Package ‘fastR2’

May 8, 2026

Type Package

Title Foundations and Applications of Statistics Using R (2nd Edition)

Version 1.2.5

Description Data sets and utilities to accompany the second edition of ``Foundations and Applications of Statistics: an Introduction using R'' (R Pruim, published by AMS, 2017), a text covering topics from probability and mathematical statistics at an advanced undergraduate level. R is integrated throughout, and access to all the R code in the book is provided via the snippet() function.

License GPL (>= 2)

Depends R (>= 3.0.0), mosaic (>= 1.3.0)

Imports maxLik, numDeriv, dplyr, ggplot2 (>= 3.0.0), lattice, miscTools

Suggests ggformula, mosaicCalc, tidyr, readr, MASS, faraway, Hmisc, DAAG, multcomp, vcd, car, alr4, corrgram, BradleyTerry2, cubature, knitr, mosaicData, rmarkdown

VignetteBuilder knitr

LazyLoad yes

LazyData yes

RoxygenNote 7.3.2

Encoding UTF-8

URL https://github.com/rpruim/fastR2, http://rpruim.github.io/fastR2/

BugReports https://github.com/rpruim/fastR2/issues

NeedsCompilation no

Author Randall Pruim [aut, cre]

Maintainer Randall Pruim <rpruim@calvin.edu>

Repository CRAN

Date/Publication 2025-07-30 16:30:15 UTC

Contents

`ACTgpa` *ACT scores and GPA*

Description

ACT scores and college GPA for a small sample of college students.

Format

A data frame with 26 observations on the following 2 variables.

ACT ACT score

GPA GPA

Examples

```
gf_point(GPA ~ ACT, data = ACTgpa)
```

AirlineArrival *Airline On-Time Arrival Data*

Description

Flights categorized by destination city, airline, and whether or not the flight was on time.

Format

A data frame with 11000 observations on the following 3 variables.

airport a factor with levels `LosAngeles`, `Phoenix`, `SanDiego`, `SanFrancisco`, `Seattle`

result a factor with levels `Delayed`, `OnTime`

airline a factor with levels `Alaska`, `AmericaWest`

Source

Barnett, Arnold. 1994. "How numbers can trick you." *Technology Review*, vol. 97, no. 7, pp. 38–45.

References

These and similar data appear in many text books under the topic of Simpson's paradox.

Examples

```
tally(
  airline ~ result, data = AirlineArrival,
  format = "perc", margins = TRUE)
tally(
  result ~ airline + airport,
  data = AirlineArrival, format = "perc", margins = TRUE)
AirlineArrival2 <-
  AirlineArrival %>%
  group_by(airport, airline, result) %>%
  summarise(count = n()) %>%
  group_by(airport, airline) %>%
  mutate(total = sum(count), percent = count/total * 100) %>%
  filter(result == "Delayed")
AirlineArrival3 <-
  AirlineArrival %>%
  group_by(airline, result) %>%
  summarise(count = n()) %>%
  group_by(airline) %>%
  mutate(total = sum(count), percent = count/total * 100) %>%
  filter(result == "Delayed")
  gf_line(percent ~ airport, color = ~ airline, group = ~ airline,
          data = AirlineArrival2) %>%
    gf_point(percent ~ airport, color = ~ airline, size = ~total,
```

```
          data = AirlineArrival2) %>%
  gf_hline(yintercept = ~ percent, color = ~airline,
          data = AirlineArrival3, linetype = "dashed") %>%
  gf_labs(y = "percent delayed")
```

`AirPollution` *Air pollution measurements*

Description

Air pollution measurements at three locations.

Format

A data frame with 6 observations on the following 2 variables.

pollution a numeric vector

location a factor with levels `Hill Suburb`, `Plains Suburb`, `Urban City`

Source

David J. Saville and Graham R. Wood, *Statistical methods: A geometric primer*, Springer, 1996.

Examples

```
data(AirPollution)
summary(lm(pollution ~ location, data = AirPollution))
```

`BallDrop` *Ball dropping data*

Description

Undergraduate students in a physics lab recorded the height from which a ball was dropped and the time it took to reach the floor.

Format

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

height height in meters

time time in seconds

Source

Steve Plath, Calvin College Physics Department

Examples

```
gf_point(time ~ height, data = BallDrop)
```

`Batting` *Major League Batting 2000-2005*

Description

Major League batting data for the seasons from 2000-2005.

Format

A data frame with 8062 observations on the following 22 variables.

player unique identifier for each player

year year

stint for players who were traded mid-season, indicates which portion of the season the data cover

team three-letter code for team

league a factor with levels `AA` `AL` `NL`

G games

AB at bats

R runs

H hits

H2B doubles

H3B triples

HR home runs

RBI runs batted in

SB stolen bases

CS caught stealing

BB bases on balls (walks)

SO strike outs

IBB intentional base on balls

HBP hit by pitch

SH a numeric vector

SF sacrifice fly

GIDP grounded into double play

Examples

```
data(Batting)
gf_histogram( ~ HR, data = Batting)
```

`Buckthorn` *Buckthorn*

Description

Data from an experiment to determine the efficacy of various methods of eradicating buckthorn, an invasive woody shrub. Buckthorn plants were chopped down and the stumps treated with various concentrations of glyphosate. The next season, researchers returned to see whether the plant had regrown.

Format

A data frame with 165 observations on the following 3 variables.

shoots number of new shoots coming from stump

conc concentration of glyphosate applied

dead weather the stump was considered dead

Source

David Dornbos, Calvin College

Examples

```
data(Buckthorn)
```

`Bugs` *Bugs*

Description

This data frame contains data from an experiment to see if insects are more attracted to some colors than to others. The researchers prepared colored cards with a sticky substance so that insects that landed on them could not escape. The cards were placed in a field of oats in July. Later the researchers returned, collected the cards, and counted the number of cereal leaf beetles trapped on each card.

Format

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

color color of card; one of `B`(lue) `G`(reen) `W`(hite) `Y`(ellow)

trapped number of insects trapped on the card

Source

M. C. Wilson and R. E. Shade, Relative attractiveness of various luminescent colors to the cereal leaf beetle and the meadow spittlebug, *Journal of Economic Entomology* 60 (1967), 578–580.

Examples

```
data(Bugs)
favstats(trapped ~ color, data = Bugs)
```

`col.fastR` *Lattice Theme*

Description

A theme for use with lattice graphics.

Usage

```
col.fastR(bw = FALSE, lty = 1:7)
```

Arguments

`bw`	whether color scheme should be "black and white"
`lty`	vector of line type codes

Value

Returns a list that can be supplied as the `theme` to [lattice::trellis.par.set()].

Note

This theme was used in the production of the book *Foundations and Applications of Statistics*

Author(s)

Randall Pruim

See Also

[lattice::trellis.par.set()], [lattice::show.settings()]

Examples

```
trellis.par.set(theme=col.fastR(bw=TRUE))
show.settings()
trellis.par.set(theme=col.fastR())
show.settings()
```

col.perc — *Row and Column Percentages*

Description

Convenience wrappers around `apply()` to compute row and column percentages of matrix-like structures, including output of `xtabs`.

Usage

```
col.perc(x)

row.perc(x)
```

Arguments

`x`	matrix-like structure

Author(s)

Randall Pruim

Examples

```
row.perc(tally(~ airline + result, data = AirlineArrival))
col.perc(tally(~ airline + result, data = AirlineArrival))
```

Concrete — *Concrete Compressive Strength Data*

Description

These data were collected by I-Cheng Yeh to determine how the compressive strength of concrete is related to its ingredients (cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate) and age.

Format

`Concrete` is a data frame with the following variables.

limestone percentage of limestone

water water-cement ratio

strength compressive strength (MPa) after 28 days

References

Appeared in Devore's "Probability and Statsistics for Engineers and the Sciences (6th ed). The variables have been renamed.

`ConcreteAll` *#' Concrete Compressive Strength Data*

Description

These data were collected by I-Cheng Yeh to determine how the compressive strength of concrete is related to its ingredients (cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate) and age.

Format

`concreteAll` is a data frame with the following 9 variables.

cement amount of cement (kg/m^3)

slag amount of blast furnace slag (kg/m^3)

ash amount of fly ash(kg/m^3)

water amount of water (kg/m^3)

superP amount of superplasticizer (kg/m^3)

coarseAg amount of coarse aggregate (kg/m^3)

fineAg amount of fine aggregate (kg/m^3)

age age of concrete in days

strength compressive strength measured in MPa

`Concrete` is a subset of `ConcreteAll`.

Source

Data were obtained from the Machine Learning Repository (https://archive.ics.uci.edu/ml/) where they were deposited by I-Cheng Yeh (<icyeh@chu.edu.tw>) who retains the copyright for these data.

References

I-Cheng Yeh (1998), "Modeling of strength of high performance concrete using artificial neural networks," *Cement and Concrete Research*, Vol. 28, No. 12, pp. 1797-1808.

