Package 'Lock5withR'

August 29, 2016

Type Package

Title Datasets for 'Statistics: Unlocking the Power of Data'

Version 1.2.2

Date 2015-12-22

Author Randall Pruim

Maintainer Randall Pruim <rpruim@calvin.edu>

Description Data sets and other utilities for
'Statistics: Unlocking the Power of Data'
by Lock, Lock, Lock, Lock and Lock
(ISBN : 978-0-470-60187-7, http://lock5stat.com/).

License GPL-2

Depends R (>= 2.10)

Suggests lubridate, lattice, mosaic

LazyLoad yes

LazyData yes

RoxygenNote 5.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2015-12-22 20:50:46

R topics documented:

Lock5Data-package   *Lock5 Datasets*

Description

Datasets for Statistics: Unlocking the Power of Data
by Lock, Lock, Lock, Lock and Lock

Author(s)

Robin Lock <rlock@stlawu.edu>

ACS   *American Community Survey*

Description

Data from a sample of individuals in the American Community Survey

Format

A dataset with 1000 observations on the following 9 variables.

- Sex a factor with levels Female and Male
- Age age in years
- Married a factor with levels Married and Not Married
- Income Wages and salary for the past 12 months (in \$1,000's)
- HoursWk Hours of work per week
- Race a factor with levels asian, black, white, or other
- USCitizen a factor wit two levels (Citizen and Noncitizen).
- HealthInsurance a factor with two levels (Insured and Uninsured
- Language a factor with two levels (English and Other)

Details

The American Community Survey, administered by the US Census Bureau, is given every year to a
random sample of about 3.5 million households (about 3% of all US households). Data on a random
sample of 1% of all US residents are made public (after ensuring anonymity), and we have selected
a random sub-sample of n = 1000 from the 2010 data for this dataset.

Source

The full public dataset can be downloaded at http://www.census.gov/programs-surveys/acs/
data/pums.html, and the full list of variables are at http://www.census.gov/programs-surveys/
acs/guidance.html.

AllCountries   *AllCountries*

Description

Data on the countries of the world

Format

A dataset with 213 observations on the following 18 variables.

- Country Name of the country
- Code Three letter country code
- LandArea Size in sq. kilometers
- Population Population in millions
- Energy Energy usage (kilotons of oil)
- Rural Percentage of population living in rural areas
- Military Percentage of government expenditures directed toward the military
- Health Percentage of government expenditures directed towards healthcare
- HIV Percentage of the population with HIV
- Internet Percentage of the population with access to the internet
- Developed A numeric code for the level of development based on kilowatt hours per capita
- kwhPerCap An ordered factor of categories for kilowatt hours per capita, under 2500, 2500 to 5000, or over 5000
- BirthRate Births per 1000 people
- ElderlyPop Percentage of the population at least 65 year old
- LifeExpectancy Average life expectancy (years)
- CO2 CO2 emissions (metric tons per capita)
- GDP Gross Domestic Product (per capita)
- Cell Cell phone subscriptions (per 100 people)
- Electricity Electric power consumption (kWh per capita)

Details

Most data from 2008 to avoid many missing values in more recent years.

Source

Data collected from the World Bank website, worldbank.org.

APMultipleChoice   *AP Multiple Choice*

Description

Correct responses on Advanced Placement multiple choice exams

Format

A dataset with 400 observations on the following variable.

C, D, or E

Details

Correct responses from multiple choice sections for a sample of released Advanced Placement exams

Source

Sample exams from several disciplines at <http://apcentral.collegeboard.com>

April14Temps   *April 14th Temperatures*

Description

Temperatures in Des Moines, IA and San Francisco, CA on April 14th

Format

A data frame with 16 observations on the following 3 variables.

- Year 1995-2010
- DesMoines Temperature in Des Moines (degrees F)
- SanFrancisco Temperature in San Francisco (degrees F)

Details

Average temperature for the day of April 14th in each of 16 years from 1995-2010

Source

The University of Dayton Average Daily Temperature Archive at <http://academic.udayton.edu/kissock/http/Weather/citylistUS.htm>

Examples

```
data(April14Temps)
```

AtmosphericCO2   *Atmospheric carbon dioxide levels by year*

Description

Carbon dioxide levels in the atmosphere over a 50 year span from 1960-2010.

Format

A dataset with 11 observations on the following 2 variables.

- Year Every five years from 1960-2010
- CO2 Carbon dioxide level in parts per million

Source

Dr. Pieter Tans, NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/). Values recorded at the Mauna Loa Observatory in Hawaii.

BaseballHits   *Baseball Hits*

Description

Number of hits and wins for Major League Baseball teams

Format

A dataset with 30 observations on the following 14 variables.

- Team Name of baseball team
- League Either AL or NL
- Wins Number of wins for the season
- Runs Number of runs scored
- Hits Number of hits
- Doubles Number of doubles
- Triples Number of triples
- HomeRuns Number of home runs

- RBI Number of runs batted in
- StolenBases Number of stolen bases
- CaughtStealing Number of times caught stealing
- Walks Number of walks
- Strikeouts Number of stikeouts
- BattingAvg Team batting average

Details

Data from the 2010 Major League Baseball regular season.

Source

http://www.baseball-reference.com/leagues/MLB/2011-standard-batting.shtml

---

BaseballTimes   *Baseball Game Times*

Description

Information for a sample of 30 Major League Baseball games played during the 2011 season.

Format

A dataset with 30 observations on the following 9 variables.

- Away Away team name
- Home Home team name
- Runs Total runs scored (both teams)
- Margin Margin of victory
- Hits Total number of hits (both teams)
- Errors Total number of errors (both teams)
- Pitchers Total number of pitchers used (both teams)
- Walks Total number of walks (both teams)
- Time Elapsed time for game (in minutes)

Details

Data from a sample of boxscores for Major League Baseball games played in August 2011.

Source

http://www.baseball-reference.com/boxes/2011.shtml

Benford *Benford data*

Description

Two examples to test Benford's Law

Format

A dataset with 9 observations on the following 4 variables.

- `Digit` Leading digit (1-9)
- `BenfordP` Expected proportion according to Benford's law
- `Address` Frequency as a first digit in an address
- `Invoices` Frequency as the first digit in invoice amounts

Details

Leading digits from 1188 addresses sampled from a phone book and 7273 amounts from invoices sampled at a company.

Source

Thanks to Prof. Richard Cleary for providing the data

BikeCommute *Bike Commute*

Description

Commute times for two kinds of bicycle

Format

A dataset with 56 observations on the following 9 variables.

- `Bike` Type of material (`Carbon` or `Steel`)
- `Date` Date of the bike commute as a string
- `DMY` Date of the bike commute as a date object
- `Distance` Length of commute (in miles)
- `Time` Total commute time (hours:minutes:seconds)
- `Minutes` Time converted to minutes
- `AvgSpeed` Average speed during the ride (miles per hour)

- TopSpeed Maximum speed (miles per hour)
- Seconds Time converted to seconds
- MonthStr Month MonthNum Numeric value of month
- Month Month coded as one of 1Jan, 2Feb, 3Mar, 4Apr, 5May, 6June, or 7July,

Details

Data from a personal experiment to compare commuting time based on a randomized selection between two bicycles made of different materials.

Source

Thanks to Dr. Groves for providing his data.

References

"Bicycle weight and commuting time: randomised trial," in British Medical Journal, BMJ 2010;341:c6801.

BodyFat   *Body Measurements*

Description

Body Measurements

Format

A data frame with 100 observations on the following 10 variables.

- Bodyfat a numeric vector
- Age a numeric vector
- Weight a numeric vector
- Height a numeric vector
- Neck a numeric vector
- Chest a numeric vector
- Abdomen a numeric vector
- Ankle a numeric vector
- Biceps a numeric vector
- Wrist a numeric vector

Examples

```
data(BodyFat)
```

BodyTemp50	*Body Temperatures*

Description

Sample of 50 body temperatures

Format

A dataset with 50 observations on the following 3 variables.

- BodyTemp Body temperatures in degrees F
- Pulse Pulse rates (beats per minute)
- Sex Female or Male

Details

Body temperatures and pulse rates for a sample of 50 healthy adults.

Source

Shoemaker, "What's Normal: Temperature, Gender and Heartrate", Journal of Statistics Education, Vol. 4, No. 2 (1996)

BootAtlantaCorr	*Bootstrap Correlations for Atlanta Commutes*

Description

Boostrap correlations between Time and Distance for 500 commuters in Atlanta

Format

A data frame with 1000 observations on the following variable.

- CorrTimeDist Correlation between Time and Distance for a bootstrap sample of Atlanta commuters

Details

Correlations for bootstrap samples of Time vs. Distance for the data on Atlanta commuters in CommuteAtlanta.

Source

Computer simulation

Examples

```
data(BootAtlantaCorr)
```

CaffeineTaps   *Caffeine Taps*

Description

Finger tap rates with and without caffeine

Format

A dataset with 20 observations on the following 2 variables.

- `Taps` Number of finger taps in one minute
- `Group` Treatment with levels `Caffeine` and `No Caffeine`
- `Caffeine` a recoding of `Group` with levels `"Yes"` and `"No"`

Details

Results from a double-blind experiment where a sample of male college students to tap their fingers at a rapid rate. The sample was then divided at random into two groups of ten students each. Each student drank the equivalent of about two cups of coffee, which included about 200 mg of caffeine for the students in one group but was decaffeinated coffee for the second group. After a two hour period, each student was tested to measure finger tapping rate (taps per minute). The goal of the experiment was to determine whether caffeine produces an increase in the average tap rate.

Source

Hand, Daly, Lund, McConway and Ostrowski, Handbook of Small Data Sets, Chapman and Hall, London (1994), pp. 40

CAOSExam   *CAOS Exam Scores*

Description

Scores on a pre-test and post-test of basic statistics concepts

Format

A datset with 10 observations on the following 3 variables.

