature (degrees F) for one building on the University of Minnesota's Twin Cities campus. for 39 months in 1988-92. The goal is to model consumption as a function of temperature. Higher temperature causes the use of air conditioning, so high temperatures should mean high consumption. This building is steam heated, so electricity is not used for heating.

### Format

This data frame contains the following columns:

**Temp** Monthly mean temperature, degrees F.

**C** Electricty consumption in KWH/day

### Source

Charles Ng

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(segreg)
```

---

## shocks *Small electric shocks in dairy cows*

---

### Description

Results of a small experiment to learn about the effects of small electric shocks on dairy cows.

### Format

A data frame with 6 observations on the following 3 variables.

**Intensity** Shock level, milliamps

**m** Number of trials

**Y** Number of times a positive reaction was observed

### Source

R. Norell

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(shocks)
```

`sleep1` *Sleep in mammals*

## Description

Includes species averages for 62 mammals.

## Format

This data frame uses spcies as row lable and contains the following columns:

**SWS** Slow wave nondreaming sleep, hrs/day

**PS** Paradoxical dreaming sleep, hrs/day

**TS** Total sleep, hrs/day

**BodyWt** Body weight in kg

**BrainWt** Brain weight in g

**Life** Maximum life span, years

**GP** Gestation time, days

**P** Predation index, 1=low,5=hi

**SE** Sleep exposure index 1=exposed, 5=protected

**D** Danger index, 1=least, 5=most

## Source

Allison, T. and Cicchetti, D. (1976). Sleep in Mammals: Ecological and Constitutional Correlates *Science*, vol. 194, pp. 732-734.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(sleep1)
```

## snake *Snake river data*

### Description

The data give the water content of snow and the water yield in inches in the Snake River watershed in Wyoming.

### Format

This data frame contains the following columns:

**X** water content of snow

**Y** water yield from April to July

### Source

Wilm, H. G. (1950). Statistical control in hydrologic forecasting. "Res. Notes", 61, Pacific Northwest Forest Range Experiment Station, Oregon.

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(snake)
```

## sniffer *Sniffer data*

### Description

When gasoline is pumped into a tank, hydrocarbon vapors are forced out and into the atmosphere. To reduce this significant source of air pollution, devices are installed to capture the vapor. In testing these vapor recovery systems, a "sniffer" measures the amount recovered. John Rice provided the data for the file sniffer.txt.

### Format

This data frame contains the following columns:

**TankTemp** Initial tank temperature (degrees F)

**GasTemp** Temperature of the dispensed gasoline (degrees F)

**TankPres** Initial vapor pressure in the tank (psi)

**GasPres** Vapor pressure of the dispensed gasoline (psi)

**Y** Hydrocarbons emitted (grams)

## Source

John Rice

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(sniffer)
```

---

# Stevens *Subjective scaling data collected by S. S. Stevens*

---

## Description

This experiment was apparently done by S. S. Stevens and colleagues in March 1962, although the exact reference is lost. 10 subjects were played tones at each of 5 loudnesses, presumably in random order. Subjects were asked to draw a line on paper whose length matched the loudness of the tone. Each subject repeated each loudness 3 times, for a total of 30 trials per subject. The original data are lost; reported here is the mean of the 3 log-lengths for each loudness, the sd of the three log-lengths, and the number of replications, which is always 3.

## Usage

```
data(Stevens)
```

## Format

A data frame with 50 observations on the following 5 variables.

`subject` a factor with unique values for each subject

`loudness` either 50, 60, 70, 80 or 90 db. Decibels are a logrithmic scale

`y` a numeric vector giving the mean of the log-lengths of three lines drawn. Exponentiating these values would give the geometric mean of the three lengths in cm.

`sd` a numeric vector, giving the sd of the three log lengths

`n` a numeric vector, equal to the constant value 3

## Details

This is a classic example of a psychophysics experiment pioneered by S. S. Stevens. The basic idea is that the phychological response y to a physical stimulus x should be proportional to x to a power. Since both the response and the loudness are already in log-scale, linear fits should be expected.

## Source

These data were obtained in the early 1970s from the data library in the Harvard University Statistics Department.

## References

Stevens, S. S. (1966). A metric for social consensus, *Science*, 151, 530-541, http://www.jstor.org/stable/1717034

## Examples

```
head(Stevens)
```

---

`stopping` *Stopping distances*

---

## Description

Ezekiel and Fox (1959) data on auto stopping distances.

## Format

This data frame contains the following columns:

**Speed** Speed (mph)

**Distance** Stopping distance (in feet)

## Source

Ezekiel, M. and Fox, K. A. (1959). *Methods of Correlation Analysis, Linear and Curvilinear*, Hoboken NJ: Wiley.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(stopping)
```

## swan96 *Black crappie study on Swan Lake, Minnesota*

### Description

Log catch per unit effect of 200 mm or longer black crappies was recored 27 times over the course of 1996 on Swan Lake, Minnesota.