Examples

```
data(Concrete)
```

`CoolingWater` *Cooling Water*

Description

Temperature of a mug of water as it cools.

Usage

```
data(CoolingWater1)

data(CoolingWater2)

data(CoolingWater3)

data(CoolingWater4)
```

Format

A data frame with the following variables.

`time` time in seconds

`temp` temperature in Celsius (`CoolingWater1`, `CoolingWater2`) or Fahrenheit (`CoolingWater3`, `CoolingWater4`)

Source

These data were collected by Stan Wagon and his students at Macelester College to explore Newton's Law of Cooling and the ways that the law fails to capture all of the physics involved in cooling water. `CoolingWater1` and `CoolingWater2` appeared in a plot in Wagon (2013) and were (approximatley) extracted from the plot. `CoolingWater3` and `CoolingWater4` appeared in a plot in Wagon (2005). The data in `CoolingWater2` and `CoolingWater4` were collected with a film of oil on the surface of the water to minimize evaporation.

References

- R. Portmann and S. Wagon. "How quickly does hot water cool?" *Mathematica in Education and Research*, 10(3):1-9, July 2005.
- R. Israel, P. Saltzman, and S. Wagon. "Cooling coffee without solving differential equations". *Mathematics Magazine*, 86(3):204-210, 2013.

Examples

```
data(CoolingWater1)
data(CoolingWater2)
data(CoolingWater3)
data(CoolingWater4)
if (require(ggformula)) {
```

```
  gf_line(
    temp ~ time, color = ~ condition,
    data = rbind(CoolingWater1, CoolingWater2))
}
if (require(ggformula)) {
  gf_line(
    temp ~ time, color = ~ condition,
    data = rbind(CoolingWater3, CoolingWater4))
}
```

Corn *Corn Yield*

Description

William Gosset analyzed data from an experiment comparing the yield of regular and kiln-dried corn.

Format

A data frame with 11 observations on the following 2 variables.

reg yield of regular corn (lbs/acre)

kiln yield of kiln-dried corn (lbs/acre)

Details

Gosset (Student) reported on the results of seeding plots with two different kinds of seed. Each type of seed (regular and kiln-dried) was planted in adjacent plots, accounting for 11 pairs of "split" plots.

Source

These data are also available at DASL, the data and story library (https://dasl.datadescription.com/).

References

W.S. Gosset, "The Probable Error of a Mean," Biometrika, 6 (1908), pp 1-25.

Examples

```
Corn2 <- stack(Corn)
names(Corn2) <- c('yield','treatment')
lm(yield ~ treatment, data = Corn2)
t.test(yield ~ treatment, data = Corn2)
t.test(Corn$reg, Corn$kiln)
```

Cuckoo *Cuckoo eggs in other birds' nests*

Description

Cuckoos are knows to lay their eggs in the nests of other (host) birds. The eggs are then adopted and hatched by the host birds. These data were originally collected by O. M. Latter in 1902 to see how the size of a cuckoo egg is related to the species of the host bird.

Format

A data frame with 120 observations on the following 2 variables.

length length of egg (mm)

species a factor with levels `hedge sparrow` `meadow pipet` `pied wagtail` `robin` `tree pipet` `wren`

Source

L.H.C. Tippett, *The Methods of Statistics*, 4th Edition, John Wiley and Sons, Inc., 1952, p. 176.

References

These data are also available from DASL, the data and story library (https://dasl.datadescription.com/).

Examples

```
data(Cuckoo)
gf_boxplot(length ~ species, data = Cuckoo)
```

DeathPenalty *Death Penalty and Race*

Description

A famous example of Simpson's paradox.

Format

A data frame with 326 observations.

id a subject id

victim a factor with levels `Bl` `Wh`)

defendant a factor with levels `Bl`, `Wh`

death a factor with levels `Yes`, `No`

penalty a factor with levels `death` `other`

Source

Radelet, M. (1981). Racial characteristics and imposition of the death penalty. *American Sociological Review*, 46:918–927.

Examples

```
tally(penalty ~ defendant, data = DeathPenalty)
tally(penalty ~ defendant + victim, data = DeathPenalty)
```

Drag — *Drag force experiment*

Description

The data come from an experiment to determine how terminal velocity depends on the mass of the falling object. A helium balloon was rigged with a small basket and just the ballast to make it neutrally buoyant. Mass was then added and the terminal velocity is calculated by measuring the time it took to fall between two sensors once terminal velocity has been reached. Larger masses were drop from higher heights and used sensors more widely spaced.

Format

A data frame with 42 observations on the following 5 variables.

time time (in seconds) to travel between two sensors

mass net mass (in kg) of falling object

height distance (in meters) between two sensors

velocity average velocity (in m/s) computed from `time` and `height`

force.drag calculated drag force (in N, `force.drag` = `mass * 9.8`) using the fact that at terminal velocity, the drag force is equal to the force of gravity

Source

Calvin College physics students under the supervision of Professor Steve Plath.

Examples

```
data(Drag)
with(Drag, force.drag / mass)
gf_point(velocity ~ mass, data = Drag)
```

Endurance *Endurance and vitamin C*

Description

The effect of a single 600 mg dose of ascorbic acid versus a sugar placebo on the muscular endurance (as measured by repetitive grip strength trials) of fifteen male volunteers (19-23 years old).

Format

A data frame with 15 observations on the following 5 variables.

vitamin number of repetitions until reaching 50 maximal grip after taking viatimin

first which treatment was done first, a factor with levels `Placebo` `Vitamin`

placebo number of repetitions until reaching 50 strength after taking placebo

Details

Three initial maximal contractions were performed for each subject, with the greatest value indicating maximal grip strength. Muscular endurance was measured by having the subjects squeeze the dynamometer, hold the contraction for three seconds, and repeat continuously until a value of 50 maximum grip strength was achieved for three consecutive contractions. Endurance was defined as the number of repetitions required to go from maximum grip strength to the initial 50 positive verbal encouragement in an effort to have them complete as many repetitions as possible.

The study was conducted in a double-blind manner with crossover.

Source

These data are available from OzDASL, the Australasian data and story library (https://dasl.datadescription.com/).

References

Keith, R. E., and Merrill, E. (1983). The effects of vitamin C on maximum grip strength and muscular endurance. *Journal of Sports Medicine and Physical Fitness*, 23, 253-256.

Examples

```
data(Endurance)
t.test(Endurance$vitamin, Endurance$placebo, paired = TRUE)
t.test(log(Endurance$vitamin), log(Endurance$placebo), paired = TRUE)
t.test(1/Endurance$vitamin, 1/Endurance$placebo, paired = TRUE)
gf_qq( ~ vitamin - placebo, data = Endurance)
gf_qq( ~ log(vitamin) - log(placebo), data = Endurance)
gf_qq( ~ 1/vitamin - 1/placebo, data = Endurance)
```

`FamilySmoking` *Family smoking*

Description

A cross-tabulation of whether a student smokes and how many of his or her parents smoke from a study conducted in the 1960's.

Format

A data frame with 5375 observations on the following 2 variables.

student a factor with levels `DoesNotSmoke` `Smokes`

parents a factor with levels `NeitherSmokes` `OneSmokesBothSmoke`

Source

S. V. Zagona (ed.), *Studies and issues in smoking behavior*, University of Arizona Press, 1967.

References

The data also appear in

Brigitte Baldi and David S. Moore, *The Practice of Statistics in the Life Sciences*, Freeman, 2009.

Examples

```
data(FamilySmoking)
xchisq.test( tally(parents ~ student, data = FamilySmoking) )
```

`Fumbles` *NCAA football fumbles*

Description

This data frame gives the number of fumbles by each NCAA FBS team for the first three weeks in November, 2010.

Format

A data frame with 120 observations on the following 7 variables.

team NCAA football team

rank rank based on fumbles per game through games on November 26, 2010

W number of wins through games on November 26, 2010

L number of losses through games on November 26, 2010

week1 number of fumbles on November 6, 2010

week2 number of fumbles on November 13, 2010

week3 number of fumbles on November 20, 2010

Details

The fumble counts listed here are total fumbles, not fumbles lost. Some of these fumbles were recovered by the team that fumbled.

Source

https://www.teamrankings.com/college-football/stat/fumbles-per-game

Examples

```
data(Fumbles)
m <- max(Fumbles$week1)
table(factor(Fumbles$week1,levels = 0:m))
favstats( ~ week1, data = Fumbles)
# compare with Poisson distribution
cbind(
  fumbles = 0:m,
  observedCount = table(factor(Fumbles$week1,levels = 0:m)),
  modelCount= 120* dpois(0:m,mean(Fumbles$week1)),
  observedPct = table(factor(Fumbles$week1,levels = 0:m))/120,
  modelPct= dpois(0:m,mean(Fumbles$week1))
) %>% signif(3)
showFumbles <- function(x, lambda = mean(x),...) {
  result <-
    gf_dhistogram( ~ week1, data = Fumbles, binwidth = 1, alpha = 0.3) %>%
    gf_dist("pois", lambda = mean( ~ week1, data = Fumbles) )
  print(result)
  return(result)
}
showFumbles(Fumbles$week1)
showFumbles(Fumbles$week2)
showFumbles(Fumbles$week3)
```

geolm	*Geometric representation of linear model*

Description

geolm create a graphical representation of the fit of a linear model.