- `Student` ID code for student
- `Pretest` CAOS Pretest score
- `Posttest` CAOS Posttest score

Details

The CAOS (Comprehensive Assessment of Outcomes in First Statistics Course) exam is designed to measure comprehension basic statistical ideas in an introductory statistics course. This dataset has scores for ten students who took the CAOS pre-test at the start of a course and the post-test during the course itself. Each exam consists of 40 multiple choice questions and the score is the percentage correct.

Source

A sample of 10 students from an introductory statisics course.

References

Find out more about the CAOS exam at `http://app.gen.umn.edu/artist/caos.html`

Cereal   *Breakfast Cereals*

Description

Nutrition information for a sample of 30 breakfast cereals

Format

A dataset with 30 observations on the following 10 variables.

- `Name` Brand name of cereal
- `Company` Manufacturer coded as `G`=General Mills, `K`=Kellog's or `Q`=Quaker
- `Serving` Serving size (in cups)
- `Calories` Calories (per cup)
- `Fat` Fat (grams per cup)
- `Sodium` Sodium (mg per cup)
- `Carbs` Carbohydrates (grams per cup)
- `Fiber` Dietary Fiber (grams per cup)
- `Sugars` Sugars (grams per cup)
- `Protein` Protein (grams per cup)

Details

Nutrition contents for a sample of breakfast cererals, derived from nutrition lables. Values are per cup of cereal (rather than per serving).

Source

Cereal data obtained from nutrition labels at http://www.nutritionresource.com/foodcomp2.cfm?id=0800

CocaineTreatment   *Coacaine Treatment*

Description

Relapse/no relapse responses to three different treatments for cocaine addiction

Format

A dataset with 72 observations on the following 2 variables.

- Drug Treatment drug: Desipramine, Lithium, or Placebo
- Relapse Did the patient relapse? no or yes

Details

Data from an experiment to investigate the effectiveness of the two drugs, desipramine and lithium, in the treatment of cocaine addiction. Subjects (cocaine addicts seeking treatment) were randomly assigned to take one of the treatment drugs or a placebo. The response variable is whether or not the subject relapsed (went back to using cocaine) after the treatment.

Source

Gawin, F., et.al., "Desipramine Facilitation of Initial Cocaine Abstinence", Archives of General Psychiatry, 1989; 46(2): 117 - 121.

ColaCalcium *Cola Calcium*

Description

Calcium excretion with diet cola and water

Format

A dataset with 16 observations on the following 2 variables.

- `Drink` Type of drink: `Diet cola` or `Water`
- `Calcium` Amount of calcium excreted (in mg.)

Details

A sample of 16 healthy women aged 18 - 40 were randomly assigned to drink 24 ounces of either diet cola or water. Their urine was collected for three hours after ingestion of the beverage and calcium excretion (in mg.) was measured . The researchers were investigating whether diet cola leaches calcium out of the system, which would increase the amount of calcium in the urine for diet cola drinkers.

Source

Larson, Amin, Olsen, and Poth, Effect of Diet Cola on Urine Calcium Excretion, Endocrine Reviews, 31[3]: S1070, June 2010. These data are recreated from the published summary statistics, and are estimates of the actual data.

CommuteAtlanta *Commute Atlanta*

Description

Commute times and distance for a sample of 500 people in Atlanta

Format

A data frame with 500 observations on the following 5 variables.

St. Louis

Details

Data from the US Census Bureau's American Housing Survey (AHS) which contains information about housing and living conditions for samples from certain metropolitan areas. These data were extracted respondents in the Atlanta metropolitan area. They include only cases where the respondent worked somewhere other than home. Values show the time (in minutes) and distance (in miles) that respondents typically traveled on their commute to work each day as well as age and sex.

Source

Sample chosen using DataFerret at http://www.thedataweb.org/index.html.

Examples

```
data(CommuteAtlanta)
```

CommuteStLouis   *Comute times in St. Louis*

Description

Commute times and distance for a sample of 500 people in St. Louis

Format

A dataset with 500 observations on the following 5 variables.

- City St. Louis
- Age Age of the respondent (in years)
- Distance Commute distance (in miles)
- Time Commute time (in minutes)
- Sex F or M

Details

Data from the US Census Bureau's American Housing Survey (AHS) which contains information about housing and living conditions for samples from certain metropolitan areas. These data were extracted respondents in the St. Louis metropolitan area. They include only cases where the respondent worked somewhere other than home. Values show the time (in minutes) and distance (in miles) that respondents typically traveled on their commute to work each day as well as age and sex.

Source

Sample chosen using DataFerret at http://www.thedataweb.org/index.html.

CompassionateRats   *Compassionate Rats*

Description

Would a rat attempt to free a trapped rat?

Format

A dataset with 30 observations on the following 2 variables.

- Sex Sex of the rat: coded as F or M
- Empathy Freed the trapped rat? no or yes

Details

In a recent study, some rats showed compassion by freeing another trapped rat, even when chocolate served as a distraction and even when the rats would then have to share the chocolate with their freed companion.

Source

Bartal I.B., Decety J., and Mason P., "Empathy and Pro-Social Behavior in Rats," Science, 2011; 224(6061):1427-1430.

CricketChirps   *Cricket Chirps*

Description

Cricket chirp rate and temperature

Format

A dataset with 7 observations on the following 2 variables.

- Temperature Air temperature in degrees F
- Chirps Cricket chirp rate (chirps per minute)

Details

The data were collected by E.A. Bessey and C.A. Bessey who measured chirp rates for crickets and temperatures during the summer of 1898.

Source

From E.A Bessey and C.A Bessey, Further Notes on Thermometer Crickets, American Naturalist, (1898) 32, 263-264.

Digits *Digit counts*

Description

Digits from social security numbers and student selected random numbers

Format

A data frame with 150 observations on the following 7 variables.

- Random Four digit random numbers picked by a sample of students
- RND1 First digit
- RND2 Second digit
- RND3 Third digit
- RND4 Fourth digit
- SSN8 Eighth digit of social security number
- SSN9 Last digit of social security number

Details

A sample of students were asked to pick a random four digit number. The numbers are given in the dataset, along with separate columns for each of the four digits. The data also show the last two digits of each student's social security number (SSN).

Source

In class-student surveys from several classes.

Examples

```
data(Digits)
```

DogOwner   *Dog/Owner matches*

Description

Experiment to match dogs with owners

Format

A data frame with 25 observations on the following variable.

- Match Was the dog correctly paired with it's owner? no or yes

Details

Pictures were taken of 25 owners and their purebred dogs, selected at random from dog parks. Study participants were shown a picture of an owner together with pictures of two dogs (the owner's dog and another random dog from the study) and asked to choose which dog most resembled the owner. Each dog-owner pair was viewed by 28 naive undergraduate judges, and the pairing was deemed "correct" (yes) if the majority of judges (more than 14) chose the correct dog to go with the owner.

Source

Roy and Christenfeld, Do Dogs Resemble their Owners?, Psychological Science, Vol. 15, No. 5, 2004, pp. 361 - 363.

Examples

```
data(DogOwner)
```

ElectionMargin   *Election Margin*

Description

Approval rating and election margin for recent presidential elections

Format

A data frame with 11 observations on the following 5 variables.

- Year Certain election years from 1940-2004
- Candidate Incumbent US president
- Approval Presidential approval rating at time of election
- Margin margin of victory/defeat (as a percentage)
- Result Outcome of the election for the imbumbent: Lost or Won

Details

Data include US Presidential elections since 1940 in which an incumbent was running for president. The approval rating for the sitting president is comapred to the margin of victory/defeat in the election.

Source

Silver, Nate, "Approval Ratings and Re-Election Odds", fivethirtyeight.com, posted January 28, 2011.

Examples

```
data(ElectionMargin)
```

EmployedACS   *Employed in American Community Survey*

Description

Employed individiuals from the American Community Survey (ACS) dataset

Format

A dataset with 431 observations on the following 9 variables.

- sex Female or Male
- Sex 0 or 1
- Age Age (years)
- married 0 or 1
- Married Not Married or Married
- Income Wages and salary for the past 12 months (in \$1,000's)
- HoursWk Hours of work per week
- Race asian, black, white, or other
- UScitizen Citizen or Noncitizen
- USCitizen 0 or 1
- healthInsurance Insured or Uninsured
- HealthInsurance 0 or 1
- codelanguage Native language: English or Other
- codeLanguage Native language: 0 or 1

Details

This is a subset of the ACS dataset including only 431 individuals who were employed.

Several variables in this data set are included in two encodings. (Watch your capitalization.) The lowercase versions have more intuitive codings and can be used to interpret the numerical codes.

Source

The full public dataset can be downloaded at http://www.census.gov/programs-surveys/acs/data/pums.html, and the full list of variables are at http://www.census.gov/programs-surveys/acs/guidance.html.

ExerciseHours   *Exercise Hours*

Description

Amount of exercise per week for students (and other variables)

Format

A data frame with 50 observations on the following 7 variables.

- Year Year in school (1=First year,..., 4=Senior)
- Gender F or M
- Hand Left (l) or Right (r) handed?
- Exercise Hours of exercise per week
- TV Hours of TV viewing per week
- Pulse Resting pulse rate (beats per minute)
- Pierces Number of body piercings

Details

Data from an in-class survey of statistics students asking about amount of exercise, TV viewing, handedness, gender, pulse rate, and number of body piercings.

Source

In-class student survey.

Examples

```
data(ExerciseHours)
```

FacebookFriends   *Facebook Friends*

Description

Data on number of Facebook friends and grey matter density in brain regions related to social perception and associative memory.

Format

A dataset with 40 observations on the following 2 variables.

- `GMdensity` Normalized z-scores of grey matter density in certain brain regions
- `FBfriends` Number of friends on Facebook

Details

A recent study in Great Britain examines the relationship between the number of friends an individual has on Facebook and grey matter density in the areas of the brain associated with social perception and associative memory. The study included 40 students at City University London.

Source

Kanai, R., Bahrami, B., Roylance, R., and Rees, G., "Online social network size is reflected in human brain structure," Proceedings of the Royal Society, 7 April 2012; 279(1732): 1327-1334. Data approximated from information in the article.

FatMice18   *Fat Mice 18*

Description

Weight gain for mice with different nighttime light conditions

Format

A dataset with 18 observations on the following 2 variables.