### Format

A data frame with 27 observations on the following 2 variables.

**Day** Number of days after June 16, 1996

**LCPUE** log of the catch of 200 mm or longer black crappies per unit effort (WHAT IS THE BASE?)

### Source

Minnesota Department of Natural Resources

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(swan96)
```

## turk0 *Turkey data, one source*

### Description

Turkey weight increase in an experiment in which the supplementation with methionine was varied.

### Format

This data frame contains the following columns:

**A** Amount of methionine supplement (percent of diet)

**Gain** Pen weight increase (g)

### Source

Cook, R. D. and Witmer, J. (1985). A note on the parameter-effects curvature. *Journal of the American Statistical Association*, 80, 872-878.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(turk0)
```

---

`turkey` *Turkey data, all sources*

---

## Description

Data from an experiment on the growth of turkeys. 60 pens of turkeys were grown with a similar diet, supplemented with a dose of methionine from one of three sources. The response is average pen weight. Recorded is dose, source, m, always 5 except for dose=0, average weight gain, and within group SS.

## Format

This data frame contains the following columns:

**A** Dose: Amount of supplement as a percent of the total diet

**Gain** Ave. weight gain, over all replications

**S** A factor for the source of methionine, three levels numbers 1, 2 and 3.

**m** Number of replications or pens

**SD** SD of the m pens with the same values of S and A.

## Source

R. D. Cook and J. Witmer (1985). A note on parameter-effects curvature. *Journal of the American Statistical Association*, 80, 872–878.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(turkey)
```

## twins *Burt's twin data*

### Description

The given data are IQ scores from identical twins; one raised in a foster home, and the other raised by birth parents.

### Format

This data frame contains the following columns:

**C** Social class, C1=high, C2=medium, C3=low, a factor

**IQb** biological

**IQf** foster

### Source

Burt, C. (1966). The genetic estimation of differences in intelligence: A study of monozygotic twins reared together and apart. *Br. J. Psych.*, 57, 147-153.

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(twins)
```

## UBSprices *UBS price data*

### Description

The international bank UBS produces a report on prices in major world cities every three years. This data.frame includes price data for a 1 kg loaf of bread, 1 kg of rice and for a Big Mac hamburger, for the years 2003 and 2009. All these prices are measured in the minutes of labor required by the typical worker in that country to buy the product, so it adjusts for currency, wages and price levels.

### Usage

```
data(UBSprices)
```

## Format

A data frame with 54 observations on the following 6 variables.

`bigmac2009` 2009 Big Mac price, in minutes of labor

`bread2009` 2009 Bread price, in minutes of labor

`rice2009` 2009 Rice price, in minutes of labor

`bigmac2003` 2003 Big Mac price, in minutes of labor

`bread2003` 2003 Bread price, in minutes of labor

`rice2003` 2003 Rice price, in minutes of labor

## Details

City names are the row labels.

## Source

Union Bank of Switzerland

## Examples

```
data(UBSprices)
## maybe str(UBSprices) ; plot(UBSprices) ...
```

---

# ufc *Height-diameter data for Upper Flat Creek, Idaho*

---

## Description

These data are forest inventory measures from the Upper Flat Creek stand of the University of Idaho Experimental Forest, dated 1991.

The file `ufc` contains all the data. `ufcwc` contains only Western red cedar. `ufcgf` contains only grand fir.

## Format

A data frame with the following 5 variables.

**Plot** Plot number

**Tree** Tree within plot

**Species** a factor with levels `DF` = Douglas-fir, `GF` = Grand fir, `SF` = Subalpine fir, `WL` = Western larch, `WC` = Western red cedar, `WP` = White pine

**Dbh** Diameter 137 cm perpendicular to the bole, mm

**Height** Height of the tree, in decimeters

## Source

Andrew Robinson

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. New York: Wiley.

## Examples

```
head(ufcgf)
```

---

# UN1 *National statistics from the United Nations*

---

## Description

Demographic data for 193 places, mostly UN members, but also other areas like Hong Kong that are not independent countries.