Usage

```
geolm(formula, data = parent.env(), type = "xz", version = 1, plot = TRUE, ...)

to2d(x, y, z, type = NULL, xas = c(0.4, -0.3), yas = c(1, 0), zas = c(0, 1))
```

Arguments

formula	a formula as used in lm.
data	a data frame as in lm.
type	character: indicating the type of projection to use to collapse multi-dimensional data space into two dimensions of the display.
version	an integer (currently 1 or 2). Which version of the plot to display.
plot	a logical: should the plot be displayed?
...	other arguments passed to lm
x, y, z	numeric.
xas, yas, zas	numeric vector of length 2 indicating the projection of c(1,0,0), c(0,1,0), and c(0,0,1).

Author(s)

Randall Pruim

See Also

lm.

Examples

```
geolm(pollution ~ location, data = AirPollution)
geolm(distance ~ projectileWt, data = Trebuchet2)
```

givenOrder *Create ordered factor with order inferred from order given*

Description

The order of the resulting factor is determined by the order in which unique labels first appear in the vector or factor x.

Usage

```
givenOrder(x)
```

Arguments

x a vector or factor to be converted into an ordered factor.

Examples

```
givenOrder(c("First", "Second", "Third", "Fourth", "Fifth", "Sixth"))
```

golfballs *Golf ball numbers*

Description

Allan Rossman used to live on a golf course in a spot where dozens of balls would come into his yard every week. He collected the balls and eventually tallied up the numbers on the first 5000 golf balls he collected. Of these 486 bore the number 1, 2, 3, or 4. The remaining 14 golf balls were omitted from the data.

Format

The format is: num [1:4] 137 138 107 104

Source

Data collected by Allan Rossman in Carlisle, PA.

Examples

```
data(golfballs)
golfballs/sum(golfballs)
chisq.test(golfballs, p = rep(.25,4))
```

GooosePermits *Goose permits*

Description

In a 1979 study by Bishop and Heberlein, 237 hunters were each offered one of 11 cash amounts (bids) ranging from $1 to $200 in return for their hunting permits. The data records how many hunters offered each bid kept or sold their permit.

Format

A data frame with 11 rows and 5 columns. Each row corresponds to a bid (in US dollars) offered for a goose permit. The colums `keep` and `sell` indicate how many hunters offered that bid kept or sold their permit, respectively. `n` is the sum of `keep` and `sell` and `prop_sell` is the proportion that sold.

References

Bishop and Heberlein (Amer. J. Agr. Econ. 61, 1979).

Examples

```
goose.mod <- glm( cbind(sell, keep) ~ log(bid), data = GoosePermits, family = binomial())
gf_point(0 ~ bid, size = ~keep, color = "gray50", data = GoosePermits) %>%
  gf_point(1 ~ bid, size = ~ sell, color = "navy") %>%
  gf_function(fun = makeFun(goose.mod)) %>%
  gf_refine(guides(size = "none"))

ggplot(data = GoosePermits) +
  geom_point( aes(x = bid, y = 0, size = keep), colour = "gray50") +
  geom_point( aes(x = bid, y = 1, size = sell), colour = "navy") +
  stat_function(fun = makeFun(goose.mod)) +
  guides( size = "none")

gf_point( (sell / (sell + keep)) ~ bid, data = GoosePermits,
    size = ~ sell + keep, color = "navy") %>%
  gf_function(fun = makeFun(goose.mod))  %>%
  gf_text(label = ~ as.character(sell + keep), colour = "white", size = 3) %>%
  gf_refine(scale_size_area()) %>%
  gf_labs(y = "probabity of selling")

ggplot(data = GoosePermits) +
  stat_function(fun = makeFun(goose.mod)) +
 geom_point( aes(x = bid, y = sell / (sell + keep), size = sell + keep), colour = "navy") +
  geom_text( aes(x = bid, y = sell / (sell + keep), label = as.character(sell + keep)),
    colour = "white", size = 3) +
  scale_size_area() +
  labs(y = "probabity of selling")
```

`GPA` *GPA, ACT, and SAT scores*

Description

GPA, ACT, and SAT scores for a sample of students.

Format

A data frame with 271 observations on the following 4 variables.

act ACT score

gpa college grade point average

satm SAT mathematics score

satv SAT verbal score

Examples

```
data(GPA)
splom(GPA)
```

`HeliumFootballs` *Punting helium- and air-filled footballs*

Description

Two identical footballs, one air-filled and one helium-filled, were used outdoors on a windless day at The Ohio State University's athletic complex. Each football was kicked 39 times and the two footballs were alternated with each kick. The experimenter recorded the distance traveled by each ball.

Format

A data frame with 39 observations on the following 3 variables.

trial trial number

air distance traveled by air-filled football (yards)

helium distance traveled by helium-filled football (yards)

Source

These data are available from DASL, the data and story library (https://dasl.datadescription.com/).

References

Lafferty, M. B. (1993), "OSU scientists get a kick out of sports controversy", *The Columbus Dispatch* (November, 21, 1993), B7.

Examples

```
data(HeliumFootballs)
gf_point(helium ~ air, data = HeliumFootballs)
gf_dhistogram(
  ~ (helium - air), data = HeliumFootballs,
  fill = ~ (helium > air),  bins = 15, boundary = 0
)
```

Ice *Cooling muscles with ice*

Description

This data set contains the results of an experiment comparing the efficacy of different forms of dry ice application in reducing the temperature of the calf muscle.

Details

The 12 subjects in this study came three times, at least four days apart, and received one of three ice treatments (cubed ice, crushed ice, or ice mixed with water). In each case, the ice was prepared in a plastic bag and applied dry to the subjects calf muscle. The temperature measurements were taken on the skin surface and inside the calf muscle (via a 4 cm long probe) every 30 seconds for 20 minutes prior to icing, for 20 minutes during icing, and for 2 hours after the ice had been removed. The temperature measurements are stored in variables that begin with `b` (baseline), `t` (treatment), or `r` (recovery) followed by a numerical code for the elapsed time formed by concatenating the number of minutes and seconds. For example, `R1230` contains the temperatures 12 minutes and 30 seconds after the ice had been removed.

Variables include

Subject identification number

sex a factor with levels `female` `male`

weight weight of subject (kg)

Height height of subject (cm)

Skinfold skinfold thickness

calf calf diameter (cm)

Age age of subject

location a factor with levels `intramuscular` `surface`

Treatment a factor with levels `crushed` `cubed` `wet`

B0 baseline temperature at time 0

b30 baseline temperature 30 seconds after start

b100 baseline temperature 1 minute after start

b1930 baseline temperature 19 minutes 30 seconds start

t0 treatment temperature at beginning of treatment

t30 treatment temperature 30 seconds after start of treatment

t100 treatment temperature 1 minute after start of treatment

t1930 treatment temperature 19 minutes 30 seconds after start of treatment

R0 recovery temperature at start of recovery

r30 recovery temperature 30 seconds after start of recovery

r100 recovery temperature 1 minute after start of recovery

r12000 recovery temperature 120 minutes after start of recovery

Source

Dykstra, J. H., Hill, H. M., Miller, M. G., Michael T. J., Cheatham, C. C., and Baker, R.J., Comparisons of cubed ice, crushed ice, and wetted ice on intramuscular and surface temperature changes, *Journal of Athletic Training* 44 (2009), no. 2, 136–141.

Examples

```
data(Ice)
gf_point(weight ~ skinfold, color = ~ sex, data = Ice)
if (require(readr) && require(tidyr)) {
  Ice2 <- Ice %>%
  gather("key", "temp", b0:r12000) %>%
  separate(key, c("phase", "time"), sep = 1) %>%
  mutate(time = parse_number(time), subject = as.character(subject))

  gf_line( temp ~ time, data = Ice2 %>% filter(phase == "t"),
           color = ~ sex,  group = ~subject, alpha = 0.6) %>%
    gf_facet_grid( treatment ~ location)
}
```

`Inflation` *Inflation data*

Description

The article developed four measures of central bank independence and explored their relation to inflation outcomes in developed and developing countries. This datafile deals with two of these measures in 23 nations.

Format

A data frame with 23 observations on the following 5 variables.

country country where data were collected

ques questionnaire index of independence

inf annual inflation rate, 1980-1989 (percent)

legal legal index of independence

dev developed (1) or developing (2) nation

Source

These data are available from OzDASL, the Australasian Data and Story Library (`https://dasl.datadescription.com/`).

References

A. Cukierman, S.B. Webb, and B. Negapi, "Measuring the Independence of Central Banks and Its Effect on Policy Outcomes," World Bank Economic Review, Vol. 6 No. 3 (Sept 1992), 353-398.