- `Light` Light treatment: `LD`= normal light/dark cycle or `LL`=bright light at night
- `WgtGain4` Weight gain (grams over a four week period)

Details

This is a subset of the LightatNight dataset, showing body mass gain in mice after 4 weeks for two of the treatment conditions: a normal light/dark cycle (LD) or a bright light on at night (LL).

Source

Fonken, L., et. al., "Light at night increases body mass by shifting time of food intake," Proceedings of the National Academy of Sciences, October 26, 2010; 107(43): 18664-18669.

FishGills12 *FishGills12*

Description

An experiment to look at fish repiration rates in water withdiffernet levels of calcium.

Format

A data frame with 360 observations on the following 2 variables.

- `Calcium` Amount of calcium in the water (mg/L)
- `GillRate` Respiration rate (beats per minute)

Details

Fish were randomly assigned to twelve tanks with different levels (measured in mg/L) levels of calcium. Respiration rate was measured as number of gill beats per minute.

Source

Thanks to Pro. Brad Baldwin for supplying the data

Examples

```
data(FishGills12)
```

FishGills3 *FishGills3*

Description

Respiration rate for fish in three levels of calcium.

Format

A data frame with 360 observations on the following 2 variables.

- `Calcium` Level of Calcium Low 0.71 mg.L Medium 5.24 mg/L High 18.24 mg/L
- `GillRate` Respiration rate (beats per minute)

Details

Fish were randomly assigned to three tanks with different levels (low, medium and high) levels of calcium. Respiration rate was measured as number of gill beats per minute.

Source

Thanks to Pro. Brad Baldwin for supplying the data

Examples

```
data(FishGills3)
```

Flight179   *Flight times*

Description

Flight times for Flight 179 (Boston-SF) and Flight 180 (SF-Boston)

Format

A data frame with 36 observations on the following 3 variables.

- `Date` Date of the flight (5th, 15th and 25th of each month in 2010) as a factor
- `MDY` Date as a date object
- `Flight179` Flying time (Boston-SF) in minutes
- `Flight180` Fllying time (SF-Boston) in minutes

Details

United Airlines Flight 179 is a daily flight from Boston to San Francisco. Flight 180 goes in the other direction (SF to Boston). The data show the airborn flying times for each flight on the three dates each month (5th, 15th and 25th) in 2010.

Source

Data collected from the Bureau of Transportation Statistics website at http://www.bts.gov/xml/ontimesummarystatistics/src/dstat/OntimeSummaryAirtime.xml

Examples

```
data(Flight179)
```

FloridaLakes	*Florida Lakes*

Description

Water quality for a sample of lakes in Florida

Format

A data frame with 53 observations on the following 12 variables.

- `ID` An identifying number for each lake
- `Lake` Name of the lake
- `Alkalinity` Concentration of calcium carbonate (in mg/L)
- `pH` Acidity
- `Calcium` Amount of calcium in water
- `Chlorophyll` Amount of chlorophyll in water
- `AvgMercury` Average mercury level for a sample of fish (large mouth bass) from each lake
- `NumSamples` Number of fish sampled at each lake
- `MinMercury` Minimum mercury level in a sampled fish
- `MaxMercury` Maximum mercury level in a sampled fish
- `ThreeYrStdMercury` Adjusted mercury level to account for the age of the fish
- `AgeData` Mean age of fish in each sample

Details

This dataset describes characteristics of water and fish samples from 53 Florida lakes. Some variables (e.g. Alkalinity, pH, and Calcium) reflect the chemistry of the water samples. Mercury levels were recorded for a sample of large mouth bass selected at each lake.

Source

Lange, Royals, and Connor, Transactions of the American Fisheries Society (1993)

Examples

```
data(FloridaLakes)
```

GlobalInternet   *Global Internet Usage*

Description

Internet usage for several countries

Format

A data frame with 9 observations on the following 3 variables.

- Country Name of country
- PercentFastConnection Percent of internet users with a fast connection
- HoursOnline Average number of hours online in February 2011

Details

The Nielsen Company measured connection speeds on home computers in nine different countries. Variables include the percent of internet users with a fast connection (defined as 2Mb/sec or faster) and the average amount of time spent online, defined as total hours connected to the web from a home computer during the month of February 2011.

Source

NielsenWire, "Swiss Lead in Speed: Comparing Global Internet Connections", April 1, 2011

Examples

```
data(GlobalInternet)
```

GPAGender   *GPA and Gender*

Description

Data from a survey of introductory statistics students.

Format

A dataset with 343 observations on the following 6 variables.

- `Exercise` Hours of exercise (per week)
- `SAT` Combined SAT scores (out of 1600)
- `GPA` Grade Point Average (0.00-4.00 scale)
- `Pulse` Pulse rate (beats per minute)
- `Piercings` Number of body piercings
- `Sex` `Female` or `Male`

Details

This is a subset of the StudentSurvey dataset where cases with missing values have been dropped and gender is coded as a 0/1 indicator variable.

Source

A first day survey over several different introductory statistics classes.

`HappyPlanetIndex` *Happy Planet Index*

Description

Measurements related to happiness and well-being for 143 countries.

Format

A dataset with 143 observations on the following 11 variables.

- `Country` Name of country
- `Region` Three-digit country code
- `Happiness` Score on a 0-10 scale for average level of happiness (10 is happiest)
- `LifeExpectancy` Average life expectancy (in years)
- `Footprint` Ecological footprint – a measure of the (per capita) ecological impact
- `HLY` Happy Life Years - combines life expectancy with well-being
- `HPI` Happy Planet Index (0-100 scale)
- `HPIRank` HPI rank for the country
- `GDPperCapita` Gross Domestic Product (per capita)
- `HDI` Human Development Index
- `Population` Population (in millions)

Details

Data for 143 countries from the Happy Planet Index Project that works to quantify indicators of happiness, well-being, and ecological footprint at a country level.

Source

Data downloaded from http://www.happyplanetindex.org/data/

References

Marks, N., "The Happy Planet Index", www.TED.com/talks, August 29, 2010.

HockeyPenalties   *Hockey Penalties*

Description

Penalty minutes (per game) for NHL teams in 2010-11

Format

A data frame with 30 observations on the following 2 variables.

- Team Name of the team
- PIMperG Average penaly minutes per game

Details

Data give the average numeber of penalty minutes for each of the 30 National HOckey League (NHL) teams during the 2010-11 regualar season.

Source

Data obtained online at www.nhl.com

Examples

```
data(HockeyPenalties)
```

HollywoodMovies2011   *Hollywood Movies in 2011*

Description

Hollywood Movies in 2011

Format

A data frame with 136 observations on the following 14 variables.

- Movie a factor with many levels

- LeadStudio a factor with many levels

- RottenTomatoes a numeric vector

- AudienceScore a numeric vector

- Story a factor with many levels

- Genre a factor with levels Action Adventure Animation Comedy Drama Fantasy Horror Romance Thriller

- TheatersOpenWeek a numeric vector

- BOAverageOpenWeek a numeric vector

- DomesticGross a numeric vector

- ForeignGross a numeric vector

- WorldGross a numeric vector

- Budget a numeric vector

- Profitability a numeric vector

- OpeningWeekend a numeric vector

Examples

```
data(HollywoodMovies2011)
## maybe str(HollywoodMovies2011) ; plot(HollywoodMovies2011) ...
```

HomesForSale *Home for Sale*

Description

Data on homes for sale in four states

Format

A data frame with 120 observations on the following 5 variables.

- State Location of the home: CA NJ NY PA
- Price Asking price (in \$1,000's)
- Size Area of all rooms (in 1,000's sq. ft.)
- Beds Number of bedrooms
- Baths Number of bathrooms

Details

Data for samples of homes for sale in each state, selected from zillow.com.

References

Data collected from www.zillow.com in 2010.

Examples

```
data(HomesForSale)
```

HomesForSaleCA *Home for Sale in California*

Description

Data for a sample of homes offered for sale in California

Format

A data frame with 30 observations on the following 5 variables.

- State Location of the home: CA
- Price Asking price (in \$1,000's)
- Size Area of all rooms (in 1,000's sq. ft.)
- Beds Number of bedrooms
- Baths Number of bathrooms

Details

This is a subset of the HomesForSale data with just information from homes in California (CA). Data were collected offerings listed on an online site.

Source

Data collected from www.zillow.com in 2010.

Examples

```
data(HomesForSaleCA)
```

HomesForSaleCanton   *Homes for sale in Canton, NY*

Description

Prices of homes for sale in Canton, NY

Format

A data frame with 10 observations on the following variable.

- `Price` Asking price for the home (in \$1,000's)

Details

Data for samples of homes for sale in Canton, NY, selected from zillow.com.

Source

Data collected from www.zillow.com in 2010.

Examples

```
data(HomesForSaleCanton)
```

HomesForSaleNY   *Home for Sale in New York*

Description

Data for a sample of homes offered for sale in New York State

Format

A data frame with 30 observations on the following 5 variables.

- State Location of the home: NY
- Price Asking price (in \$1,000's)
- Size Area of all rooms (in 1,000's sq. ft.)
- Beds Number of bedrooms
- Baths Number of bathrooms

Details

This is a subset of the HomesForSale data with just information from homes in New York State (NY). Data were collected offerings listed on an online site.

Source

Data collected from www.zillow.com in 2010.

Examples

```
data(HomesForSaleNY)
```

HoneybeeCircuits   *Honeybee Circuits*

Description

Number of circuits for honeybee dances and nest quality

Format

A dataset with 78 observations on the following 2 variables.

- Circuits Number of waggle dance circuits for a returning scout bee
- Quality Quality of the nest site: High or Low

Details

When honeybees are looking for a new home, they send out scouts to explore options. When a scout returns, she does a "waggle dance" with multiple circuit repetitions to tell the swarm about the option she found. The bees then decide between the options and pick the best one. Scientists wanted to find out how honeybees decide which is the best option, so they took a swarm of honeybees to an island with only two possible options for new homes: one of very high honeybee quality and one of low quality. They then kept track of the scouts who visited each option and counted the number of waggle dance circuits each scout bee did when describing the option.