## Format

This data frame uses the locality name as a row label. In some cases the geographic area is smaller than a country; for example Hong Kong. The file contains the following columns:

**Fertility** Expected number of live births per female, 2000

**PPgdp** Per capita 2001 GDP, in US \$

## Details

These data were collected at published by the UN from a variety of sources. See original source for additional footnotes concerning values for individual countries. Country names are given in the first column of the data file.

## Source

http://unstats.un.org/unsd/demographic

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(UN1)
```

UN11 *National statistics from the United Nations mostly from 2009-2011*

## Description

National health, welfare, and education statistics for 210 places, mostly UN members, but also other areas like Hong Kong that are not independent countries.

## Usage

```
data(UN11)
```

## Format

A data frame with 237 observations on the following 32 variables.

`region` region of the world

`group` a factor with levels `oecd` for countries that are members of the OECD, the Organization for Economic Co-operation and Development, as of May 2012, `africa` for countries on the African continent, and `other` for all other countries. No OECD countries are located in Africa

`fertility` number of children per woman

`ppgdp` Per capita gross domestic product in US dollars

`lifeExpF` Female life expectancy, years

`pctUrban` Percent Urban

## Details

Similar data, from the period 2000-2003, appears in the `alr3` package under the name `UN3`.

## Source

All data were collected from UN tables accessed at http://unstats.un.org/unsd/demographic/products/socind/ on April 23, 2012. OECD membership is from www.oecd.org, accessed May 25, 2012..

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
data(UN11)
## maybe str(UN11) ; plot(UN11) ...
```

## walleye *Walleye length at age*

### Description

These data give length and age for over 3000 walleye (a type of fish) captured in Butternut Lake, Wisconsin, in three periods with different management method in place.

### Format

A data frame with 3198 observations on the following 3 variables.

**age** Age of the fish, years

**length** Length, mm

**period** 1 = pre 1990, 2 = 1991-1996, 3=1997-2000

### Source

Michelle LeBeau

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(walleye)
```

## water *California water*

### Description

Can Southern California's water supply in future years be predicted from past data? One factor affecting water availability is stream runoff. If runoff could be predicted, engineers, planners and policy makers could do their jobs more efficiently. Multiple linear regression models have been used in this regard. This dataset contains 43 years worth of precipitation measurements taken at six sites in the Owens Valley ( labeled APMAM, APSAB, APSLAKE, OPBPC, OPRC, and OPSLAKE), and stream runoff volume at a site near Bishop, California.

## Format

This data frame contains the following columns:

**Year** collection year

**APMAM** Snowfall in inches measurement site

**APSAB** Snowfall in inches measurement site

**APSLAKE** Snowfall in inches measurement site

**OPBPC** Snowfall in inches measurement site

**OPRC** Snowfall in inches measurement site

**OPSLAKE** Snowfall in inches measurement site

**BSAAM** Stream runoff near Bishop, CA, in acre-feet

## Source

Source: http://www.stat.ucla.edu.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(water)
```

---

# wblake *West Bearskin Lake small mouth bass data.*

---

## Description

Data on samples of small mouth bass collected in West Bearskin Lake, Minnesota, in 1991. The file `wblake` includes only fish of ages 8 or younger.

## Format

This data frame contains the following columns:

**Age** Age at capture (yrs)

**Length** Length at capture (mm)

**Scale** radius of a key scale, mm

## Source

Minnesta Department of Natural Resources

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(wblake)  # excludes fish age 9 or older
```

---

`Whitestar` *Titanic Survivors*

---

## Description

For each person on board the fatal maiden voyage of the ocean liner Titanic, this dataset records sex, age (adult/child), economic status (first/second/third class, or crew) and whether or not that person survived. The name of the company that owned the Titanic was White Star. Several versions of these data exist in the R universe.

## Format

This data frame contains the following columns:

**surv** Number of survivors

**m** survivors + deaths

**class** Crew or passanger class

**age** adult or child

**sex** male or female

## Source

Report on the Loss of the 'Titanic' (S.S.) (1990), *British Board of Trade Inquiry Report* (reprint), Gloucester, UK: Allan Sutton Publishing. Taken from the *Journal on Statistical Education Archive*, submitted by rdawson@husky1.stmarys.ca.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(Whitestar)
```

## wm1 *Simple windmill data*

### Description

Windspeed data collected at a test site for a windmill, and also at a nearby long-term weather site, in Northern South Dakota. Data collected every six hours for all of 2002, except that all of the month of May and a few other observations are missing.