Examples

```
data(Inflation)
```

`information` *Information*

Description

Extract information from a maxLik object

Usage

```
information(object, ...)
```

Arguments

`object`	an object of class "`maxLik`".
`...`	additional arguments

`Jordan8687` *Michael Jordan personal scoring*

Description

The number of points scored by Michael Jordan in each game of the 1986-87 regular season.

Format

A data frame with 82 observations on the following 2 variables.

game a numeric vector

points a numeric vector

Examples

```
data(Jordan8687)
gf_qq(~ points, data = Jordan8687)
```

`Kids` *Goals and popularity factors for school kids*

Description

Subjects were students in grades 4-6 from three school districts in Michigan. Students were selected from urban, suburban, and rural school districts with approximately 1/3 of their sample coming from each district. Students indicated whether good grades, athletic ability, or popularity was most important to them. They also ranked four factors: grades, sports, looks, and money, in order of their importance for popularity. The questionnaire also asked for gender, grade level, and other demographic information.

Format

A data frame with 478 observations on the following 11 variables.

gender a factor with levels boy `girl`

grade grade in school

age student age

race a factor with levels `other` `White`

urban.rural a factor with levels `Rural` `Suburban` `Urban`

school a factor with levels `Brentwood Elementary` `Brentwood Middle` `Brown Middle` `Elm` `Main` `Portage` `Ridge` `Sand` `Westdale Middle`

goals a factor with levels `Grades` `Popular` `Sports`

grades rank of 'make good grades' (1 = most important for popularity; 4 = least important)

sports rank of 'beging good at sports' (1 = most important for popularity; 4 = least important)

looks rank of 'beging handsome or pretty' (1 = most important for popularity; 4 = least important)

money rank of 'having lots of money' (1 = most important for popularity; 4 = least important)

Source

These data are available at DASL, the data and story library (https://dasl.datadescription.com/).

References

Chase, M. A., and Dummer, G. M. (1992), "The Role of Sports as a Social Determinant for Children," Research Quarterly for Exercise and Sport, 63, 418-424.

Examples

```
data(Kids)
tally(goals ~ urban.rural, data = Kids)
chisq.test(tally(~ goals + urban.rural, data = Kids))
```

`LittleSurvey` *Results from a little survey*

Description

These data are from a little survey given to a number of students in introductory statistics courses. Several of the items were prepared in multiple versions and distributed randomly to the students.

Format

A data frame with 279 observations on the following 20 variables.

number a number between 1 and 30

colorver which version of the 'favorite color' question was on the survey. A factor with levels `v1` `v2`

color favorite color if among predefined choices. A factor with levels `black` `green` `other` `purple` `red`

othercolor favorite color if not among choices above.

animalver which version of the 'favorite color' question was on the survey. A factor with levels `v1` `v2`

animal favorite animal if among predefined choices. A factor with levels `elephant` `giraffe` `lion` `other`.

otheranimal favorite animal if not among the predefined choices.

pulsever which version of the 'pulse' question was on the survey

pulse self-reported pulse

tvver which of three versions of the TV question was on the survey

tvbox a factor with levels `<1` `>4` `>8` `1-2` `2-4` `4-8` `none` `other`

tvhours a numeric vector

surprisever which of two versions of the 'surprise' question was on the survey

surprise a factor with levels `no` `yes`

playver which of two versions of the 'play' question was on the survey

play a factor with levels `no` `yes`

diseasever which of two versions of the 'play' question was on the survey

disease a factor with levels `A` `B`

homeworkver which of two versions of the 'homework' question was on the survey

homework a factor with levels `A` `B`

Question Wording

1.1. Write down any number between 1 and 30 (inclusive).

2.1. What is your favorite color? Choices: black red; green; purple; other

2.2. What is your favorite color?

3.1. What is your favorite zoo animal? Choices: giraffe; lion; elephant; other

3.2. What is your favorite zoo animal?

4.1. Measure and record your pulse.

5.1. How much time have you spent watching TV in the last week?

5.2. How much time have you spent watching TV in the last week? Choises: none; under 1; hour 1-2 hours; 2-4 hours; more than 4 hours

5.3. How much time have you spent watching TV in the last week? Choises: under 1 hour; 1-2 hours; 2-4 hours; 4-8 hours; more than 8 hours

6.1. Social science researchers have conducted extensive empirical studies and concluded that the expression "absence makes the heart grow fonder" is generally true. Do you find this result surprising or not surprising?

6.2. Social science researchers have conducted extensive empirical studies and concluded that the expression "out of sight out of mind" is generally true. Do you find this result surprising or not surprising?

7.1. Suppose that you have decided to see a play for which the admission charge is $20 per ticket. As you prepare to purchase the ticket, you discover that you have lost a $20 bill. Would you still pay $20 for a ticket to see the play?

7.2. Suppose that you have decided to see a play for which the admission charge is $20 per ticket. As you prepare to enter the theater, you discover that you have lost your ticket. Would you pay $20 to buy a new ticket to see the play?

8.1. suppose that the United States is preparing for the outbreak of an unusual Asian disease that is expected to kill 600 people. Two alternative programs to combat the disease have been proposed.

Assume that the exact scientific estimates of the consequences of the programs are as follows: If program A is adopted, 200 people will be saved. If program B is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that nobody will be saved. Which of the two programs would you favor?

8.2. Suppose that the United States is preparing for the outbreak of an unusual Asian disease that is expected to kill 600 people. two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If program A is adopted, 400 people will die. If program B is adopted, there is a 1/3 probability that no one will die and a 2/3 probability that all 600 people will die. Which of the two programs would you favor? A or B

9.1. A national survey of college students revealed that professors at this college assign "significantly more homework that the nationwide average for an institution of its type." How does this finding compare with your experience? Choises: a. That sounds about right to me; b that doesn't sound right to me.

9.2. A national survey of college students revealed that professors at this college assign an amount of homework that "is fairly typical for institutions of its type." How does this finding compare with your experience? Choices: A that sounds about right to me; b that doesn't sound right to me.

Examples

```
data(LittleSurvey)
tally(surprise ~ surprisever, data = LittleSurvey)
tally(disease ~ diseasever, data = LittleSurvey)
```

`MathNoise` *Test performance and noise*

Description

In this experiment, hyperactive and control students were given a mathematics test in either a quiet or loud testing environment.

Format

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

score score on a mathematics test

noise a factor with levels `hi` `lo`

group a factor with levels `control` `hyper`

Source

Sydney S. Zentall and Jandira H. Shaw, Effects of classroom noise on perfor- mance and activity of second-grade hyperactive and control children, *Journal of Educational Psychology* 72 (1980), no. 6, 830.

Examples

```
data(MathNoise)
xyplot (score ~ noise, data = MathNoise, group = group, type = 'a',
auto.key = list(columns = 2, lines = TRUE, points = FALSE))

gf_jitter(score ~ noise, data = MathNoise, color = ~ group, alpha = 0.4,
          width = 0.1, height = 0) %>%
  gf_line(score ~ noise, data = MathNoise, color = ~ group, group = ~ group,
        stat = "summary")
```

---

maxLik2 *Augmented version of maxLik*

Description

This version of [max::maxLik()] stores additional information in the returned object enabling a plot method.

Usage

```
maxLik2(loglik, ..., env = parent.frame())
```

Arguments

`loglik`	a log-likelihood function as for [maxLik::maxLik()].
`...`	additional arguments passed to [maxLik::maxLik()].
`env`	an environment in which to evaluate `loglik`.

MIAA05 *MIAA basketball 2004-2005 season*

Description

Individual player statistics for the 2004-2005 Michigan Intercollegiate Athletic Association basketball season.

Format

A data frame with 134 observations on the following 27 variables.

number jersey number

player player's name

GP games played

GS games started

Min minutes played

AvgMin average minutes played per game

FG field goals made

FGA field goals attempted

FGPct field goal percentage

FG3 3-point field goals made

FG3A 3-point field goals attempted

FG3Pct 3-point field goal percentage

FT free throws made

FTA free throws attempted

FTPct free throw percentage

Off offensive rebounds

Def defensive rebounds

Tot total rebounds

RBG rebounds per game

PF personal fouls

FO games fouled out

A assists

TO turn overs

Blk blocked shots

Stl steals

Pts points scored

PTSG points per game

Source

MIAA sports archives (https://miaa.org/)

Examples

```
data(MIAA05)
gf_histogram(~ FTPct, data = MIAA05)
```

MLB2004 *Major League Baseball 2004 team data*

Description

Team batting statistics, runs allowed, and runs scored for the 2004 Major League Baseball season.