Source

Seeley, T., Honeybee Democracy, Princeton University Press, Princeton, NJ, 2010, p. 128

HoneybeeWaggle   *Honeybee Waggle*

Description

Honeybee dance duration and distance to nesting site

Format

A dataset with 7 observations on the following 2 variables.

- `Distance` Distance to the potential nest site (in meters)
- `Duration` Duration of the waggle dance (in seconds)

Details

When honeybee scouts find a food source or a nice site for a new home, they communicate the location to the rest of the swarm by doing a "waggle dance." They point in the direction of the site and dance longer for sites farther away. The rest of the bees use the duration of the dance to predict distance to the site.

Source

Seeley, T., Honeybee Democracy, Princeton University Press, Princeton, NJ, 2010, p. 128

HotDogs   *Hot Dog Eating Contest*

Description

Winning number of hot dogs consumed in an eating contest

Format

A dataset with 10 observations on the following 2 variables.

- Year Year of the contest: 2002-2011
- HotDogs Winning number of hot dogs consumed

Details

Every Fourth of July, Nathan's Famous in New York City holds a hot dog eating contest, in which contestants try to eat as many hot dogs (with buns) as possible in ten minutes. The winning number of hot dogs are given for each year from 2002-2011.

Source

Nathan's Famous webste at http://nathansfamous.com/contest/hall_of_fame

---

ICUAdmissions   *Intensive Care Unit Admissions*

Description

data from patients admitted to an intensive care unit

Format

A data frame with 200 observations on the following 21 variables.

- ID Patient ID number
- status Patient status: Lived or Died
- Status numerical code for Status
- Age Patient's age (in years)
- sex Male or Female
- Sex numerical code for sex
- race Patient's race: White, Black, or Other
- Race numerical code for race
- service Type of service: Medical or Surgical

- Service numerical code for `service`
- `cancer` Is cancer involved? `No` or `Yes`
- `Cancer` Is cancer involved? `0` or `1`
- `renal` Is chronic renal failure involved? `No` or `Yes`
- `Renal` Is chronic renal failure involved? `0` or `1`
- `infection` Is infection involved? `No` or `Yes`
- `Infection` Is infection involved? `0` or `1`
- `cpr` Patient gets CPR prior to admission? `No` or `Yes`
- `CPR` Patient gets CPR prior to admission? `0` or `1`
- `Systolic` Systolic blood pressure (in mm of Hg)
- `HeartRate` Pulse rate (beats per minute)
- `previous` Previous admission to ICU wihtin 6 months? `No` or `Yes`
- `Previous` Previous admission to ICU wihtin 6 months? `0` or `1`
- `type` Admission type: `Elective` or `Emergency`
- `Type` Admission type: `0` or `1`
- `fracture` Fractured bone involved? `No` or `Yes`
- `Fracture` Fractured bone involved? `0` or `1`
- `pO2` Partial oxygen level from blood gases under 60? `No` or `Yes`
- `PO2` Partial oxygen level from blood gases under 60? `0` or `1`
- `pHlow` pH from blood gas under 7.25? `No` or `Yes`
- `pH` pH from blood gas under or over 7.25? `Low` or `Hi`
- `PH` pH from blood gas under or over 7.25? `0` or `1`
- `pCO2hi` Partial carbon dioxide level from blood gas over 45? `No` or `Yes`
- `pCO2` Partial carbon dioxide level from blood gas over or under 45? `Low` or `Hi`
- `PCO2` Partial carbon dioxide level from blood gas over or under 45? `0` or `1`
- `bicarbonateLow` Bicarbonate from blood gas under 18? `No` or `Yes`
- `bicarbonate` Bicarbonate from blood gas under or over 18? `Low` or `Hi`
- `Bicarbonate` Bicarbonate from blood gas under or over 18? `0` or `1`
- `creatinineHi` Creatinine from blood gas over 2.0? `No` or `Yes`
- `creatinine` Creatinine from blood gas over or under 2.0? `Low` or `Hi`
- `Creatinine` Creatinine from blood gas over or under 2.0? `0` or `1`
- `consciousness` Levels: `Conscious`, `Deep Stupor`, or `Coma`
- `Consciousness` Levels: 0, 1, or 2

Details

Data from a sample of 200 patients following admission to an adult intensive care unit (ICU).

Source

DASL dataset downloaded from <http://lib.stat.cmu.edu/DASL/Datafiles/ICU.html>

Examples

```
data(ICUAdmissions)
```

ImmuneTea   *Immune Tea*

Description

Interferon gamma production and tea drinking

Format

A data frame with 21 observations on the following 2 variables.

- `InterferonGamma` Measure of interferon gamma production
- `Drink` Type of drink: `Coffee` or `Tea`

Details

Eleven healthy non-tea-drinking individuals were asked to drink five or six cups of tea a day, while ten healthy non-tea and non-coffee-drinkers were asked to drink the same amount of coffee, which has caffeine but not the L-theanine that is in tea. The groups were randomly assigned. After two weeks, blood samples were exposed to an antigen and production of interferon gamma was measured.

Source

Adapted from Kamath, et.al., "Antigens in tea-Beverage prime human V 2V2 T cells in vitro and in vivo for memory and non-memory antibacterial cytokine responses", Proceedings of the National Academy of Sciences, May 13, 2003.

Examples

```
data(ImmuneTea)
```

```
InkjetPrinters          Inkjet Printers
```

Description

Data from online reviews of inkjet printers

Format

A dataset with 20 observations on the following 6 variables.

- `Model` Model name of printer
- `PPM` Printing rate (pages per minute) for a benchmark set of print jobs
- `PhotoTime` Time (in seconds) to print 4x6 color photos
- `Price` Typical retail price (in dollars)
- `CostBW` Cost per page (in cents) for printing in black & white
- `CostColor` Cost per page (in cents) for printing in color

Details

Information from reviews of inkjet printers at PCMag.com in August 2011.

Source

Inkjet printer reviews found at <http://www.pcmag.com/reviews/printers>, August 2011.

```
LifeExpectancyVehicles
                    Life Expectancy and Vehicle Registrations
```

Description

Life Expectancy and Vehicle Registrations

Format

A data frame with 40 observations on the following 3 variables.

- `Year` a numeric vector
- `LifeExpectancy` a numeric vector
- `Vehicles` a numeric vector

Examples

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

`LightatNight`	*Light at Night for Mice*

Description

Data from an experiment with mice having different nighttime light conditions

Format

A dataset with 27 observations on the following 9 variables.

- `Light` DM=dim light at night, LD=dark at night, or LL=bright light at night
- `BMGain` Body mass gain (in grams over a four week period)
- `Corticosterone` Blood corticosterene level (a measure of stress)
- `DayPct` Percent of calories eaten during the day
- `Consumption` Daily food consumption (grams)
- `GlucoseInt` Glucose intolerant? No or Yes
- `GTT15` Glucose level in the blood 15 minutes after a glucose injection
- `GTT120` Glucose level in the blood 120 minutes after a glucose injection
- `Activity` A measure of physical activity level

Details

In this study, 27 mice were randomly split into three groups. One group was on a normal light/dark cycle (LD), one group had bright light on all the time (LL), and one group had light during the day and dim light at night (DM). The dim light was equivalent to having a television set on in a room. The mice in darkness ate most of their food during their active (nighttime) period, matching the behavior of mice in the wild. The mice in both dim light and bright light, however, consumed more than half of their food during the well-lit rest period, when most mice are sleeping.

Source

Fonken, L., et. al., "Light at night increases body mass by shifting time of food intake," Proceedings of the National Academy of Sciences, October 26, 2010; 107(43): 18664-18669.

---

locket   *Display or execute a snippet of R code*

Description

This command will display and/or execute small snippets of R code from the book *Statistics: Un-Locking the Power of Data*. locket works much like [demo](), but the interface is simplified.

Usage

```
locket(name, execute = TRUE, view = !execute, echo = TRUE,
  ask = getOption("demo.ask"), verbose = getOption("verbose"),
  lib.loc = NULL, character.only = FALSE)
```

Arguments

name	name of snippet
execute	a logical. If TRUE, snippet code is executed. (The code and the results of the execution will be visible if echo is TRUE.)
view	a logical. If TRUE, snippet code is displayed 'as is'.
echo	a logical. If TRUE, show the R input when executing.
ask	a logical (or "default") indicating if devAskNewPage(ask=TRUE) should be called before graphical output happens from the snippet code. The value "default" (the factory-fresh default) means to ask if echo == TRUE and the graphics device appears to be interactive. This parameter applies both to any currently opened device and to any devices opened by the demo code. If this is evaluated to TRUE and the session is interactive, the user is asked to press RETURN to start.
verbose	a logical. If TRUE, additional diagnostics are printed.
lib.loc	character vector of directory names of R libraries, or NULL. The default value of NULL corresponds to all libraries currently known.
character.only	logical. If TRUE, use nameas character string.

See Also

[demo](), [source]().

MalevolentUniformsNFL   *Malevolent Uniforms NFL*

Description

Perceived malevolence of uniforms and penalies for National Football League (NFL) teams

Format

A data frame with 28 observations on the following 7 variables.

- NFLTeam Team name
- NFL_Malevolence Score reflecting the "malevolence" of a team's uniform
- ZPenYds Z-score for penalty yards

Details

Participants with no knowledge of the teams rated the jerseys on characteristics such as timid/aggressive, nice/mean and good/bad. The averages of these responses produced a "malevolence" index with higher scores signifying impressions of more malevolent uniforms. To measure aggressiveness, the authors used the amount of penalty yards converted to z-scores and averaged for each team over the seasons from 1970-1986.

Source

Frank and Gilovich, "The Dark Side of Self- and Social Perception: Black Uniforms and Aggression in Professional Sports", Journal of Personality and Social Psychology, Vol. 54, No. 1, 1988, p. 74-85.

Examples

```
data(MalevolentUniformsNFL)
```

MalevolentUniformsNHL   *Malevolent Uniforms NHL*

Description

Perceived malevolence of uniforms and penalies for National Hockey League (NHL) teams

Format

A data frame with 28 observations on the following 3 variables.