### Format

A data frame with 1116 observations on the following 3 variables.

**Date** A text variable with values like "2002/1/2/6" meaning the reading at 6AM on January 2, 2002

**CSpd** Windspeed in m/s at the candidate site

**RSpd** Windspeed for the reference site

### Source

Mark Ahlstrom and Rolf Miller, WindLogics, Inc.

### References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

### Examples

```
head(wm1)
```

## wm2 *Windmill data with direction information*

### Description

Windspeed data collected at a test site for a windmill, and also at a nearby long-term weather site, in Northern South Dakota. Data collected every six hours for all of 2002, except that all of the month of May and a few other observations missing.

### Format

A data frame with 1116 observations on the following 5 variables.

**Date** A text variable with values like "2002/1/2/6" meaning the reading at 6AM on January 2, 2002

**CSpd** Windspeed in m/s at the candidate site

**RSpd** Windspeed for the reference site

**RDir** Wind direction, in degrees, at the reference site

**Bin** Wind direction binned into 16 equal width bins

## Source

Mark Ahlstrom and Rolf Miller, WindLogics, Inc.

## References

Weisberg, S. (2014). *Applied Linear Regression*, 4th edition. Hoboken NJ: Wiley.

## Examples

```
head(wm2)
```

# Index


∗Topic **datasets**
ais, 3
allshoots, 4
baeskel, 5
BGSall, 6
BigMac2003, 7
BlowBS, 8
Blowdown, 8
brains, 9
cakes, 10
cathedral, 10
caution, 11
Challeng, 12
cloud, 13
domedata, 13
Donner, 14
Downer, 15
drugcost, 16
dwaste, 17
florida, 18
Forbes, 18
ftcollinssnow, 19
ftcollinstemp, 20
fuel2001, 20
galapagos, 21
galtonpeas, 22
Heights, 23
Highway, 23
Hooker, 24
Htwt, 25
jevons, 26
lakemary, 26
lakes, 27
landrent, 28
lathe1, 29
mantel, 29
mile, 30
MinnLand, 31
MinnWater, 32
Mitchell, 33
MWwords, 34
npdata, 35
oldfaith, 35
physics, 36
pipeline, 37
prodscore, 37
rat, 38
Rateprof, 39
Rpdata, 40
salary, 41
salarygov, 42
segreg, 42
shocks, 43
sleep1, 44
snake, 45
sniffer, 45
Stevens, 46
stopping, 47
swan96, 48
turk0, 48
turkey, 49
twins, 50
UBSprices, 50
ufc, 51
UN1, 52
UN11, 53
walleye, 54
water, 54
wblake, 55
Whitestar, 56
wm1, 57
wm2, 57
∗Topic **interface**
alr4Web, 5

ais, 3
allshoots, 4
alr4Web, 5

baeskel, 5
BGSall, 6
BGSboys (BGSall), 6
BGSgirls (BGSall), 6
BigMac2003, 7
BlowBS, 8
Blowdown, 8
brains, 9

cakes, 10
cathedral, 10
caution, 11
Challeng, 12
cloud, 13

domedata, 13
domedata1 (domedata), 13
Donner, 14
Downer, 15
drugcost, 16
dwaste, 17

florida, 18
Forbes, 18
ftcollinssnow, 19
ftcollinstemp, 20
fuel2001, 20

galapagos, 21
galtonpeas, 22

Heights, 23
Highway, 23
Hooker, 24
Htwt, 25

jevons, 26

lakemary, 26
lakes, 27
landrent, 28
lathe1, 29
longshoots (allshoots), 4

mantel, 29
mile, 30
MinnLand, 31
MinnWater, 32
Mitchell, 33
MWwords, 34

npdata, 35

oldfaith, 35

physics, 36
physics1 (physics), 36
pipeline, 37
prodscore, 37

rat, 38
Rateprof, 39
Rpdata, 40

salary, 41
salarygov, 42
segreg, 42
shocks, 43
shortshoots (allshoots), 4
sleep1, 44
snake, 45
sniffer, 45
Stevens, 46
stopping, 47
swan96, 48

turk0, 48
turkey, 49
twins, 50

UBSprices, 50
ufc, 51
ufcdf (ufc), 51
ufcgf (ufc), 51
ufcwc (ufc), 51
UN1, 52
UN11, 53

walleye, 54
water, 54
wblake, 55
Whitestar, 56
wm1, 57
wm2, 57