Format

A data frame with 30 observations on the following 20 variables.

team team city, a factor

league League, a factor with levels `AL` `NL`

W number of wins

L number of losses

G number of games

R number of runs scored

OR oppnents' runs – number of runs allowed

Rdiff run difference – `R - OR`

AB number of at bats

H number of hits

DBL number of doubles

TPL number of triples

HR number of home runs

BB number of walks (bases on balls)

SO number of strike outs

SB number of stolen bases

CS number of times caught stealing

BA batting average

SLG slugging percentage

OBA on base average

Examples

```
data(MLB2004)
gf_point(W ~ Rdiff, data = MLB2004)
```

NCAAbb *NCAA Division I Basketball Results*

Description

Results of NCAA basketball games

Format

Nine variables describing NCAA Division I basketball games.

date date on which game was played

away visiting team

ascore visiting team's score

home home team

hscore home team's score

notes code indicting games played at neutral sites (n or N) or in tournaments (T)

location where game was played

season a character indicating which season the game belonged to

postseason a logical indicating whether the game is a postseason game

Source

https://kenpom.com

Examples

```
data(NCAAbb)
# select one year and add some additional variables to the data frame
NCAA2010 <-
  NCAAbb %>%
  filter(season == "2009-10") %>%
  mutate(
    dscore = hscore - ascore,
    homeTeamWon = dscore > 0,
    numHomeTeamWon <- -1 + 2 * as.numeric(homeTeamWon),
    winner = ifelse(homeTeamWon, home, away),
    loser  = ifelse(homeTeamWon, away, home),
    wscore = ifelse(homeTeamWon, hscore, ascore),
    lscore = ifelse(homeTeamWon, ascore, hscore)
  )
NCAA2010 %>% select(date, winner, loser, wscore, lscore, dscore, homeTeamWon) %>% head()
```

NFL2007 *NFL 2007 season*

Description

Results of National Football League games (2007 season, including playoffs)

Format

A data frame with 267 observations on the following 7 variables.

date date on which game was played

visitor visiting team

visitorScore score for visiting team

home home team

homeScore score for home team

line ‘betting line’

totalLine ’over/under’ line (for combined score of both teams)

Examples

```
data(NFL2007)
NFL <- NFL2007
NFL$dscore <- NFL$homeScore - NFL$visitorScore
w <- which(NFL$dscore > 0)
NFL$winner <- NFL$visitor; NFL$winner[w] <- NFL$home[w]
NFL$loser <- NFL$home; NFL$loser[w] <- NFL$visitor[w]
# did the home team win?
NFL$homeTeamWon <- NFL$dscore > 0
table(NFL$homeTeamWon)
table(NFL$dscore > NFL$line)
```

nlmax *Nonlinear maximization and minimization*

Description

`nlmin` and `nlmax` are thin wrappers around `nlm`, a non-linear minimizer. `nlmax` avoids the necessity of modifying the function to construct a minimization problem from a problem that is naturally a maximization problem. The `summary` method for the resulting objects provides output that is easier for humnans to read.

Usage

```
nlmax(f, ...)

nlmin(f, ...)

## S3 method for class 'nlmax'
summary(object, nsmall = 4, ...)

## S3 method for class 'nlmin'
summary(object, nsmall = 4, ...)
```

Arguments

`f`	a function to optimize
`...`	additional arguments passed to `nlm`. Note that `p` is a required argument for `nlm`. See the help for `nlm` for details.
`object`	an object returned from `nlmin` or `nlmax`
`nsmall`	a numeric passed through to `format`

Examples

```
summary( nlmax( function(x) 5 - 3*x - 5*x^2, p=0 ) )
```

`Noise` *Noise*

Description

In order to test the effect of room noise, subjects were given a test under 5 different sets of conditions: 1) no noise, 2) intermittent low volume, 3) intermittent high volume, 4) continuous low volume, and 5) continuous high volume.

Format

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

id subject identifier

score score on the test

condition numeric code for condition

volume a factor with levels `high` `low` `none`

frequency a factor with levels `continuous` `intermittent` `none`

Examples

```
data(Noise)
Noise2 <- Noise %>% filter(volume != 'none')
model <- lm(score ~ volume * frequency, data = Noise2)
anova(model)
gf_jitter(score ~ volume, data = Noise2, color = ~ frequency,
          alpha = 0.4, width = 0.1, height = 0) %>%
  gf_line(score ~ volume, data = Noise2, group = ~frequency, color = ~ frequency,
          stat = "summary")

gf_jitter(score ~ frequency, data = Noise2, color = ~ volume,
          alpha = 0.4, width = 0.1, height = 0) %>%
  gf_line(score ~ frequency, data = Noise2, group = ~ volume, color = ~ volume,
          stat = "summary")
```

`Pallets` *Pallet repair data*

Description

The paletts data set contains data from a firm that recycles paletts. Paletts from warehouses are bought, repaired, and resold. (Repairing a palette typically involves replacing one or two boards.) The company has four employees who do the repairs. The employer sampled five days for each employee and recorded the number of pallets repaired.

Format

A data frame with 20 observations on the following 3 variables.

pallets number of pallets repaired

employee a factor with levels `A` `B` `C` `D`

day a factor with levels `day1` `day2` `day3` `day4` `day5`

Source

Michael Stob, Calvin College

Examples

```
data(Pallets)
# Do the employees differ in the rate at which they repair pallets?
pal.lm1 <- lm(pallets ~ employee, data = Pallets)
anova(pal.lm1)
# Now using day as a blocking variable
pal.lm2 <- lm(pallets ~ employee + day, data = Pallets)
anova(pal.lm2)
gf_line(pallets ~ day, data = Pallets,
```

```
group = ~employee,
color = ~employee) %>%
  gf_point() %>%
  gf_labs(title = "Productivity by day and employee")
```

`PaperPlanes` *Paper airplanes*

Description

Student-collected data from an experiment investigating the design of paper airplanes.

Format

A data frame with 16 observations on the following 5 variables.

distance distance plane traveled (cm)

paper type of paper used

angle a numeric vector

design design of plane (`hi performance` or `simple`)

order order in which planes were thrown

Details

These data were collected by Stewart Fischer and David Tippetts, statistics students at the Queensland University of Technology in a subject taught by Dr. Margaret Mackisack. Here is their description of the data and its collection:

The experiment decided upon was to see if by using two different designs of paper aeroplane, how far the plane would travel. In considering this, the question arose, whether different types of paper and different angles of release would have any effect on the distance travelled. Knowing that paper aeroplanes are greatly influenced by wind, we had to find a way to eliminate this factor. We decided to perform the experiment in a hallway of the University, where the effects of wind can be controlled to some extent by closing doors.

In order to make the experimental units as homogeneous as possible we allocated one person to a task, so person 1 folded and threw all planes, person 2 calculated the random order assignment, measured all the distances, checked that the angles of flight were right, and checked that the plane release was the same each time.

The factors that we considered each had two levels as follows:

Paper: A4 size, 80g and 50g

Design: High Performance Dual Glider, and Incredibly Simple Glider (patterns attached to original report)

Angle of release: Horizontal, or 45 degrees upward.

The random order assignment was calculated using the random number function of a calculator. Each combination of factors was assigned a number from one to eight, the random numbers were generated and accordingly the order of the experiment was found.

Source

These data are also available at OzDASL, the Australasian Data and Story Library (https://dasl.datadescription.com/).

References

Mackisack, M. S. (1994). What is the use of experiments conducted by statistics students? *Journal of Statistics Education*, 2, no 1.

Examples

```
data(PaperPlanes)
```

`Pendulum` *Pendulum data*

Description

Period and pendulum length for a number of string and mass pendulums constructed by physics students. The same mass was used throughout, but the length of the string was varied from 10cm to 16 m.

Format

A data frame with 27 observations on the following 3 variables.

length length of the pendulum (in meters)

period average time of period (in seconds) over several swings of the pendulum

delta.length an estimate of the accuracy of the length measurement

Source

Calvin College physics students under the direction of Professor Steve Plath.

Examples

```
data(Pendulum)
gf_point(period ~ length, data = Pendulum)
```

PetStress — *Pets and stress*

Description

Does having a pet or a friend cause more stress?

Format

A data frame with 45 observations on the following 2 variables.

group a factor with levels Control, Friend, or Pet

rate average heart rate while performing a stressful task

Details

Fourty-five women, all self-proclaimed dog-lovers, were randomly divided into three groups of subjects. Each performed a stressful task either alone, with a friend present, or with their dog present. The average heart rate during the task was used as a measure of stress.

Source

K. M. Allen, J. Blascovich, J. Tomaka, and R. M. Kelsey, Presence of human friends and pet dogs as moderators of autonomic responses to stress in women, *Journal of Personality and Social Psychology* 61 (1991), no. 4, 582–589.

References

These data also appear in

Brigitte Baldi and David S. Moore, *The Practice of Statistics in the Life Sciences*, Freeman, 2009.

Examples

```
data(PetStress)
xyplot(rate ~ group, data = PetStress, jitter.x = TRUE, type = c('p', 'a'))
gf_jitter(rate ~ group, data = PetStress, width = 0.1, height = 0) %>%
  gf_line(group = 1, stat = "summary", color = "red")
```

Pheno *FUSION type 2 diabetes study*

Description

Phenotype and genotype data from the Finland United States Investigation of NIDDM (type 2) Diabetes (FUSION) study.