- `NHLTeam` Team name
- `NHL_Malevolence` Score reflecting the "malevolence" of a team's uniform
- `ZPenMin` Z-score for penalty minutes

Details

Participants with no knowledge of the teams rated the jerseys on characteristics such as timid/aggressive, nice/mean and good/bad. The averages of these responses produced a "malevolence" index with higher scores signifying impressions of more malevolent uniforms. To measure aggressiveness, the authors used the amount of penalty minutes converted to z-scores and averaged for each team over the seasons from 1970-1986.

Source

Frank and Gilovich, "The Dark Side of Self- and Social Perception: Black Uniforms and Aggression in Professional Sports", Journal of Personality and Social Psychology, Vol. 54, No. 1, 1988, p. 74-85.

Examples

```
data(MalevolentUniformsNHL)
```

MammalLongevity   *Mammal Longevity*

Description

Longevity and gestation period for mammals

Format

A data frame with 40 observations on the following 3 variables.

- `Animal` Species of mammal
- `Gestation` Time from fertilization until birth (in days)
- `Longevity` Average lifespan (in years)

Details

Dataset with average lifespan (in years) and typical gestation period (in days) for 40 different species of mammals.

Source

2010 World Almanac, pg. 292.

Examples

```
data(MammalLongevity)
```

ManhattanApartments   *Manhattan Apartment Prices*

Description

Monthly rent for one-bedroom apartments in Manhattan, NY

Format

A dataset with 20 observations on the following variable.

   • Rent Montly rent in dollars

Details

Monthly rents for a sample of 20 one-bedroom apartments in Manhattan, NY that were advertised
on Craig's List in July, 2011.

Source

Apartments advertised on Craig's List at http://newyork.craigslist.org, July 5, 2011.

MarriageAges   *Marriage Ages*

Description

Ages for husbands and wives from marriage licenses

Format

A data frame with 100 observations on the following 2 variables.

   • Husband Age of husband at marriage
   • Wife Age of wife at marriage

Details

Data from a sample of 100 marriage licences in St. Lawrence County, NY gives the ages of husbands and wives for newly married couples.

Source

Thanks to Linda Casserly, St. Lawrence County Clerk's Office

Examples

```
data(MarriageAges)
```

MastersGolf   *Masters Golf Scores*

Description

Scores from the 2011 Masters golf tournament

Format

A dataset with 20 observations on the following 2 variables.

- First First round score (in relation to par)
- Final Final four round score (in relation to par)

Details

Data for a random sample of 20 golfers who made the cut at the 2011 Masters golf tournament.

MentalMuscle   *Mental Muscle*

Description

Comparing actual movements to mental imaging movements

Format

A dataset with 32 observations on the following 3 variables.

- Action Treatment: Actual motions or Mental imaging motions
- PreFatigue Time (in seconds) to complete motions before fatigue
- PostFatigue Time (in seconds) to complete motions after fatigue

Details

In this study, participants were asked to either perform actual arm pointing motions or to mentally imagine equivalent arm pointing motions. Participants then developed muscle fatigue by holding a heavy weight out horizontally as long as they could. After becoming fatigued, they were asked to repeat the previous mental or actual motions. Eight participants were assigned to each group, and the time in seconds to complete the motions was measured before and after fatigue.

Source

Data approximated from summary statistics in: Demougeot L. and Papaxanthis C., "Muscle Fatigue Affects Mental Simulation of Action," The Journal of Neuroscience, July 20, 2011, 31(29):10712-10720.

MiamiHeat   *Miami Heat basketball*

Description

Game log data for the Miami Heat basketball team in 2010-11

Format

A data frame with 82 observations on the following 33 variables.

- Game ID number for each game
- MDY Data the game was played as a date object
- Date Data the game was played as a character string
- Location Away or Home
- Opp Opponent tream
- Win Game result: L or W
- FG Field goals made
- FGA Field goals attempted
- FG3 Three-point field goals made
- FG3A Three-point field goals attempted
- FT Free throws made
- FTA Free throws attempted
- Rebounds Total rebounds
- OffReb Offensive rebounds
- Assists Number of assists
- Steals Number of steals
- Blocks Number of shots blocked
- Trunovers Number of turnovers

- `Fouls` Number of fouls

- `Points` Number of points scored

- `OppFG` Opponet's field goals made

- `OppFGA` Opponent's Field goals attempted

- `OppFG3` Opponent's Three-point field goals made

- `OppFG3A` Opponent's Three-point field goals attempted

- `OppFT` Opponent's Free throws made

- `OppFTA` Opponent's Free throws attempted

- `OppOffReb` Opponent's Total rebounds

- `OppRebounds` Opponent's Offensive rebounds

- `OppAssists` Opponent's assists

- `OppSteals` Opponent's steals

- `OppBlocks` Opponent's shots blocked

- `OppTurnovers` Opponent's turnovers

- `OppFouls` Opponent's fouls

- `OppPoints` Opponent's points scored

Details

Information from online boxscores for all 82 regular season games payed by the Miami Heat basketball team during the 2010-11 regular season.

Source

Data for the 2010-11 Miami games downloaded from http://www.basketball-reference.com/teams/MIA/2011/gamelog/

Examples

```
data(MiamiHeat)
```

MindsetMatters *Mindset Matters*

Description

Data from a study of perceived exercise with maids

Format

A data frame with 75 observations on the following 14 variables.

- `Condition` Treatment condition: `uninformed` or `informed`
- `Cond` Treatment condition: `0`=uninformed or `1`=informed
- `Age` Age (in years)
- `Wt` Original weight (in pounds)
- `Wt2` Weight after 4 weeks (in pounds
- `BMI` Original body mass index
- `BMI2` Body mass index after 4 weeks
- `Fat` Original body fat percentage
- `Fat2` Body fat percentage after 4 weeks
- `WHR` Original waist to hip ratio
- `WHR2` Waist to hip ratio
- `Syst` Original systolic blood pressure
- `Syst2` Systolic blood pressure after 4 weeks
- `Diast` Original diastolic blood pressure
- `Diast2` Diastolic blood pressure after 4 weeks

Details

In 2007 a Harvard psychologist recruited 75 female maids working in different hotels to participate in a study. She informed 41 maids (randomly chosen) that the work they do satisfies the Surgeon General's recommendations for an active lifestyle (which is true), giving them examples for how why their work is good exercise. The other 34 maids were told nothing (uninformed). Various chacteristics (weight, body mass index, ...) were recorded for each subject at the start of the experiment and again four weeks later. Maids with missing values for weight change have been removed.

Source

Crum, A.J. and Langer, E.J. (2007). Mind-Set Matters: Exercise and the Placebo Effect, Psychological Science, 18:165-171. Thanks to the authors for supplying the data.

Examples

```
data(MindsetMatters)
```

MinistersRum   *Ministers and Rum*

Description

Number of ministers and rum imports to New England between 1860 and 1940

Format

A data frame with 17 observations on the following 3 variables.

- `Year` Every five years from 1860-1940
- `Ministers` Number of Methodist ministers in New England
- `Rum` Number of barrels of rum imported into Boston

Details

Data every five years from 1860 to 1940 on he number of Methodist ministers working in New England and the annual rum imports into Boston.

Examples

```
data(MinistersRum)
```

MustangPrice   *Mustang Prices*

Description

Price, age, and mileage for used Mustangs at an internet website

Format

A data frame with 25 observations on the following 3 variables.

- `Age` Age of the car (in years)
- `Miles` Mileage on the car (in 1,000's)
- `Price` Asking price in (\$1,000's)

Details

A statistics student, Gabe McBride, was interested in prices for used Mustang cars being offered for sale on an internet site. He sampled 25 cars from the website and recorded the age (in years), mileage (in thousands of miles) and asking price (in \$1,000's) for each car in his sample.

Source

Student project with data collected from autotrader.com in 2008.

Examples

```
data(MustangPrice)
```

NBAPlayers2011   *NBA Players data for 2010-11 Season*

Description

Data from the 2010-2011 regular season for 176 NBA basketball players.

Format

A data frame with 176 observations on the following 25 variables.

- Player Name of player
- Age Age in years)
- Team Team name
- Games Games played (out of 82)
- Starts Games started
- Mins Minutes played
- MinPerGame Minutes per game
- FGMade Field goals made
- FGAttempt Field goals attempted
- FGPct Three-point field goal percentage
- FG3Made Three-point field goals made
- FG3Attempt Three-point field goals attempted
- FG3Pct Field goal percentage
- FTMade Free throws made
- FTAttempt Free throws attempted
- FTPct Free throw percentage
- OffRebound Offensive rebounds
- DefRebound Defensive rebounds
- Rebounds Total rebounds
- Assists Number of assists
- Steals Number of steals

- Blocks Number of blocked shots
- Turnovers Number of turnovers
- Fouls Number of personal fouls
- Points Number of points scored

Source

http://www.basketball-reference.com/leagues/NBA_2011_stats.html

Examples

```
data(NBAPlayers2011)
```

NBAStandings   *NBA 2010-11 Regular Season Standings*

Description

Won-Loss record and statistics for NBA Teams

Format

A data frame with 30 observations on the following 5 variables.

- Team Team name
- Wins Number of wins in an 82 game regular season
- Losses Number of losses
- WinPct Proportion of games won
- PtsFor Average points scored per game
- PtsAgainst Average points allowed per game

Details

Won-Loss record and regular season statistics for 30 teams in the National Basketball Association for the 2010-2011 season

Source

Data downloaded from http://www.basketball-reference.com/leagues/NBA_2011_games.html

Examples

```
data(NBAStandings)
```

NFLScores2011   *NFL Game Scores in 2011*

Description

Results for all NFL games for the 2011 regular season

Format

A dataset with 256 observations on the following 11 variables.