Format

Data frames with the following variables.

id subject ID number for matching between data sets

t2d a factor with levels `case` `control`

bmi body mass index

sex a factor with levels `F` `M`

age age of subject at time phenotypes were colelcted

smoker a factor with levels `former` `never` `occasional` `regular`

chol total cholesterol

waist waist circumference (cm)

weight weight (kg)

height height (cm)

whr waist hip ratio

sbp systolic blood pressure

dbp diastolic blood pressure

marker RS name of SNP

markerID numeric ID for SNP

allele1 first allele coded as 1 = A, 2 = C, 3 = G, 4 = T

allele2 second allele coded as 1 = A, 2 = C, 3 = G, 4 = T

genotype both alleles coded as a factor

Adose number of A alleles

Cdose number of C alleles

Gdose number of G alleles

Tdose number of T alleles

Source

Similar to the data presented in

Laura J. Scott, Karen L. Mohlke, Lori L. Bonnycastle, Cristen J. Willer, Yun Li, William L. Duren, Michael R. Erdos, Heather M. Stringham, Pe- ter S. Chines, Anne U. Jackson, Ludmila Prokunina-Olsson, Chia-Jen J. Ding, Amy J. Swift, Narisu Narisu, Tianle Hu, Randall Pruim, Rui Xiao, Xiao- Yi Y. Li, Karen N. Conneely, Nancy L. Riebow, Andrew G. Sprau, Maurine Tong, Peggy P. White, Kurt N. Hetrick, Michael W. Barnhart, Craig W. Bark, Janet L. Goldstein, Lee Watkins, Fang Xiang, Jouko Saramies, Thomas A. Buchanan, Richard M. Watanabe, Timo T. Valle, Leena Kinnunen, Goncalo R. Abecasis, Elizabeth W. Pugh, Kimberly F. Doheny, Richard N. Bergman, Jaakko Tuomilehto, Francis S. Collins, and Michael Boehnke, A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility vari- ants, *Science* (2007).

Examples

```
data(Pheno); data(FUSION1); data(FUSION2)
FUSION1m <- merge(FUSION1, Pheno, by = "id", all.x = FALSE, all.y = FALSE)
xtabs( ~ t2d + genotype, data = FUSION1m)
xtabs( ~ t2d + Gdose, data = FUSION1m)
chisq.test( xtabs( ~ t2d + genotype, data = FUSION1m ) )
f1.glm <- glm( factor(t2d) ~ Gdose, data = FUSION1m, family = binomial)
summary(f1.glm)
```

Pigs *Pass the Pigs*

Description

This data set contains information collected from rolling the pair of pigs (found in the game "Pass the Pigs") 6000 times.

Format

A data frame with 6000 observations on the following 6 variables.

roll roll number (1-6000)

blackScore numerical code for position of black pig

black position of black pig coded as a factor

pinkScore numerical code for position of pink pig

pink position of pink pig coded as a factor

score score of the roll

height height from which pigs were rolled (5 or 8 inches)

start starting position of the pigs (0 = both pigs backwards, 1 = one bacwards one forwards, 2 = both forwards)

Details

In "Pass the Pigs", players roll two pig-shaped rubber dice and earn or lose points depending on the configuration of the rolled pigs. Players compete individually to earn 100 points. On each turn, a player rolls he or she decides to stop or until "pigging out" or

The pig configurations and their associated scores are

1 = Dot Up (0)

2 = Dot Down (0)

3 = Trotter (5)

4 = Razorback (5)

5 = Snouter (10)

6 = Leaning Jowler (15)

7 = Pigs are touching one another (-1; lose all points)

One pig Dot Up and one Dot Down ends the turn (a "pig out") and results in 0 points for the turn. If the pigs touch, the turn is ended and all points for the game must be forfeited. Two pigs in the Dot Up or Dot Down configuration score 1 point. Otherwise, The scores of the two pigs in different configurations are added together. The score is doubled if both both pigs have the same configuration, so, for example, two Snouters are worth 40 rather than 20.

Source

John C. Kern II, Duquesne University (`<kern@mathcs.duq.edu>`)

Examples

```
data(Pigs)
tally( ~ black, data = Pigs )
if (require(tidyr)) {
  Pigs %>%
  select(roll, black, pink) %>%
  gather(pig, state, black, pink) %>%
  tally( state ~ pig, data = ., format = "prop", margins = TRUE)
}
```

`Pitching2005` *Major League Baseball 2005 pitching*

Description

Major League Baseball pitching statistics for the 2005 season.

Format

A data frame with 653 observations on the following 26 variables.

playerID unique identifier for each player

yearID year

stint for players who played with multiple teams in the same season, `stint` is increased by one each time the player joins a new team

teamID three-letter identifier for team

lgID league team plays in, coded as `AL` or `NL`

W wins

L losses

G games played in

GS games started

CG complete games

SHO shut outs

SV saves recorded

IPouts outs recorded (innings pitched, measured in outs rather than innings)

H hits allowed

ER earned runs allowed

HR home runs allowed

BB walks (bases on balls) allowed

SO strike outs

ERA earned run average

IBB intentional walks

WP wild pitches

HBP number of batters hit by pitch

BK balks

BFP batters faced pitching

GF ratio of ground balls to fly balls

R runs allowed

Examples

```
data(Pitching2005)
gf_point(IPouts/3 ~ W, data = Pitching2005, ylab = "innings pitched", xlab = "wins")
```

`plot.maxLik2` *plot method for augment maxLik objects*

Description

See [maxLik2()] and [maxLik::maxLik()] for how to create the objects this method prints.

Usage

```
## S3 method for class 'maxLik2'
plot(x, y, ci = "Wald", hline = FALSE, ...)
```

Arguments

`x`	an object of class "maxLik2"
`y`	ignored
`ci`	a character vector with values among "Wald" and "likelihood" specifying the type of interval to display
`hline`	a logical indicating whether a horizontal line should be added
`...`	additional arguments, currently ignored.

`Poison` *Poison data*

Description

The data give the survival times (in hours) in a 3 x 4 factorial experiment, the factors being (a) three poisons and (b) four treatments. Each combination of the two factors is used for four animals. The allocation to animals is completely randomized.

Format

A data frame with 48 observations on the following 3 variables.

poison type of poison (1, 2, or 3)

treatment manner of treatment (1, 2, 3, or 4)

time time until death (hours)

Source

These data are also available from OzDASL, the Australian Data and Story Library (https://dasl.datadescription.com/). (Note: The time measurements of the data at OzDASL are in units of tens of hours.)

References

Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations (with Discussion). J. R. Statist. Soc. B, 26, 211-252.

Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using GLIM. Appl. Statist., 36, 332-339.

Smyth, G. K., and Verbyla, A. P. (1999). Adjusted likelihood methods for modelling dispersion in generalized linear models. Environmetrics 10, 696-709. http://www.statsci.org/smyth/pubs/ties98tr.html.

Examples

```
data(poison)
poison.lm <- lm(time~factor(poison) * factor(treatment), data = Poison)
plot(poison.lm,w = c(4,2))
anova(poison.lm)
# improved fit using a transformation
poison.lm2 <- lm(1/time ~ factor(poison) * factor(treatment), data = Poison)
plot(poison.lm2,w = c(4,2))
anova(poison.lm)
```

`Punting` *American football punting*

Description

Investigators studied physical characteristics and ability in 13 football punters. Each volunteer punted a football ten times. The investigators recorded the average distance for the ten punts, in feet. They also recorded the average hang time (time the ball is in the air before the receiver catches it), and a number of measures of leg strength and flexibility.

Format

A data frame with 13 observations on the following 7 variables.

distance mean distance for 10 punts (feet)

hang mean hang time (seconds)

rStrength right leg strength (pounds)

lStrength left leg strength (pounds)

rFlexibility right leg flexibility (degrees)

lFlexibility left leg flexibility (degrees)

oStrength overall leg strength (foot-pounds)

Source

These data are also available at OzDASL (https://dasl.datadescription.com/).

References

"The relationship between selected physical performance variables and football punting ability" by the Department of Health, Physical Education and Recreation at the Virginia Polytechnic Institute and State University, 1983.

Examples

```
data(Punting)
gf_point(hang ~ distance, data = Punting)
```

`RatPoison` *Rat poison – unfinished documentation*

Description

Data from an experiment to see whether flavor and location of rat poison influence the consumption by rats.

Format

A data frame with 20 observations on the following 3 variables.

consumption a numeric vector

flavor a factor with levels `bread` `butter-vanilla` `plain` `roast` `beef`

location a factor with levels `A` `B` `C` `D` `E`

Examples

```
data(RatPoison)
gf_line(consumption ~ flavor, group = ~ location, color = ~ location, data = RatPoison) %>%
  gf_point()
```

`rgolfballs` *Simulated golf ball data*

Description

A matrix of random golf ball numbers simulated using `rmultinom(n = 10000,size = 486,prob = rep(0.25,4))`.