- Week a numeric vector
- HomeTeam Home team name
- AwayTeam Visiting team name
- HomeScore Points scored by the home team
- AwayScore Points scored by the visiting team
- HomeYards Yards gained by the home team
- AwayYards Yards gained by the visiting team
- HomeTO Turnovers lost by the home team
- AwayTO Turnovers lost by the visiting team
- Date Date of the game (as a character string)
- YDM Date of the game (as a date object)
- Day Day of the week: Mon, Sat, Sun, or Thu

Details

Data for all 256 regular season games in the National Football League (NFL) for the 2011 season.

Source

NFL scores and game statistics found at http://www.pro-football-reference.com/years/2011/games.htm.

`NutritionStudy`   *Nutrition Study*

Description

Variables related to nutrition and health for 315 individuals

Format

A data frame with 315 observations on the following 17 variables.

- `ID` ID number for each subject in this sample
- `Age` Subject's age (in years)
- `Smoke` a factor with levels `No` `Yes`
- `Quetelet` Weight/(Height^2)
- `Vitamin` Vitamin use: 1=Regular, 2=Occasional, or 3=No
- `Calories` Number of calories consumed per day
- `Fat` Grams of fat consumed per day
- `Fiber` Grams of fiber consumed per day
- `Alcohol` Number of alcoholic drinks consumed per week
- `Cholesterol` Cholesterol consumed (mg per day)
- `BetaDiet` Dietary beta-carotene consumed (mcg per day)
- `RetinolDiet` Dietary retinol consumed (mcg per day)
- `BetaPlasma` Plasma beta-carotene (ng/ml)
- `RetinolPlasma` Plasma retinol (ng/ml)
- `Sex` Cosed as `Female` or `Male`
- `VitaminUse` Coded as `No` `Occasional` `Regular`
- `EverSmoke` Smoking status: `Never`, `Former`, or `Current`
- `PriorSmoke` Smoking status: 1, 2, or 3

Details

Data from a cross-sectional study to investigate the relationship between personal characteristics and dietary factors, and plasma concentrations of retinol, beta-carotene and other carotenoids. Study subjects were patients who had an elective surgical procedure during a three-year period to biopsy or remove a lesion of the lung, colon, breast, skin, ovary or uterus that was found to be non-cancerous.

Source

Nierenberg, Stukel, Baron, Dain, and Greenberg, "Determinants of plasma levels of beta-carotene and retinol", American Journal of Epidemiology (1989).

References

Data downloaded from <http://lib.stat.cmu.edu/DASL/>.

Examples

```
data(NutritionStudy)
```

OlympicMarathon   *2008 Olympic Men's Marathon*

Description

Times for all finishers in the men's marathon at the 2008 Olympics

Format

A data frame with 76 observations on the following 5 variables.

- Rank Order of finish
- Athlete Name of marathoner
- Nationality Country of marathoner
- Time Time as H:MM:SS
- Minutes Time in minutes

Details

Results for all finishers in the 2008 Men's Olympic marathon in Beijing, China.

Source

<http://2008olympics.runnersworld.com/2008/08/mens-marathon-results.html>

Examples

```
data(OlympicMarathon)
```

OttawaSenators   *Ottaw Senators hockey team*

Description

Data for 24 players on the 2009-10 Ottawa Senators

Format

A data frame with 24 observations on the following 2 variables.

- Points Number of points (goals + assists) scored
- PenMins Number of penalty minutes

Details

Points scored and penaly minutes for 24 players (excluding goalies) playing ice hockey for the Ottawa Senators during the 2009-10 NHL regular season.

Source

Data obtained from http://senators.nhl.com/club/stats.htm.

Examples

```
data(OttawaSenators)
```

PizzaGirl   *Pizza Girl Tips*

Description

Data on tips for pizza deliveries

Format

A dataset with 24 observations on the following 2 variables.

- Tip Amount of tip (in dollars)
- Shift Which of three different shifts

Details

"Pizza Girl" collected data on her deliveries and tips over three different evening shifts.

Source

Pizza Girl: Statistical Analysis at `http://slice.seriouseats.com/archives/2010/04/statistical-analysis-of-a-`
`html`.

QuizPulse10	*Quiz vs Lecture Pulse Rates*

Description

Paired data with pulse rates in a lecture and during a quiz for 10 students

Format

A data frame with 10 observations on the following 3 variables.

- Student ID number for the student
- Quiz Pulse rate (beats per minute) during a quiz
- Lecture Pulse rate (beats per minute) during a lecture

Details

Ten students in an introductory statistics class measured their pulse rate in two settings: in the middle of a regular class lecture and again while taking a quiz.

Source

In-class data collection

Examples

```
data(QuizPulse10)
```

RandomP50N200	*Simulated proportions*

Description

Counts and proportions for 5000 simulated samples with n=200 and p=0.50

Format

A data frame with 5000 observations on the following 2 variables.

- Count Number of simulated "yes" responses in 200 trials
- Phat Sample proportion (Count/200)

Details

Results from 5000 simulations of samples of size n=200 from a population with proportoin of "yes" responses at p=0.50.

Source

Computer simulation

Examples

```
data(RandomP50N200)
```

RestaurantTips	*Restaurant Tips*

Description

Tip data from the First Crush Bistro

Format

A data frame with 157 observations on the following 7 variables.

- Bill Size of the bill (in dollars)
- Tip Size of the tip (in dollars)
- CreditCard Paid with a credit card? No or Yes
- Credit Paid with a credit card? n or y
- Guests Number of peole in the group
- Day Day of the week: m=Monday, t=Tuesday, w=Wednesday, th=Thursday, or f=Friday
- Server Code for waiter/waitress: A, B, or C
- PctTip Tip as a percentage of the bill

Details

The owner of a bistro called First Crush in Potsdam, NY was interested in studying the tipping patterns of his customers. He collected restaurant bills over a two week period that he believes provide a good sample of his customers. The data recorded from 157 bills include the amount of the bill, size of the tip, percentage tip, number of customers in the group, whether or not a credit card was used, day of the week, and a coded identity of the server.

Source

Thanks to Tom DeRosa for providing the tipping data.

Examples

```
data(RestaurantTips)
```

RetailSales   *Retail Sales*

Description

Monthly U.S. Retail Sales (in billions)

Format

A data frame with 144 observations on the following 3 variables.

- Month Month of the year
- Year Year (from 2002 to 2011)
- Date Date in date format (day of month is meaningless)
- Sales U.S. retail sales (in billions of dollars)

Details

Data show the monthly retail sales (in billions) for the U.S. economy in each month from 2002 through 2011.

Source

http://www.census.gov/retail/

Examples

```
data(RetailSales)
if (require(lattice)) {
  xyplot(Sales ~ Date, RetailSales, type='l')
  xyplot(Sales ~ Date, RetailSales, type='l', groups=Month)
}
```

RockandRoll *Rock & Roll Hall of fame*

Description

Groups and Individuals in the Rock and Roll hall of Fame

Format

A data frame with 273 observations on the following 4 variables.

- Inductee Name of he group or individual
- FemaleMembers Yes if individual or membr of the group is female, otherwise No
- Category Type of indiviudal or group: Performer, Non-performer, Early INfluence, Lifetime Achievement, Sideman
- People Number of people in the group

Details

All inductees of the Rock & Roll Hall of Fame as of 2012.

Source

Rock & Roll Hall of Fame website, <http://rockhall.com/inductees/alphabetical/>

Examples

```
data(RockandRoll)
```

SalaryGender *Salary and Gender*

Description

Salaries for college teachers

Format

A dataset with 100 observations on the following 4 variables.

- Salary Annual salary in \$1,000's
- Gender 0=female or 1=male
- Sex Female or Male
- Age Age in years
- phD No or Yes
- PhD 0 or 1

Details

A random sample of college teachers taken from the 2010 American Community Survey (ACS) 1-year Public Use Microdata Sample (PUMS).

Source

SampCountries   *AllCountries*

Description

Data on a sample of countries of the world

Format

A data frame with 50 observations on the following 13 variables.

- Country Name of the country
- LandArea Size in sq. kilometers
- Population Population in millions
- Energy Energy usage (kilotons of oil)
- Rural Percentage of population living in rural areas
- Military Percentage of government expenditures directed toward the military
- Health Percentage of government expenditures directed towards healthcare
- HIV Percentage of the population with HIV
- Internet Percentage of the population with access to the internet
- kwhPerCap An ordered factor of categories for kilowatt hours per capita, under 2500, 2500 to 5000, or over 5000
- Developed A numerical code for kwhPerCap
- BirthRate Births per 1000 people
- ElderlyPop Percentage of the population at least 65 year old
- LifeExpectancy Average life expectancy (in years)

Details

A subset of data from AllCountries for a random sample of 50 countries. Data for 2008 to avoid many missing values in more recent years.

Source

Data collected from the World Bank website, worldbank.org.

Examples

```
data(SampCountries)
```

SandP500	*S \& P 500 Prices*

Description

Daily data for S\& P 500 Stock Index

Format

A data frame with 252 observations on the following 6 variables.

- Date Date as a character string
- MDY Date as a date object
- Open Opening value
- High High point for the day
- Low Low point for the day
- Close Closing value
- Volume Shares traded (in millions)

Details

Daily prices for the S\& P 500 Stock Index for trading days in 2010.

Source

Downladed from http://finance.yahoo.com/q/hp?s=^GSPC+Historical+Prices

Examples

```
data(SandP500)
if (require(lattice)) {
  xyplot( High + Low ~ Date, data=SandP500, type="l",
    main="S and P 500",
    auto.key=list(lines=TRUE, points=FALSE))
 }
```

SandwichAnts *Sandwich Ants*

Description

Ant Counts on samples of different sandwiches

Format

A data frame with 24 observations on the following 5 variables.

- Butter Butter on the sandwich? no
- Filling Type of filling: Ham & Pickles, Peanut Butter, or Vegemite
- Bread Type of bread: Multigrain, Rye, White, or Wholemeal
- Ants Number of ants on the sandwich
- Order Trial number

Details

As young students, Dominic Kelly and his friends enjoyed watching ants gather on pieces of sandwiches. Later, as a university student, Dominic decided to study this with a more formal experiment. He chose three types of sandwich fillings (vegemite, peanut butter, and ham \& pickles), four types of bread (multigrain, rye, white, and wholemeal), and put butter on some of the sandwiches.