Examples

```
data(rgolfballs)
```

RubberBand *Rubber band launching – unfinished documentation*

Description

Results of an experiment comparing a rubber band travels to the amount it was stretched prior to launch.

Format

A data frame with 16 observations on the following 2 variables.

stretch amount rubber band was stretched before launch

distance distance rubber band traveled

Examples

```
data(RubberBand)
gf_point(distance ~ stretch, data = RubberBand) %>%
  gf_lm(interval = "confidence")
```

Scent *Maze tracing and scents*

Description

Subjects were asked to to complete a pencil and paper maze when they were smelling a floral scent and when they were not.

Format

A data frame with 21 observations on the following 12 variables.

id ID number

sex a factor with levels `F` and`M`

smoker a factor with levels `N`, `Y`

opinion opinion of the odor (`indiff`, `neg`, or `pos`)

age age of subject (in years)

first which treatment was first, `scented` or `unscented`

u1 time (in seconds) in first unscented trial

u2 time (in seconds) in second unscented trial

u3 time (in seconds) in third unscented trial

s1 time (in seconds) in first scented trial

s2 time (in seconds) in second scented trial

s3 time (in seconds) in third scented trial

Source

These data are also available at DASL, the data and story library (https://dasl.datadescription.com/).

References

Hirsch, A. R., and Johnston, L. H. "Odors and Learning," Smell & Taste Treatment and Research Foundation, Chicago.

Examples

```
data(Scent)
summary(Scent)
```

---

snippet *Display or execute a snippet of R code*

Description

This command will display and/or execute small snippets of R code from the book *Foundations and Applications of Statistics: An Introduction Using R*.

Usage

```
snippet(
  name,
  eval = TRUE,
  execute = eval,
  view = !execute,
  echo = TRUE,
  ask = getOption("demo.ask"),
  verbose = getOption("verbose"),
  lib.loc = NULL,
  character.only = FALSE,
  regex = NULL,
  max.files = 10L
)
```

Arguments

`name`	name of snippet
`eval`	a logical. An alias for ‘execute‘.
`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 `name`as character string.
`regex`	ignored. Retained for backwards compatibility.
`max.files`	an integer limiting the number of files retrieved.

Details

`snippet` works much like `demo`, but the interface is simplified. Partial matching is used to select snippets, so any unique prefix is sufficient to specify a snippet. Sequenced snippets (identified by trailing 2-digit numbers) will be executed in sequence if a unique prefix to the non-numeric portion is given. To run just one of a sequence of snippets, provide the full snippet name. See the examples.

Author(s)

Randall Pruim

See Also

`demo`, `source`.

Examples

```
snippet("normal01")
# prefix works
snippet("normal")
# this prefix is ambiguous
snippet("norm")
# sequence of "histogram" snippets
snippet("hist", eval = FALSE, echo = TRUE, view = FALSE)
# just one of the "histogram" snippets
snippet("histogram04", eval = FALSE, echo = TRUE, view = FALSE)
# Prefix too short, but a helpful message is displayed
snippet("h", eval = FALSE, echo = TRUE, view = FALSE)
```

Soap *Dwindling soap*

Description

A bar of soap was weighed after showering to see how much soap was used each shower.

Format

A data frame with 15 observations on the following 3 variables.

date

day days since start of soap usage and data collection

weight weight of bar of soap (in grams)

Details

According to Rex Boggs:

I had a hypothesis that the daily weight of my bar of soap [in grams] in my shower wasn't a linear function, the reason being that the tiny little bar of soap at the end of its life seemed to hang around for just about ever. I wanted to throw it out, but I felt I shouldn't do so until it became unusable. And that seemed to take weeks.

Also I had recently bought some digital kitchen scales and felt I needed to use them to justify the cost. I hypothesized that the daily weight of a bar of soap might be dependent upon surface area, and hence would be a quadratic function . . . .

The data ends at day 22. On day 23 the soap broke into two pieces and one piece went down the plughole.

Source

Data collected by Rex Boggs and available from OzDASL (https://dasl.datadescription.com/).

Examples

```
data(Soap)
gf_point(weight ~ day, data = Soap)
```

Spheres *Measuring spheres*

Description

Measurements of the diameter (in meters) and mass (in kilograms) of a set of steel ball bearings.

Format

A data frame with 12 observations on the following 2 variables.

diameter diameter of bearing (m)

mass mass of the bearing (kg)

Source

These data were collected by Calvin College physics students under the direction of Steve Plath.

Examples

```
data(Spheres)
gf_point(mass ~ diameter, data = Spheres)
gf_point(mass ~ diameter, data = Spheres) %>%
  gf_refine(scale_x_log10(), scale_y_log10())
```

SSplot *Sum of Squares Plots*

Description

This function creates plots showing the "consumption" of residual sum of squares resulting from adding predictors to a model.

Usage

```
SSplot(
  model1,
  model2,
  n = 1,
  col1 = "gray50",
  size1 = 0.6,
  col2 = "navy",
  size2 = 1,
  col3 = "red",
  size3 = 1,
  ...,
  env = parent.frame()
)
```

Arguments

`model1`	a linear model
`model2`	a linear model, often using [mosaic::rand()].
`n`	an integer specifying how many times to regenerate `model2`.
`col1, col2, col3`	Colors for the line segments in the plot
`size1, size2, size3`	Sizes of the line segments in the plot
`...`	additional arguments (currently ignored)
`env`	an environment in which to evaluate the models.

Examples

```
SSplot(
  lm(strength ~ limestone + water, data = Concrete),
  lm(strength ~ limestone + rand(7), data = Concrete),
  n = 50)
## Not run:
SSplot(
  lm(strength ~ water + limestone, data = Concrete),
  lm(strength ~ water + rand(7), data = Concrete),
  n = 1000)

## End(Not run)
```

`Step` *Stepping experiment*

Description

An experiment was conducted by students at The Ohio State University in the fall of 1993 to explore the nature of the relationship between a person's heart rate and the frequency at which that person stepped up and down on steps of various heights.

Format

A data frame with 30 observations on the following 7 variables.

order performance order

block number of experimenter block

restHR resting heart rate (beats per minute)

HR final heart rate

height height of step (`hi` or `lo`)

freq whether subject stepped `fast`, `medium`, or `slow`

Details

An experiment was conducted by students at The Ohio State University in the fall of 1993 to explore the nature of the relationship between a person's heart rate and the frequency at which that person stepped up and down on steps of various heights. The response variable, heart rate, was measured in beats per minute. There were two different step heights: 5.75 inches (coded as `lo`), and 11.5 inches (coded as `hi`). There were three rates of stepping: 14 steps/min. (coded as `slow`), 21 steps/min. (coded as `medium`), and 28 steps/min. (coded as `fast`). This resulted in six possible height/frequency combinations. Each subject performed the activity for three minutes. Subjects were kept on pace by the beat of an electric metronome. One experimenter counted the subject's pulse for 20 seconds before and after each trial. The subject always rested between trials until her or his heart rate returned to close to the beginning rate. Another experimenter kept track of the time spent stepping. Each subject was always measured and timed by the same pair of experimenters to reduce variability in the experiment. Each pair of experimenters was treated as a block.

Source

These data are available at DASL, the data and story library (https://dasl.datadescription.com/).

Examples

```
data(Step)
gf_jitter(HR-restHR ~ freq, color = ~height, data = Step, group = ~height,
          height = 0, width = 0.1) %>%
  gf_line(stat = "summary", group = ~height)
gf_jitter(HR-restHR ~ height, color = ~freq, data = Step, group = ~freq,
          height = 0, width = 0.1) %>%
  gf_line(stat = "summary", group = ~freq)
```

`Stereogram` *Stereogram fusion*

Description

Results of an experiment on the effect of prior information on the time to fuse random dot stereograms. One group (NV) was given either no information or just verbal information about the shape of the embedded object. A second group (group VV) received both verbal information and visual information (e.g., a drawing of the object).

Format

A data frame with 78 observations on the following 2 variables.

time time until subject was able to fuse a random dot stereogram

group treatment group: `NV`(no visual instructions) `VV` (visual instructions)

Source

These data are available at DASL, the data and story library (https://dasl.datadescription.com/).

References

Frisby, J. P. and Clatworthy, J. L., "Learning to see complex random-dot stereograms," *Perception*, 4, (1975), pp. 173-178.

Cleveland, W. S. *Visualizing Data*. 1993.

Examples

```
data(Stereogram)
favstats(time ~ group, data = Stereogram)
gf_violin(time ~ group, data = Stereogram, alpha = 0.2, fill = "skyblue") %>%
gf_jitter(time ~ group, data = Stereogram, height = 0, width = 0.25)
```

Students — *Standardized test scores and GPAs*

Description

Standardized test scores and GPAs for 1000 students.

Format

A data frame with 1000 observations on the following 6 variables.