To conduct the experiment he randomly chose a sandwich, broke off a piece, and left it on the ground near an ant hill. After several minutes he placed a jar over the sandwich bit and counteed the number of ants. He repeated the process, allowing time for ants to return to the hill after each trial, until he had two samples for each combination of the factors.

This dataset has only sandwiches with no butter. The data in SandwichAnts2 adds information for samples with butter.

Source

Margaret Mackisack, "Favourite Experiments: An Addendum to What is the Use of Experiments Conducted by Statistics Students?", Journal of Statistics Education (1994) http://www.amstat.org/publications/jse/v2n1/mackisack.supp.html

Examples

```
data(SandwichAnts)
```

SandwichAnts2	*Sandwich Ants - Part 2*

Description

Ant counts on samples of different sandwiches

Format

A data frame with 48 observations on the following 5 variables.

- `Butter` Butter on the sandwich? `no`
- `Filling` Type of filling: `Ham & Pickles`, `Peanut Butter`, or `Vegemite`
- `Bread` Type of bread: `Multigrain`, `Rye`, `White`, or `Wholemeal`
- `Ants` Number of ants on the sandwich
- `Order` Trial number

Details

As young students, Dominic Kelly and his friends enjoyed watching ants gather on pieces of sandwiches. Later, as a university student, Dominic decided to study this with a more formal experiment. He chose three types of sandwich fillings (vegemite, peanut butter, and ham \& pickles), four types of bread (multigrain, rye, white, and wholemeal), and put butter on some of the sandwiches.

To conduct the experiment he randomly chose a sandwich, broke off a piece, and left it on the ground near an ant hill. After several minutes he placed a jar over the sandwich bit and counteed the number of ants. He repeated the process, allowing time for ants to return to the hill after each trial, until he had two samples for each combination of the three factors.

Source

Margaret Mackisack, "Favourite Experiments: An Addendum to What is the Use of Experiments Conducted by Statistics Students?", Journal of Statistics Education (1994) http://www.amstat.org/publications/jse/v2n1/mackisack.supp.html

Examples

```
data(SandwichAnts2)
```

SkateboardPrices   *Skateboard Prices*

Description

Prices of skateboards for sale online

Format

A dataset with 20 observations on the following variable.

- Price Selling price in dollars

Details

Prices for skateboards offered for sale on eBay.

Source

Random sample taken from all skateboards available for sale on eBay on February 12, 2012.

SleepCaffeine   *Sleep Caffeine*

Description

Experimentn to compare word recall after sleep or caffeine

Format

A data frame with 24 observations on the following 2 variables.

- Group Treatment: Caffeine Sleep
- Words Number of words recalled

Details

A random sample of 24 adults were divided equally into two groups and given a list of 24 words to memorize. During a break, one group takes a 90 minute nap while another group is given a caffeine pill. The response variable is the number of words participants are able to recall following the break.

Source

Mednick, Cai, Kanady, and Drummond, "Comparing the benefits of caffeine, naps and placebo on verbal, motor and perceptual memory", Behavioural Brain Research, 193 (2008), 79-86.

Examples

```
data(SleepCaffeine)
```

SleepStudy *Sleep Study*

Description

Data from a study of sleep patterns for college students.

Format

A dataset with 253 observations on the following 27 variables.

- Gender 1=male, 0=female
- Sex Female or Male
- ClassYear Year in school, 1=first year, ..., 4=senior
- LarkOwl Early riser or night owl? Lark, Neither, or Owl
- NumEarlyClass Number of classes per week before 9 am
- earlyClass Indicator for any early classes
- EarlyClass Indicator for any early classes
- GPA Grade point average (0-4 scale)
- ClassesMissed Number of classes missed in a semester
- CognitionZscore Z-score on a test of cognitive skills
- PoorSleepQuality Measure of sleep quality (higher values are poorer sleep)
- DepressionScore Measure of degree of depression
- AnxietyScore Measure of amount of anxiety
- StressScore Measure of amount of stress
- DepressionStatus Coded depression score: normal, moderate, or severe
- AnxietyStatus Coded anxiety score: normal, moderate, or severe
- Stress Coded stress score: normal or high
- DASScore Combined score for depression, anxiety and stress
- Happiness Measure of degree of happiness
- AlcoholUse Self-reported: Abstain, Light, Moderate, or Heavy
- Drinks Number of alcoholic drinks per week
- WeekdayBed Average weekday bedtime (24.0=midnight)
- WeekdayRise Average weekday rise time (8.0=8 am)
- WeekdaySleep Average hours of sleep on weekdays

- `WeekendBed` Average weekend bedtime (24.0=midnight)
- `WeekendRise` Average weekend rise time (8.0=8 am)
- `WeekendSleep` Average weekend bedtime (24.0=midnight)
- `AverageSleep` Average hours of sleep for all days
- `allNighter` Had an all-nighter this semester? `Yes` or `No`
- `AllNighter` Had an all-nighter this semester? `0` or `1`

Details

The data were obtained from a sample of students who did skills tests to measure cognitive function, completed a survey that asked many questions about attitudes and habits, and kept a sleep diary to record time and quality of sleep over a two week period.

Source

Onyper, S., Thacher, P., Gilbert, J., Gradess, S., "Class Start Times, Sleep, and Academic Performance in College: A Path Analysis," April 2012; 29(3): 318-335. Thanks to the authors for supplying the data.

Smiles   *Smiles*

Description

Experiment to study effect of smiling on leniency in judicial matters

Format

A data frame with 68 observations on the following 2 variables.

- `Leniency` Score assigned by a judgement panel (higher is more lenient)
- `Group` Treatment group: `neutral` or `smile`

Details

Hecht and LeFrance conducted a study examining the effect of a smile on the leniency of disciplinary action for wrongdoers. Participants in the experiment took on the role of members of a college disciplinary panel judging students accused of cheating. For each suspect, along with a description of the offense, a picture was provided with either a smile or neutral facial expression. A leniency score was calculated based on the disciplinary decisions made by the participants.

Source

LaFrance, M., & Hecht, M. A., "Why smiles generate leniency", Personality and Social Psychology Bulletin, 21, 1995, 207-214.

Examples

```
data(Smiles)
```

SpeedDating *Speed Dating*

Description

Data from a sample of four minute speed dates.

Format

A dataset with 276 observations on the following 22 variables.

- DecisionMale Would the male like another date? Yes or No
- DecisionM Would the male like another date? 0 or 1
- DecisionFemale Would the female like another date? Yes or No
- DecisionF Would the female like another date? 0 or 1
- LikeM How much the male likes his partner (1-10 scale)
- LikeF How much the female likes her partner (1-10 scale)
- PartnerYesM Male's estimate of chance the female wants another date (1-10 scale)
- PartnerYesF Female's estimate of chance the male wants another date (1-10 scale)
- AgeM Male's age (in years)
- AgeF Females age (in years)
- RaceM Male's race: Asian, Black, Caucasian, Latino, or Other
- RaceF Female's race: Asian, Black, Caucasian, Latino, or Other
- AttractiveM Male's rating of female's attractiveness (1-10 scale)
- AttractiveF Female's rating of male's attractiveness (1-10 scale)
- SincereM Male's rating of female's sincerity (1-10 scale)
- SincereF Female's rating of male's sincerity (1-10 scale)
- IntelligentM Male's rating of female's intelligence (1-10 scale)
- IntelligentF Female's rating of male's intelligence (1-10 scale)
- FunM Male's rating of female as fun (1-10 scale)
- FunF Female's rating of male as fun (1-10 scale)
- AmbitiousM Male's rating of female's ambition (1-10 scale)
- AmbitiousF Female's rating of male's ambition (1-10 scale)
- SharedInterestsM Male's rating of female's shared interests (1-10 scale)
- SharedInterestsF Female's rating of male's shared interests (1-10 scale)

Details

Participants were students at Columbia's graduate and professional schools, recruited by mass email, posted fliers, and fliers handed out by research assistants. Each participant attended one speed dating session, in which they met with each participant of the opposite sex for four minutes. Order and session assignments were randomly determined. After each four minute "speed date," participants filled out a form rating their date on a scale of 1-10 on various attributes. Only data from the first date in each session is recorded here.

Source

Gelman, A. and Hill, J., Data analysis using regression and multilevel/hierarchical models, Cambridge University Press: New York, 2007

StatGrades   *Statistics Exam Grades*

Description

Grades on statistics exams

Format

A dataset with 50 observations on the following 3 variables.

- Exam1 Score (out of 100 points) on the first exam
- Exam2 Score (out of 100 points) on the first exam
- Final Score (out of 100 points) on the final exam

Details

Exam scores for a sample of students who completed a course using Statistics: Unlocking the Power of Data as a text. The dataset contains scores on Exam 1 (Chapters 1 to 4), Exam 2 (Chapters 5 to 8), and the Final exam (entire book).

Source

Random selection of students in an introductory statistics course.

StatisticsPhD *Statistics PhD Programs*

Description

Enrollments in Statistics PhD Programs

Format

A data frame with 82 observations on the following 3 variables.

- University Name of the school
- Department Type of department: Biostatistics or Statistics
- FTGradEnrollment Full time graduate student enrollment

Details

Graduate student enrollments in Statistics and Biostatistics departments in 2009. The list does not include combined departments of mathematics and statistics and does not include departments that did not reply to the AMS survey.

Source

The full list of the 82 Group IV departments was obtained at http://www.ams.org/profession/data/annual-survey/group_iv. Data on enrollment were obtained primarily from Assistantships and Graduate Fellowships in the Mathematical Sciences, 2009, American Mathematical Society.

Examples

```
data(StatisticsPhD)
```

StockChanges *Stock Changes*

Description

Stock price change for a smaple of stocks from the S \& P 500 (August 2-6, 2010)

Format

A data frame with 50 observations on the following variable.

- SPChange Change in sock price (in dollars)

Details

A random sample of 50 companies from Standard \& Poor's index of 500 companies was selected. The change in the price of the stock (in dollars) over the 5-day period from August 2 - 6, 2010 was recorded for each company in the sample.