ACT ACT score

SAT SAT score

grad has the student graduated from college?

gradGPA college GPA at graduation

hsGPA high school GPA

cohort year of graduation or expected graduation

Examples

```
data(Students)
gf_point(ACT ~ SAT, data = Students)
gf_point(gradGPA ~ hsGPA, data = Students)
```

`TasteTest` *Taste test data*

Description

The results from a study comparing different preparation methods for taste test samples.

Format

A data frame with 16 observations on 2 (`taste1`) or 4 (`tastetest`) variables.

score taste score from a group of 50 testers

scr a factor with levels `coarse` `fine`

liq a factor with levels `hi` `lo`

type a factor with levels `A` `B` `C` `D`

Details

The samples were prepared for tasting using either a coarse screen or a fine screen, and with either a high or low liquid content. A total taste score is recorded for each of 16 groups of 50 testers each. Each group had 25 men and 25 women, each of whom scored the samples on a scale from -3 (terrible) to 3 (excellent). The sum of these individual scores is the overall taste score for the group.

Source

E. Street and M. G. Carroll, *Preliminary evaluation of a food product*, Statistics: A Guide to the Unknown (Judith M. Tanur et al., eds.), Holden-Day, 1972, pp. 220-238.

Examples

```
data(TasteTest)
data(Taste1)
gf_jitter(score ~ scr, data = TasteTest, color = ~liq, width = 0.2, height =0) %>%
  gf_line(stat = "summary", group = ~liq)
df_stats(score ~ scr | liq, data = TasteTest)
```

tdf — *Compute degrees of freedom for a 2-sample t-test*

Description

This function computes degrees of freedom for a 2-sample t-test from the standard deviations and sample sizes of the two samples.

Usage

```
tdf(sd1, sd2, n1, n2)
```

Arguments

sd1	standard deviation of the sample 1
sd2	standard deviation of the sample 2
n1	size of sample 1
n2	size of sample 2

Value

estimated degrees of freedom for 2-sample t-test

Examples

```
data(KidsFeet, package="mosaicData")
fs <- favstats( length ~ sex, data=KidsFeet ); fs
t.test( length ~ sex, data=KidsFeet )
tdf( fs[1,'sd'], fs[2,'sd'], fs[1,'n'], fs[2,'n'])
```

TireWear — *Estimating tire wear*

Description

Tread wear is estimated by two methods: weight loss and groove wear.

Format

A data frame with 16 observations on the following 2 variables.

weight estimated wear (1000's of miles) base on weight loss

groove estimated wear (1000's of miles) based on groove wear

Source

These data are available at DASL, the Data and Story Library (https://dasl.datadescription.com/).

References

R. D. Stichler, G. G. Richey, and J. Mandel, "Measurement of Treadware of Commercial Tires", *Rubber Age*, 73:2 (May 1953).

Examples

```
data(TireWear)
gf_point(weight ~ groove, data = TireWear)
```

Traffic *New England traffic fatalities (1951-1959)*

Description

Used by Tufte as an example of the importance of context, these data show the traffic fatality rates in New England in the 1950s. Connecticut increased enforcement of speed limits in 1956. In their full context, it is difficult to say if the decline in Connecticut traffic fatalities from 1955 to 1956 can be attributed to the stricter enforcement.

Format

A data frame with 9 observations on the following 6 variables.

year a year from 1951 to 1959

cn.deaths number of traffic deaths in Connecticut

ny deaths per 100,000 in New York

cn deaths per 100,000 in Connecticut

ma deaths per 100,000 in Massachusetts

ri deaths per 100,000 in in Rhode Island

Source

Tufte, E. R. *The Visual Display of Quantitative Information*, 2nd ed. Graphics Press, 2001.

References

Donald T. Campbell and H. Laurence Ross. "The Connecticut Crackdown on Speeding: Time-Series Data in Quasi-Experimental Analysis", *Law & Society Review* Vol. 3, No. 1 (Aug., 1968), pp. 33-54.

Gene V. Glass. "Analysis of Data on the Connecticut Speeding Crackdown as a Time-Series Quasi-Experiment" *Law & Society Review*, Vol. 3, No. 1 (Aug., 1968), pp. 55-76.

Examples

```
data(Traffic)
gf_line(cn.deaths ~ year, data = Traffic)
if (require(tidyr)) {
  TrafficLong <-
    Traffic %>%
    select(-2) %>%
    gather(state, fatality.rate, ny:ri)
   gf_line(fatality.rate ~ year, group = ~state, color = ~state, data = TrafficLong) %>%
    gf_point(fatality.rate ~ year, group = ~state, color = ~state, data = TrafficLong) %>%
  gf_lims(y = c(0, NA))
}
```

`Trebuchet` *Trebuchet data*

Description

Measurements from an experiment that involved firing projectiles with a small trebuchet under different conditions.

Format

Data frames with the following variables.

object the object serving as projectilebean `big washerb` `bigWash` `BWB` `foose` `golf` `MWB` `SWB` `tennis ball` `wood`

projectileWt weight of projectile (in grams)

counterWt weight of counter weight (in kg)

distance distance projectile traveled (in cm)

form a factor with levels `a` `b` `B` `c` describing the configuration of the trebuchet.

Details

`Trebuchet1` and `Trebuchet2` are subsets of `Trebuchet` restricted to a single value of `counterWt`

Source

Data collected by Andrew Pruim as part of a Science Olympiad competition.

Examples

```
data(Trebuchet); data(Trebuchet1); data(Trebuchet2)
gf_point(distance ~ projectileWt, data = Trebuchet1)
gf_point(distance ~ projectileWt, data = Trebuchet2)
gf_point(distance ~ projectileWt, color = ~ factor(counterWt), data = Trebuchet) %>%
  gf_smooth(alpha = 0.2, fill = ~factor(counterWt))
```

`undocumented` *Undocumented functions*

Description

These objects are undocumented.

Details

Some are left-overs from a previous version of the book and package. In other cases, the functions are of limited suitability for general use.

Author(s)

Randall Pruim

`Unemployment` *Unemployment data*

Description

Unemployment data

Usage

```
data(Unemployment)
```

Format

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

- `unemp` Millions of unemployed people
- `production` Federal Reserve Board index of industrial production
- `year`
- `iyear` indexed year

Source

Paul F. Velleman and Roy E. Welsch. "Efficient Computing of Regression Diagnostics", The American Statistician, Vol. 35, No. 4 (Nov., 1981), pp. 234-242. (https://www.jstor.org/stable/2683296)

Examples

```
data(Unemployment)
```

vaov	*ANOVA vectors*

Description

Compute vectors associated with 1-way ANOVA

Usage

```
vaov(x, ...)

## S3 method for class 'formula'
vaov(x, data = parent.frame(), ...)
```

Arguments

`x`	a formula.
`...`	additional arguments.
`data`	a data frame.

Details

This is primarily designed for demonstration purposes to show how 1-way ANOVA models partition variance. It may not work properly for more complicated models.

Value

A data frame with variables including `grandMean`, `groupMean`, `ObsVsGrand`, `STotal`, `ObsVsGroup`, `SError`, `GroupVsGrand`, and `STreatment`. The usual SS terms can be computed from these by summing.

Examples

```
aov(pollution ~ location, data = AirPollution)
vaov(pollution ~ location, data = AirPollution)
```

wilson.ci — *Confidence Intervals for Proportions*

Description

Alternatives to `prop.test` and `binom.test`.

Usage

```
wilson.ci(x, n = 100, conf.level = 0.95)
```

Arguments

`x`	number of 'successes'
`n`	number of trials
`conf.level`	confidence level

Details

`wald.ci` produces Wald confidence intervals. `wilson.ci` produces Wilson confidence intervals (also called "plus-4" confidence intervals) which are Wald intervals computed from data formed by adding 2 successes and 2 failures. The Wilson confidence intervals have better coverage rates for small samples.

Value

Lower and upper bounds of a two-sided confidence interval.

Author(s)

Randall Pruim

References

A. Agresti and B. A. Coull, Approximate is better then 'exact' for interval estimation of binomial proportions, *American Statistician* 52 (1998), 119–126.

Examples

```
prop.test(12,30)
prop.test(12,30, correct=FALSE)
wald.ci(12,30)
wilson.ci(12,30)
wald.ci(12+2,30+4)
```

`WorkingWomen` *Women in the workforce*

Description

The labor force participation rate of women in each of 19 U.S. cities in each of two years. # Reference: United States Department of Labor Statistics # # Authorization: free use # # Description: # # Variable Names: # # 1. City: City in the United States # 2. labor72: Labor Force Participation rate of women in 1972 # 3. labor68: Labor Force Participation rate of women in 1968 # # The Data: #

Format

A data frame with 19 observations on the following 3 variables.

city name of a U.S. city (coded as a factor with 19 levels)

labor72 percent of women in labor force in 1972

labor68 percent of women in labor force in 1968

Source

These data are from the United States Department of Labor Statistics and are also available at DASL, the Data and Story Library (https://dasl.datadescription.com/).

Examples

```
data(WorkingWomen)
gf_point(labor72 ~ labor68, data = WorkingWomen)
```

Index