Examples

```
data(StockChanges)
```

StorySpoilers *Story Spoilers*

Description

Raitngs for stories with and without spoilers

Format

A dataset with 12 observations on the following 3 variables.

- Story ID for story
- Spoiler Average (0-10) rating for spoiler version
- Original Average (0-10) rating for original version

Details

This study investigated whether a story spoiler that gives away the ending early diminishes suspense and hurts enjoyment? For twelve different short stories, the study's authors created a second version in which a spoiler paragraph at the beginning discussed the story and revealed the outcome. Each version of the twelve stories was read by at least 30 people and rated on a 1 to 10 scale to create an overall rating for the story, with higher ratings indicating greater enjoyment of the story. Stories 1 to 4 were ironic twist stories, stories 5 to 8 were mysteries, and stories 9 to 12 were literary stories.

Source

Leavitt, J. and Christenfeld, N., "Story Spoilers Don't Spoil Stories," Psychological Science, published OnlineFirst, August 12, 2011.

StressedMice   *Stressed Mice*

Description

Time in darkness for mice in different environments

Format

A data frame with 14 observations on the following 2 variables.

- Time Time spent in darkness (in seconds)
- Environment Type of environment: Enriched or Standard

Details

In the recent study, mice were randomly assigned to either an enriched environment where there was an exercise wheel available, or a standard environment with no exercise options. After three weeks in the specified environment, for five minutes a day for two weeks, the mice were each exposed to a "mouse bully" - a mouse who was very strong, aggressive, and territorial. one measure of mouse anxiety is amount of time hiding in a dark compartment, with mice who are more anxious spending more time in darkness. The amount of time spent in darkness is recorded for each of the mice.

Source

Data approximated from summary statistics in: Lehmann and Herkenham, "Environmental Enrichment Confers Stress Resiliency to Social Defeat through an Infralimbic Cortex-Dependent Neuroanatomical Pathway", The Journal of Neuroscience, April 20, 2011, 31(16):61596173.

Examples

```
data(StressedMice)
```

StudentSurvey   *Student Survey Data*

Description

Data from a survey of students in introductory statistics courses

Format

A data frame with 362 observations on the following 17 variables.

- Year Year in school
- Gender Student's gender: F or M
- Smoke Smoker? No or Yes
- Award Prefered award: Academy Nobel Olympic
- HigherSAT Which SAT is higher? Math or Verbal
- Exercise Hours of exercsie per week
- TV Hours of TV viewing per week
- Height Height (in inches)
- Weight Weight (in pounds)
- Siblings Number of siblings
- BirthOrder Birth order, 1=oldest
- VerbalSAT Verbal SAT score
- MathSAT Math SAT score
- SAT Combined Verbal + Math SAT
- GPA College grade point average
- Pulse Pulse rate (beats per minute)
- Piercings Number of body piercings

Details

Data from an in-class survey given to introductory statistics students over several years.

Source

In-class student survey

Examples

```
data(StudentSurvey)
```

TenCountries   *Ten Countries*

Description

A subset of the ALLCountries data for a random sample of ten countries

Format

A data frame with 10 observations on the following 4 variables.

- `Country` a factor with levels `Armenia` `Bahamas` `Lebanon` `Macedonia` `North Korea` `Romania` `Serbia` `Slovenia` `Tunisia` `Uzbekistan`
- `Code` a factor with levels `ARM` `BHS` `LBN` `MKD` `PRK` `ROU` `SRB` `SVN` `TUN` `UZB`
- `Area` a numeric vector
- `PctRural` a numeric vector

Examples

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

TextbookCosts   *Textbook Costs*

Description

Prices for textbooks for different courses

Format

A data frame with 40 observations on the following 3 variables.

- `Field` General discipline of the course: `Arts`, `Humanities`, `NaturalScience`, or `SocialScience`
- `Books` Number of books requiired
- `Cost` Total cost (in dollars) for required books

Details

Data are from samples of ten courses in each of four disciplines at a liberal arts college. For each course the bookstore's website lists the required texts(s) and costs. Data were collected for the Fall 2011 semester.

Source

Bookstore online site

Examples

```
data(TextbookCosts)
```

ToenailArsenic   *Toenail Arsenic*

Description

Arsenic in toenails of 19 people using private wells in New Hampshire

Format

A data frame with 19 observations on the following variable.

- Arsenic Level of arsenic found in toenails

Details

Level of arsenic was measured in toenails of 19 subjects from New Hampshire, all with private wells as their main water source.

Source

Adapted from Karagas, et.al.,"Toenail Samples as an Indicator of Drinking Water Arsenic Exposure", Cancer Epidemiology, Biomarkers and Prevention 1996;5:849-852.

Examples

```
data(ToenailArsenic)
```

TrafficFlow	*Traffic Flow*

Description

Traffic flow times from a simulation with timed and flexible traffic lights

Format

A data frame with 24 observations on the following 3 variables.

- Timed Delay time (in minutes) for fixed timed lights
- Flexible Delay time (in minutes) for flexible communicating lights.
- Difference Difference (Timed-Flexible) for each simulation

Details

Engineers in Dresden, Germany were looking at ways to improve traffic flow by enabling traffic lights to communicate information about traffic flow with nearby traffic lights. The data in show results of one experiment where they simulated buses moving along a street and recorded the delay time (in seconds) for both a fixed time and a flexible system of lights. The process was repeated under both conditions for a sample of 24 simulated scenarios.

Source

Lammer and Helbing, "Self-Stabilizing decentralized signal control of realistic, saturated network traffic", Santa Fe Institute working paper \# 10-09-019, September 2010.

Examples

```
data(TrafficFlow)
```

USStates	*US State Data*

Description

Various data for all 50 US States

Format

A data frame with 50 observations on the following 17 variables.

- State Name of state
- HouseholdIncome Mean household income (in dollars)
- IQ Mean IQ score of residents
- McCainVote Percentage of votes for John McCain in 2008 Presidential election
- Region Area of the country: MW=Midwest, NE=Northeast, S=South, or W=West
- ObamaMcCain Which 2008 Presidential candidate won state? M=McCain or O=Obama
- Pres2008 Which 2008 Presidential candidate won state? M=McCain or O=Obama
- Population Number of residents (in millions)
- EighthGradeMath a numeric vector
- HighSchool Percentage of high school graduates
- GSP Gross State Product (dollars per capita)
- FiveVegetables Percentage of residents who eat at least five servings of fruits/vegetables per day
- Smokers Percentage of residents who smoke
- PhysicalActivity Percentage of residents who have competed in a physical activity in past month
- Obese Percentage of residents classified as obese
- College Percentage of residents with college degrees
- NonWhite Percentage of residents who are not white
- HeavyDrinkers Percentage of residents who drink heavily

Source

Various online sources, mostly at www.census.gov

Examples

```
data(USStates)
```

WaterStriders   *Water Striders*

Description

Mating activity for water striders

Format

A dataset with 10 observations on the following 3 variables.

- `AggressiveMale` Hyper-aggressive male in group? No or Yes
- `FemalesHiding` Proportion of time the female water striders were in hiding
- `MatingActivity` Measure of mean mating activity (higher numbers meaning more mating)

Details

Water striders are common bugs that skate across the surface of water. Water striders have different personalities and some of the males are hyper-aggressive, meaning they jump on and wrestle with any other water strider near them. Individually, because hyper-aggressive males are much more active, they tend to have better mating success than more inactive striders. This study examined the effect they have on a group. Four males and three females were put in each of ten pools of water. Half of the groups had a hyper-aggressive male as one of the males and half did not. The proportion of time females are in hiding was measured for each of the 10 groups, and a measure of mean mating activity was also measured with higher numbers meaning more mating.

Source

Sih, A. and Watters, J., "The mix matters: behavioural types and group dynamics in water striders," Behaviour, 2005; 142(9-10): 1423.

WaterTaste   *WaterTaste*

Description

Blind taste test to compare brands of bottled water

Format

A data frame with 100 observations on the following 10 variables.

- `Gender` Gender of respondent: F=Female M=Male
- `Age` Age (in years)
- `Class` Year in school F=First year J=Junior O=Other P SO=Sophomore SR=Senior

- UsuallyDrink Usual source of drinking water: `Bottled`, `Filtered`, or `Tap`

- FavBotWatBrand Favorite brand of bottled water

- Preference Order of perference: `A`=Sams Choice, `B`=Aquafina, `C`=Fiji, and `D`=Tap water

- First Top choice among `Aquafina`, `Fiji`, `SamsChoice`, or `Tap`

- Second Second choice

- Third Third choice

- Fourth Fourth choice

Details

Result from a blind taste test comparing different four different types of water (Sam's Choice, Aqufina, Fiji, and tap water). Participants rank ordered waters when presented in a random order.

Source

"Water Taste Test Data" by M. Leigh Lunsford and Alix D. Dowling Finch in the Journal of Statistics Education (Vol 18, No, 1) 2010

References

http://www.amstat.org/publications/jse/v18n1/lunsford.pdf

Examples

```
data(WaterTaste)
```

Wetsuits   *Wetsuits*

Description

Swim velocity (for 1500 meters) with and without wearing a wetsuit

Format

A data frame with 12 observations on the following 4 variables.

- Wetsuit Maximum swim velocity (m/sec) when wearing a wetsuit

- NoWetsuit Maximum swim velocity (m/sec) when wearing a regular bathing suit

- Gender Gender of swimmer: `F` or `M`

- Type Type of athlete: `swimmer` or `triathlete`

Details

A study tested whether wearing wetsuits influences swimming velocity. Twelve competitive swimmers and triathletes swam 1500m at maximum speed twice each; once wearing a wetsuit and once wearing a regular bathing suit. The order of the trials was randomized. Each time, the maximum velocity in meters/sec of the swimmer was recorded.

Source

de Lucas, R.D., Balildan, P., Neiva, C.M., Greco, C.C., Denadai, B.S. (2000). "The effects of wetsuits on physiological and biomechanical indices during swimming," Journal of Science and Medicine in Sport, 3 (1): 1-8.

Examples

```
data(Wetsuits)
```

Index